VOLUME 5 □ NUMBER 10 OCTOBER 1968

IEEE spectrum

features

Departments: *please turn to the next page*

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

the cover

Real-time acoustic holography offers intriguing prospects for application in many fields. The Bragg diffraction technique—one of the methods described in the article beginning on page 45—has enabled instantaneous acoustic images to be recorded successfully. The acoustic image of a small wire cube is shown on this month's cover.

versity, Ithaca, N.Y. He has been with the university since 1964.

H. G. Tasker (SM) has been named chairman emeritus of the board of Tasker Industries, Van Nuys, Calif.

R. R. Unterberger (M), formerly senior research associate at the Chevron Research Company, La Habra, Calif., has been named professor of geophysics at the College of Geosciences, Texas A&M University, College Station, Tex.

George Wells (M) has been appointed director of engineering for Babcock Aerospace Division, Babcock Electronics Corporation, Costa Mesa, Calif.

H. K. Weiss (SM) has been named chief scientist of the Analysis Laboratory, Data Systems Division, Litton Systems, Inc., Van Nuys, Calif.

Obituaries

Harold Rorden dies; was high-voltage specialist

Harold L. Rorden (M '30, SM '36, F '57, L), a consulting engineer with the American Electric Power Service Corporation, New York, N.Y., died July 27. He lived in Nutley, N.J.

Mr. Rorden was noted for devising a technique for the barehanded repair and manipulation of extra-high-voltage transmission lines. Linemen using this technique are supported in fiberglass buckets and are "grounded" to the same potential as the line being serviced.

He was born in The Dalles, Oreg., in 1900, and received the A.B. degree from Stanford University in 1925 and the E.E. degree from Stanford in 1926. He joined the General Electric Company as a student engineer and later went to Pittsfield, Mass., as a junior engineer in high-voltage laboratory development. In 1931 he became engineer, high-voltage research and transformer design.

Mr. Rorden joined the Ohio Brass Company, Barberton, Ohio, in 1933 as research engineer, high-voltage development. In 1942 he was employed by the Bonneville Power Administration in Portland, Oreg., as high-voltage engineer responsible for lightning and surge protection on BPA high-voltage transmission systems. He remained with this company until 1954, when he joined the American Gas and Electric Service Corporation, where he remained until his retirement in 1965.

Gordon William Gerell, expert on protective relays

Gordon William Gerell (M '27, SM '37, F' 54, L) of St Louis, Mo., died in April. He was formerly a system superintendent for the Union Electric Company of Missouri, St. Louis, and was noted for his contributions to improvements in electric service by application of protective relaying and in preventive maintenance of oil-insulated apparatus.

He was born in Toronto, Ont., Canada, in 1901 and was graduated from Washington University in 1924 with a bachelor of science degree in electrical engineering. Upon graduation, he joined the Union Electric Company.

In 1930 Mr. Gerell became interested in system relay protection, and he became company system relay engineer in 1942. In 1943 he instituted a program of original research on methods for determining the sludging characterisitics of transformer insulating oil. A direct result of the investigations was the development of the interfacial tension method of testing samples of insulating oil removed from transformers while in operation.

Ernest Aldor (A '52, M '57) of New York, N.Y., died recently.

Cole Alexander Armstrong (A '43, SM '46) of Washington, D.C., June 2.

Frank David Binns (M '57) of Nashville, Tenn., died June 28.

Conner Calhoun Clardy (M '08) of San Diego, Calif., January 10.

Lee Rowe Etheridge (M '55) of College Park, Md., June 15.

George W. Franks (M '47) of Annandale, Va.

James J. Gilmartin, Jr. (A '47, M '47) of Huntington Station, N.Y.

Richard M. Harrigan (S '67, M '67) of Buffalo, N.Y., June 26.

Dorothy J. H. Harrison (A '45) of Red Bank, N.J., May 6.

Joseph I. Heller (A '30, SM '45) of Brooklyn, N.Y., May 23.

Henry George Howard (M '08, SM '25, L) of Funchal Madeira, Portugal, August 2.

John H. Howard (SM '50) of Mountain Lakes, N.J., July 29.

Onni I. Hyvonen (M '58) of Brooklyn N.Y., August 1.

Eugene Robert Ingersoll (A '49, M '55, SM '56) of Woodland Hills, Calif.

George Milton Knauf (M '56) of Bellevue, Wash., June 19.

Douglas W. Matthews (A '47) of Santa Clara, Calif., July 15.

Irwin Meyer (M '46) of Chicago, Ill, May 6.

Carl G. Miller (M '30) of Chicago, Ill., May 23.

Raymond E. Morgan, (M '43, SM '49) of Schenectady, N.Y., March 30.

James E. Mulavey (M '55) of Southfield, Mich., March 21.

Dennis P. Quimby (S '66, M '68) of Dover, N.H., recently.

O. B. Rudolph (M '46, SM '63) of Knoxville, Tenn., October 1, 1967.

Frank Albert Samuelson (M '49) of Salem, Oreg.

Donald F. Stafford (S '64, M '67) of Austin, Tex., in July.

William H. Taylor (M '64) of Wynnewood, Pa., July 9.

Paul C. Thompson (M '24, L) of Apache Junction, Ariz., March 12.

Earl J. Tuttle (M '48, SM '54) of Oklahoma City, Okla., November 3.

Phyllis P. Whalen (A '54) of Cleveland, Ohio.

William Joseph Williams (A '51) of Morris Plains, N.J., October 18, 1967.

Charles Parkinson Wood (M '11, SM '13, L) of New York, N.Y., June 29.

Spectral lines

New IEEE information services. In January 1968 the IEEE became copublisher of *Electrical and Electronics Abstracts* (*EEA*)* with the Institution of Electrical Engineers (IEE) of London. Thus the IEEE began to participate in the editorial processing, indexing, and publishing of the approximately 30 000 abstracts of papers and reports in the electrical/electronics field included in the 1968 *EEA*. Negotiations have just been completed that extend the copublishing program of the IEE and IEEE into the future, and also extend it considerably in scope.

Several major changes in the services available are planned for 1969. First, IEEE will participate not only in the publication of *EEA* but also in Part C of *Science Abstracts*, which will change its name from *Control Abstracts* to *Computer and Control Abstracts* (*CCA*). This monthly periodical will be extended to include both the hardware and software aspects of computers and its content will be more than doubled to provide effective coverage of the rapidly expanding computer literature. The overlap of the coverage of this field in *EEA* will be reduced. Preliminary discussions are under way that may well result in close cooperation in this area with the Association for Computing Machinery and other members of the American Federation of Information Processing Societies, as well as with the International Federation for Information Processing. Since the International Federation on Automatic Control is already closely associated with the present *Control Abstracts*, it appears probable that *CCA* will become the most comprehensive, complete, and universally supported abstracts journal in this field.

Unfortunately, the cost of publishing these monthly journals (*EEA* and *CCA*) is such that they are outside the range of most individual members of the IEEE. (The 1969 prices have been set at $156 for *EEA*, $84 for *CCA*, and $192 for a combined order for both publications.) For coverage of the entire field of interest of the IEEE, both journals are needed. They will contain approximately 50 000 abstracts during 1969—34 000 in *EEA* and 17 000 in *CCA*, with an overlap of 1000.

In order to provide comprehensive current awareness service for individual members of the IEEE in these areas, it is planned to modify and expand *Current Papers in Electrotechnology* and *Current Papers in Control* to considerably larger monthly periodicals, which will contain titles, authors and affiliations, and complete bibliographical reference for each article abstracted in *EEA* and *CCA*.

* Part B of Science Abstracts, which has also included *Physics Abstracts* as Part A and *Control Abstracts* as Part C.

These publications will be renamed *Current Papers in Electrical and Electronics Engineering* (*CPE*) and *Current Papers in Computers and Control* (*CPC*), to indicate their extended coverage. They will be available to members at special prices of $9 for *CPE* and $6 for *CPC*, one half the public prices. These publications are designed primarily for the current awareness needs of individual members but are not convenient for retrospective retrieval purposes since they are not indexed. The contents will be classified according to the same categories used in *EEA* and *CCA*, however.

In order to provide for the information retrieval needs of the individual member, the feasibility of providing in a single volume year-end indexes of all the articles published in *EEA* and in *CCA* will be studied. Whether this possibility is economically viable if priced in the individual member range is still being investigated. The prices would depend primarily on the number of volumes sold. Of considerable importance in estimating the number of member purchases will be the number of individual subscribers to *CPE* and *CPC* and also the volume of sales of a combined index of the IEEE's publications, which will be published in March 1969 and will cover all papers appearing in IEEE technical periodicals or conference proceedings in 1968. A subscription to the "CP" publications and the single-volume year-end annual indexes would give the individual member both a current awareness service and a retrieval capability. It is anticipated that libraries and other information-handling agencies will continue to provide their users with complete abstract service included in *EEA* and *CCA*, which are indexed on a monthly basis and also provide a cumulative index every six months.

In 1969, all of these publications will be photocomposed for printing under computer control, from bibliographic information stored in a computer file. Plans are now being made to make magnetic tapes of that file available under suitable licensing arrangements, beginning in July 1969. These tapes should be an excellent source of input data for mechanized information services that cover any part of our field.

In developing these services in conjunction with the IEE, the IEEE is attempting to anticipate and supply the information needs of its members and the profession in the most effective and economical way. If you can see additional ways that we should be following, please exercise your prerogative as a member of the IEEE, and let us have your advice.

F. Karl Willenbrock

Authors

Acoustic imaging and holography (page 45)

Adrianus Korpel (SM) was born in Rotterdam, Netherlands, and received the M.S. degree in electrical engineering from the University of Delft in 1955. From 1956 through 1960, he was employed by the Research Laboratories of the Postmaster General's Department in Melbourne, Australia, where his main activities were in the field of information and communication theory as applied to television and later in the field of parametric amplifiers. He joined Zenith Radio Corporation in 1960 and, since 1963, has been head of the light-modulation group. This group is concerned with image-type applications of lasers such as television displays, sound visualization, and acoustic holography. Mr. Korpel is a member of the Royal Dutch Institute of Engineers, the Institution of Engineers of Australia, and the Physical Society of America.

Transportation—1993 (page 53)

William W. Seifert (SM) is director of Project Transport for Massachusetts Institute of Technology in Cambridge, Mass. He has held this position since 1964. He is also professor of engineering and electrical engineering. Prior to his association with M.I.T., Dr. Seifert was an instructor in electrical engineering at Rensselaer Polytechnic Institute in Troy, N.Y., the city of his birth. He received the B.E.E. degree from R.P.I. in 1941, the S.M. degree from M.I.T. in 1947, and the Sc.D. degree from M.I.T. in 1951. Dr. Seifert has been associated with many interesting projects at M.I.T. in addition to Project Transport. For several years he served in various responsible positions with the Dynamic Analyses and Control Laboratory. He is a member of ASEE and Sigma Xi and is the author of numerous technical papers and books. He served on the editorial committee for the special issue on transportation of the PROCEEDINGS OF THE IEEE in April of 1968.

Privacy and the computer age (page 58)

Charles P. Lickson is a member of the law firm of King, Plotkin, and Lickson in Stamford, Conn. He entered into the private practice of law in association with the firm in June of this year. Prior to this association, he was with the Bunker-Ramo Corporation in the capacity of counsel for the business and industry division in Stamford with the title of assistant general attorney and assistant secretary. He was born in New York City and received the B.A. degree in political science in 1961 from Johns Hopkins University. He entered Georgetown University Law Center in 1961 and received the Juris Doctorate in 1964. Dr. Lickson worked with the Division of Volunteer Field Support at the Peace Corps in Washington and was deputy clerk at the United States District Court for the District of Columbia. His work there included training as court clerk in civil, criminal, and motions courts. In 1964, he was appointed law clerk to the Honorable Alexander Holtzoff, United States District Judge for the District of Columbia. In 1965, Dr. Lickson was ordered to active duty in the United States Army and during his period of service was trial and defense counsel for Special Courts Martial, a member of Special and General Courts Martial, and Security Office with the rank of Captain U.S.A.R. He is a member of the American Bar Association, the World Peace Through Law Center, and the Stamford Bar Association.

Electronic halftones (page 64)


R. L. Hallows, Jr., received the B.S.E.E. degree in 1952 from the Missouri School of Mines and joined the Radio Corporation of America, specializing in developmental work and studies on television and imaging systems. His assignments have included work with RCA's Electron Tube Division and Astro-Electronics Division, and the RCA Laboratories. Mr. Hallows is currently associated with the RCA Graphic Systems Division, where he has worked on problems concerned with electronic halftone generation and color reproduction. He is a member of The Society of Motion Picture and Television Engineers.

R. J. Klensch received the B.S.E.E. degree from the University of Illinois in 1951 and did graduate work at Princeton University from 1952 to 1958. Mr. Klensch joined RCA Laboratories in 1952. From 1954 to 1956 he served in the U.S. Army and returned to RCA Laboratories in 1956. Subsequently, he did research in the areas of microwave scanning antennas, time-division multiplex systems, color television, and digital and analog communication systems. Recently Mr. Klensch has been investigating new electronic halftone generation techniques and CRT display systems.

Optical heterodyne detection (page 77)

O. E. DeLange (F) received the B.S. degree in electrical engineering from the University of Utah in 1930 and the M.A. degree in physics from Columbia University in 1937. From 1930 to the present time he has been employed by Bell Telephone Laboratories with most of this period devoted to radio research. He was involved in studies of FM up to the start of World War II. The war years were spent on the design and development of naval radar. A number of years following the war were devoted to studies of broadband amplifiers and pulse systems with emphasis on PCM. This work included a short-pulse, microwave propagation experiment and the devising of broadband oscilloscopes for viewing short pulses. He was responsible for the satellite-tracking radar employed at the Holmdel, N.J., laboratories for the Echo I experiment. In recent years he has worked on studies of light propagation and light transmission systems.

Complex waves (page 86)

James R. Melcher (M), associate professor of electrical engineering at the Massachusetts Institute of Technology, is a native of Iowa. In 1957 he received the B.S. degree in electrical engineering from Iowa State University, followed by the M.S. in nuclear engineering, and the Ph.D. degree from M.I.T. Since joining the M.I.T. faculty in 1962, his research and consulting interests have centered around continuum electromechanics, with major emphasis on a range of electrohydrodynamic topics, magnetohydrodynamics, continuum feedback control, and energy conversion. Basically, his research is aimed at understanding interactions between electric fields and fluids. His engineering projects include the orientation of cryogenics in zero-gravity environments. During the past six years, he has been heavily involved with the development of an undergraduate program in the broadly defined area of electromechanics.

Acoustic imaging and holography

Applications of acoustic holography are beginning to have far-reaching effects in more and more fields; especially with the increasing use of computers

Adrianus Korpel *Zenith Radio Corporation*

Applications of active and passive devices employing the characteristics of sound to obtain results have been known to investigators for many years. However, in addition to these sonar or pulse-echo methods of analysis, there exist a variety of techniques in which a sound image is obtained in a way more analogous to optics. This analogy has, in the last few years, been extended to include holographic techniques, and has led to the inception of a new discipline—acoustic holography. In this article, we trace the development of this new field as it originated from a fusion between "conventional" acoustic imaging and optical holography; and we see how, in the few years of its existence, it has already generated its own unique methods and problems.

The ability to "see" with sound has always been attractive to workers in such widely varying fields as, for example, geology, oceanography, structural engineering, and medicine. Thus, the geophysicist observes pulses of sound generated by explosive charges or even by earthquakes. These pulses travel along and through the earth's crust and, by their relative time of arrival, tell him about its composition and structure.[1] The oceanographer[2] sends his "sonar" pulses down to the sea bottom in a continuous stream and, by processing their echo returns, constructs maps and profiles in much the same way as the radar engineer. The same technique may be used by the physician[3] to map a patient's interior and see structures missed by X rays, and again by the structural engineer[4] for the location of voids and cracks in materials.

More recently, techniques have been developed to generate and record sound images, and the emergence of acoustic holography appears to be playing an increasingly vital role in understanding and solving a wide array of problems.

Holography may be described as a method for recording and reconstructing the amplitude and phase distribution of a propagating field in a given plane. Imaging, on the other hand, refers to a system that concerns itself only with the amplitude, or rather the power—that is, the square of the amplitude. The preservation of the phase in holography is of crucial importance in the sense that, upon reconstructing, the field is automatically reproduced faithfully everywhere in space and not only in the plane of the recording (hologram plane).[5–7] Thus, in the field of optics, holography has become identified with three-dimensional reconstruction, as vividly illustrated by the striking spectacle of toy trains, chessmen, statuettes, etc., suspended in space somewhere behind a hologram. It may be well to state right here that no such vistas await us in acoustic holography. The reason for this can be understood without knowing in detail how acoustic holograms are made.

The heart of the matter is that acoustic holograms are recorded at the wavelength of sound (Λ), but are then reconstructed at the wavelength of visible light (λ). This scaling down in wavelength makes exact three-dimensional reproduction impossible. Sound and light fields are described by similar equations, the so-called wave equations. In fact, in a scalar theory, the equations are identical if the corresponding wavelengths are used as units of length. Hence, a faithful three-dimensional reconstruction of the sound field (at Λ) in visible light (at λ) is only possible if all three dimensions are scaled by a factor λ/Λ. For the usual range of sound frequencies, this implies quite a considerable demagnification (e.g., a factor of 3000 for 1-MHz sound in water).

If we are not satisfied with a tiny model of the sound field, but prefer to reconstruct the field in such a way that, at least in the two lateral dimensions, there is no net change of length from the original, a longitudinal excess magnification by a factor Λ/λ is unavoidably introduced. (A familiar example of the same effect is the exaggerated depth of field we notice when using binoculars or taking pictures with a telephoto lens.) This distortion is a consequence of the basic wave equations and independent of the ingenuity of the experimenter.

It must be said in all fairness that not all my colleagues share this pessimistic view, and various schemes to reduce the distortions inherent in the reconstruction have been proposed.[8] In the meantime, however, we are faced with these sad alternatives: either we tolerate this longitudinal distortion and sacrifice three-dimensional viewing, or we scale everything down according to the ratio λ/Λ and use a microscope if the reconstruction becomes too small. In the latter, there is no perception of depth.

At this point, one may well ask whether acoustic holography is of any use whatsoever, since the best we can seemingly accomplish is to focus on certain cross sections and inspect them one by one rather than view the object in its entirety. Why not, in that case, be satisfied with a conventional acoustic imaging device, which, in combination with acoustic lenses, would permit us to do exactly the same thing?

To a certain extent (i.e., excluding the toy train spectacle), the field of optical holography has been haunted by this same embarrassing question, and only in the last few years have some unique applications become apparent. In acoustic holography, perhaps the most important consideration is simply the fact that the use of lenses is avoided. This is important, since the quality of acoustic lenses is generally inferior to that of optical ones[9]; moreover, they must of necessity be situated inside the medium through which the sound propagates. This makes focusing cumbersome if the sound medium is a liquid, and nearly impossible if we are dealing with a solid. Let us, therefore, for the time being simply stipulate the usefulness of acoustic holography and proceed with the details of the actual technique. It is convenient to divide the various methods into two groups: those that are similar in nature to optical holography, and those that are unique to the acoustic region.

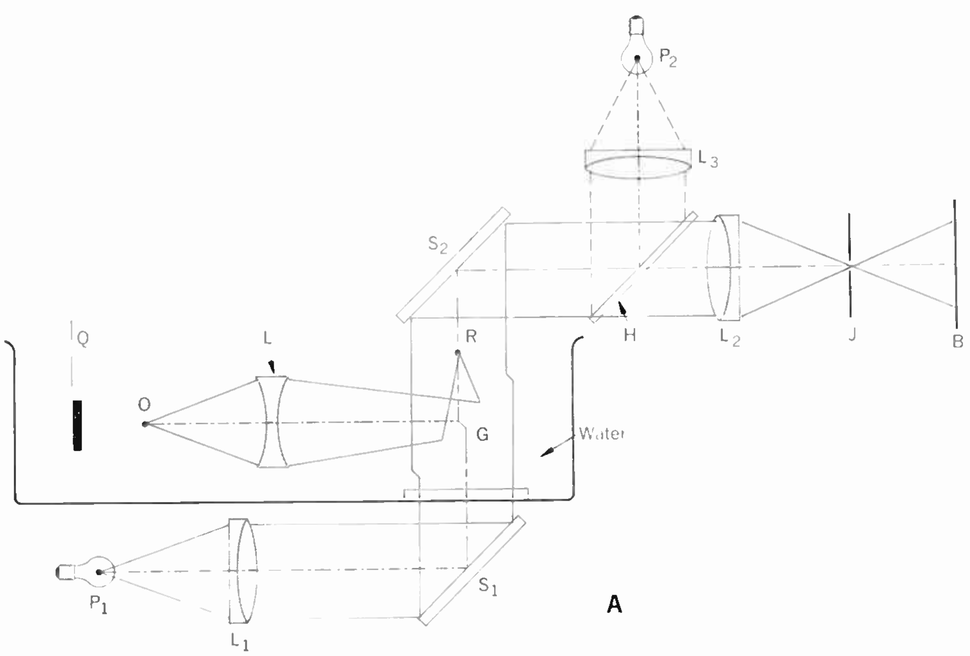
In optical holography, an image field is made to interfere with a so-called reference beam and the resulting interference pattern is recorded on photographic, thermoplastic, or photochromic film. This pattern consists of a system of fine fringes varying both in contrast and fringe spacing. The contrast at any particular point is a measure of the amplitude of the image field at that point, whereas the positions of the fringes relate to the phase, with their spacing being determined by the slope of the image-field wavefront relative to that of the reference beam. Thus, although the recording medium is basically responsive only to power, it is nevertheless possible to record both amplitude and phase by using a reference beam. There is a direct analogy to communication engineering[10] if the fringes are compared with the waves of a radio-frequency carrier modulated both in phase (fringe position) and amplitude (fringe contrast).

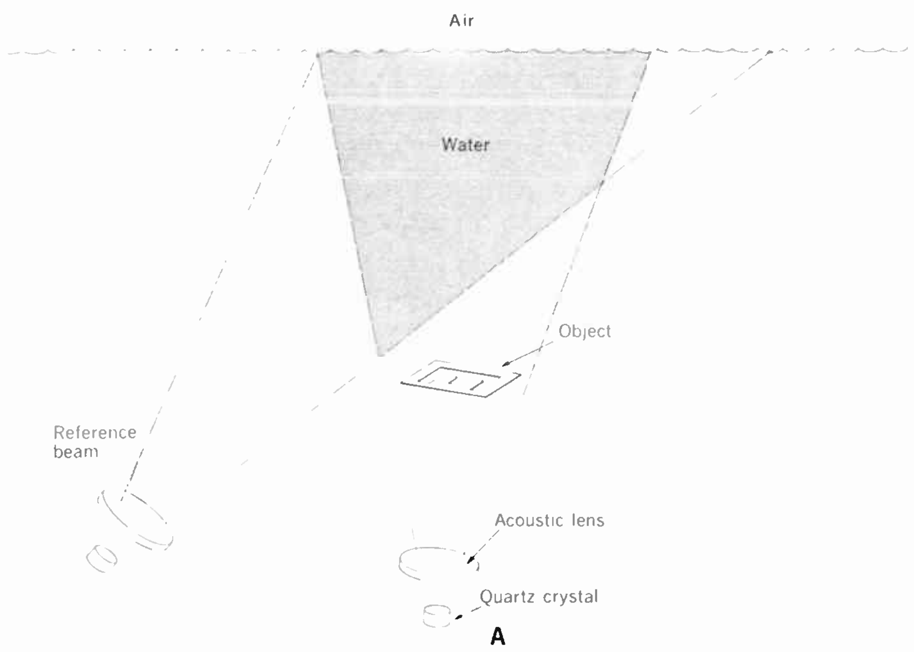
Reconstruction of the image field is accomplished by illuminating the recorded interference pattern with the original reference beam. Strictly speaking, this generates two related fields (conjugate images), which propagate in different directions and may be separated by spatial filters, a process analogous to the separation of sidebands by electrical filters.

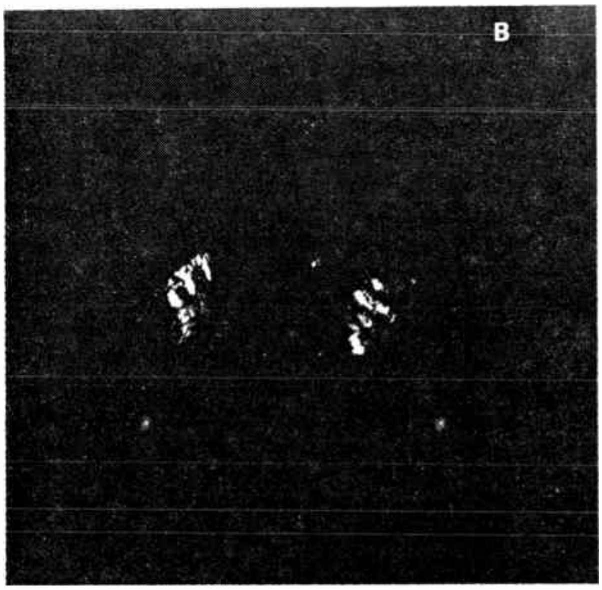
It is very straightforward to adapt this system to acoustic holography: We select a conventional acoustic image-conversion device and merely add an acoustic reference beam to the sound field. A pattern of fringes appears on the image-conversion device. We photograph this fringe pattern and illuminate the developed negative with a laser beam. Depending on the scale of the hologram, we can then inspect various cross sections of the sound field by one of the methods mentioned before. For instance, if the hologram is scaled down by λ/Λ, we obtain a faithful demagnified three-dimensional reproduction requiring the use of a microscope. Various technical refinements may be added to eliminate the twin image and reduce scatter due to irregularities in film thickness.[11,12]

The relative merits of the various methods in this group depend mainly on the properties of the particular acoustic image-conversion technique that they employ. Most of these techniques predate holography. An excellent survey may be found in Refs. 13 and 14. Not all

FIGURE 1. Surface-relief method of visualization (after Schuster).







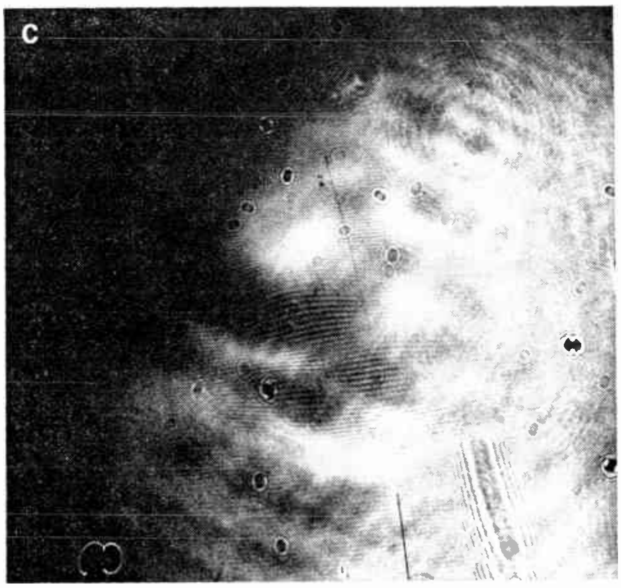


FIGURE 2. A—Holographic surface-relief method (after Mueller and Sheridan). B—Acoustic hologram of the letter "E." C—Reconstruction showing the letter "E" and the out-of-focus conjugate image.

of the techniques have been adapted to holography, but two that have been moderately successful are the surface-relief method[15] and a technique involving the so-called Sokolov tube.[16] In the surface-relief method (Fig. 1), a cross section of a sound field is imaged onto the surface of a liquid by means of an acoustic lens. The image exists in the form of a stationary pattern of surface perturbations caused by the radiation pressure of the sound, with gravity and surface tension acting as restoring forces. This may be viewed and photographed by transmission or reflection Schlieren techniques.

In such techniques, objects are photographed by only making use of the light that is scattered over large angles while discarding light scattered over narrow angles, or vice versa. Scattering centers such as surface perturbations then show up very clearly. Thus, in Fig. 1, the object O, illuminated by transducer Q, is imaged in R on the surface of the liquid via acoustic lens L and acoustic mirror G. Condenser lens L_1 forms a collimated beam of the light emitted by P_1. This beam is used to illuminate the liquid surface at R via mirror S_1 (acoustic mirror G is transparent for light). Lens L_2 collects the beam after it has traversed the ripple pattern near R and forms an image of the liquid surface on screen B. The Schlieren aperture J, placed in the focal plane of L_2, blocks light scattered at large angles. With this arrangement, the unperturbed water surface shows up as a uniformly illuminated area. Any subsequent ripples appear as dark lines against this bright background. If desired, reflection from the liquid surface may be used, rather than transmission through it, by the combination of light source P_2, condenser lens L_3, and half-silvered mirror H.

For the purposes of photographing the surface perturbation, it is often sufficient merely to use a camera with a lens opening small enough to serve as a Schlieren aperture. Mueller and Sheridan[17] adapted this method to holography by adding a sound reference beam as

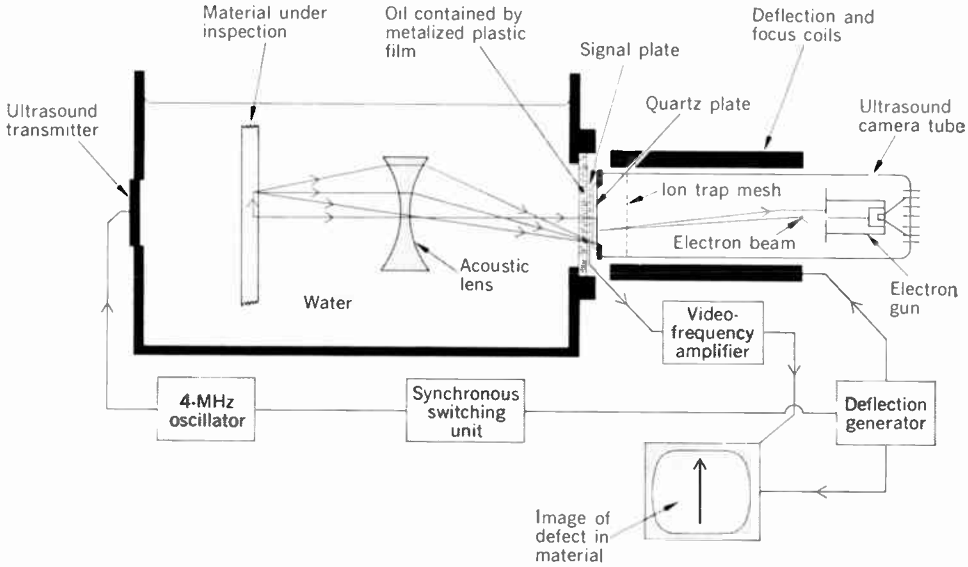
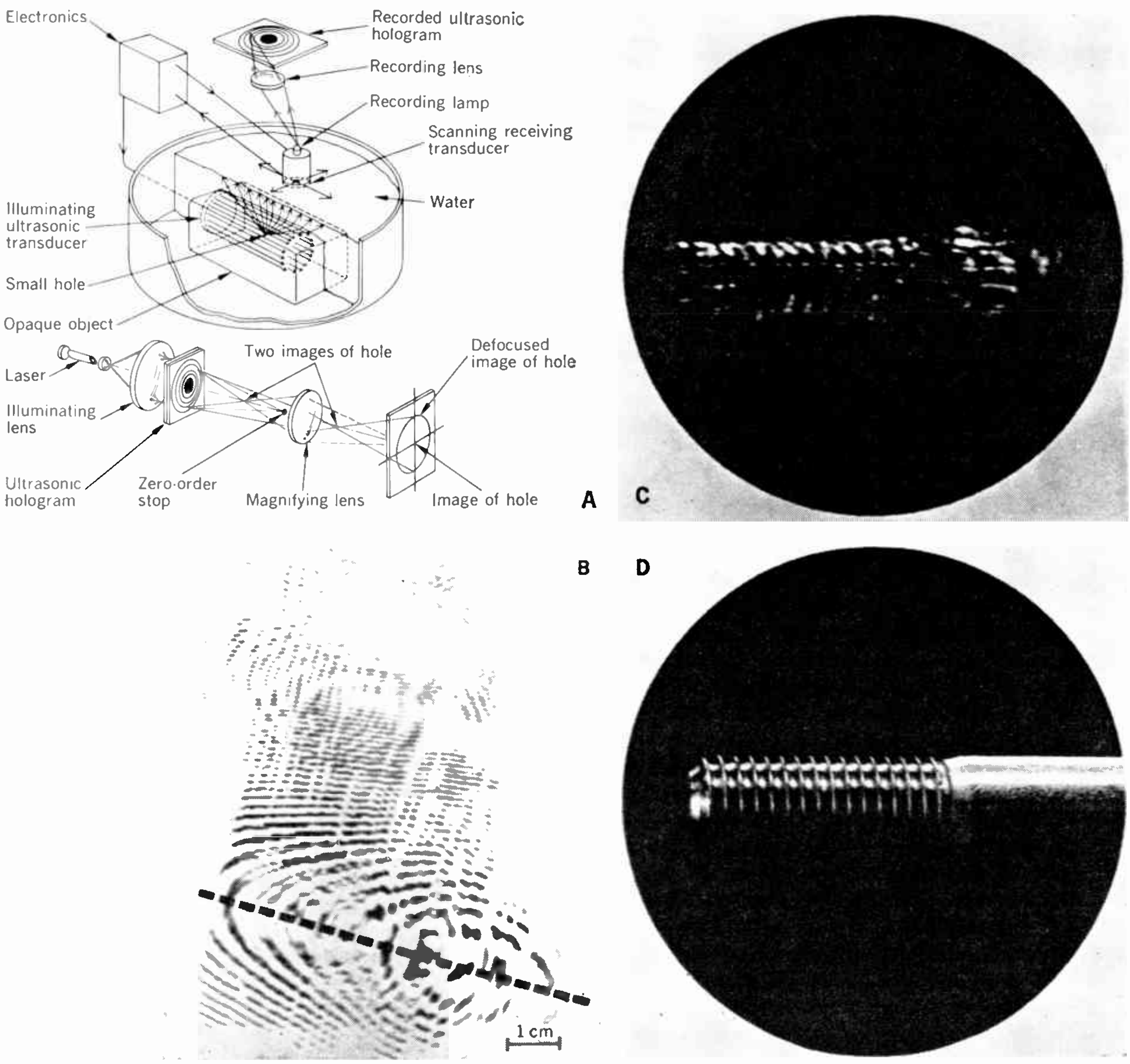


FIGURE 3. Sokolov-tube visualization technique (after Smyth).

FIGURE 4. A—Ultrasonic hologram scanning and reconstruction technique. B—Recorded hologram. C—Reconstruction. D—Comparison with optical photograph. (After Preston and Kreuzer)



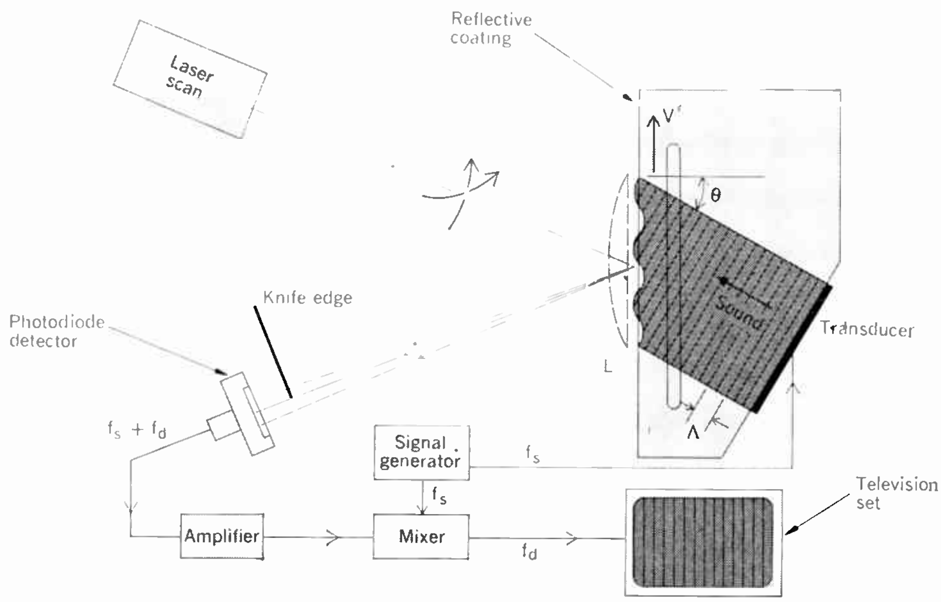


FIGURE 5. Rapid sampling of acoustic hologram by scanning laser beam (after Korpel and Desmares).

shown in Fig. 2(A). Note that the two acoustic lenses in this illustration are condenser lenses; no acoustic imaging lens is used in this experiment. Figure 2(B) shows the ripple pattern on the water surface (i.e., the hologram); and Fig. 2(C), the reconstruction of the letter "E" used as an object. No spatial filtering was used to remove the twin image, and it is visible (although out of focus) in the photograph. The frequency of the sound used in this experiment was 7 MHz.

In the Sokolov-tube method, one uses what is basically an image orthicon with a piezoelectric (quartz) rather than a photoemissive surface (Fig. 3). The sound field incident on the tube induces charges on the piezoelectric surface, which are read off by a scanning electron beam. The resulting electric signal is fed to a monitor, which then displays a picture of the incident sound field. In comparison with the surface-relief method, the Sokolov technique is more sensitive (about 10^{-7} W/cm^2 as against 10^{-3} W/cm^2) and less critical with respect to turbulence of the liquid. By leaving out the acoustic lens in Fig. 3 and adding a sound reference beam, a hologram may be recorded. Such an experiment (at 7 MHz) was performed by Marom *et al.*,[18] who obtained images of a quality comparable to those shown in Fig. 2.

Let us now consider some techniques that have no direct analogy in optical holography—the so-called sampling methods. They depend on the fact that, in acoustics, it is actually possible to measure amplitude and phase directly, in contrast to optics, where one is limited to power measurements. In a typical sampling experiment, the acoustic field to be recorded is scanned by a small microphone or transducer. The phase of the electric output signal is determined by comparison with an electronic reference signal.[19,20] Usually, both signals are added and their sum applied to a square-law detector, the dc output of which determines the brightness of a light bulb moved in synchronism with the scanner. An integrated exposure of the pattern traced out by the bulb results in a hologram quite similar to that obtained by using an acoustic reference beam.

Figure 4(A) shows such a system used by Kreuzer and Preston[21] at a sound frequency of 5 MHz. The top of the illustration depicts the sampling system, with the recording lamp mounted on the top of the scanning transducer. In contrast to the example of Fig. 2, where a recording was made of the sound field transmitted through the object, Fig. 4(A) gives an example of how the sound field scattered off an object (a small hole in a block of material) is recorded. The bottom of Fig. 4(A) shows a typical method of hologram reconstruction. The illuminating lens focuses the incident laser beam through the hologram to focus on a small disk (zero-order stop). The function of this disk is to block all light not diffracted by the ripple pattern on the hologram. The diffracted light forms two conjugate images located at points away from the zero stop. One of these images is magnified and projected on the screen. The other one appears out of focus because of its different location on the axis. As indicated before, it is completely possible to eliminate the other image by more complicated filtering techniques.

Figure 4(B) shows the hologram of a ½-inch-diameter bolt that was taken with this technique. The bolt's axis is parallel to the dashed lines. The reconstructed image is shown in Fig. 4(C) and, for comparison, a conventional photograph (illuminated in a similar manner, but with light rather than with sound) is shown in Fig. 4(D). Notice the similarity in highlights and the much poorer resolution of the acoustic image. The latter shows a resolution of about 0.5 mm, which is very close to what would be expected with a sound wavelength of 0.3 mm and an effective numerical aperture of 0.4 for the scanning transducer. Similar mechanical sampling experiments have been performed with lower freqencies (20 kHz) in air using ordinary microphones as receiving transducers.[22,23] Of necessity, such mechanical scanning systems are relatively slow and of limited use in the many important applications that involve moving objects, especially since movements of the order of only a wavelength during exposure already start to cause blurring of the interference fringes in the hologram.

Fast sampling is possible by the use of either scanning electron beams or laser beams to read out the acoustic

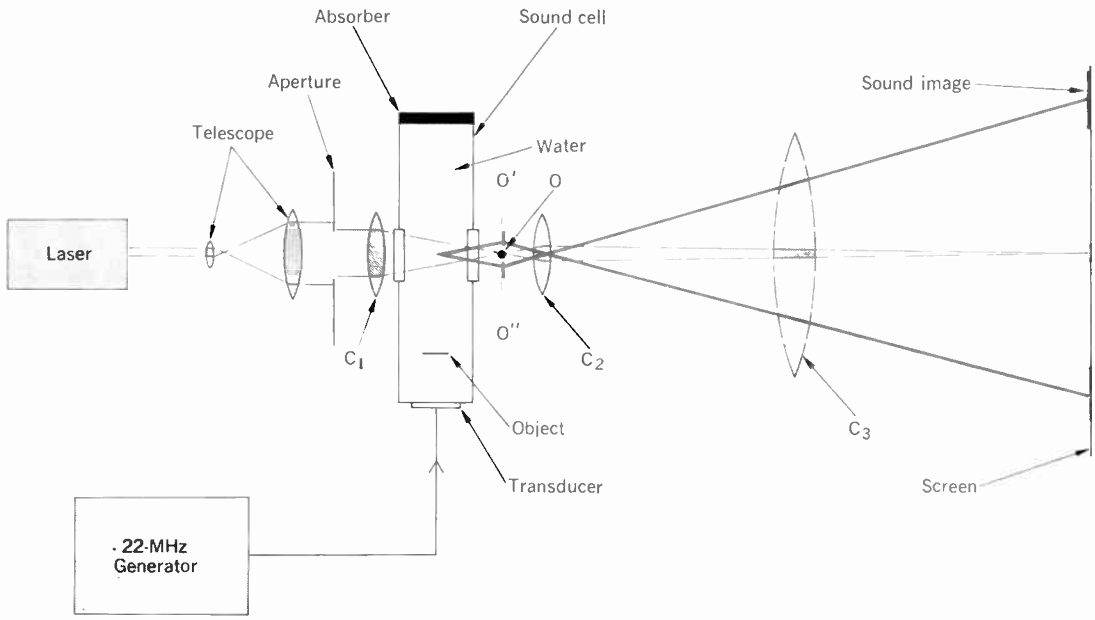


FIGURE 6. Experimental setup for Bragg diffraction imaging (after Korpel).

field. The Sokolov tube mentioned before is an example of the former technique. In our own laboratory, we have pursued the latter technique, using a scanning laser system developed earlier for television display[21] and later adapted to acoustic surface-wave visualization.[25] The principle of our method[26] is indicated in Fig. 5. As can be seen, a beam of sound is launched into a block of plastic and strikes the front surface of the block at an angle, thereby causing a surface ripple to move upward. This, in turn, causes the scanning laser beam that is reflected from the surface to be periodically deflected by a small amount. The deflection is detected by means of a knife edge, which intercepts half of the beam before it strikes a photodiode. The lens L serves to image the exit pupil of the scanning system onto the knife edge.

The output signal from the diode is basically at the sound frequency. Its phase is determined by mixing with an electronic reference source (derived from the same signal generator that produced the sound beam), and the mixer output is fed into a television monitor, which displays the final result—a hologram of the sound field incident upon the surface of the block. In this particular example, the sound field consists of a plane wave and the resulting hologram is composed of parallel fringes. In actual practice, a slot is milled into the block (solid lines) and filled with water, into which the objects to be visualized are placed. The incident plane wave of sound is then scattered by these objects, and the resulting hologram represents this scattered sound field. A photograph can be taken of the pattern on the screen and the negative then used to reconstruct the sound field in much the same way as demonstrated in Fig. 4(A).

The method just described seems a rather roundabout way to achieve final reconstruction, considering the fact that an actual hologram already exists in the pattern of sound vibrations on the surface of the block itself. It is, in principle, possible to reconstruct an image of the sound field directly from such a (moving) perturbation pattern by illuminating it with a stroboscopic readout beam. This would accomplish real-time holography. Unfortunately, the perturbations are so weak that so far only electronic methods (which make it easy to filter out unwanted signals and noise and amplify the desired signal) have enabled us to display them. Nevertheless, real-time holography is an intriguing prospect and one that may be realized in various ways. It has, for instance, been proposed[27] to display a sampled hologram on a so-called "Titus" tube.[28] This is basically a television picture tube in which the phosphor screen is replaced by a crystal plate made of an electrooptic material such as KDP. In such a tube, the hologram would be written as a pattern of induced birefringence on the transparent crystal plate. By illuminating the plate with a laser beam and using polarizing optics, instantaneous reconstruction would be achieved.

Another method for rendering acoustic images instantaneously visible was developed in our own laboratory.[29] This makes use of Bragg diffraction[30] of light by sound waves. Strictly speaking, perhaps this technique should not be called holographic, since it uses neither an acoustic nor an electronic phase reference. It does, however, provide instantaneous phase and amplitude imaging under certain restrictive conditions. Figure 6 shows an experimental arrangement. The sound field travels in a transparent medium (in this case, water), and a portion of this medium is illuminated by a laser beam focused by cylinder lens C_1 into a line O. Some of the light is scattered by the sound field. It can be shown[31] that this scattered light forms two identical spatial images of the sound field that are situated at both sides of the line focus (indicated schematically by O' and O''). In the plane of the illustration, these images are demagnified by a ratio λ/Λ. By adjusting the position on the axis of cylinder lens C_2, any desired cross section of the image field (hence, of the sound field) can be remagnified and projected on the screen. In the direction perpendicular to the plane of the illustration, there is no demagnification and another cylinder lens C_3 is inserted

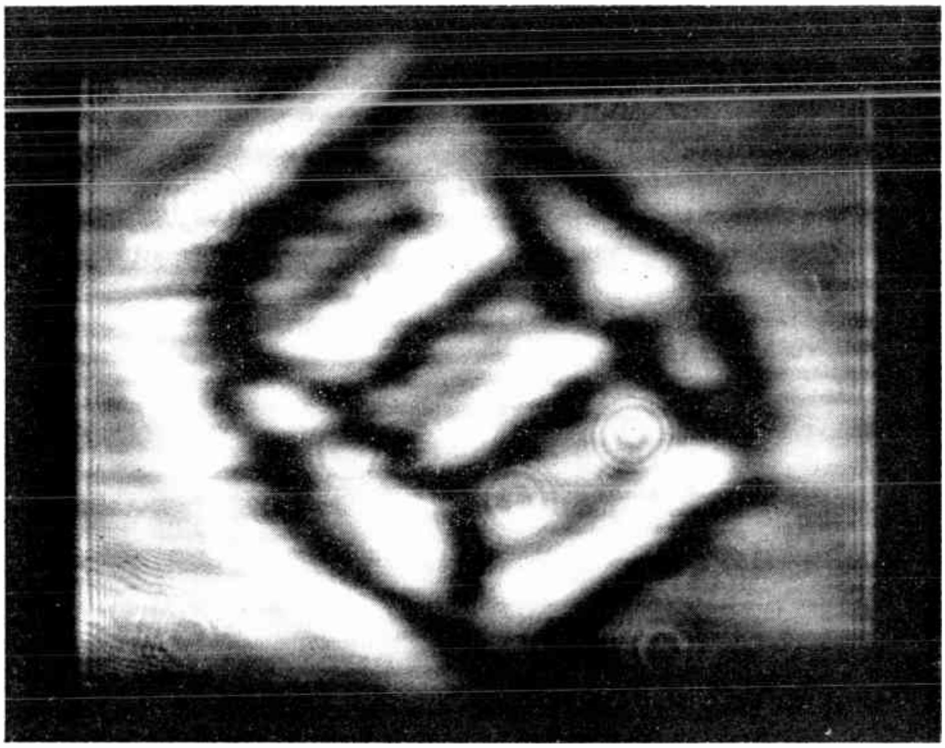


FIGURE 7. Bragg diffraction sound image of a small cube (5 × 5 × 5 mm) made of wire and illuminated by a 22-MHz sound field (after Korpel).

to provide imaging in this direction. The process may be thought of as one in which the sound field serves as its own (thick) hologram. More specifically, the individual plane waves into which the sound field may be resolved serve the same function as the individual grating patterns in a normal hologram. Thus, each one scatters the light into a particular direction—the superposition of all the scattered light waves resulting in the images O' and O''.

Figure 7 shows an example of acoustic imaging achieved by this method. The object is a 5 × 5 × 5 mm cube made of thin wire and illuminated by a rectangular beam of sound at 22 MHz. Note the highlights where the sound is scattered off the wires, and the shadows where it is obstructed.

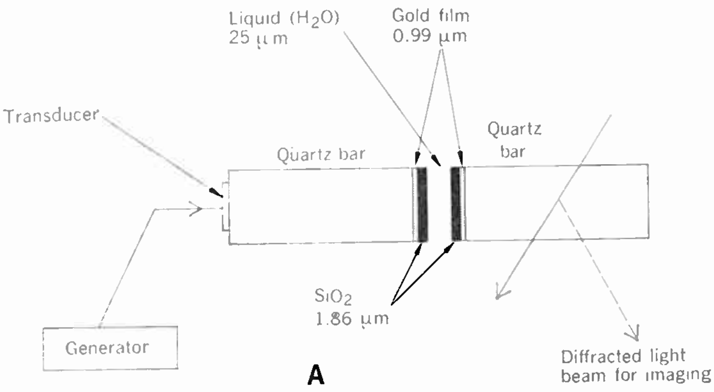
It may be worth emphasizing that, in spite of the wide variety of available acoustic holographic techniques, this field has not yet become established as a practical tool. The reasons for this perhaps lie mainly in the peculiar appearance of the world as seen by sound, rather than in the failings of any one technique. For instance, most objects are very smooth in the range of a wavelength of sound at, say, 10 MHz. Also, most experimental sound sources produce highly collimated beams of sound. The combined effect of these two factors perhaps is best compared with the case of illuminating shiny objects with a collimated laser beam. Shadows and highlights will show, but in general it will be very difficult to recognize anything clearly. Even if surfaces are rough enough to prevent specular reflection, there will still be the problem of "speckle" that is inherent in coherent illumination.[32] To solve these problems, many more experiments are needed, perhaps using diffused sound illumination with multiple frequency techniques.

In some potentially very useful applications, such as those in medicine, there is the added difficulty that the objects to be analyzed (parts of the human body) are extremely inhomogeneous to sound waves, and even a perfect hologram will not enable us to look "inside." In this case, it may be necessary to record several holograms under various conditions of illumination in order to synthetize a particular internal cross section of the object. This is reminiscent of the reconstruction techniques used in X-ray diffraction studies of organic molecules,[33] with the important difference that in acoustic diffraction we can measure phase, which could simplify matters a great deal.

If, then, acoustic holography is seen in this more general context, i.e., as a technique for recording the information necessary to reconstruct the object that caused the scattering, it is clear that computer evaluation of holograms may become very important. Some preliminary experiments on hologram reconstruction by fast Fourier transform methods have already been performed with encouraging results.[34,35]

A particular application that could become very interesting involves sound microscopy. This field, which as yet is in its very infancy, may be described as acoustic visualization at very high sound frequencies, where the sound wavelength approaches that of light. Of the various techniques that have been reviewed, it appears that the electron- or laser-beam scanning method, as well as the Bragg diffraction technique, would in principle lend itself for use in this area. In fact, the Bragg technique has been used by Quate and others[36] to visualize objects at 800 MHz, the highest frequency used for this purpose so far.

The experimental technique is similar to that described in Fig. 6. However, because of the extremely large attenuation of high-frequency sound in liquids (≈ 130 dB/mm at 800 MHz for water), it no longer becomes possible to use a liquid-filled tank to immerse the object to be visualized. As shown in Fig. 8(A), this tank is now replaced by a thin (25-μm) layer of water between



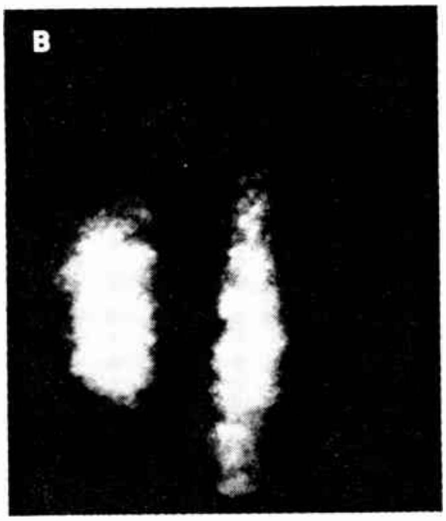


FIGURE 8. A—Object holder for Bragg diffraction imaging at 800 MHz (after Quate). B—Image of a 0.75-mm-wide gold strip at 800 MHz (after Quate).

two quartz bars, the ends coated with impedance-matching layers of gold and SiO_2. The actual diffraction of light by the sound waves takes place in the quartz. Figure 8(B) shows the sound shadow thrown by a thin (0.75-mm-wide) gold strip. The irregular bright background is the sound field emanating from the transducer.

It is often said that the present state of acoustic holography resembles the situation in which optical holography was several years ago. To a certain extent this is true, especially with respect to techniques. Nevertheless, acoustic holography seems more relevant and necessary, because, in the final analysis, it may not be just another useful method, but the *only* method that will provide a solution to certain well-defined problems. Above all, it is a new and exciting way of looking at things, both in the literal and the metaphorical sense.

REFERENCES

1. Oliver, J., "Long earthquake waves," *Sci. Am.*, vol. 18, pp. 131–138, Mar. 1959.

2. Klein, M., and Edgerton, H., "Sonar—a modern technique for ocean exploitation," *IEEE Spectrum*, vol. 5, pp. 40–46, June 1968.

3. *Diagnostic Ultrasound*, C. C. Grossmann, J. H. Holmes, C. Joyner, and E. W. Purnell, eds. New York: Plenum Press, 1966.

4. *Non-destructive Testing Handbook*, vol. 2, R. C. McMaster, ed. New York: Ronald, 1963.

5. Leith, E. N., and Upatnieks, J., "Photography by laser," *Sci. Am.*, vol. 24, p. 212, 1965.

6. Stroke, G. W., *An Introduction to Coherent Optics and Holography*. New York: Academic Press, 1966.

7. Collier, R. J., "Some current views on holography," *IEEE Spectrum*, vol. 3, pp. 67–74, July 1966.

8. Thurstone, F. L., "On holographic imaging with long wavelength fields," *Proc. IEEE*, vol. 56, pp. 768–769, Apr. 1968.

9. Sette, D., "Ultrasonic lenses of plastic materials," *J. Acoust. Soc. Am.*, vol. 21, pp. 375–381, 1949.

10. Leith, E. N., and Upatnieks, J., "Reconstructed wavefronts and communication theory," *J. Opt. Soc. Am.*, vol. 52, pp. 1123–1130, Oct. 1962.

11. Thurstone, F. L., "Ultrasound holography and visual reconstruction," *Proc. Symp. on Biomedical Eng.*, vol. 1, pp. 12–15, 1966.

12. Metherell, A. F., El-Sum, H. M. A., Dreher, J. J., and Larmore, L., "Introduction to acoustical holography," *J. Acoust. Soc. Am.*, vol. 42, no. 4, pp. 733–742, 1967.

13. Bergman, L., *Der Ultraschall*. Stuttgart: S. Hirzel Verlag, 1954, pp. 248–335.

14. Rozenberg, L. D., "Survey of methods used for the visualization of ultrasonic fields," *Soviet Phys. Acoust.*, vols. 1 and 2, 1955–1956.

15. Schuster, K., "Ultra-schall Optische Abbildung nach dem Reliefbild Verfahren," *Jenaer Jahrb.*, pp. 217–228, 1951.

16. Smyth, C. N., "The ultrasound camera—recent considerations," *Ultrasonics*, pp. 15–20, Jan. 1966.

17. Mueller, R. K., and Sheridan, N. K., "Sound holograms and optical reconstruction," *Appl. Phys. Letters*, vol. 9, Nov. 1, 1966.

18. Marom, E., Fritzler, D., and Mueller, R. K., "Ultrasonic holography by electronic scanning of a piezoelectric crystal," *Appl. Phys. Letters*, vol. 12, pp. 26–28, Jan. 15, 1968.

19. Deschamps, G. A., "Some remarks on radio-frequency holography," *Proc. IEEE*, vol. 55, pp. 570–571, Apr. 1967.

20. Metherell, A. F., and El-Sum, H. M. A., "Simulated reference in a coarsely sampled acoustical hologram," *Appl. Phys. Letters*, vol. 11, pp. 20–22, July 1, 1967.

21. Preston, H., and Kreuzer, J. L., "Ultrasonic imaging using a synthetic holographic technique," *Appl. Phys. Letters*, vol. 10, pp. 150–152, Mar. 1, 1967.

22. Metherell, A. F., El-Sum, H. M. A., Dreher, J. J., and Larmore, L., "Image reconstruction from sampled acoustical holograms," *Appl. Phys. Letters*, vol. 10, pp. 277–279, May 15, 1967.

23. Massey, G. A., "Acoustic holography in air with an electronic reference," *Proc. IEEE*, vol. 55, pp. 115–117, June 1967.

24. Korpel, A., Adler, R., Desmares, P., and Watson, W., "A television display using acoustic deflection and modulation of coherent light," *Proc. IEEE*, vol. 54, pp. 1429–1437, Oct. 1966; and *Appl. Opt.*, vol. 5, pp. 1667–1675, Oct. 1966 (joint issue).

25. Adler, R., Korpel, A., and Desmares, P., "An instrument for making surface waves visible," *IEEE Trans. Sonics and Ultrasonics*, vol. SU-15, pp. 157–161, July 1968.

26. Korpel, A., and Desmares, P., "Rapid sampling of acoustic holograms by laser scanning techniques," to be published.

27. Marom, E., Boutin, H., and Mueller, R. K., "Real-time display of sound holograms by KD*P modulation of a coherent light source," *J. Acoust. Soc. Am.*, vol. 42, p. 1169A, 1967.

28. Marie, G., "Un nouveau dispositif de restitution d'images utilisant un effet electro-optique: le tube Titus," *Philips Res. Rept.*, vol. 22, pp. 110–132, 1967.

29. Korpel, A., "Visualization of the cross section of a sound beam by Bragg diffraction," *Appl. Phys. Letters*, vol. 9, pp. 425–427, Dec. 15, 1966.

30. Adler, R., "Interaction between light and sound," *IEEE Spectrum*, vol. 4, pp. 42–54, May 1967.

31. Korpel, A., "Acoustic imaging by diffracted light. I—Two-dimensional interaction," *IEEE Trans. Sonics and Ultrasonics*, vol. SU-15, pp. 153–157, July 1968.

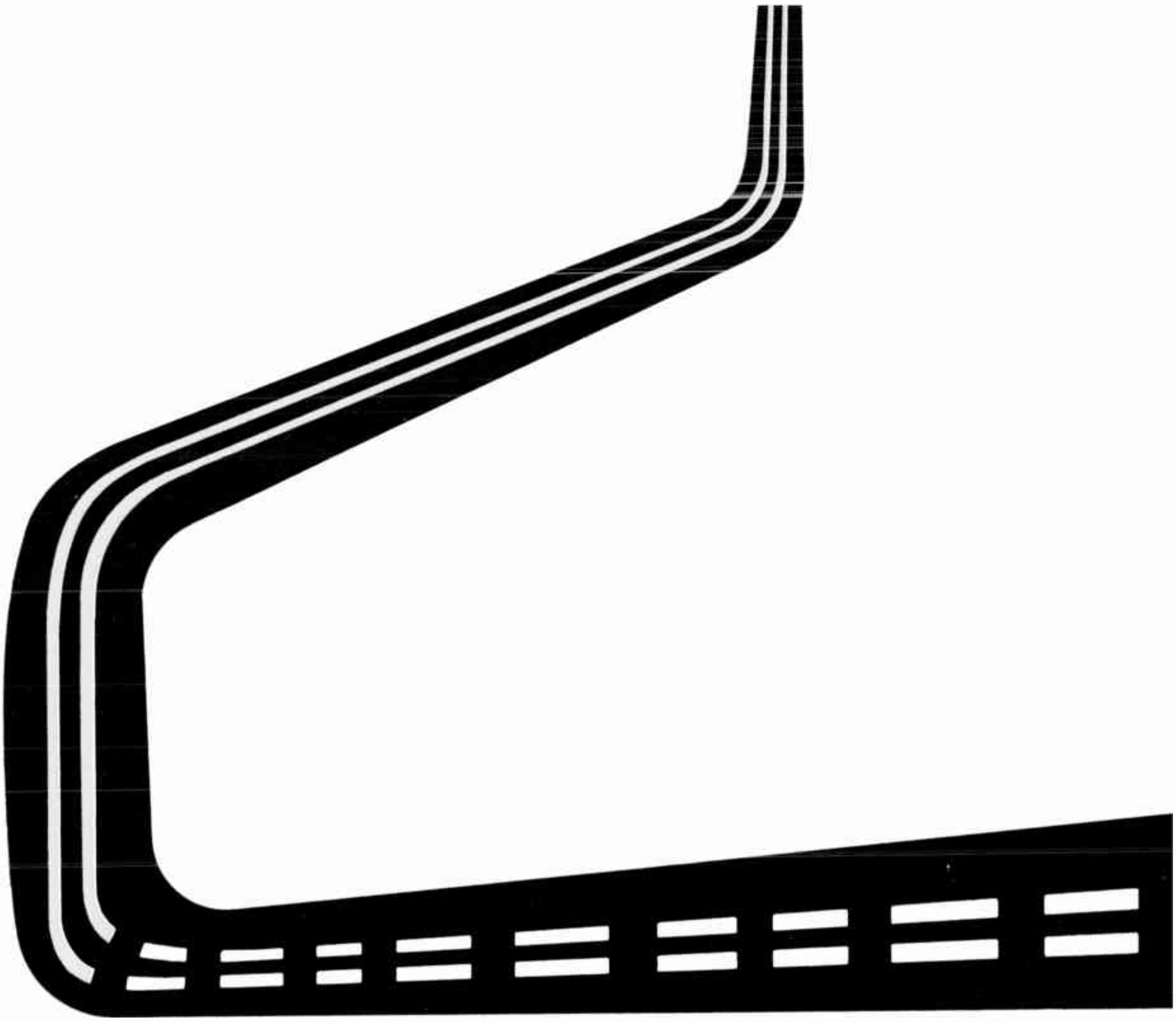
32. Rigden, J. D., and Gordon, E. I., "The granularity of scattered optical maser light," *Proc. IEEE*, vol. 50, pp. 2367–2368, Nov. 1962.

33. Kendrew, J. C., "The three-dimensional structure of a protein molecule," *Sci. Am.*, vol. 20, pp. 96–110, Dec. 1961.

34. Goodman, J. W., and Lawrence, R. W., "Digital image formation from electronically detected holograms," *Appl. Phys. Letters*, vol. 11, pp. 77–79, Aug. 1, 1967.

35. Goodman, J. W., "Digital image formation from detected holographic data," *Proc. First Internat'l Symp. on Acoustical Holography*, to be published by Plenum Press.

36. Havlice, J., Quate, C. F., and Richardson, B., "Visualization of sound beams in quartz and sapphire near 1 GHz," presented at the IEEE Symp. on Sonics and Ultrasonics, Vancouver, B.C., Canada, Oct. 1967.



Transportation—1993

As we look at the transportation scene 25 years hence we still see the private automobile dominating the field, although dramatic changes have occurred in both the vehicles and the manner in which they operate

William W. Seifert *Massachusetts Institute of Technology*

The Director of M.I.T.'s Project Transport takes us to the year 1993 to see what is expected to happen in the field of transportation in the next 25 years. From our vantage point in the future, we observe that the individual vehicle is still the most important mode of transport, although it is quite different from the automobile of 1968. For mass transportation there has been an attempt to upgrade conventional rail service but it has met with only moderate success, whereas the growth of the airlines has been phenomenal, as have been their problems.

To put ourselves in the proper frame of mind, let us project ourselves to 1993 and imagine that the IEEE is celebrating the tenth anniversary of its name change, from the Institute of Electrical and Electronics Engineers to the Institute of Environmental and Ekistical* Engineers. As part of that celebration, the Institute is holding a series of seminars to review and evaluate some of the events the stand out as specific landmarks leading to the change in name. The program of this seminar series includes discussions of the vastly expanded role of computers in our society, of strides in the biomedical area, of the greatly increased use of our ocean resources, and of the enormous increase in the use and uses of electric power. However, more important than the developments in such specific areas is the degree to which, in the past few decades, engineering and engineers have become involved with the broad interrelationships of their efforts with those of their colleagues in the social sciences, and with the implications of their work for the development of our society—in short, with their total environment. One seminar deals with transportation, and one of the specific landmarks to be discussed is the 1968 publication of a special issue of the PROCEEDINGS OF THE IEEE devoted to transportation.

* Dealing with the development of human settlements.

Imagine that you are participating in this seminar and that one of the speakers is summarizing for you the status of transportation in the United States as of 1993. In his review, he is recalling developments that brought the country from the chaotic state prevailing in transportation in the late 1960s to the more advanced, but still turbulent, state of 1993.

He points out that the U.S. gross national product is now 2.3 trillion dollars compared with a mere 750 billion dollars in 1968; the population of the United States has reached approximately 300 million people, with 210 million or 70 percent of the populace living in metropolitan areas (these are urban concentrations with central cities of 50 000 population and over). A mere 25 years ago, in 1968, the population of the U.S. was only 200 million, with approximately 65 percent living in metropolitan areas. (Others have quoted slightly higher percentages based on a less strict definition of metropolitan areas.) The important point is that there has been a continued growth of urban areas relative to the rest of the U.S. It was not until about 1960 that the region extending from Boston to Washington was identified as the country's first megalopolis.[1] Today, four of these huge megalopolitan areas contain nearly 60 percent of the people on less than 8 percent of the land area. The Northeast Corridor, or Bos-Wash, continues to be the most highly developed of these areas; but San-San in California, Mi-Jack in Florida, and New-Mil, extending from New York City westward along the lower Great Lakes to Milwaukee, have each developed rapidly.

The increased population and gross national product have resulted in the country's making vast expenditures for transportation over the past 25 years in order to meet the demands created by a greatly enlarged population enjoying a considerable increase in personal disposable income and in leisure time. As we near the end of the 20th century, the individual vehicle still dominates the passenger transportation field as it did in 1968, but during these past 25 years, dramatic changes have occurred in both the vehicles and the manner in which they are operated. In purely statistical terms, there are now some 200 million motor vehicles registered in the United States as compared with a mere 100 million in 1968. Predictions based on data available 25 years ago indicated that motor vehicle registrations might be as high as 300 million by today. However, several factors caused the actual growth rate to be somewhat below that then anticipated. The changes that have taken place have their roots in the increasing traffic congestion in our cities and in the rising number of deaths due to automobile accidents.

The role of the automobile

During the late 1960s, the Federal government initiated an extensive motor vehicle safety program, but it is sadly recounted that, in spite of a modest absolute decrease in motor vehicle fatalities during the years immediately following this effort and a significant longer-term reduction in the number of fatalities per 100 million vehicle miles traveled, the growth in traffic was so large that, by 1980, the number of deaths due to motor vehicles had risen from the 50 000 level of the late 1960s to nearly 70 000 per year.

During the late 1960s and early 1970s, several U.S. major cities, with support from the Federal government, invested amounts ranging from a few tens of millions of dollars to several billion dollars to extend existing rail rapid transit systems or to build new systems. However, it became clear as early as the mid-1970s that the U.S. public valued individualized ground transportation so highly that it could not be lured out of its automobiles unless an alternative was provided that clearly offered the door-to-door convenience and privacy of that mode. People clung to their cars even though, by 1975, the major cities had become so choked with traffic that under normal conditions average vehicle speed had dropped to 8–10 mi/h (12–16 km/h) and not infrequently, entire cities came to a complete standstill for several hours as poor weather or the simultaneous occurrence of several accidents created serious bottlenecks.

Meanwhile, almost unnoticed, help was materializing from a different quarter. During the late 1950s, a major automobile manufacturer had teamed up with one of the major electronics firms and had developed and demonstrated a scheme for automatically guiding an automobile along a highway in which a guidance cable was embedded. At that time, the system attracted little, if any, serious attention as a means for improving highway transportation. Certainly, by today's standards, it was not a technically advanced system, but, more important, the public had not yet come to realize the very high social cost it was paying for relying mainly on manually controlled vehicles and conventional highways for passenger transportation.

Fortunately, during the late 1960s, the Federal government began to take a much more active part in fostering the development of transportation. Under the Urban Mass Transportation Act of 1964, it began to support projects that would demonstrate innovations in urban mass transit. A 1966 amendment to that Act authorized money specifically for the preparation of a program to develop innovative forms of urban transportation. During 1966, the U.S. Congress also passed legislation establishing a Department of Transportation and, in 1968, the responsibility for urban transportation, which had formerly been vested in the Department of Housing and Urban Development, was transferred to the Department of Transportation.

For the first time, the Federal government was in a position to begin looking at the entire problem of transportation as a system and to begin moving toward the development of an interrelated group of transport systems, each handling that part of the total demand for which it was best suited and each providing convenient interfacing with other modes.

The stage was then set for a period of rapid development. By 1968, a number of groups working in the fields of highway transportation and urban development had concluded that the development of a system that would permit individuals to drive manually on conventional roads and streets but to operate under complete automatic control over major links of the system offered the only really effective, and socially acceptable, means for reducing automobile accidents.

Although a substantial research and development effort was required before the technical community was willing to agree to put all its efforts behind a single line of development, the technical problems involved were not nearly as complex as the problem of mobilizing the institutional and financial support necessary to permit the building of the first few demonstration installations. Although we accept automated highways as commonplace today and now build almost no major new highways that are not automated, it must be remembered that, as recently as 1968, the first automated highway had yet to be built.

However, there were a number of individuals in government, in industry, and in the academic community who became deeply concerned with the rapidity with which urban congestion was strangling the cities and restricting even long-distance travel. They realized the extent of the problem and, more important, they were convinced that something could be done about it and were willing to work to make improvements. As more and more people began to share their feelings, significant innovations began to appear in the business and organizational aspects of transportation as well as in the technical areas.

During the early 1970s, several different public–private corporations were established as a means for assembling the 100–200 million dollars needed to permit the construction of a segment of automated highway system sufficiently large to obtain a realistic appraisal of its acceptance by the traveling public. But why, one wonders, was this development obliged to wait for the efforts of the defense industries, who were searching for diversification when peace in Vietnam threatened to reduce their market? These industries now have a good half of the vehicle and guideway market, because the then big-three automobile manufacturers did not participate aggressively enough in developing transportation improvements to serve the social needs of the U.S.

Following the drastic shakeup of a basic industry, several large industrial consortiums were established, with the blessing of the Federal government, for the purpose of undertaking the development of complete new cities or building major additions to existing cities. These consortiums were established on the premise that a period of 15–20 years would elapse between the time the initial capital was committed to such a project and the time when the investors began to realize some return. They further recognized that for a new city to provide a really efficient and desirable environment in which to live and work, it should be designed as a total system, and that it should seek, from the outset, to attract a balanced mixture of residential, commercial, industrial, and recreational activities. Likewise, the transportation systems that provide circulation within the city and between the new city and the rest of the country should be an integral part of the city's design. Taking this broad systems approach, several groups built complete new cities during the late 1970s. The opportunity to take a systems approach for the construction of an entire city, coupled with the spreading of investment risk achieved by participation of several large companies, and with federal money made available for the specific purpose of carrying out demonstrations of really innovative new schemes, permitted a much bolder approach than had ever been taken in the past. The wisdom of recognizing the long period that must elapse between the initial investment and demonstration of financial success in such a project is borne out by the fact that only during the past few years has the earliest venture of this kind begun to show real financial returns.

Romulus, the subcity constructed on two islands in Boston Harbor during the late 1970s, was the first effort carried out in this way; it is now approaching the population level of 100 000 that its designers envisioned and has proved to be an extremely successful venture.

Many have already visited this city and are familiar with its principal features; however, I think you will be interested to see how closely the actual development has followed the ideas visualized in 1967 by those who conceived it.

A transportation plane extends under the entire central portion of the city. On this level, we find an automated guideway system, a conventional street system on which cars and trucks can be driven manually, and ample space for parking. Above the transportation plane is the ground level, which, throughout this area, is freed for pedestrian malls, shops, and a specialized low-speed passenger transport system. This last system provides movement both along the ground and vertically up special outside elevator shafts on all of the major high-rise buildings in the area. Residents can thus carry bundles directly from the shops to their apartments with no need to transfer or walk long distances. Several areas in the central city are domed to permit pedestrians to roam about freely during inclement weather. During the spring and summer, portions of the roof are rolled back to take full advantage of nice days. Since this city is built on a pair of islands, extensive advantage has been taken of its water location for seaside parks and marinas.

The concepts demonstrated in Romulus were enthusiastically accepted and most of the new cities that have been developed during the past 10–15 years have utilized similar ideas insofar as their geography and other physical conditions have permitted. Eureka in northern California and Clinton in Connecticut are examples of these. Some of our older cities, which have undertaken extensive redevelopment projects, have also been able to incorporate a number of the concepts demonstrated in Romulus. Unfortunately, major portions of even our more progressive older cities still remain as congested, noisy, and dangerous as they were 25 years ago in spite of extensive expenditures for incremental improvement in land usage, transportation, and general pedestrian circulation schemes. We can only conclude that bringing about major improvement in an urban area is an enormous task and one which requires time and money. Unfortunately, in spite of the advances in planning techniques that have taken place during the past three decades, our society still persists, all too often, in patching up existing systems and living with partial solutions when it should recognize that only bold, innovative approaches carried out on a major scale offer any long-term solution.

Let us now look briefly at the progress that has occurred in interurban transportation utilizing personalized vehicles. Whereas the vehicles were manually driven automobiles exclusively in 1968, today there is a rapid development of automated highways. You will recall that the first interurban automated highway was opened between New York City and Washington just ten years ago. The links between New York and Boston and between Chicago and Detroit were completed by 1988 and, within a few months, one will be able to drive on automated highways all the way from Washington, D.C., to Milwaukee, Wis., and the length of the San-San Corridor. We can certainly look forward to the completion of a complete transcontinental automated guideway system and of systems extending from Miami to Portland, Me., and from San Diego to Seattle by the close of the century. On these automated guideways, we now cruise safely and comfortably at 135 km/h (85 mi/h) compared with 70 mi/h (112 km/h) on conventional highways.

The construction of these systems has helped a great deal not only in reducing highway accidents but also in the reducing of air pollution that was plaguing all our major cities 25 years ago. Automobile fatalities have again been reduced from the high levels of the late 1970s to the level of approximately 50 000 per year experienced during the late 1960s. Although the present level is still much higher than we would hope for, it must be remembered that this reduction has been achieved in the face of a doubling in the number of vehicles registered and a 20 percent increase in the average number of miles each vehicle is driven per year. The rate of fatalities is actually less than half that of 25 years ago.

The airlines' problem

Although the motor vehicle represents the principal mode of transport that has shaped U.S. cities for the past 75 years, this discussion would not be complete without at least a brief mention of the roles of air and rail transport. This past year, the domestic trunk airlines of the U.S. carried some 700 million revenue passengers over 300 billion miles (480×10^9 kilometers). Growth of the air line industry over the past three decades has been both phenomenal and sustained, as is clear by comparison with 1968 when the domestic carriers transported a mere 130 million passengers.

This growth, of course, has been accompanied by severe problems. It has been necessary to plan and construct new airports, to develop new aircraft, to introduce a completely new air-traffic-control system, and to develop improved means for moving passengers within airports and between airports and their ground destinations. You all recall the 15-year struggle required to rally the support of the government, the airlines, and the airport operators behind the concept of having the airlines take responsibility for providing door-to-door service for its patrons. And you remember that it took the emergence of a powerful airline-users' group to focus the needed public attention on this issue. Although the airlines were initially very cool to this development, they all now hail it as the greatest recent breakthrough in air transportation.

During the 1970s the airlines introduced airbuses capable of carrying 400 or more people. Unfortunately, it was not until 1976, when two of these huge vehicles collided in mid-air killing all 790 people aboard, that the U.S. became sufficiently incensed over the inadequacy of the air-traffic-control system as it then existed, to force an immediate broad-scale investigation of the air-traffic-control issue. As a result of this investigation, a much more advanced computer-based system was implemented and responsibility for its operation was shifted from the FAA to Air-Traffic Operations, which is a private corporation.

By the mid-1970s, the airlines began to use VTOL and STOL aircraft for trips of 50 to 300 miles (80 to 480 km). The introduction of these latter types opened new possibilities for alleviating the crisis facing every major airport in the U.S. in the early 1970s. Gradually, following the pattern agreed upon for Boston in 1970, the major cities began to establish a series of V/STOL ports dispersed strategically to serve the metropolitan area. Now all flights of less than 250 miles (400 km) from Boston are served directly from these terminals, which are interconnected by a high-speed ground transportation system. An airport capable of handling subsonic planes as well as supersonic transports is located in the outer part of Boston Harbor. This location minimizes the disturbance of aircraft taking off and landing. It handles all long-haul domestic and overseas flights and also has connections to the V/STOL ports through the high-speed ground transportation system.

The noise problem that we bemoan today was as bad 25 years ago, and before that. Even then, it was decided that nothing could be done about the noise generated by the engines or by an aircraft moving through the atmosphere at great speed. Here it took the initiative of a group outside the U.S., this time French, to show that help is possible. Even if we cannot yet provide a satisfactory scientific explanation for the noise-reducing capabilities of the interference probes pioneered during the 1980s, these offer hope for greater reduction in aircraft noise than any scheme yet developed in the U.S.

Unfortunately, even now, many major cities have not yet resolved their airport problems. With the increasing population and its growing concern for the social costs resulting from such factors as aircraft noise, it has become increasingly difficult for cities to find satisfactory locations for long-haul airports. Other cities on the coasts and along the Great Lakes have followed Boston's lead, but several inland cities, after a dozen years or more of study, have not yet found satisfactory sites for new airports. They have expanded existing airports to the maximum but still face years of delay and outlays of hundreds of millions of dollars before they have facilities to enable them to retain their positions in this air age.

Rail transport

Finally, let me touch briefly on rail transport. You will recall that during the late 1960s, as a result of the growing congestion in the Northeast Corridor and the success achieved by the Japanese with their New Tokaido high-speed rail system, the Federal government initiated a program to upgrade rail passenger service in the Corridor and to undertake research into developing a 300-mi/h (480-km/h) ground transportation system.

Upgraded conventional rail service operating at speeds up to 125 mi/h (200 km/h) was initiated between Boston and Washington in 1968 but met with only a moderate degree of success—as measured in terms of patronage. Furthermore, the research program demonstrated that several different types of systems could provide service at speeds of 300 mi/h. During the 1970s schemes using linear induction motors for propulsion and air cushions for suspension were demonstrated to be technically feasible. However, progress with V/STOL aircraft and automated highways during this period, coupled with the high cost associated with building a 300-mi/h ground transport system, led to the decision to postpone actual construction until the demand for such a system was clearly demonstrated. But, as you know, research on HSGT systems has continued throughout the entire last 25 years. Within the past few years it has become apparent that even with the construction of a number of V/STOL ports throughout the Corridor, and with an automated highway system, it will be necessary to build an HSGT system to link Boston and Washington. Just this year a new public-private corporation involving the Penn Central Transportation Company, the U.S. Steel Corporation, the North Atlantic Electric Power System, and federal and state agencies was formed to construct a line to link Boston and Washington. It is hoped that service can begin during the year 2000. This new system will utilize vehicles that hold a maximum of six persons, or a single automobile. They will be propelled by linear induction motors and supported by air cushions. It is planned that nearly one half of the total length of the system will be in tunnels. Small vehicles were chosen so that the advantages of a high maximum speed could be realized by permitting demand scheduling between any points on the system with no requirement for intermediate stops.

The last regular rail passenger service in the U.S. was discontinued three years ago and the railroads are finding that air freight has attracted most of the high-value-per-kilogram long-haul shipments and that, during the past few years, the shipment of standard containers over automated guideways is beginning to make even greater inroads into freight hauling. Furthermore, now that the production of electric power relies almost entirely on nuclear plants and on thermal mine-mouth units, the hauling of coal by rail has all but ceased.

Fortunately, the Department of Transportation and the railroads recognized that such changes would occur and were able to make major readjustments. As you know, the number of railroads in the U.S. was very markedly reduced through a series of mergers in the era extending from the 1960s through the 1980s. In addition, beginning in the 1970s the Federal government began to promote the view of transportation as a total system and to permit companies to engage in multimodal operations. We see companies whose former activities were exclusively restricted to the rail transport business now becoming deeply involved in developing, building, and operating automated guideway systems in the air freight business, and, most recently, in plans to construct the HSGT system for the Northeast Corridor.

One inescapable fact that emerges as we look back over these past 25 years is that any form of transportation that offers the lowest door-to-door travel time will always drive out lower-speed competing modes unless the economics of the higher-speed system are grossly unfavorable. As we look ahead to further expansion and improvement of our transportation system during the 21st century there is no reason to believe that this rule will change.

The second big lesson we have learned in these last 25 years is the importance of developing institutions that not only can cope with the broader problems of today, but also will foster a much more aggressive and socially responsive role for private enterprise.

Conclusion

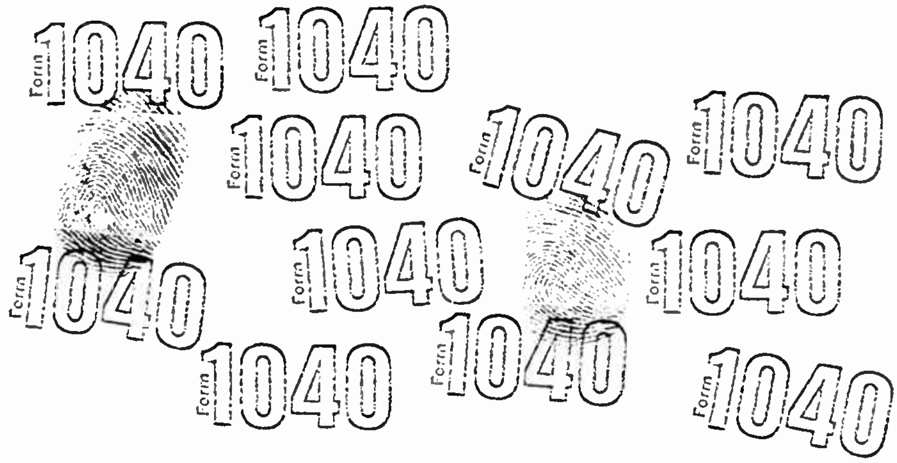
In closing, let us return to 1968 and look briefly at the lessons we can learn from these events that might have taken place. After all, 25 years is not a longer period than that involved in many long-range planning efforts, and many may already have considered the possibility of events such as I have outlined. The institutional consequences I have sketched could occur if the present U.S. government and private interests fail to face squarely the challenges now clearly before them.

You may feel that in this flight of fancy I have been unduly harsh in relating every advance in transportation to a disaster big enough to jar the conservatism of either government or private enterprise. But in pointing up shortcomings of our present transportation modes and the penalties of reacting too slowly to changing requirements, I only want to show that our institutions might be organized more effectively to create and implement innovations. They react more slowly than we might hope; but they do react, and I am convinced that the problems I have cited will be overcome without inviting the serious dislocations I have outlined. The development of means whereby private enterprise, with the blessing of government, can play a more socially responsible role while protecting its future would lead to beneficial effects such as providing complete services rather than just products. A change in the focus of business from concern with its year-to-year profits to long-term considerations of financial viability could give private enterprise the flexibility needed to enable it to assume the leadership in transportation and urban development which we so sorely need.

Essentially full text of a paper presented at the IEEE Transportation Seminar, New York, N.Y., May 28, 1968.

REFERENCE

1. Gottman, J., *Megalopolis, the Urbanized Northeastern Seaboard of the United States.* Cambridge, Mass.: M.I.T. Press, 1961.



Privacy and the computer age

Charles P. Lickson *Attorney at Law*

By the year 2000, Americans could have computers and robots in the home—and virtually no privacy. This prediction is part of the discussions of the Commission On The Year 2000, which were published in the summer issue of *Daedalus*, the Journal of the American Academy of Arts and Sciences.[1] The implications of this technological attack on privacy are reflected in the increasing interest in the protection of this vital right. The computer–electronics industry has recognized this issue, and has addressed itself to it on many occasions, including many professional gatherings.[2] Businessmen are also concerned with the inherent implications,[3] and many governmental agencies are studying the possibility of controls of some type. The Federal Communications Commission, for example, has initiated a public inquiry into the computer–communications interface.[4] One of the specific items of response to the FCC inquiry is "privacy and security."

Privacy is one of those concepts that lend themselves to a variety of definitions.

The various professions supply their own definitions; for example, psychology,[5] engineering,[6] political science,[7] and law[8,9] each defines "privacy" according to its particular frame of reference.

In this article the term "privacy" means protection of the individual (or organization) from (1) unreasonable observation by audio, visual, or psychological means; (2) unreasonable usurpation of his name or likeness by another for unauthorized use in trade or advertising; (3) unauthorized interference or interception of private communications through any means; and (4) the unauthorized access to personal or confidential information.

Professor Alan F. Westin of Columbia University, in analyzing privacy in individual life, recognizes four significant functions that privacy performs: (1) personal autonomy; (2) emotional release; (3) self-evaluation; and (4) protected communication.[10,11] Privacy, then, is a necessary part of an individual's way of life and is vital to his psychological well-being; as such, it must be considered a "basic" or "human" right. Mankind has recognized certain of these rights as being common to all individuals throughout the world. In the realm of international society, "matters have now reached the stage of development at which international guarantees of human rights must be regarded as one of the main substantive divisions of international law. The term 'basic rights' includes all aspects of civil rights, important political, economic, and social rights, and a number of collective rights, such as freedom of association, in addition to rights of a purely individual character."[12]

Unfortunately, the formal body of law is not so readily adaptable to some new legal concepts. It is a fair statement that the law has been reluctant to grant privacy the same status as other protected rights, such as freedom of the press, speech, or religion.

Privacy becomes a protected right

Court recognition of the right to privacy is a recent development. Formerly, courts looked to some other theory of the law to redress personal wrongs where there had been an invasion of privacy; e.g., the earliest cases dealt with written communications as personal property. Private property protection has been an integral part of Anglo-Saxon jurisprudence for many years. Even the early Roman law had firm rules regarding another man's property.[13] No English or U.S. court specifically ruled on an issue of privacy as a protected right of its own until after 1890.

The year 1890 has great significance for the Law of Privacy because it was in that year the Messrs. Warren and Brandeis (the latter a noted Supreme Court Justice who served from 1916 until 1939) published an article that had a profound effect upon the development of this new legal theory. The article was the first composite analysis of the cases decided up to that time. The authors called for acknowledgment of the existence of

Nationwide links to a centralized computer seem to be the inevitable method of handling cascading volumes of information accumulated about people and their personal activities. Perhaps the computer will relegate the concept of personal privacy to the annals of the past

a separate right of privacy as a reaction to lost solitude: "The intensity and complexity of life, attending upon advancing civilization, have rendered necessary some retreat from the world; and man, under the refining influence of culture, has become more sensitive to publicity, so that solitude and privacy have become essential to the individual; but modern enterprise and invention have, through invasions upon his privacy, subjected him to mental pain and distress far greater than could be inflicted by mere bodily injury."

Soon after the Warren and Brandeis article, other authors commented on the subject, advocating the existence of the right. In 1902, the New York Court of Appeals specifically rejected the views of the Warren and Brandeis article.[14] Immediately thereafter, the New York State Legislature enacted a public law recognizing the right of privacy. The statute, first of its kind, prohibited use of the name, likeness, etc., of a living person without consent for "advertising" or "purposes of trade." Although the statute was restricted in scope, it broke the ground for emergence of this right as one guaranteed by legislation at the state level.

Court decisions also reinforced the legislation or adopted the rule on their own; however, most of the cases in the early years after the Warren–Brandeis article involved the appropriation of some element of the individual's identity. This appropriation was likened to interference with a property right and was even called, at times, "the right of publicity."[15]

In 1928, the Supreme Court of the United States was confronted with its first wiretapping case (Olmstead vs. United States).[16] Faced with the necessity for adapting strict legal principles to new technological achievements of that era, the Court took the view that wiretapping did not violate the Fourth or Fifth Amendments of the Constitution. A claim had been made that the evidence seized was obtained by unlawful search and seizure. Chief Justice Taft concluded that there had been neither a search nor a seizure by the wiretap. The Court would not, at that time, enlarge the constitutional "language employed beyond the possible practical meaning of houses, persons, papers and effects, or so to apply the words search and seizure as to forbid hearing and sight." Justice Taft went on to add: "Congress may of course protect the secrecy of telephone messages . . . ," but it had not done so as of that time.

The state courts also had views on wiretapping. For example, a 1931 Kentucky decision reached a conclusion opposite to that of the U.S. Supreme Court. The issue was a civil suit for damages incurred by wiretapping. In the course of the opinion, Justice Rees clearly set forth the essence of this new civil cause of action: "Whenever a telephone line is tapped the privacy of those talking over the line is invaded and conversations, wholly proper and confidential may be overhead. Wiretapping is akin to eavesdropping, which was an indictable offense at common law, and while it has not been made a punishable offense by statute in this state, we conclude that the facts alleged in the petition in this case constitute a wrong done to appellant for which the law affords a remedy by an action for damages."[17]

One year later, in 1932, the fourth edition of Judge Cooley's landmark *Treatise on Torts* was published. His comments on the right of privacy were contained in a subtopic of the chapter "Wrongs Affecting Personal Security." Whereas he seemed to accept the right as part of the contemporary law, he added: "The right of privacy, conceding it to exist, is a purely personal one, that is, it is a right of each individual to be let alone, or not to be dragged into publicity."

Next, a series of Georgia state cases recognized the right under a variety of factual situations such as conversation overheard via a microphone, physical intrusion upon an individual's solitude, and unlawful entry into a building.[18] It should be pointed out that the Georgia cases illustrate protection by common law (court- or judge-made), whereas New York protects by statute.

Subsequently, the U.S. Supreme Court had further occasion to deal with privacy. In Goldman vs. United States,[19] involving use of a detectaphone on a wall adjoining the defendant's office, the Court refused to clarify or overrule the Olmstead case. Wiretap evidence was excluded from Federal courts, via Section 605 of the Federal Communications Act. The ban included use of knowledge gained from wiretaps, but it must be emphasized that the ruling prohibiting wiretap evidence was based upon a specific Federal statute and not upon the Court's adoption or acceptance of privacy of communications as a constitutional right. As late as 1961, the Supreme Court refused to prevent use of wiretapping evidence in state court proceedings. This prompted the dissenting opinion to call for recognition of this "Federal right."[20]

Although the right of privacy has gained recognition, it has been a struggle; and the rule has often varied according to the whim of a particular court or legislative body. A review of the current state of the law is given in the following section.

The right of privacy today (since 1966)

Contemporary history has witnessed countless intrusions into privacy. Investigators keep tabs on their victims through sophisticated technological means that often resemble Buck Rogers paraphernalia. Phones are tapped. Rooms are bugged. Employment interviews seek the most personal information. Data on all aspects of individual and corporate existence are being centralized in huge data banks. Computers are being shared. Government is increasing its size and centralization of activities.

Privacy exists today in spite of the fantastic attempts by human ingenuity and technology to eliminate it. Public awareness of this point is attested to by the recent upsurge in literary comment on the subject.[21-23] This comment served a necessary public purpose; it has made the public and its servants in the legislative assemblies, the bench, and the bar aware of the situation.

Today, the right to a civil suit for damages for violation of privacy is specifically recognized in the following states: Alabama, Alaska, Arizona, California, Connecticut, Florida, Georgia, Illinois, Indiana, Kansas, Kentucky, Louisiana, Michigan, Missouri, Montana, Nevada, New Jersey, North Carolina, Oregon, and South Carolina, and in the District of Columbia. Only a few states persist in rejecting the right. The right exists, but is limited by statute, in the states of New York, Utah, and Virginia.[24] Ironically, in England, the land from which the United States has drawn her common law, the right of privacy per se does not exist today. In order to protect privacy in terms of freedom from unwarranted publicity, English courts have had to resort to what is known as a "legal fiction" (conforming a factual situation, where there has been no ruling, into an analogous factual situation, in which there is an existing rule of law).

Dean Prosser, in *The Law of Torts*, categorizes civil U.S. privacy cases according to type. His first category is called the tort of "intrusion upon solitude." Second is the tort that Dean Prosser refers to as "disclosing private information that violates ordinary decency but which falls short of defamation." An example of this would be publishing details of a humiliating illness. A third privacy tort in this scheme is "placing an individual in a false but not defamatory position in the public eye." This differs from the second tort in that the information is true but is of a private or personal nature. The fourth tort consists of "appropriation of some element of the plaintiff's personality for a commercial use."

When Dean Prosser discusses the invasion of privacy as a tort, he is referring to the ability to recover damages for what amounts to a noncontractual legal wrong. But, as we have seen, the right of privacy encompasses more than a tort action. Perhaps it deserves constitutional protection as argued by Warren and Brandeis in 1890.

Since the Supreme Court is the final arbiter of constitutional rights, a look at several of its recent cases may clarify the present view of the law regarding things such as electronic surveillance and privacy. Silverman vs. United States[25] involved the use of a "spike mike" inserted under a baseboard of a neighboring house to pick up reflected sound. Police overheard conversations that contributed to the conviction of the defendants. The Supreme Court reversed lower-court decisions on the theory of "unauthorized physical penetration into the premises." The Supreme Court relied upon the actual intrusion into a constitutionally protected area. The Court was trapped by precedent into using the Law of Trespass as rationale, and did not consider present technological realities.

In 1962, the Court decided that a jail was not, physically, a constitutionally protected area; therefore, a conversation in prison that was overheard via electronic means could not vitiate the state conviction.[26]

There has been considerable comment about the old view that once overheard a conversation no longer deserves protection. The theory was that by permitting, even passively, someone to overhear, a person waived his right to privacy. But, in view of present-day technology, this waiver theory fails.

In a 1966 criminal case the Supreme Court commented that "our sense of fair play which dictates a fair state–individual balance requires the government to leave the individual alone until good cause is shown for disturbing him . . ."[27] This remark is considered only "dicta," as the case was not decided on the privacy issue.

In a significant case decided in June 1967, the Supreme Court finally commented directly on the Right of Privacy. The occasion was an interpretation of the New York civil rights statute protecting privacy. In this case, there was an account of the play "The Desperate Hours," which was likened by *Life* magazine to the experiences of a private family.[28] The Court virtually eliminated the existence of the right of privacy without regard to motive of the writer, at least as it pertains to a "public person." Regarding "private persons" (persons not considered subjects of news report or public interest), it is still unclear whether or not there could be recovery under New York's privacy law without the plaintiff showing some falsity or malice in the article.

In June 1967, the Supreme Court handed down a decision that involved the use of electronic eavesdropping by the planting of hidden microphones on a person's premises. The Court, basing its decision on the Fourth Amendment, threw out a New York statute that permitted "bugging" by police officers after they had obtained court approval. The Court felt that New York's statute was vague and broad. It is noteworthy that the decision addressed itself only to electronic eavesdropping and not wiretapping, although final interpretation of the ruling is not clear in view of the Federal antiwiretapping statute.[29]

It can be seen that the present status of privacy as a protected right is unclear. The Supreme Court has recognized the duty to protect against invasions into "constitutionally protected areas" and has refused to permit use of evidence seized in contravention of the Constitution, but it has clearly refused specifically to incorporate a right to privacy into the Bill of Rights as a correlative right to other freedoms, such as speech.

Several Federal statutes are of importance today in that they afford recognition of a limited right to privacy. The provisions of the Federal Communications Act, which outlaw use of information obtained by wiretapping (Section 605) or from "bugging" via certain transmission devices (Section 301), provide examples of such statutes. Also important to privacy protection is Title 18 U.S. Code, Section 1905, prohibiting Federal employees from disclosing confidential information.

Today, professional societies in the industries are formulating their own codes or standards of action, some of which would carefully protect privacy.[30] Present technology has also attempted to provide security within the system without sacrificing economic considerations of cost. Within the computer community, the ability to offer security and privacy in a shared system is a necessity for conducting business. In addition to the foregoing recognition of privacy, public hearings are being held by various Congressional and Executive bodies. The list of agencies interested in this area or specifically considering regulation includes the Federal Communication Commission, Labor Department, Senate Judiciary Committees, House Anti-Trust and Census and Statistics Subcommittee, General Services Administration, General Accounting Office, Comptroller of the Currency, National Bureau of Standards, Federal Reserve System, Internal Revenue Service, Bureau of the Budget, and Justice Department.

It is significant to note that an important barrier has been broken by a Federal agency. On June 16, 1967, the Department of Justice made an important announcement that reflected Federal Executive concern over widespread derogation of an individual's right to privacy. It was made clear that the interception by all Federal personnel of telephone, and certain nontelephone, conversations is prohibited by Presidential directive, whether or not the information was to be used within or outside the agency (a direct reference to the wording of Section 605 of the Communications Act, which is very generalized and has caused problems). Further, the Department stated that any question regarding use of a particular device as involving a prohibitive interception when a party has consented was left up to each agency. Eavesdropping by any means that trespasses into a constitutionally protected area is in violation of the Fourth Amendment.

The announcement also warned that there is growing support for the view that electronic eavesdropping on any conversation in a constitutionally protected area is a violation of Amendment Four, even if there is no trespass or entry. The Department listed homes, private offices, hotel rooms, and cars as examples.

Further, the Justice Department recognized that there is also a violation of constitutional rights when there is surreptitious electronic surveillance involving a privileged relationship such as attorney-client. As to use of mechanical or electronic devices ("bugs") by Federal employees to overhear or record nontelephone conversations, the Department prohibited these where there is a violation of the Constitution or a statute. A firm procedural guideline was established regarding any wiretapping or other overhearing of conversations. The Department added that the foregoing prohibition does not apply to installations directly related to the protection of the national security. These cases "shall continue to be taken up directly with Attorney General."[31]

The inroads into privacy mentioned in this article are traced not only to scientific progress, but also to other different contemporary pressures, such as security and crime fighting. Several commentators have pointed out that preoccupation with national security in an "age of anxiety" can only lead to destruction "in the individual (of) all sense of personal honor, decency and good faith."[32] The National Computerized Crime-Communications Network and other technological crime-fighting aids nurtured much comment. One contemporary jurist views the situation as a balancing of interests: ". . . there is an imperative need to authorize Federal and state use of electronic surveillance in the solution of private crimes; provided, of course, that this is done under suitable supervision and authorized by court order after a finding that there is a likelihood that evidence of crime will be obtained."[33] Compare Judge Lombard's views with the recent announcement by at least one state official involved in the administration of criminal justice that he would continue to use electronic eavesdropping devices in criminal investigations in spite of a recent Supreme Court decision declaring New York State's eavesdropping law unconstitutional.[34]

Other factors that are working against the concept of privacy and individualism are growth of government and governmental functions at local and national levels, necessity for high-speed sophisticated technological operations by commerce and industry, new concepts of personnel analysis, widespread use of high-speed communications for voice and data in many applications, and research at governmental and university levels aimed at further breakthroughs in high-speed, high-efficiency microminiature technology. Underlying many of the foregoing general areas is the competitive pressure of the market place. In a capitalistic system it is the market place that ultimately dictates those things which technology must come up with to please the consumer—be he private citizen, industry, university, or government.

The future of the right of privacy

We have witnessed the evolution of the right of privacy and have studied its present applications and limitations. What about the state of the law in the next 10 or 15 years?

Clearly, technology is placing a great strain on the capacity of present jurisprudence to cope with these unique problems. Some commentators have urged the necessity for establishing national policies governing the interactions of science and technology with the rest of society. Scientists and nonscientists must do this together. There is a need for comprehensive Congressional action on science and technology. Congress never has formulated a national science policy—only a collection of diverse policies on scientific and technological responsibility.[35]

New legislation is required, and its arrival is inevitable—the great number of official governmental entities interested in the subject attests to this fact. Technological advance requires creation of new economic, political, and social organizations, or readaptation of old ones to new functions. New technology, urbanization, and modern social needs have had a resounding impact on the United States' constitutional system. One concrete example is the effect of modern technology on the country's Federal system of government. The alteration in government has made the central government dominant and apportioned much activity concerning management of technology to Executive agencies (e.g., the FCC). The result, according to one view, has made Congress, and not the Supreme Court, the final arbiter in government.[36]

Professor Westin, an expert on privacy, recently foresaw three different types of approach to the issue of regulation. One view would be to oppose completely creation of pooled data centers and certain intelligence systems. In this view, the growing power of government and invasions of privacy are not worth the improved technology. Presumably this group would encourage prohibitive legislation outlawing or tightly restricting such activity. A second view would encourage the status quo in legislation on the rationale that legal safeguards are now adequate, and that the industry would use restraint or police itself. Finally, there is the moderate group, which is visualized as assuming that neither the "total ban" nor "the additional restraint" position is desirable. This group (and Professor Westin) urges "a new legal approach to the processing of personal information by authorities in a free society and a new set of legal, administrative, and systemic protections to accomplish this objective."[37]

It is feasible that in the near future regulations may be promulgated that would apply the same safeguards to data communications channels as those enjoyed now by telephone wires. This could be done through application of a similar type of regulation. Problems arise, however, in determining where the data banks end and where channels of communication begin, especially as transmission capabilities get into the wide-band range. Other problems appear with the use of extremely high data rates that are imminent, such as microwave and lasers. The current FCC inquiry is certain to recommend some type of regulation regarding data communication.

Recently, the State of Pennsylvania announced it had completed the first step in a program to establish a state computer center for collection, storage, and retrieval of vital government statistics. Systems of this type are the forerunner of state data banks. The Federal government already has much data in banks and is considering pooling them (e.g., the National Data Center). This type of application raises serious privacy problems, to which future laws are bound to respond.[38] Combining of files by businesses is also bound to get legislative reaction in the future. Professor Westin's suggested approaches are mentioned above. Other possible Federal legislation could take the form of (1) a statute authorizing the person whose name is mentioned to have a copy of the contents of his "dossier" (other than classified information) to be able to protect himself against inaccuracies; and (2) legislation to control information that may be entered in the file and to limit the persons permitted to retrieve such information.[39]

In addition to this type of Federal legislation, it is not beyond the realm of possibility to foresee the adoption of similar restrictive statutes by the various states. Also, in view of the increased technical knowledge required by regulatory agencies at the state and Federal level, it is quite possible that future boards, agencies, and hearing examiners might have scientific members or expert assistance at all levels.

We've mentioned the future legislative courses of action. Next to be considered are the Federal Court views regarding privacy. Slowly, the Supreme Court has been approaching the issue of privacy and has been attempting to curb invasion of privacy by electronic means.* Within the next few terms, this august group of jurists will undoubtedly view the right of privacy as a direct corollary to the right of free speech. This will mean the adoption of the rule protecting privacy, first for Federal courts as part of the Bill of Rights. Then, after some delay, the Supreme Court will probably require state compliance through the 14th Amendment. Other courts, state and Federal, will keep in line with the nation's highest Court. Perhaps even before the Supreme Court rules, the few remaining states that have refused to recognize the right will do so. Those states that restrict applicability of the right by statute will gradually be drawn away from the limits of statutory protection to the freedom of common law protection. One example of this can be seen in the Supreme Court's recent interpretation of New York's privacy law.[28] All court decisions will, no doubt, reflect greater willingness by courts not only to permit damages for the tort of invasions of privacy by publicity, electronic means, use of confidential data, etc., but also to grant injunctions to prevent invasions.

In addition to the growth of more comprehensive statutes and firmer court decisions, elements within the industries and governmental agencies will formulate codes of ethics and set operating standards to minimize undesirable conduct. As mentioned previously, the Justice Department has attempted to regulate certain agency activity, and has left it to the various agencies to police themselves in certain matters. Some elements of the private sector of the industry are already working on professional codes of ethics, and are even requiring certification showing familiarity with these codes as a prerequisite for official professional recognition. Such codes or private regulation should help avoid litigation and, in the event of litigation, could be important to courts in determining a standard of conduct. Eventually, industries could even have their own ethical fact-finding boards.[40] These methods are perhaps more dignified and optimistic approaches to the same issue.

Finally, in discussing the future of the right of privacy, it is necessary to discuss what technology is doing to protect it. A detailed discussion must be left to the engi-

*The Supreme Court announced Dec. 18, 1967, its decision in Catz vs. U.S. (19 L ed. 2nd 576). The particular case related to the use of an electronic device on the top of a telephone booth to overhear the conversation within the booth. The Court reversed Catz's conviction, holding that judicial authorization, which had not been given in this case, was a precondition to this type of eavesdropping.

neer. For our purposes, suffice it to say that there is a need for creation of fail-safe and secure systems that are economically feasible. There are vulnerabilities throughout most data systems that make a system accessible to accidental or intentional violations of security or privacy. "Security cannot be attained in the absolute sense. Every security system seeks to attain a probability of loss which is commensurate with the value returned by the operation being secured. For each activity which exposes private, valuable, or classified information to possible loss, it is necessary that reasonable steps be taken to reduce the probability of loss."[41] Development of low-cost technological security features could include the use of scrambling devices, new methods of transmission from one secure communications site to another, better physical site security, and randomization of codes. An example of the latter currently being demonstrated for future use by law-enforcement agencies to avoid monitoring is "programmable coding." This new technique uses an encoder decoder developed by private industry. The codes may be changed frequently and the number of possible combinations is so large that chances of breaking a code are remote. Also, the system's security is not compromised if a unit is stolen or copied.[42]

Because of the lack of clear-cut protective legislation in the area of privacy and the lack of universality of approach by the courts, much of the burden of providing security for data and assuring privacy rests with private industry. Parties have traditionally protected themselves by contract.[43] This type of protection, which is common in the market place, has afforded some degree of security in the past, and will undoubtedly continue to offer private legal protection in the future. Presently, it is through technology that private industry is able to offer reasonably secure systems. This technological development must continue to fill the serious void in privacy protection until legislation and court decisions can catch up to the state of the art.

Essentially full text of an article appearing in the January 1968 issue of *Computer Group News*, the newsletter of the IEEE Computer Group.

REFERENCES

1. *The New York Times*, p. 30, Aug. 13, 1967.

2. *AFIPS Conf. Proc.*, vol. 30, 1967.

3. Bigelow, "Legal security issues posed by computer utilities," *Harvard Business Rev.*, p. 150, Sept. Oct. 1967.

4. FCC Docket No. 16979. See also *IEEE Computer Group News*, May 1967.

5. Michael, D. M., "Speculations on the relation of the computer to individual freedom and the right to privacy," *George Washington University Law Rev.*, vol. 33, p. 270, 1964.

6. Ware, W., "Security and privacy; similarities and differences," *AFIPS Conf. Proc.*, vol. 30, p. 279, 1967.

7. Lippman, Walter, *Public Opinion*. New York: Macmillan, 1960, pp. 44–45.

8. Warren, S. D., and Brandeis, L., "The right to privacy," *HarvardL aw Rev.*, vol. 4, p. 193, 1890.

9. McCarthy, F. J., "Privacy: Burgeoning rights and remedies," *Connecticut Bar J.*, vol. 38, p. 555, 1964.

10. Westin, A. F., "Science, privacy, and freedom: issues and proposals for the 1970's," *Columbia Law Rev.*, vol. 66, p. 1003, 1966.

11. Westin, A. F., *Privacy and Freedom*. New York: Atheneum, 1967.

12. Jenks, C. W., *Common Law of Mankind*. London: Stevens & Sons, 1958, p. 45.

13. Rudin, Max, *Handbook of Roman Law*. St. Paul: West Publishing Co., 1927, p. 128.

14. Roberson vs. Rochester Folding Box Co., 171, N.Y. 538, 64 N.E., 442, 1902.

15. Prosser, W. L., *Law of Torts*, 2nd ed. St. Paul: West Publishing Co., 1955, p. 639.

16. Olmstead vs. United States, 277, U.S. 438, 1928.

17. 238 Ky. 225, 337 SW2d 46, 1931.

18. Cases are collected in Prosser, *op. cit.*, pp. 635–644.

19. 316, U.S. 129, 1942.

20. Pugach vs. Dollinger, 365 U.S. 458, 1961. (For an excellent analysis of the Supreme Court's view of this issue, see King, D. B., "Electronic surveillance and constitutional rights: some recent developments and observations," *George Washington University Law Rev.*, vol. 33, p. 240.)

21. Packard, V. O., *The Naked Society*. New York: McKay.

22. Benton, M., *The Privacy Invaders*. New York: Coward McCann, 1964.

23. Ernst, M. L., and Schwartz, A. V., *Privacy—The Right to Be Let Alone*. New York: Macmillan, 1962.

24. Prosser, *op. cit.*, pp. 635–637.

25. 365, U.S. 505, 1961.

26. Lanza vs. New York, 379, U.S. 140, p. 143.

27. Tehan vs. Shott, 382, U.S. 406.

28. Time, Inc., vs. Hill, 87, U.S. 534. 1967.

29. Berger vs. New York, 87, U.S. 1873. 1967.

30. *Sci. Am.*, Sept. 1966 (in this issue a Computer Bill of Rights is mentioned as part of a consideration of privacy).

31. U.S. Dept. of Justice Announcement. June 16, 1967, as reported in *U.S. Law Week*, vol. 36, p. 2033.

32. O'Brian, J. L., *National Security and Individual Freedom*. Cambridge, Mass.: Harvard University Press, 1955, pp. 17–18.

33. Remarks of Hon. Edward Lombard, Chief Judge of the U.S. Court of Appeals for the Second Circuit, presented at the 90th Annual American Bar Association Convention, 1967, as reported in *U.S. Law Week*, vol. 36, p. 2122.

34. Remarks of Aaron E. Koota, Brooklyn, N.Y., District Attorney, as reported in *The New York Times*, Aug. 23, 1967.

35. Beresford, S. M., "Lawyers, science and the Government," *George Washington University Law Rev.*, vol. 33, p. 181.

36. Miller, A. S., "Technology, social change, and the Constitution," *George Washington University Law Rev.*, vol. 33, p. 17.

37. Westin, A. F., "Legal safeguards to insure privacy in a computer society," presented at the AFIPS Spring Joint Computer Conference, 1967.

38. For a legal discussion of the problems and a study of the National Data Center, see "The computer and invasion of privacy," *Hearings before a Subcommittee of the Committee on Government Operations, House of Representatives, 88th Congress*. U.S. Government Printing Office, Washington, D.C., 1966.

39. Bigelow, "Legal & security issues posed by computer utilities," *Harvard Business Rev.*, p. 150, Sept.–Oct. 1967.

40. Wessel, "Legal protection of computer programs," *Harvard Business Rev.*, Mar.–Apr. 1967.

41. Peters, B., "Security considerations in a multiprogrammed computer system," presented at the AFIPS Spring Joint Computer Conference, 1967.

42. "Electronics newsletter," *Electronics*, p. 25, Sept. 4, 1967.

43 Lickson, C. P. "Protecting privacy of data communications by contract," *Business Lawyer*, vol. 23, p. 971, 1968.

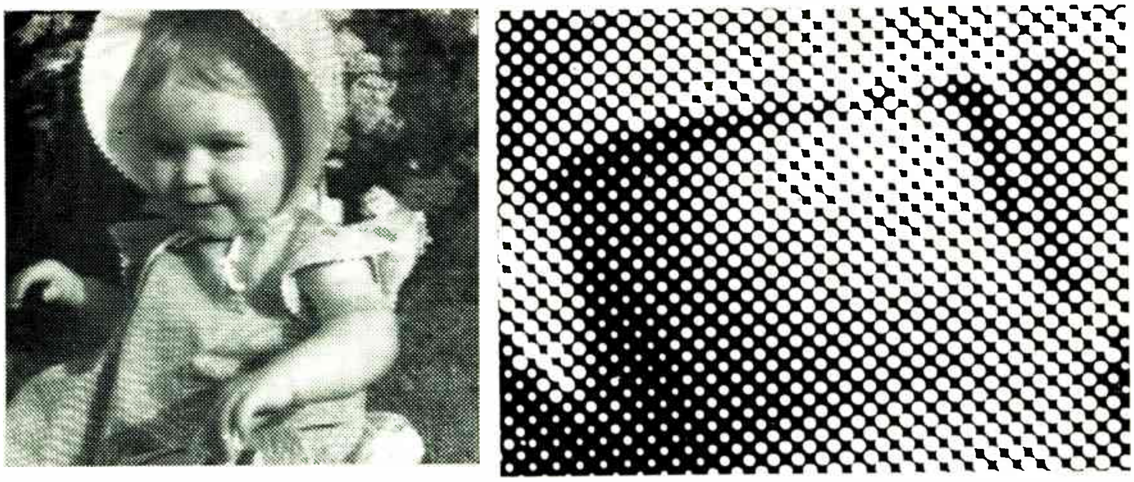


FIGURE 1. Electronically reproduced newspaper scan 65-line halftone with portion enlarged 5.5×, using white-dot halftone system.

Electronic halftones

As a further step in the automation of the publishing industry, halftone images, which are now produced by an optical method, can be created by utilizing new electronic techniques

R. L. Hallows, Jr., R. J. Klensch RCA

FIGURE 2. Block diagram of developmental halftone system for spirally scanned dots.

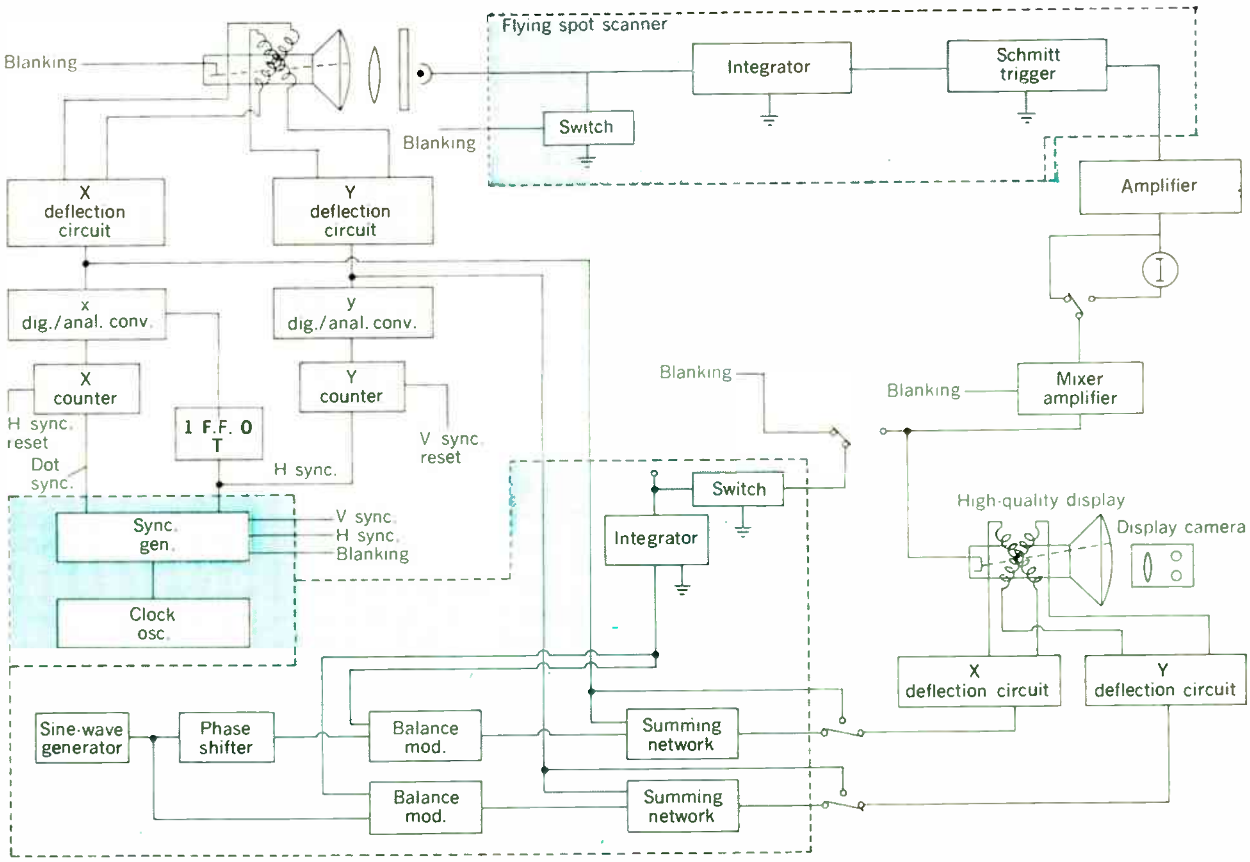


FIGURE 3. Hex-format white-dot structure—110-line screening in original; this reproduction magnified approximately 3.15× for clarity in illustrating the hex structure.

To derive the greatest advantage from automation in the production of newspapers and books the present optical method of producing halftone dot patterns must be replaced by an electronic method. This article discusses developmental electronic-halftone techniques that use a high-resolution cathode-ray tube as the display device.

To the casual reader, the printed word is the important thing; the techniques of producing it are of little interest. However, with the information explosion going on about us, it is becoming more and more difficult to get that word into print—and automation seems the only solution. And, since a picture proverbially is worth a thousand words, illustration methods are of equal concern.

The term "halftone" as used in the graphic arts industry refers to an image composed only of black and white elements, of variable fractional area (commonly small dots), that represent the local integrated reflectance of the continuous-tone illustration being produced.

The halftone shown in Fig. 1 is an electronically produced *newspaper scan* with the rows of dots running at 45 degrees to the horizontal. This type of dot orientation is achieved by displacing alternate rows of dots by a distance equal to one half the horizontal interdot spacing.

The area of the individual dots is a function of the light transmittance of the sampled input picture, and is controlled by spiral scanning of the cathode-ray tube (CRT) beam until the desired dot size is achieved, and then blanking the electron beam.

Figure 2 is a block diagram of the developmental system for producing halftones electronically. Either the 45-degree news scan, or the hex structure comprising the halftone in Fig. 3, can be obtained from the system by using different sampling-raster configurations.

In using spiral scanning to produce circular halftone dots, the scan is achieved by modulating cosine and sine waves, which are added in deflection-amplifier input-summing networks. This cosine sine relationship produces a circular shape when plotted orthogonally.

The density loss in the exposed film due to the increased scanning velocity with increased radius should be roughly compensated by writing a higher number of scanned circles per unit radius increment as the dot is scanned; that is, a constant area of the dot per unit time must be scanned, and therefore the radius must increase as the square root of time. In our evaluation the required half-power law was approximated by a simple exponential curve, as shown in Fig. 4. The exponential was made adjustable to obtain the closest match by providing adjustment of the voltage toward which the capacitor charged in a simple *RC* charging circuit, and adjustment of the series resistance of the *RC* network. The capacitor voltage can be returned to zero at the end of the dot-unblanking interval or by the variable-pulse-width signal itself.

Uniformity of dot intensity as a function of dot radius is adequate to expose the high-contrast films used for halftone recordings to a density that is high enough to prevent significant additional density in the regions of dot overlap. The half-power radial modulation requirement can thus be seen to be fairly noncritical.

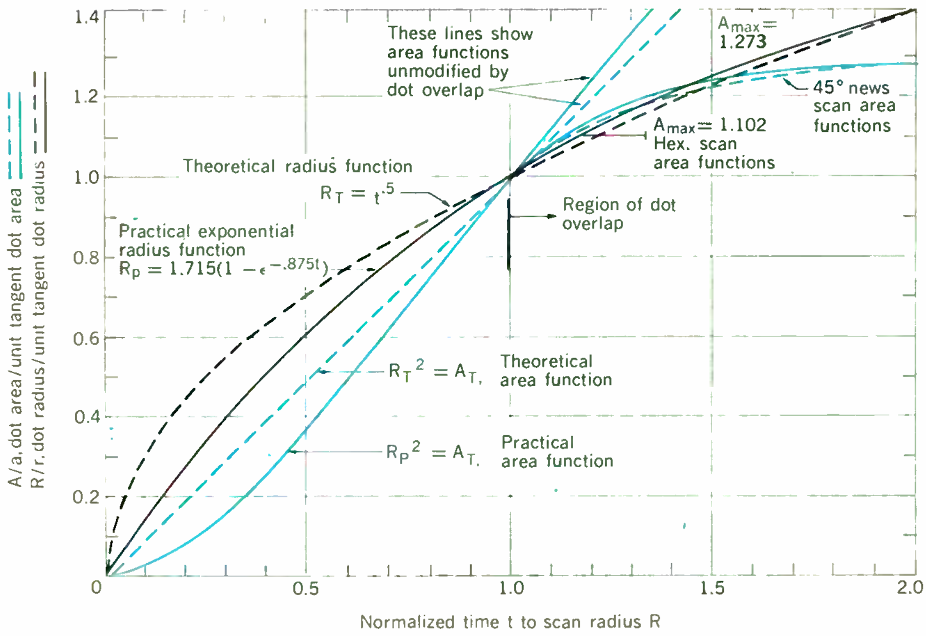


FIGURE 4. Radius and area functions for spiral-scanned halftone dots.

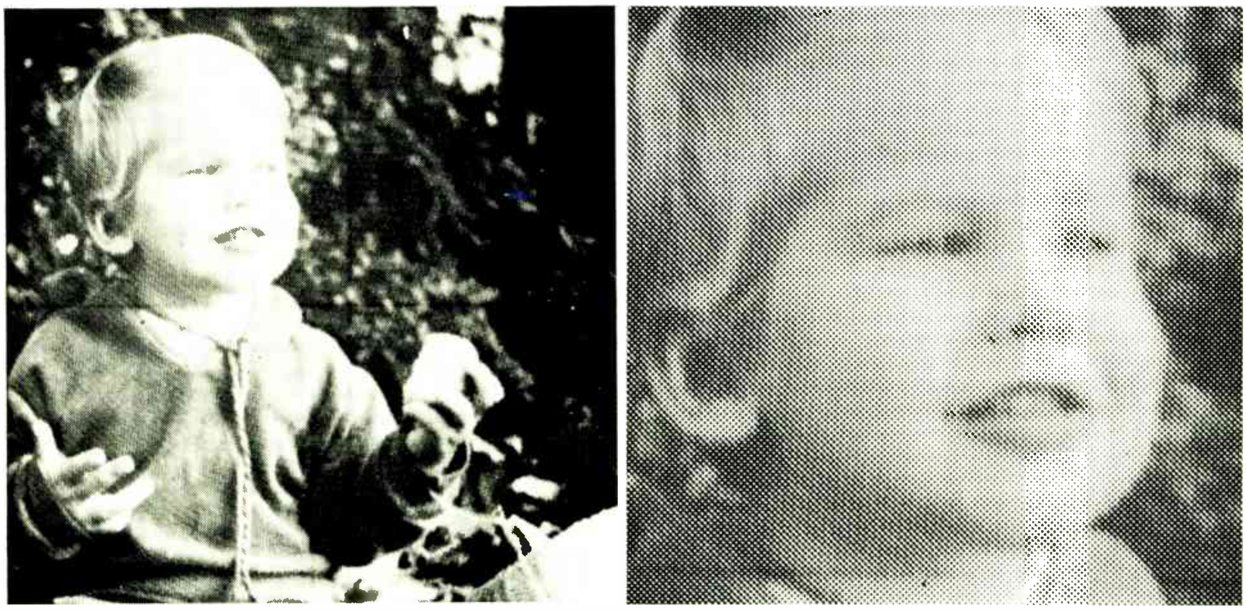


FIGURE 5. Black-dot 105-line news dot structure, actual size. Face enlargement is approximately 2.6×. Note concave-upward hum bars in both reproductions. Enlarged section is shown for clarity in illustrating the black-dot structure.

The modified rate of dot area increase resulting from dot overlap is plotted in Fig. 4 for the theoretical half-power radius function and for an assumed practical *RC* charging function. Note that the practical function requires black and white stretch of the transfer characteristic to restore a linear area–time function.

Dot polarity

A close examination of the halftones in Figs. 1 and 3 will show that the circular portions—the dots—are white, being delineated by the surrounding black printing ink. In the example shown in Fig. 5, the black areas are round dots, more nearly in keeping with photographically screened halftone structures.

The white-dot structures of Figs. 1 and 3 were produced by configuring the pulse-width-modulation circuitry to produce longer unblanking time when the lighter portions of the positive transparencies are scanned. The halftone image of the positive transparencies thus appeared as positive on the CRT faceplate and a photographic reversal process was used to provide photographic positives as the first reproduction film records. The next step was the preparation of high-contrast photocopy negatives of the halftones. These negatives were then used to make the plates used to print the pages on which the halftones appear.

The black-dot photograph of Fig. 5 was made by presenting a negative image of the same positive transparency on the CRT by electronic inversion of the pulse-width-modulation process. The negative CRT image was photographed onto high-contrast (negative) film. This resulted in a black-dot positive image as the first reproduction film record. As in the case of "white-dot" halftones, an intermediate photographic negative was used to transfer this image to the printing plate. The white-dot process is more analogous to the television process because the video voltage is proportional to the reproduced light level; that is, the signal "creates" the white dots that produce the visual sensation. The white-dot system may be thought of as being additive whereas the black-dot system can be considered subtractive, because the video signal is proportional to the amount of ink that creates the spaces between the sensation-producing light areas.

The black-dot process would probably be preferred by the segment of the printing industry that uses letterpress printing because of its superior printability by the letterpress process and because of its closer resemblance to usual halftone structure. However, there is a disadvantage inherent in black-dot generation by electrical processes, which arises because it is difficult to maintain a uniformly small difference between two large quantities. If we observe that the black-dot process requires large signals to produce adjacent black halftone dots with relatively small white areas, we can see that a small percentage of perturbation in the diameter or position of one black dot with respect to its neighbor will cause a substantial percentage change in the area of included white space. This constitutes a sizable modulation of the represented brightness level.

The type of perturbation likely to appear in an electronic system is hum modulation from the ac line. It can be seen in the black-dot halftone, Fig. 5, as the semicircular swirl pattern that appears convex-upward over a large portion of the picture. It was found that by locking the halftone generator to the ac line and by counting an integral number of cycles per line of halftone dots, the pattern could be immobilized at a minimum spatial frequency to lie along straight rows perpendicular to the horizontal-dot row. Threshold tests indicated that the hum-pattern visibility was lowered by a factor of only about two by this technique. For these tests, the spiral modulators and all other video and CRT geometry and the focus-correcting equipment were temporarily disconnected to eliminate possible ground loops that might introduce hum. Then hum displacement was inserted through a calibrated attenuator into the deflection amplifier input. Dot areas of about 50 percent were achievable for this test, not by spiral scanning but by defocusing the apertured CRT to form a fairly uniform dot cross section. The 50 percent dot blank-field structure has a threshold visibility for displacement modulation of about 1.0 percent of the 9-mil interdot spacing used in the test. This threshold was extrapolated to be about one third as large, or 0.33 percent (0.03 mil), for the assumption of a 95 percent dot, the usually permissible maximum, especially with the use of letterpress printing. This observation is fairly consistent with calculations based on the assumption that the brightness threshold ΔB is about one percent for large contiguous fields having a sharp line of demarcation between them. For the large dot percentages of interest, the allowed dot perturbation ΔS_{max} is proportional to the brightness threshold $\Delta B = 0.01$ times the square root of the fractional white area T_{min} times the interline spacing S:

$$\Delta S_{max} = 0.01 \sqrt{T_{min}}\, S$$

This is plotted for a few values of contrast ratio

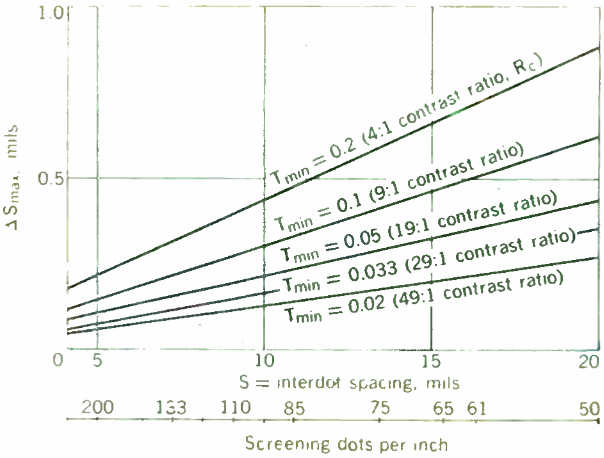


FIGURE 6. Permissible displacement ΔS_{max} of a halftone dot row in V or H direction for assumed brightness threshold $\Delta B = 0.01$, as a function of interdot spacing and dots per inch.

$$R_c = \frac{1 - T_{min}}{T_{min}}$$

in terms of the screened dots per inch and the interdot spacing in Fig. 6.

For the same case of 110-line screening (9-mil interdot spacing) as in the foregoing measurements with a 19:1 contrast ratio, the allowable dot displacement from the normal is about 0.021 mil, or 0.231 percent of the interdot spacing. Since the measured or calculated allowable displacement error amounts to about two or three parts in one million, the dynamic range and hum-rejection ratio of the deflection circuitry to produce such a 10-inch final picture (regardless of the demagnification on the CRT) must be about 120 dB. This stringent requirement was met in a stripped-down version of the halftone CRT display.

An important observation to make, however, is that the "white-dot" system is, by its nature, orders of magnitude less sensitive to hum disturbance. This is true because a hum signal in the deflection system does not modulate the area of the small white dots but just moves them slightly in a black field. Hum pickup in the video pulse-width-modulation system is to be avoided of course because it has the same effect on dot area as the video modulating signal.

Video pulse-width-modulation considerations

Pulse-width unblanking to control the recorded size of the spirally scanned dot has been produced by driving an integrator from the photomultiplier-tube output of the flying spot scanner, and allowing the integrated voltage to activate a Schmitt-trigger circuit. This form of pulse-width modulator uses a constant-current source to fire the Schmitt trigger just before blanking resets the integrator circuit. This is the "black-level" control. The electrical system using the pulse-width modulator is described as negative-acting because increased light transmission in the subject produces decreased dot area on the CRT. The first halftone experiments were made with this pulse-width converter and produced surpris-

I. Summary of operating conditions for various positive and negative halftone processes

Input Subject	Film Output Process	Electrical Sense	Nature of Required Transfer Characteristic to Give Constant, ΔD per Output Step: Referred to PEC Video	Referred to Output Copy	Process Results: Halftone Dot	Hard-Copy Display
Case I, pos	a. pos (reversed)	1 pos	linear	linear	white	pos
	pos (reversed)	2 neg	black stretch	black stretch	white	neg
	b. photo-neg	1 pos	white stretch	black stretch	black	neg
	photo-neg	2 neg	linear	linear	black	pos
Case II, neg	a. pos (reversed)	1 pos	linear	linear	white	neg
	pos (reversed)	2 neg	black stretch	black stretch	white	pos
	b. photo-neg	1 pos	white stretch	black stretch	black	pos
	photo-neg	2 neg	linear	linear	black	neg

ingly good results. Analysis showed that a "black stretch" ($\gamma < 1$) characteristic is needed to produce approximately equal unit density changes in the output for unit density changes in the input transparency. The integrator had such a characteristic by virtue of a relatively low voltage with respect to the Schmitt-trigger level, to which the integrator was reset by the dot-blanking pulse. The voltage level to which the integrator was resettable was adjustable over a limited range to provide some control of the black-stretch characteristic.

A more versatile pulse-width-modulation method was evaluated—one in which a sawtooth voltage was added to the video signal. In this device the video signal caused the sawtooth voltage to trigger the Schmitt-trigger circuit at varying points on its slope, depending upon the extent to which the video level raised the sawtooth into the Schmitt threshold. This device made a linear transform of video voltage to pulse duration for a linear sawtooth. It had an advantage in that the sense of the pulse-width modulation could be made negative or positive for a given video polarity simply by inverting the sawtooth voltage waveform. The modulation characteristic could be distorted to produce black stretch by using a nonlinear sawtooth, which was produced by using a relatively large portion of the exponential *RC* charging characteristic.

Gray-scale requirements

In reproducing the gray scale of a photographic negative as a positive image, it is desirable that the transmittance of the reproduction be proportional to some power of the reciprocal of the object transmittance.

The polarities of the electric and photographic processes that constitute the halftone system can be arranged to produce negative or positive copy from either negative or positive input. The conditions necessary to produce the possible combinations are summarized in Table I, which indicates the nature of the halftone dot and the sense (black or white stretch) of the electrical distortion of the transfer characteristic needed to give the desired unit-density correspondence in the output copy. The black-stretch or white-stretch definition must be referred to the addition of differential gain in those respective portions of the photomultiplier-tube video-output signal, since this circuitry has no "knowledge" of the polarity of the output and input film processes that affect whether black stretch or white stretch occurs with respect to the output process. Table I also shows

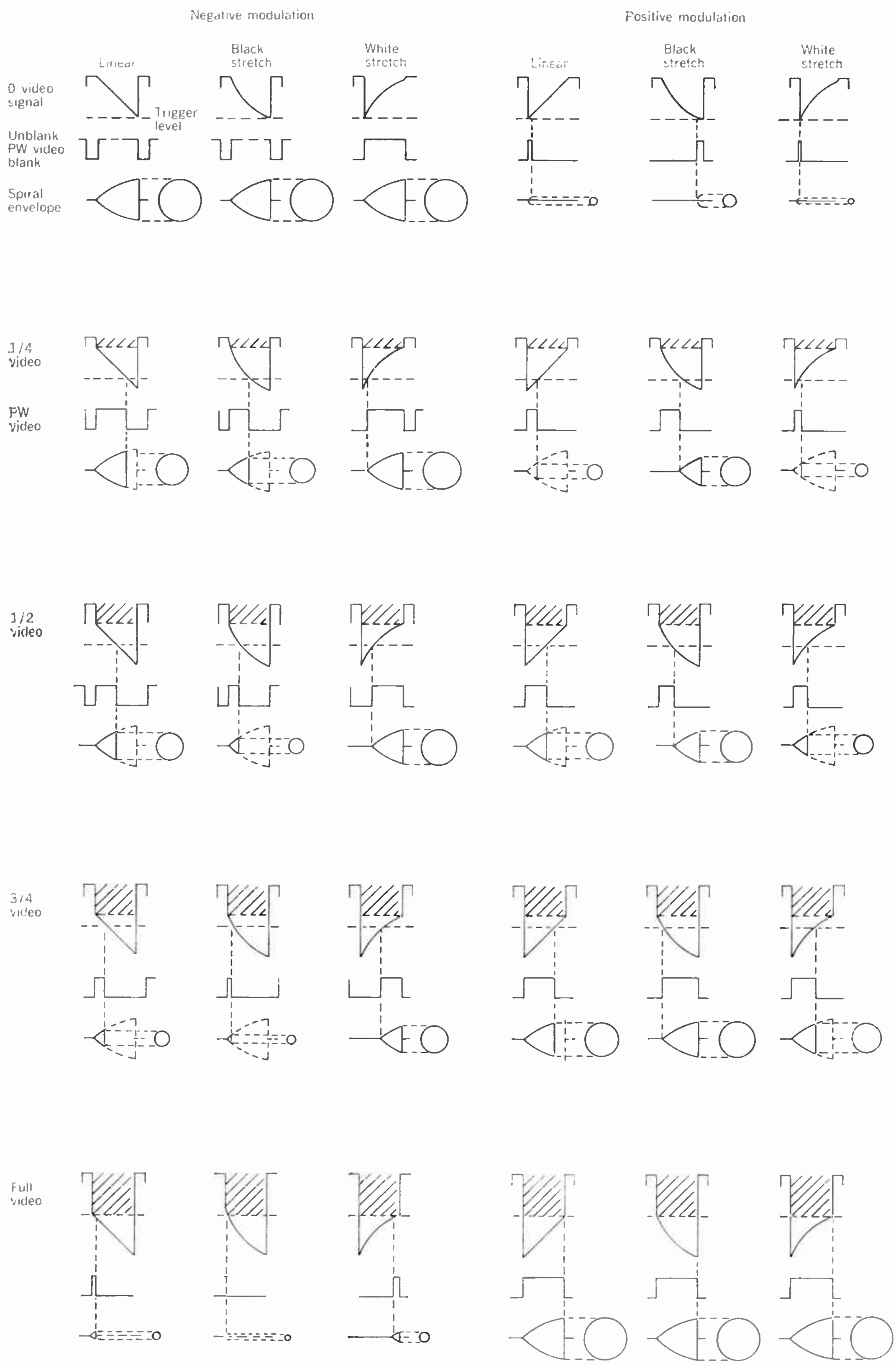
FIGURE 7 (right). Illustration of pulse-width modulation concept for producing positive and negative halftones with gray-scale compensation.

the sense of the required transfer characteristic referred to the output copy. This is seen in every case to be a black-stretch requirement, indicating that the need for white stretch within the system is only an internal artifice of implementation.

Figure 7 illustrates the generation of video-pulse-width waveforms for the linear black-stretch and white-stretch conditions in producing both positives and negatives, for various degrees of video modulation. The active portion of the spiral-scan modulation envelope is sketched within the unblanking interval, and the corresponding dot size appears adjacent to the envelope.

The required electrical black-stretch waveform for negative modulaton is made by using a negative-sawtooth slope with exponential curvature and by starting the spiral scan at the end of the normal dot-blanking period. For electrical white stretch with negative modulation, the exponentially shaped sawtooth is used as it is generated, in its noninverted form. The spiral scan is started from the residual of the pulse-width-modulation waveform by inverting it for input into the spiral scan generator. With positive modulation, the exponential sawtooth for black stretch is used with negative slope with the residual pulse-width-modulation mode of operation. Electrical white stretch for positive modulation uses the positive curved-sawtooth slope and the normal, nonresidual, pulse-width-modulation mode.

The developmental equipment used in making the dot halftones shown with this article utilized the superimposed-sawtooth-type modulator to produce the pulse-width-modulation video signal. The white-dot examples correspond to case I-a-1 in Table I, and the black-dot example to case I-b-2. The theoretically required linear setting of the pulse-width modulator was modified to provide some black stretch. In addition, it was found advantageous to use some black and white stretch from a separate gradation correction amplifier, which used diode-break circuits to synthesize greater differential gain in those video signal regions. This can be explained by noting the S-shaped response of the practical dot area versus time curve shown in Fig. 4. Control of the transfer


Negative modulation
Positive modulation
Linear
Black stretch
White stretch
Linear
Black stretch
White stretch
0 video signal
Trigger level
Unblank PW video blank
Spiral envelope
1/4 video
PW video
1/2 video
3/4 video
Full video

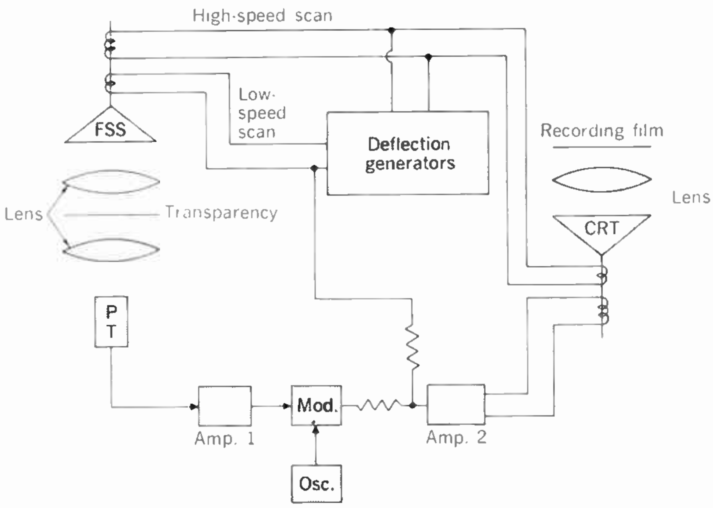


FIGURE 8. Simple line-width halftone generator.

FIGURE 9. Simple FM halftone generator.

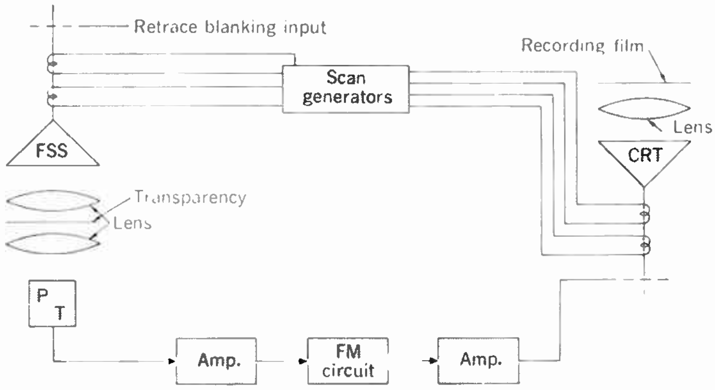
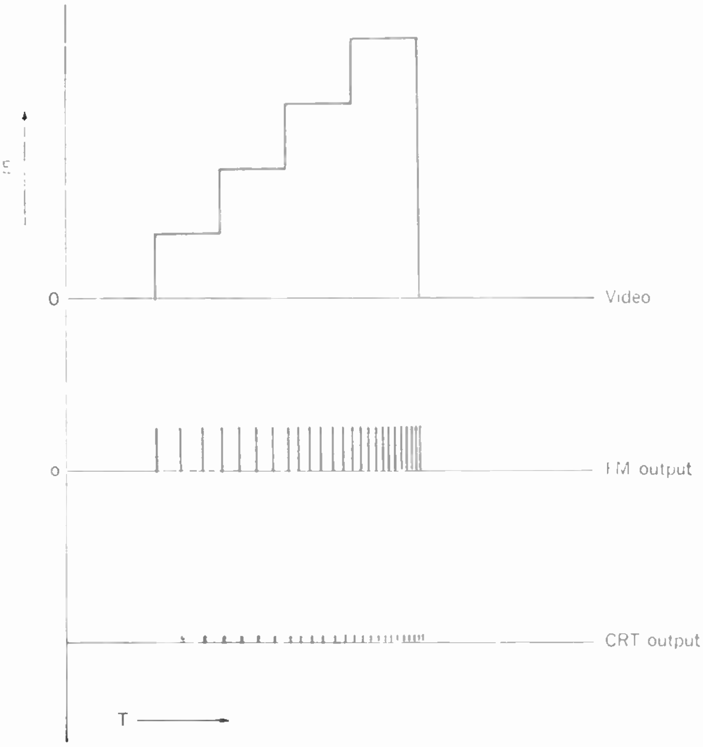


FIGURE 10. FM system waveforms.



characteristic to compensate for the printing process used for this article has not been optimized. Therefore, these figures do not represent the ultimate quality believed possible with electronic methods.

Variable-line-width halftones

Although the use of variable-size dots can be considered the standard means for generating halftones, other systems have been devised and experimentally implemented. In the system described here, a television-type scan is used and the width of the scan lines is made to be dependent on a video-control signal. Figure 8 is a simplified block diagram of this system. A flying spot scanner scans a continuous-tone transparency in television fashion and the photomultiplier tube detects the amount of transmitted light through the required lenses. The phototube output signal is amplified and fed to a balanced modulator, modulating the sine-wave carrier generated by the oscillator. The modulator output is summed with the low-speed scan, amplified, and fed as a deflection current to the low-speed yoke. The high-speed scan system is not modified. Also, the scanner and readout CRT are synchronously scanned by use of common deflection generators.

There are a number of systems that can be used to generate electronic line-width-modulation halftones. Some of those systems and the various constraints and limitations of each were outlined in the preceding paragraphs. Modifications such as curved, rather than straight, width-modulated scan lines were not discussed, but these can indeed yield at least interesting halftones. A major newspaper uses such optically screened halftones for some eye-catching advertising, e.g., concentric line-width-modulated circles in lieu of straight lines.

FM halftone systems

The systems that have been described produce halftones by varying the size of a dot, or line, while keeping the general structure fixed. The FM system discussed in the following generates a halftone using small, fixed-size dots of adjustable spacing. The system is described as an FM system because the spatial dot frequency is made to vary with the desired optical density.

The standard variable-dot-size technique can be called an AM system, since the spatial frequency of the dots is constant whereas their size, or amplitude, varies as a function of optical density. In the FM system a television-type scan is used to generate the raster and the unblanking waveform determines which spot locations within the raster will be used. The dots, when producing an FM halftone, are all the same size, which means that the unblanking waveform must be a fixed-width pulse of a rate dependent on the desired optical density. Figure 9 is a simplified block diagram indicating a possible configuration for such a system.

The flying spot scanner and readout CRT are driven from synchronized scan or raster generators. A continuous-tone transparency is scanned in television fashion with the photomultiplier tube gathering the light transmitted through the transparency by means of imaging and collecting optics. The amplified photomultiplier-tube output is fed to an FM circuit that converts the amplitude-varying signal into a period-varying fixed-width pulse train. After amplification and buffering, the pulse train is applied to the control grid of the CRT

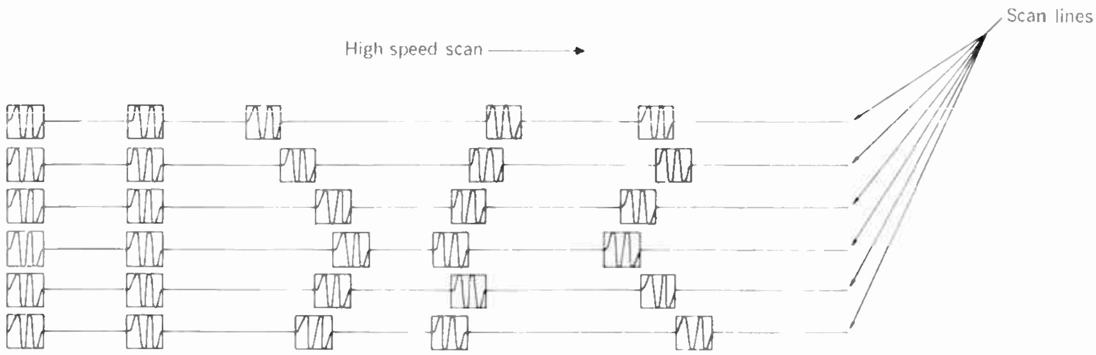


FIGURE 11. Possible FM dot pattern.

for unblanking. Sample waveforms at various circuit points in the system are shown in Fig. 10.

The first waveform represents the video signal on the scan line from the photomultiplier using a step-wedge transparency. The second waveform represents the pulse-train output from the FM circuit as a result of the foregoing video waveform. The third line indicates the dot pattern as seen on the CRT faceplate. The display consists of high-optical-density dots surrounded by unexposed or low-optical-density film base. In the example, if the video waveform corresponds to light transmitted through the transparency then the picture taken off the CRT faceplate will be a negative of the original transparency. A print of this negative will, of course, yield a positive. It is also possible to change the polarity of the image electrically by inverting and offsetting the gamma-corrected original video signal prior to application to the FM circuit. Therefore, positives or negatives of the original are possible with this system.

The operation speed of this or other systems is limited to various constraints imposed by CRT-brightness capabilities vs. spot size and recording-film speed or sensitivity. A reasonable maximum spot velocity on the film is approximately 5 inches per millisecond. The required "on" time for the FM circuit is determined using the following equation:

$$T_{ON} = \frac{\text{Desired spot width} - \text{actual spot width}}{\text{Spot velocity}} \quad (1)$$

where it is assumed that the actual spot width is smaller than the desired spot width and the spot velocity can have a maximum value of 5 in/ms. The "off" time requirement for the FM circuit when producing maximum optical density is given by

$$T_{OFF}\Big|_{D = MAX} = \frac{\text{Actual spot width}}{\text{Spot velocity}} \quad (2)$$

and the "off" time for minimum optical density is given by

$$T_{OFF}\Big|_{D = MIN} = \frac{\text{Desired spot width} - \text{actual spot width}}{(\text{spot vel.})(\log^{-1} \text{density}_{MIN}) - 1}$$

$$= \frac{T_{ON}}{(\log^{-1} D_{MIN}) - 1} \quad (3)$$

The low-optical-density end of the picture can be set by a constraint that gives a maximum dot-to-dot spacing for reasons of objectionable structure discernibility. From the preceding remarks and equations it should be apparent that the density-range capability is determined by the spot width and maximum interdot spacing. The spot shape, for ease of discussion, is assumed to be square, and the spacing between successive scan lines is set equal to a spot width.

To illustrate the use of the foregoing equations, consider the following example. Assume that the readout device produces a square spot 1 1/2 mils on a side and that a square spot 3 mils on a side is desired in order to be compatible with the smallest reliable printable spot in the final inked process. Also, assume a spot of velocity of 0.15 in/ms, which yields an "on" time requirement for the FM circuit of 10 μs. (To increase the spot dimension from 38 to 76 μm in the other direction, such techniques as spot wobble or an asymmetric aperture may be used.) The "off" time for maximum density will then also be 10 μs. If 10 mils (or 160 lines/cm) is the maximum allowed interdot spacing then the minimum density can be calculated using Eq. 3 with the knowledge that

$$T_{OFF} = \frac{\text{Spot spacing} - \text{actual spot width}}{\text{Spot velocity}}$$

$$= \frac{8.5 \text{ mils}}{150 \text{ mils/ms}}$$

$$= 56.7\ \mu s$$

The required "off" time variation is therefore 10 μs to 56.7 μs or a ratio of 5.67 to 1. Thus, the density D is

$$D = \log\left(\frac{T_{ON} + T_{OFF}}{T_{OFF}}\right)$$

$$\approx 0.07$$

and is the theoretical minimum density using the assumptions stated above. Also assumed is that the optical density of the spot has a value of one or greater. If the optical density were exactly one, the density figure of 0.07 would be reduced to approximately 0.062 due to the light transmission through the spot itself. The maximum density is, for this example, just equal to the optical density of the spot itself. A paper fully covered with ink can, for example, produce an optical density of 1.4 and consequently a "tone" or density range of 1.4 − 0.07, or 1.33, would be possible.

The FM circuit has a threshold characteristic that

FIGURE 12. An FM electronically screened halftone.

FIGURE 13. Various patterns versus optical density for the bidirectional FM system.

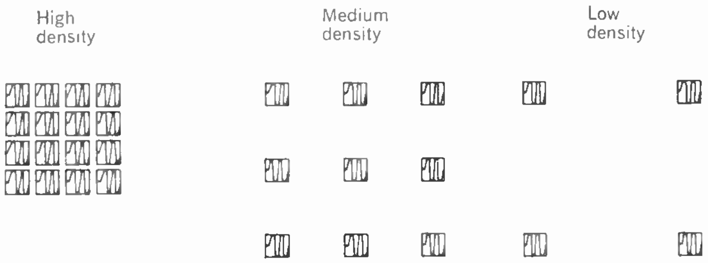
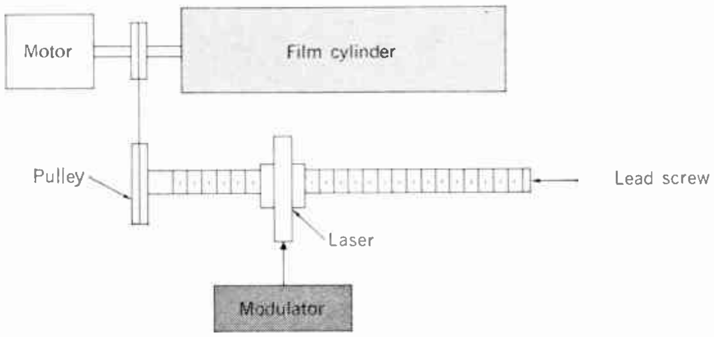


FIGURE 14. Possible implementation of FM system using a laser as a modulated light source.



allows the tone range to be extended in the highlight areas of the picture. That is, at video levels below the adjustable threshold the FM circuit ceases to oscillate. Since no spots are then produced, the density theoretically goes to zero. A halftone made in that fashion has a continuous density range from about 1.4 down to 0.07 and then a step from 0.07 to approximately zero.

When a field of constant density is produced with the FM system, a structure becomes apparent to the eye. The structure consists of lines running in a direction parallel to the low-speed scan. Figure 11 illustrates the pattern formed. The first two columns are shown as if the dot rate were synchronized to the high-speed scan rate and the last three columns indicate a possible pattern if no synchronization exists.

For normal pictures, however, the information content is high. Consequently there are no large areas of constant level and the picture information obscures the structure by virtue of the constantly changing modulation. It is possible to obscure the structure, even for large areas of constant density, by the addition of random noise to the video-signal input to the FM circuit. The amount of noise added should be inversely proportional to the video-signal level to allow deviations of plus or minus one half the average interdot spacing for optimum structure obscurity. Although the original structure can be made to vanish in the noise, the unremovable low-frequency noise components tend to produce a slightly blotchy effect. Normal pictures of high information content are generally relatively free of interference patterns (Fig. 12).

A still more complex FM system, called the bidirectional FM system, is currently being investigated. In this system, the dot spacing is made to vary not only in the direction of the high-speed scan, but also in the direction orthogonal to it. Figure 13 shows the pattern obtained from this system for various optical densities. It should be noted that the pattern is always regular and, in the case shown. approximately square.

No noise will be introduced into the bidirectional system because the regular pattern should not be objectionable (ordinary halftones currently used in industry have a regular pattern) and added noise will only tend to introduce a slight blotchiness, as noted previously.

Advantages of FM systems

One advantage of FM systems is that they lend themselves to incorporation with lasers. Consider a rotating drum wrapped with a sheet of film and, as shown in Fig. 14, a laser mounted on a carriage driven by a lead screw. Scanning is accomplished by rotation and translation. The laser is "on–off" modulated by the FM circuit with a suitable driver. This avoids the laser-deflection problem, and could lead to a direct printing-plate production method, using a laser of sufficient power to generate gravure cavities directly in the final printing plate. This would eliminate the intermediate steps of optical screening and photographic plate making.

Another advantage of the FM system is the absence of the nonlinear transfer function that relates the dot size on the printing plate to the dot size on the final inked copy, because, in the FM system, all dots are the same size. It should be noted that both the FM and AM systems are still in the experimental stage and further exploration is required for an operational model.

Optical heterodyne detection

Once the stringent requirements of optical heterodyne detection are met, this technique provides numerous advantages over other types of detection, including high discrimination against noise

O. E. DeLange *Bell Telephone Laboratories, Inc.*

Optical heterodyne detection presents some difficult problems, but for certain applications it offers considerable advantage over the more conventional detection methods. Some of these advantages and problems, and how they are affected by the transmission medium involved, are discussed in this article. Experimental evidence has shown that although heterodyne reception is not desirable at the end of long atmospheric transmission paths, it is entirely feasible for long paths that are enclosed to isolate them from atmospheric disturbances.

The heterodyne—or, as it is often called, superheterodyne—circuit is used very extensively in modern receivers operating at conventional frequencies. This use covers the range from broadcast home receivers to transcontinental microwave systems. We therefore ask, "Will the advantages of the heterodyne receiver give it an equally prominent place in optical communications systems of the future?"

The heterodyne circuit was adopted for use at radio frequencies because it provided a means of moving an incoming signal down to some lower frequency at which processing was easier. The optical situation is somewhat analogous in that heterodyning makes it possible to translate the information-bearing optical signals down to some much lower frequency range, for which more satisfactory amplifiers, filters, and detectors are available.

Strictly speaking, one should refer to these devices not as "heterodyne detectors" but as "translators" or "frequency converters." The process of detection—that is, recovery of the modulation frequencies—comes at a later stage in the heterodyne receiver. However, the term "heterodyne detector" has come into rather general usage, probably because the same device that is used for frequency translation can also serve as a direct detector. Since there is no other satisfactory generic name for these nonlinear devices, they are often referred to as optical detectors, regardless of how they are employed. In this article we shall try to differentiate between the process of detection and that of conversion; of necessity, however, we may refer to optical detectors even when these devices are used as frequency converters.

Although it is not possible at the moment to predict whether or not the heterodyne receiver will play as prominent a part at optical frequencies as it has played at lower frequencies, it is known to provide definite advantages for some applications. It also presents some special problems. These advantages and problems are discussed in the following paragraphs.

Optical heterodyning

In principle, heterodyning at light frequencies is the same as at much lower frequencies. For the conventional receiver, voltages at two different frequencies are applied to a nonlinear circuit element, such as a diode, which acts as the frequency translator. Current at the difference frequency is recovered from the output of the device and is amplified and processed. Similarly, at optical wavelengths two electric fields at different frequencies are applied to a photodetector, and current at the difference frequency is recovered from the output. This technique provides a signal that is an exact replica of the original but moved down to some lower frequency range in which amplification, filtering, and detection are easier to achieve.

As might be expected, the greatest difference between the optical heterodyne circuit and its conventional counterpart results from the very large difference in the frequency of operation. At lower frequencies the signal and beating oscillator voltages are somehow applied to the conductors that connect to the converter. This is essentially a lumped-circuit device, even though it may have distributed inductance and capacitance. For the light receiver, the local-oscillator power is contained in one beam and the signal in another. These two beams are superimposed on a detector where, in a typical case, they cover an area greater than 1000 wavelengths in diameter. Each element of area contributes a current, and for maximum response these currents must be exactly in phase. Therefore, the two waves must have the same relative phase relationship over the total area. It will be shown later that it is rather difficult to maintain a proper phase relationship over the complete area of superposition.

Direct detectors

Before proceeding to the special case of the heterodyne receiver we shall devote some attention to optical detectors in general. Although a number of types of light detectors are available, we shall give only a brief discussion of two of them. The first is the photocathode, which consists of an active surface that emits electrons when exposed to light. Since low-level light produces very few electrons, this detector is often used in conjunction with an electron multiplier, in which the emitted electrons are accelerated and applied to a dynode. This is an electrode that has the property of emitting a number of secondary electrons for each high-velocity electron impinging upon it, thereby producing current multiplication. The process is repeated many times to produce large current gains; a factor of 10^6 is not unusual. This gain can be accomplished without adding any appreciable noise. Moreover, photocathode devices have been constructed having output circuits exhibiting flat frequency response from dc to several gigahertz. Photomultipliers and similar devices make very satisfactory light detectors over part of the visible spectrum. However, emission from a photocathode decreases so rapidly with increasing wavelength that the best cathodes known have a detection efficiency of only a few percent at the red end of the visible spectrum and are completely impractical for wavelengths much longer than one micrometer.

Another type of detector is the photodiode, which can have high detection efficiencies well out into the infrared region. These devices, which are very similar to the semiconductor diodes used as microwave detectors, are operated back-biased so as to be nonconducting when not illuminated. Light acting on the material releases electrons, which are collected at the terminals to provide the signal current. The signal current is directly proportional to the light power input. As with the photocathode, the photodiode produces very small output for low light level; unfortunately, it has not as yet been possible to get large amounts of noise-free current gain from such a diode.

B. M. Oliver[1] has analyzed light detectors in some detail; we shall discuss them only very briefly. For all detectors that we shall consider, the current output is directly proportional to the light power input; this property implies a characteristic of the form $i = \alpha E^2$, where α is a constant. If the electric field set up by the light is given by $E = A \cos \omega_s t$, then the detector current i is as follows:

$$i = \alpha(A \cos \omega_s t)^2 = \alpha A^2 \cos^2 \omega_s t$$

By trigonometry,

$$i = \alpha A^2(\tfrac{1}{2} + \tfrac{1}{2} \cos 2\omega_s t)$$

The $\cos 2\omega_s t$ term represents a frequency much too high to be passed by the detector and can be ignored, leaving $i = \alpha A^2/2$. This can also be written $i = \alpha P_s$, where P_s is the power in the light signal.

We next need to evaluate the constant α. For an ideal optical detector, each quantum of light releases one electron, which is collected as useful output. Most practical detectors require some number N of photons, or light quanta, for each electron; thus, the average number of electrons per photon is $1/N$. The reciprocal of N is called the quantum efficiency and is usually designated by η.* The current produced by M photons per second is simply $i = M\eta q$, where q is the electronic charge. One quantum of light $= hf$ joules, where h is Planck's constant and f is the light frequency ($\omega_s = 2\pi f_s$). The number of quanta per second is given by $M = P_s/hf$, and $i = \eta q P_s/hf$. It is evident from comparing the two expressions for current that $\alpha = \eta q/hf$. The signal power is given by

$$W_s = i^2 R = \alpha^2 P_s^2 R \qquad (1)$$

The detector just discussed is sometimes called a "quantum counter" or "direct detector." Since the current obtained from the detector is composed of discrete electrons, it will fluctuate in time. The smaller the current the greater the percentage fluctuation. This noise is similar to the shot noise encountered in the operation of electron tubes. If we call the mean-square fluctuation or shot-noise current i_{sN}, it is well known[1] that, for a bandwidth b,

$$(i_{sN})^2 = 2qbi_{dc} = \frac{2q\eta b P_s}{hf} \qquad (2)$$

The ratio of signal power to shot-noise power,

$$\frac{S}{N} = \frac{i^2}{(i_{sN})^2} = \frac{\eta P_s}{2hfb} \qquad (2a)$$

represents the maximum signal-to-noise ratio obtainable from a direct detector.

For a practical receiver there will be other sources of noise in addition to the signal shot noise. Two examples are thermal noise at the input to the signal amplifier and detector dark-current noise. These noise powers are independent of input signal level whereas the signal shot-noise power increases directly with signal input power as shown by Eq. (2). The output signal power increases as the square of the input signal power; see Eq. (1). From these considerations it is evident that the performance of a direct detector, as expressed by Eq. (2a), approaches the ideal when its quantum efficiency is near unity and the optical power is great enough that shot noise predominates.

To illustrate the low signal level obtained from a photodiode, without current amplification, consider such a detector operating in the visible red and having a quantum efficiency of 50 percent. For a space communication system operating in a range of interest and employing reasonable power and antennas, one might expect a signal level of approximately 10^{-10} watt at the receiver. From Eq. (1), if the diode load resistance is 1000 ohms, the signal power

$$W_s = \alpha^2 P_s^2 R = \left(\frac{\eta q}{hf}\right)^2 P_s^2 R$$

$$= \left[\frac{0.5 \times 1.6 \times 10^{-19} \times 10^{-10}}{6.6 \times 10^{-34} \times 4.75 \times 10^{14}}\right]^2 \times 10^3$$

$$= 6.5 \times 10^{-19} \text{ watt}$$

For a 10-MHz bandwidth this signal power is smaller than the thermal-noise power at the input to the signal amplifier by a factor of about 10^5, or 50 dB.

A multiplier-type photodetector could provide suffi-

*To stress the importance of quantum efficiency let us point out that a detector with one percent efficiency is equivalent to a detector with unity efficiency having in its input an attenuator that absorbs 99 percent of the incident light.

cient gain to make the thermal noise negligible. However, the best photocathode known has a quantum efficiency of only about 5 percent at the 0.63-μm wavelength; with such a device the signal-to-noise ratio obtainable would be only five percent of the ideal as expressed by Eq. 2(a).

Figure 1 shows a direct-detection receiver in block-diagram form. The signal P_s is passed through an optical bandpass filter, having bandwidth B, to the detector. At the output of the detector, there is a current amplifier followed by a signal amplifier having some smaller bandwidth b. The thermal noise at the input to the signal amplifier is represented by N_T.

Another source of noise may arise from background light that reaches the detector. This background-light power, spread over a band B (as shown in Fig. 1), is represented by P_B at the input to the receiver. Although the background is usually of low intensity, it may be spread over an extremely wide frequency range. Optical filtering can provide considerable improvement, but even a good filter may pass a band 100 GHz wide. If the total background-light power reaching the detector is comparable to the signal power, the shot noise produced by background will be comparable to that produced by the signal. Therefore, the obtainable signal-to-noise ratio will be reduced. It is not unusual, in some applications, for total background-light power to exceed signal power—for example, in space communications, where the receiver is subjected to light scattered from the atmosphere or may be looking directly at some bright object such as a planet.

The direct detector is not responsive to the phase of light waves; therefore, it does not provide satisfactory detection of light that has been modulated in phase or frequency.

Heterodyne receivers

Figure 2 illustrates the arrangement of a simple optical heterodyne receiver.* The signal and local-oscillator beams are combined by means of an optical hybrid, such as a beam splitter. The beam splitter would be designed to transmit one half of the light applied to it and reflect the remaining half. In this way, half of the signal power would be applied to each detector along with half the power of the local oscillator. Two detectors are employed so that all of the light will be utilized. The beams would be superimposed on the detectors and so arranged that their electric fields add. The intermediate-frequency outputs of the two detectors would be combined and passed through an amplifier having bandwidth b, as illustrated by Fig. 3(B). This bandwidth needs to be only great enough to pass the signal and can be much less than the optical bandwidth B.

Figure 3(A) represents the light spectrum at the input to one of the detectors. The large input at f_o represents the local oscillator and f_s the signal. The background light is represented by many small components, spread over the band B, and each having some amplitude A_k and frequency f_k. Local oscillator and signal can be represented as $E_o = A_o \cos \omega_o t$ and $E_s = A_s \cos \omega_s t$, respectively. The total background light can be represented as $E_B = \Sigma A_k \cos \omega_k t$. A_k and ω_k have values for every component of light present in the background. The current at the output of the square-law detector is given by

$$i = \alpha(A_o \cos \omega_o t + A_s \cos \omega_s t + \Sigma A_k \cos \omega_k t)^2 \quad (3)$$

As a result of the squaring process indicated by Eq. (3) there are cross products between the various terms, and we obtain components involving the sums and differences of these frequencies. The sum frequencies are much too high to be passed by the detector and can, therefore, be neglected. Using the difference terms yields

$$i = \alpha[\tfrac{1}{2}(A_o^2 + A_s^2 + \Sigma A_k^2) + A_o A_s \cos(\omega_o - \omega_s)t] \quad (4)$$

In Eq. (4) we have also neglected all terms resulting from cross products, or beats, between the $\Sigma A_k \cos \omega_k t$ components and other components. Because of the excellent

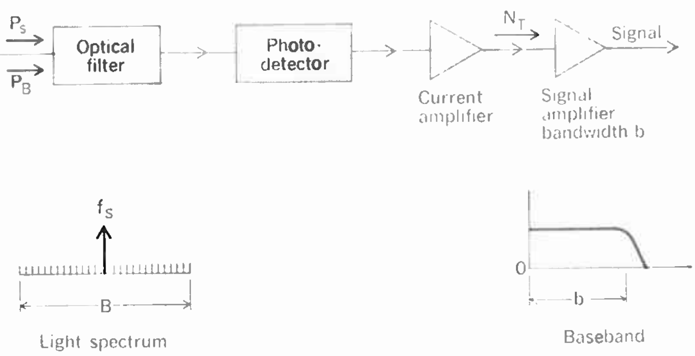


FIGURE 1. Direct-detection receiver and spectrum.

FIGURE 2. Simple optical heterodyne receiver.

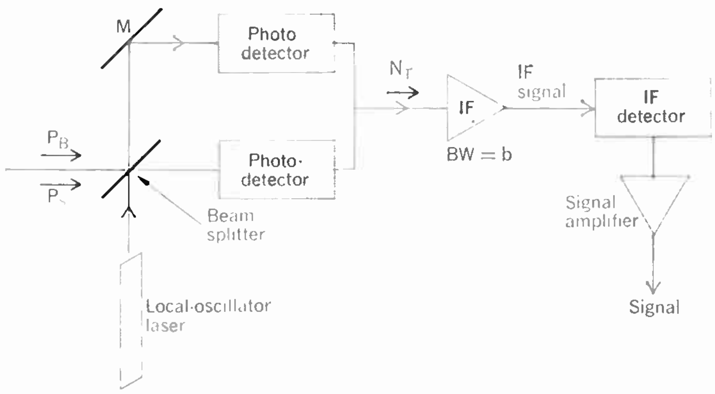
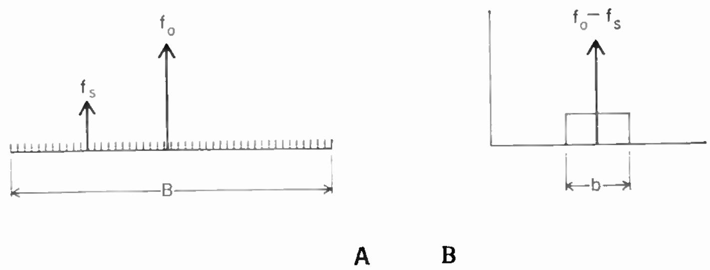


FIGURE 3. (A) Light spectrum and (B) IF spectrum of heterodyne receiver.



*Optical heterodyne receivers are sometimes referred to as "coherent light detectors" to distinguish them from other detectors, which are independent of the phase of the light signal.

frequency filtering that the heterodyne circuit can provide, as well as its critical spatial discrimination, these beats are negligibly small except in unusual circumstances and will not be considered further. The first sum in parentheses in Eq. (4) represents the direct current resulting from the local oscillator, the signal, and the total background light. This is equivalent to $i_{dc} = \alpha(P_o + P_s + P_B)$ where P_o is local-oscillator power, P_s is signal power, and P_B is total background-light power. The $A_oA_s \cos(\omega_o - \omega_s)t$ term represents the beat between the local oscillator and the signal and is the desired intermediate-frequency carrier.

In addition to the direct current and desired IF current there is the shot noise that accompanies the direct current; it is characterized by

$$(i_{SN})^2 = 2\alpha q(P_o + P_s + P_B)b = \frac{2\eta q^2 b}{hf}(P_o + P_s + P_B) \tag{5}$$

If we compare Eqs. (2) and (5), we note that when $P_o \gg P_s$ one effect of heterodyning is a large increase in resultant shot-noise power relative to that generated in a direct detector. Fortunately, as will be shown later, there is a corresponding increase in IF signal power.

With the heterodyne receiver, thermal noise at the input to the IF amplifier will be important in the same way that such noise is important at the input to the signal amplifier when direct detection is employed. This noise power can be specified by the equation $N_T = FkTb$, where F is the noise figure of the amplifier when connected to the photodiode, k is Boltzmann's constant, T is the absolute temperature, and b is the IF bandwidth. It is evident that the best performance will result when the noise figure F is the lowest attainable.

From Eq. (4) we obtain, for the intermediate-frequency current,

$$i_{IF} = \alpha A_o A_s \cos(\omega_o - \omega_s)t$$

The IF or carrier power P_c is given by the relation

$$P_c = (i_{IF})^2 R = \frac{\alpha^2}{2} A_o^2 A_s^2 R = 2\alpha^2 P_o P_s R \tag{6}$$

The carrier-to-noise ratio C/N is as follows:

$$\frac{C}{N} = \frac{(i_{IF})^2}{i_N^2} = \frac{2\alpha^2 P_o P_s}{2\alpha q(P_o + P_s + P_B)b + FkTb} \tag{7}$$

The noise includes total shot noise, as expressed by Eq. (5), and thermal noise. Product noise is neglected, as it can be for most practical situations, and for practical receivers P_o can usually be made much greater than $P_s + P_B$ and can be large enough to make the shot noise override the thermal noise. Under these conditions the only noise of consequence is the shot noise produced by the local oscillator. Thus,

$$\frac{C}{N} = \frac{\alpha P_s}{qb} = \frac{\eta P_s}{hfb} \qquad P_o \gg P_s + P_B \tag{8}$$

The IF signal power is seen from Eq. (6) to depend upon the product of local-oscillator light power and signal light power. For the direct detector the power of the recovered signal is proportional to the square of the signal light power alone. If we compare the output of the heterodyne detector with that of the direct detector by comparing Eq. (6) with Eq. (1) we find that the converter has a gain $G_{het} = KP_o/P_s$ where G_{het} is the power gain of the converter compared with that of the direct detector and K is a constant; that is, the gain is equal to the ratio of local-oscillator power to signal power. For the purpose of comparison let us again consider the space receiver with a light-signal input power of 10^{-10} watt but now operated as a heterodyne receiver with local-oscillator power of 10^{-3} watt. Since $P_o/P_s = 10^7$, this receiver should provide an output signal well above thermal noise. Furthermore, photodiodes with quantum efficiencies of 50 percent are available to serve as the detector, and the signal-to-noise ratio could be within a factor of two of the ideal. This is ten times better than the SNR provided by the photomultiplier in the previous example.

From these considerations we see that the heterodyne receiver provides three major advantages. First, the conversion process provides gain so that the signal output of the detector may be made large enough to override thermal and detector noise. Second, the heterodyne receiver provides high discrimination against background light and other unwanted radiation. For a direct detector the latter becomes serious when the total background power at the detector is comparable to the signal power. For the heterodyne receiver, background is bothersome only when the total background power is comparable to the local-oscillator power, which can usually be made orders of magnitude greater than signal power. Third, the heterodyne receiver is more satisfactory than the direct detector for the recovery of phase- or frequency-modulated signals. It does not follow that the heterodyne receiver is always superior to other receivers. Under certain conditions a direct detector can provide equal or even superior performance. The performance of each type depends largely on the light wavelength, signal power, detector development, background environment, and other factors.

The homodyne receiver is a special form of heterodyne receiver in which the local-oscillator frequency is the same as that of the received-signal carrier and must have the proper phase relationship to this carrier. Aside from these factors, operation of the homodyne is no different from that of the more usual heterodyne circuit. A homodyne receiver can provide an SNR 3 dB higher than that obtained from the conventional heterodyne configuration.[1]

Geometrical considerations

As stated previously, heterodyne conversion is a distributed phenomenon; that is, it takes place over an area extremely large in comparison with a light wavelength. This is illustrated in Fig. 4, which shows a signal beam with diameter D_s partially superimposed on a beating-oscillator beam of diameter D_o. If we consider a differential area da in the region of overlap, it is evident that this area will generate a differential amount of current at the intermediate frequency. If the two beams have the same polarization, the magnitude of this current is given by

$$|di| = \alpha\sqrt{2p_s p_o}$$

where p_s is the amount of signal power falling upon the differential area and p_o is the amount of beating-oscillator power in the same area. The phase of di depends upon the phase of both optical signals. The total current from the converter is the sum of the currents from all of the differential elements; since these currents can have different phase angles, they must be added vectorially.

From the foregoing discussion it is evident that the maximum IF, or beat, signal is obtained only when the signal and beating-oscillator beams bear the same phase relationship over the complete area of coincidence. It is therefore implied that the optical phase must be uniform over the complete wavefront of each beam. These requirements are met only under these conditions:

1. The two beams must have the same mode structure, which usually means limiting operation to the fundamental mode, as will be discussed later.

2. The two beams must be coincident and, to provide maximum signal-to-noise ratio, their diameters must be equal.

3. The beams must propagate in the same direction; that is, their Poynting vectors must be coincident.

4. The wavefronts must have the same curvature; that is, both must be plane waves or, if curved, must have the same radius of curvature.

5. The beams must be identically polarized, so that their electric vectors will be coincident.

Modes. Although a general discussion of optical modes is beyond the scope of this article, the reader may be given some feel for the problem by a brief discussion of a simple case—that of modes in an optical resonator, such as a laser.

Because their dimensions are large in comparison with the wavelength, optical resonators are usually able to support oscillations in many transverse modes. Optical transmission lines, which will be discussed later, can propagate some of these modes. A few simple, low-order modes such as those obtained from lasers are shown in Fig. 5. The simplest of these, and the one for which losses are usually lowest, is the fundamental, or dominant, mode. This mode, which is shown in Figs. 5(A) and 6(A), has the appearance of a single bright spot of light. The intensity distribution is Gaussian, with a maximum intensity at the center of the beam. What is more important from the standpoint of heterodyning is that the optical phase is constant over the whole wavefront. The TEM_{10} mode, Fig. 5(B), is seen to have the appearance of two spots of light. The intensity goes to zero and the optical phase reverses at the center of this beam. The TEM_{11} mode, Fig. 5(C), appears as four spots of light, with the intensity going to zero between them. For this mode the phase reverses when crossing the center of the beam in either of two orthogonal directions, only one of which is shown in the figure.

For a local-oscillator beam in the TEM_{00} mode and a signal in either the TEM_{10} or TEM_{11} mode it is evident that there will be no output from the heterodyne receiver if the beams are symmetrically aligned, since the beat-note current produced by one area will be canceled by that produced at a corresponding area on the opposite side of the axis. The same situation prevails for many other modes. Some output could be obtained by offsetting the local-oscillator beam with respect to the signal beam, but it would never be possible to obtain efficient conversion. It might be possible under certain conditions to match the mode pattern of the local-oscillator beam to a complicated signal mode pattern, but this matching would be very difficult to achieve.

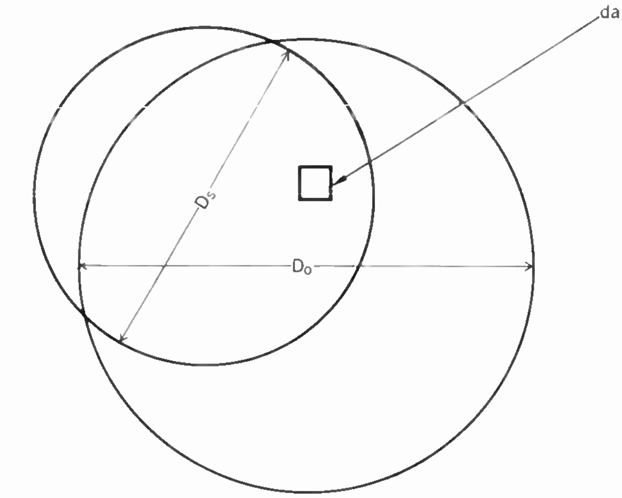


FIGURE 4. Heterodyne beams. A—Dominant mode. B—Complicated mode pattern.

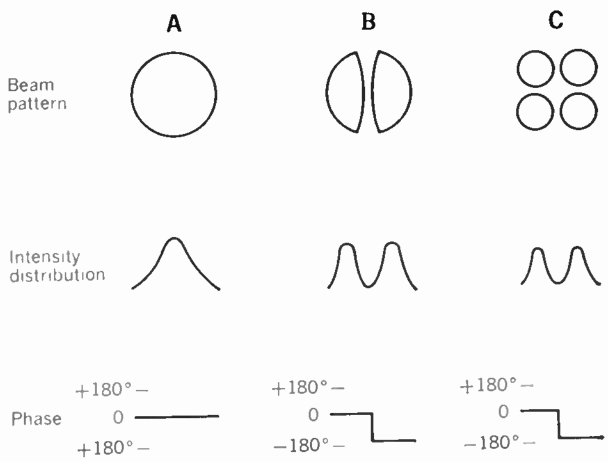
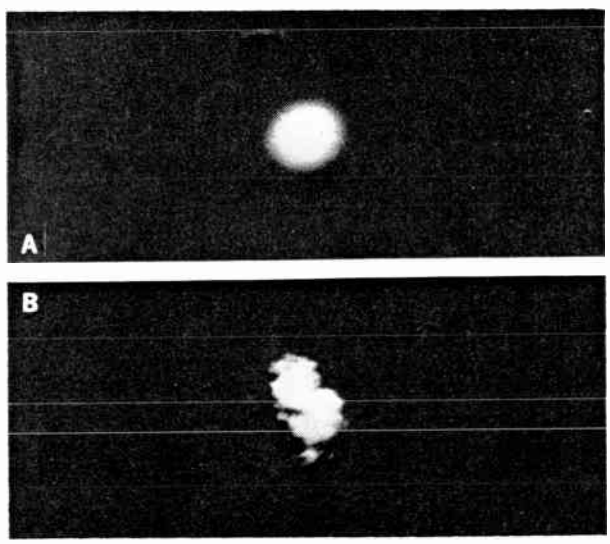


FIGURE 5. Some simple optical modes. A—TEM_{00} (dominant). B—TEM_{10}. C—TEM_{11}.

FIGURE 6. Optical mode patterns. A—Fundamental mode. B—Complicated mode pattern resulting from transmission through a turbulent atmosphere.



Hence, efficient conversion is usually limited to operation in the fundamental mode.

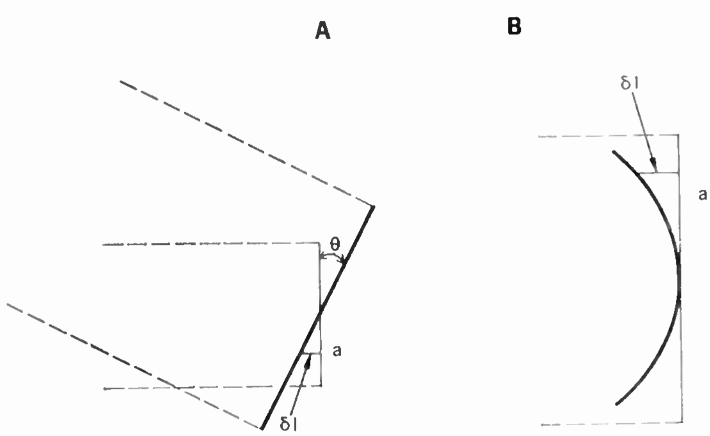
The preceding discussion has been limited to pure modes. In actual transmission the beam will most likely be launched in the fundamental mode but may have some of its energy converted to higher-order modes by wavefront distortions during transmission. As a result, the received beam will consist of a mixture of fundamental and higher-order modes and be more complicated than any situation shown in Fig. 5. This is illustrated in Fig. 6, which shows a beam with all of its energy concentrated in the fundamental mode, in addition to a beam that has additional modes resulting from transmission through 2.6 km of clear but turbulent atmosphere. For transmission systems employing optical components of limited size, losses are greater for the higher-order modes than for the dominant mode. For a more complete discussion of optical modes, see Ref. 2.

Beam coincidence. It can be seen from Fig. 4 that the energy contained in any part of the signal beam that does not overlap the oscillator beam is lost as far as conversion is concerned. Moreover, any part of the local-oscillator beam that falls outside the signal beam contributes to the receiver noise but not to useful signal.

Beam direction. Figure 7(A) shows two plane waves propagating in slightly different directions, as indicated by the angle θ. Consider the differential beat-frequency current di generated on an area da at the center of the beams, and take the phase of this current as the reference phase. At some distance a from the center the wavefronts will be separated by a distance $\delta l = a \tan \theta$ and the phase will be different by an amount $\delta\varphi = 2\pi \tan \theta\, a/\lambda$. We are interested only in very small values of θ. For such values, $\tan \theta$ can be taken equal to θ and $\delta\phi \doteq 2\pi\theta a/\lambda$. When $\theta a = \lambda/2$, $\delta\varphi = \pi$ and the differential current from this area is exactly out of phase with the reference current. The maximum value the distance a can have is the radius R of the beam, and $\delta\varphi_{max} = 2\pi\theta R/\lambda$.

The curves of Fig. 8 show beat-frequency current as a function of $\theta R/\lambda$. The dashed curve represents the function when the light intensity is uniformly distributed over the beam. The quantity $\theta R/\lambda$ represents the distance δl of Fig. 7, expressed in wavelengths, and the

FIGURE 7. Mismatched beams. A—Tilted-plane wavefronts. B—Mismatched curvature.



current goes to zero whenever this distance is an exact multiple of a half wavelength. The solid curve of Fig. 8 is for the more usual case, where the light intensity has a Gaussian distribution. The current never goes completely to zero but approaches this value as $\theta R/\lambda$ is made large.

As a practical example, consider a receiver operating at the common visible-red wavelength of 0.63 μm and with a beam diameter of 1 mm. The data of Fig. 8 can now be plotted in terms of θ alone. The scale at the top of the figure represents this angle in minutes of arc. Employing this scale we find, for example, that for a detection efficiency greater than 90 percent the angle θ must be less than 0.3 minute or 18 seconds of arc.

Wavefront curvature. Figure 7(B) shows the situation when a plane wave and a curved wavefront are superimposed. As in the case for the tilted beams, the phase of the signal recovered from the center of the beam differs from that derived from other areas, and thus the conversion efficiency is decreased.

Polarization. Heterodyne conversion results from the effect of squaring the instantaneous sum of two electric fields by the nonlinear detector. Only the component of the signal field that is spatially aligned with the local-oscillator field can add to it and thereby be effective in producing signals. This implies that the beams must be of the same polarization for maximum conversion.

The heterodyne receiver may be considered disadvantageous in that, for optimum performance, it must meet all of the requirements we have set forth, some of which are very severe. However, these critical requirements aid in discriminating against background light, which is usually incoherent, exists in many modes, comes from a wide angle, is randomly polarized, and covers a frequency spectrum very wide in comparison with the required bandwidth of the receiver. From Fig. 8 we can determine that the angular discrimination of the 0.63-μm receiver described there is the same as that of an antenna with a half-power beamwidth of ± 0.54 minute or 32 seconds of arc—an extremely sharp antenna.[3]

Another problem encountered in heterodyne reception is that of providing a local oscillator maintained at the proper frequency. By employing an IF discriminator in an automatic-frequency-control circuit, the local-oscillator frequency can be made to track the incoming signal frequency; for most applications, this problem is readily solved. For some space communication systems, Doppler shifts of the optical-signal frequency may become large enough to make frequency tracking difficult.

Transmission medium

As indicated in the preceding section, efficient heterodyne reception depends upon providing an accurate match between the wavefronts of the signal and local oscillator. If the wavefront of the signal beam is altered to any extent by transmission through some medium, it may not be possible to obtain such a match. The transmission medium involved may, therefore, determine whether or not heterodyne reception is feasible.

Atmospheric propagation. In addition to large fluctuations of attenuation produced by the presence of fog, rain, smoke, etc., the atmosphere is subject to other difficulties, which may be present even when the attenuation is not significant. The atmosphere often becomes

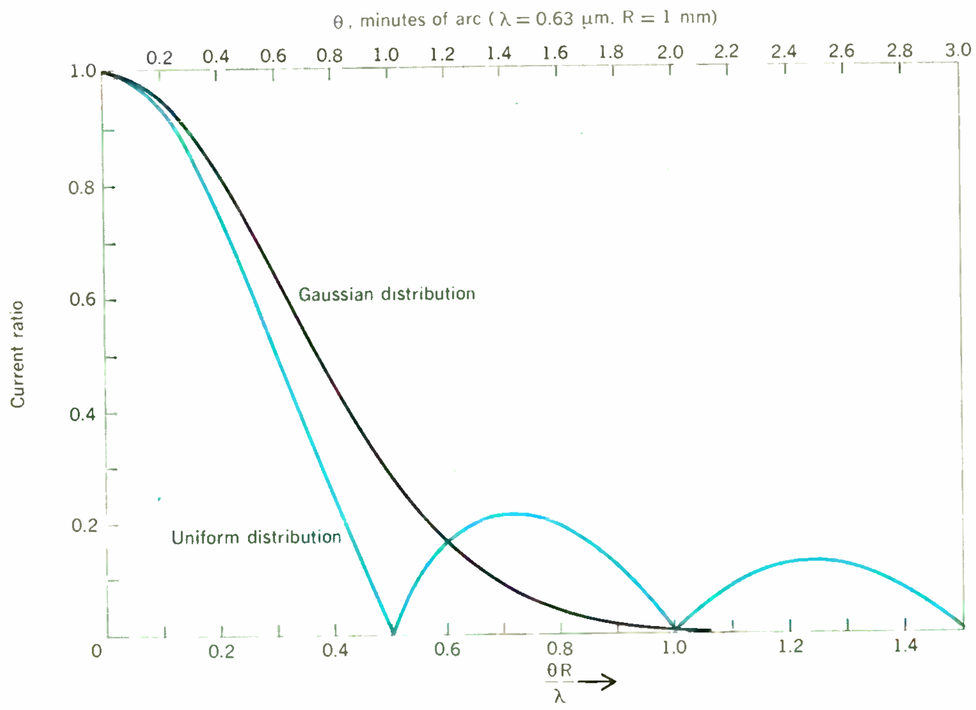
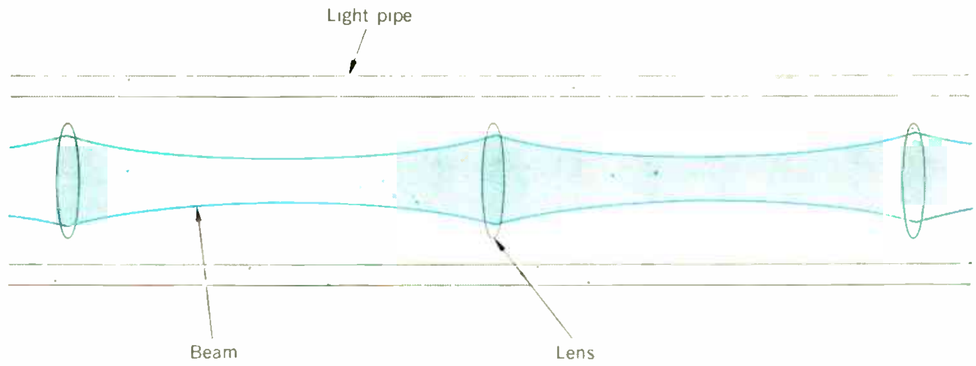


FIGURE 8. Effect of beam tilt on recovered current.

FIGURE 9. Enclosed optical transmission path.



a very inhomogeneous medium composed of many cells of air, each at a slightly different temperature and, therefore, at a different index of refraction. These cells are usually in motion and may pass through the beam, thus producing bending and wavefront distortion. This is the cause of the shimmering that one sees when looking out over a desert or a hot parking lot.

As a result of the wavefront distortion produced by air turbulence, a considerable portion of the light power can be converted into higher-order modes (see Fig. 6). In addition to being complicated, the resultant mode pattern is usually changing with time. It is therefore very difficult to match with the pattern of the local oscillator. Except for unusual circumstances, heterodyne reception of light that has been transmitted for any considerable distance through the atmosphere is inefficient and very unstable; thus, some other form of detector should be employed. Reference 4 describes some experiments on transmission through the atmosphere.

Space propagation. Propagation through free space is ideal for heterodyne, as well as for other types of reception, since there is nothing to produce wavefront distortions. One promising application of optical heterodyne reception is for space communications. Here the receiver may be looking directly at some bright object—the moon or a planet, for example—and may require a degree of discrimination against background light that only a heterodyne detector can provide.

The heterodyne receiver currently has another advantage for space communications. The most efficient, and at the same time high-power, source of continuous coherent light is the CO_2 laser, which operates at 10.6 μm. However, at the present time there is no satisfactory direct detector for this wavelength range; the diode detectors that have been devised are very noisy. As pointed out previously, for the heterodyne receiver the only noise of consequence is the shot noise resulting from the local-oscillator power, provided this power is large enough. Employing the heterodyne principle should make possible an efficient receiver at this wavelength, where an efficient, high-power transmitter also exists.

One major advantage of coherent light transmission results from the fact that, at the wavelengths involved, the energy can be confined to very narrow beams. However, this advantage is obtained only if one can provide very accurate pointing of the beams. This problem is only slightly more difficult with heterodyne reception than with other types of detection.

Unfortunately, however, if the earth's atmosphere is included in the transmission path of a space system the situation described in the preceding section will prevail, and heterodyne reception may not be desirable.

Enclosed transmission path. The most likely method of reliable light transmission over long terrestrial routes would involve an enclosed path, to preclude contamination and reduce turbulence. It is therefore important to determine whether heterodyne reception will be advantageous at the end of such a path. Since light beams travel only in straight lines and tend to diverge, this path would need to be equipped with some means for directing and refocusing the beam. Lenses, or pairs of mirrors, placed at intervals along the line as shown in Fig. 9 can perform the functions of directing and refocusing. Since most of the losses in such a line are produced by the directing elements, the total loss depends upon the number of these elements. Total loss should not exceed 0.3 to 0.6 dB/km.[5] If the line is placed some distance underground, or is well insulated, turbulence can be reduced to a neglfgible level. This is borne out by some experiments performed by D. Gloge[6] and by one that will be described in the next section.

The enclosed line is not without problems. Changes of the mechanical position of the directing elements or temperature gradients in the line can cause displacements of the beam. In view of the critical requirements on a heterodyne receiver, as stated earlier, one might expect beam shifts caused by temperature changes or vibrations of the guiding elements to be very serious. It has been found that the beam position can be accurately controlled by means of a simple servo control system applied to some of the directing elements and that the vibrations of these elements are not serious.

One might also expect some distortion of the beam wavefront when the transmission is through a long line. Assuming a spacing of 100 meters between beam directors and a repeater spacing of 64 km, the beam would encounter (and be modified by) 640 optical components in traversing one repeater link. If each reflection, or transmission, produced only a small distortion of the wavefront, the accumulated distortion might be expected to become important. Such distortions convert some of the light power into higher-order modes. Because of the limited aperture of the optical components in a practical line, the number of modes that can propagate in such a line is limited. However, for some of the lower-order modes, such as TEM_{10}, line loss is only slightly higher than for the dominant mode. These modes can, therefore, propagate and be detected by a direct detector but not by a heterodyne receiver. As a result, one might expect line losses to be higher for heterodyne reception than for direct detection if there is any serious wavefront distortion during transmission. To determine the feasibility of heterodyne reception at the end of a long, enclosed path of this type, an experiment was set up, as described briefly in the next section. A more detailed discussion of this experiment can be found elsewhere.[7]

Experimental performance

The transmission path, shown in Fig. 10, was enclosed in a horizontal pipe 15.3 cm in diameter and 100 meters long. At the wavelength employed (0.63 μm), the maximum beam diameter was only 1 cm, and thus multiple transmission through the pipe was possible. Eight concave mirrors, each 2.5 cm in diameter, were placed at one end of the line; seven similar mirrors were placed at the other end. By reflecting the beam back and forth between these mirrors it was possible to accomplish eight round trips for a total distance of 1.6 km.

The output of the laser was modulated a few percent at an 84-kHz rate before it was launched into the line. The output of the line was divided into two beams by means of a beam splitter, as shown in Fig. 10. One of these beams was applied to a photomultiplier, which acted as a direct detector to recover the 84-kHz modulation. The second beam was applied to a second beam splitter, where it was combined with the output of the local-oscillator laser. One output of this beam splitter was applied to a solid-state photodiode, which served as the detector for frequency conversion. The two lasers were kept tuned 70 MHz apart by means of an automatic-frequency-control circuit. The 70-MHz output of the photodiode was amplified and applied to an IF detector, which recovered the 84-kHz modulation.[8]

By comparing the performance of the heterodyne receiver with that of the direct detector we could determine the relative efficiencies of the two. We found that, within the accuracy of measurement, the efficiencies were the same. The path length was extended to 6.4 km and then by a pulse technique to 64 km. Again there was no measurable difference in efficiency.

These experiments indicate that for distances up to at least 64 km, and involving over 600 reflections, very satisfactory heterodyne reception can be established. Equally important, stability is such that this performance can be maintained over long periods of time. There was little noticeable difference between the stability of the heterodyne and of the direct-detector receiver.

Our experiments have shown that heterodyne reception is satisfactory at the end of long enclosed paths. But does it provide any advantages? For an enclosed system there is usually no background light or other interference, and hence no need for discrimination against such radiation. However, because of its excellent discrimination, the heterodyne receiver should provide considerable advantage for systems employing frequency-division or space-division multiplexing of many channels. It is also simpler and more efficient than the direct detector for the reception of frequency-modulated or phase-modulated light signals. In the infrared region heterodyning appears, at least at present, to be the only way of obtaining sufficient gain to bring low-power signals above the detector noise and thermal noise levels. There appears to be no great advantage over the direct detector for the reception of a single channel of amplitude-modulated light in the blue and green spectral regions where photomultipliers, and similar devices, with high quantum efficiencies and large current gains

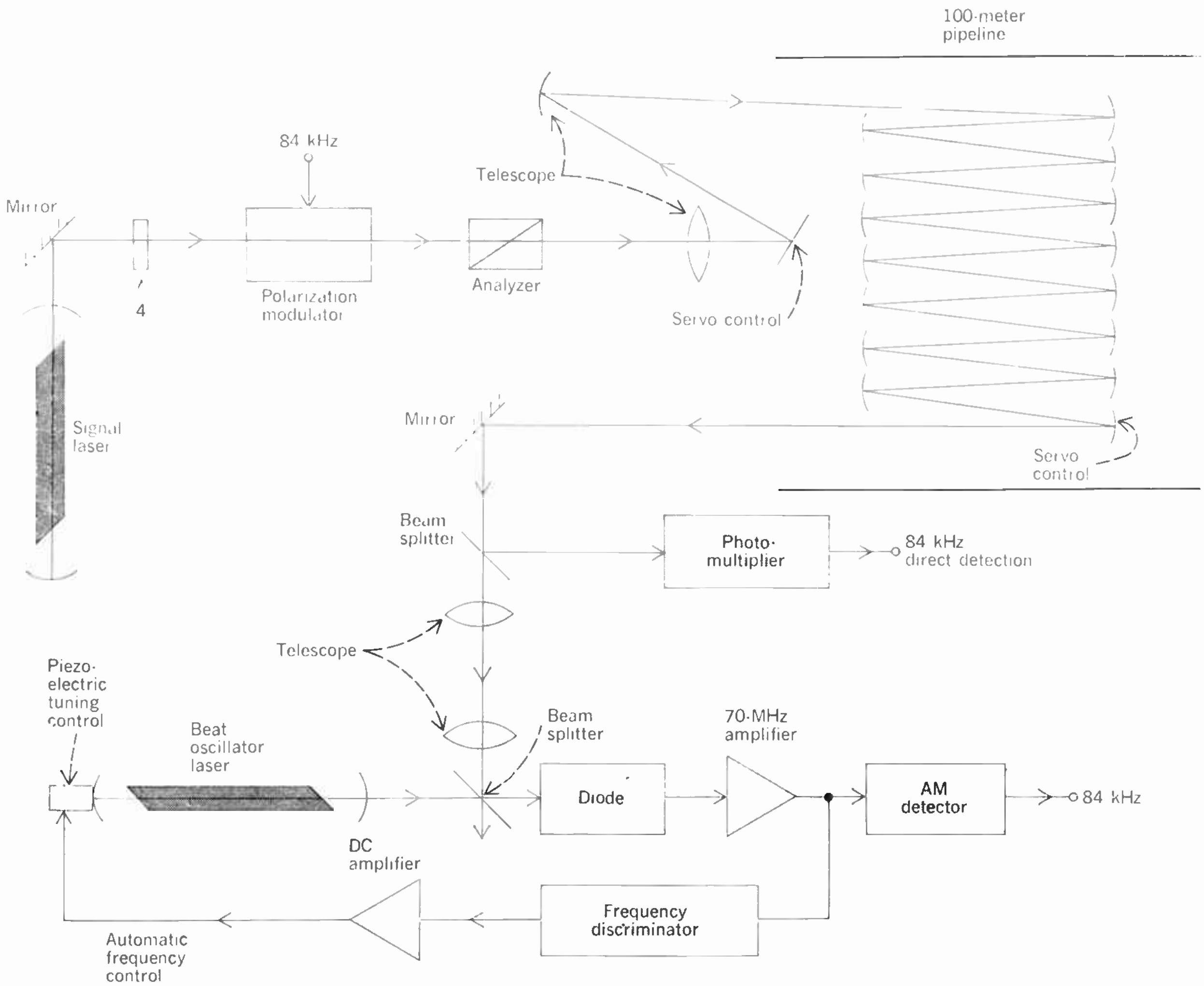


FIGURE 10. Optical heterodyne transmission experiment.

are available. The heterodyne receiver provides less advantage when used to receive higher-power signals.

Conclusions

Both analysis and experiment indicate that rather stringent requirements must be met in order to obtain efficient optical heterodyne detection. There is considerable experimental evidence that these requirements can be met by employing an enclosed transmission path, the so-called optical pipeline, and that from a practical standpoint the difficulties are not much greater than for other types of detection. Such a line with servo control of beam position should provide a very satisfactory transmission medium for any type of receiver. The heterodyne receiver appears to be more satisfactory than the direct detector for the reception of phase- or frequency-modulated light or multiplexed optical signals. At some wavelengths it may provide the only means of overcoming thermal-noise and detector-noise problems. The operation of Doppler radars depends upon the heterodyning process. For most applications, supplying a properly tuned local oscillator presents no great difficulty.

The coherent receiver provides high discrimination against background light. It also provides efficient detection in the infrared region, in which other detectors are deficient. For these reasons and since free space is an ideal transmission medium, the optical heterodyne receiver may have considerable advantage over others for space communication applications.

Because of inhomogeneities in its index of refraction, the atmosphere is a very unsatisfactory transmission medium if heterodyne reception is employed. For such transmissions other methods of detection are usually preferable.

REFERENCES

1. Oliver, B. M., "Thermal and quantum noise," *Proc. IEEE*, vol. 53, pp. 436-454, May 1965.

2. Kogelnik, H., and Li, T., "Laser beams and resonators," *Appl. Opt.*, vol. 5, Oct. 1966.

3. Siegman, A. E., "The antenna properties of optical heterodyne receivers," *Proc. IEEE*, vol. 54, pp. 1350-1356, Oct. 1966.

4. Goldstein, I., Miles, P. A., and Chabot, A., "Heterodyne measurements of light propagated through atmospheric turbulence," *Proc. IEEE*, vol. 53, pp. 1172-1180, Sept. 1965.

5. DeLange, O. E., "Losses suffered by coherent light redirected and refocused many times in an enclosed medium," *Bell System Tech. J.*, vol. 44, pp. 283-302, Feb. 1965.

6. Gloge, D., "Experiments with an underground lens waveguide," *Bell System Tech. J.*, vol. 46, pp. 721-735, Apr. 1967.

7. DeLange, O. E., and Dietrich, A. F., "Optical heterodyne experiments with enclosed transmission paths," *Bell System Tech. J.*, vol 47, pp. 161-178, Feb. 1968.

8. Teich, M. C., "Infrared heterodyne detection," *Proc. IEEE*, vol. 56, pp. 37-46, Jan. 1968.

Complex waves

As a means of describing distributed or continuum linear systems, complex waves have become a persistent, unifying theme in electrical engineering. Moreover, they are expected to become increasingly important in the future

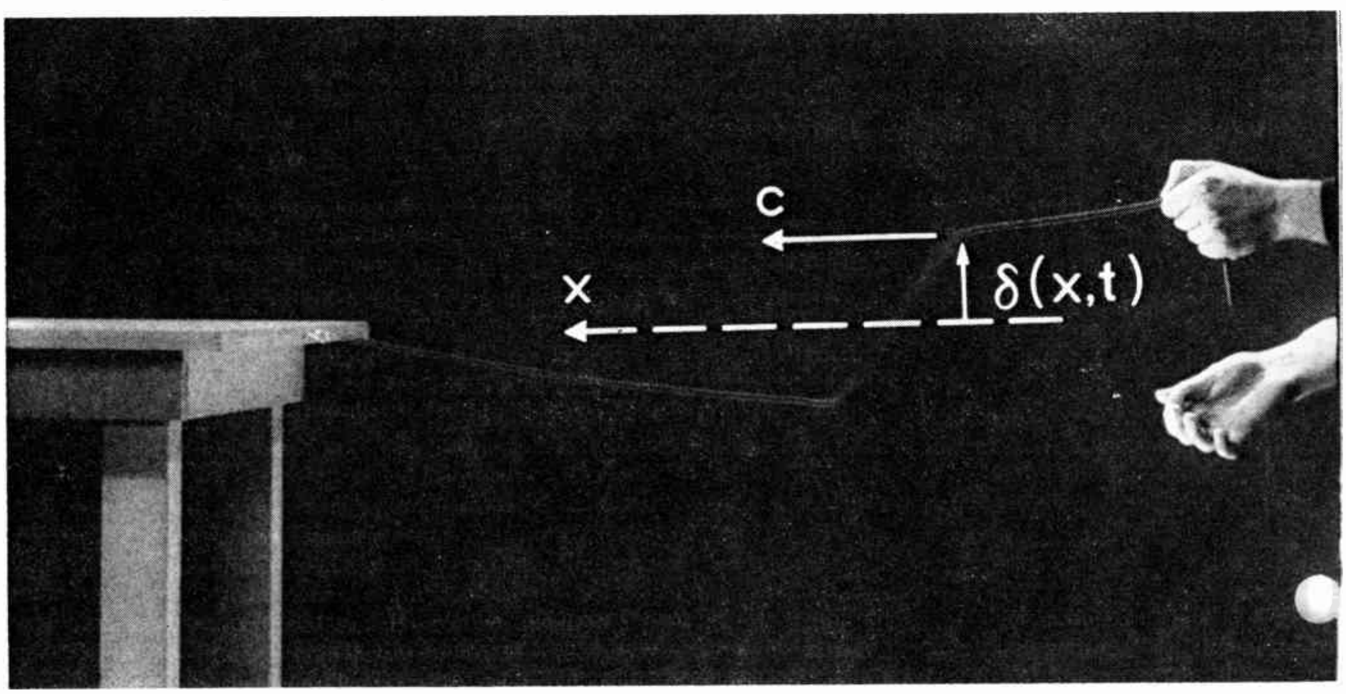
James R. Melcher *Massachusetts Institute of Technology*

Plotting the frequency versus the wave number provides a unifying way of permitting visualization of the linear dynamics of distributed systems. A series of simple examples is presented in order to illustrate the physical significance of ordinary and evanescent waves, absolute instability, waves on moving media, and convective instability or wave amplification. The examples give a rare opportunity to develop a complete physical and mathematical picture of the dynamics, since the presentation can serve as a guide to two films—produced by the Education Development Center for the National Committee on Electrical Engineering Films—in which the dispersion relation evolves by computer animation on the bottom half screen in synchronism with changes in the physical phenomenon in the top half screen.

New areas—born as disciplines interact—continually diversify and redirect engineering interests. During the past decade, engineers concerned with continuum dynamics have turned their attention to gaseous and solid-state plasmas, microwave magnetics, laser technology, and biological communications. A basic understanding of wave dynamics, acquired in earlier fields of interest, often carries over into these fields. For example, close familiarity with electron-beam dynamics serves the engineer who studies plasma dynamics. The electron beam as an electromechanical continuum can be a case study that has lasting usefulness for the engineer; this is possible because he sees it in physical and also in symbolic, or analytical, terms, and can use his synthesis of these views as a starting point in creating useful new devices and models.

For a case study to be useful this way, it requires

FIGURE 1. The "string" as a helically wound spring under tension and supporting a traveling wave initiated at the right.



the simplest possible mathematical model, and the physical situation should be one for which a physical "feel" can easily be established, preferably one in which the dynamics can be visually associated with the analytical model.

In the following discussion, the basic electromechanical continuum is a simple "string" with transverse deflections, which can be readily visualized in time and space. The presentation has the same outline as that of a pair of films,[1] which have been produced for the National Committee on Electrical Engineering Films by the Education Development Center under the sponsorship of the National Science Foundation.* The films are aimed at fusing the simplest mathematical and physical ideas concerned with complex waves, as a basis for gaining insight into more complicated and less physically accessible problems. Typically, the physical phenomenon is shown on the screen, together with a geometric representation of the analytical model (the plot of frequency vs. wave number). Changes in physical parameters such as voltage or current are then reflected both in the phenomena and in the animation of the plots.†

The purpose of the examples can be understood by considering the central role that simple case studies play in understanding lumped parameter systems. The *RLC* circuit represents a case study that the electrical engineer finds useful in working not only with more complicated electric circuits, but also with mechanical or electromechanical systems. The case is useful because it is so simple that a complete picture of the interrelationship of the physics and the mathematics can be retained. In lumped-parameter, time-invariant, linear circuits, the dynamical response is a linear combination of solutions having the form

$$v(t) = \text{Re}\,[\hat{v}e^{st}] \tag{1}$$

where $v(t)$ is a dependent variable, such as a voltage, $\hat{v}$ is a complex amplitude, and s is the complex frequency.

Because of his knowledge of specific examples, a sketch of the system function poles (natural frequencies) in the complex s plane is a geometric representation of the mathematics sufficient to give the electrical engineer a qualitative picture of both the transient and driven responses of a circuit.

In dealing with linear, time-invariant continuum systems, the response is conventionally represented by a linear combination of solutions expressed in the general form

$$\delta(x, t) = \text{Re}\,[\hat{\delta}e^{j(\omega t - kx)}] \tag{2}$$

These solutions are to linear continuum theory what solutions having the form of Eq. (1) are to the theory of linear, lumped-parameter systems.

Like s, the (angular) frequency ω and wave number k can be complex; $\omega = \omega_r + j\omega_i$ and $k = k_r + jk_i$. There is a unity to many areas of electrical engineering and physics brought about because of the common language that results when continuum dynamics are pictured in terms of these *complex waves*. Just as the s plane has become a universal picture of the linear lumped-parameter dynamics, the ω–k plot affords a representation of linear continuum dynamics that transcends the boundaries of any given specialty.

What we shall call "ordinary" waves are those described by the wave equation

$$\frac{\partial^2\delta}{\partial t^2} = c^2\frac{\partial^2\delta}{\partial x^2} \tag{3}$$

where c is a real constant. A $\delta(x, t)$ of the form of Eq. (2) is found, by substitution, to be a solution of Eq. (3) only if

$$\omega = \pm ck \tag{4}$$

When t increases by Δt, the value of $\delta(x, t)$ in Eq. (2) is not changed, provided x changes by an amount Δx such that $\omega\Delta t = k\Delta x$. Therefore, the disturbance represented by δ in Eq. (2) travels with a velocity $\Delta x/\Delta t = \omega/k$ and Eq. (4) shows that $\omega/k = c$. Because c is a constant, waves of all frequencies have the same velocity; all Fourier components of a disturbance travel at the same rate, and therefore any disturbance is propagated without change in its waveform. In other cases, when the relation between ω and k is not linear, the Fourier components of the disturbance travel at different rates, and the disturbance changes shape as it progresses; sinusoidal components of the disturbance disperse. The relation between ω and k is called the "dispersion equation," and the name remains appropriate in our broad context, even though its origin goes back to the special case of refraction of light in glass.

Ordinary waves that are periodic in space are also periodic in time; a real k implies a real ω. By contrast, a dispersive wave might involve a dispersion equation in which complex values of ω are found for real values of k or complex values of k are found for real values of ω; in these cases we refer to the waves as being "complex." In the following examples we will be concerned with extreme cases of dispersion, often encountered when electromagnetic fields interact with matter.

The subject of complex waves is more than a simple extension of circuit theory. A superficial comparison of Eqs. (1) and (2) might lead to the conclusion that the spatial dependence introduced for the continuum system is little more than a complication, handled by the same techniques as those developed for circuit theory. But causality requires that we deal with space and time quite differently, and there are subtle and significant issues at stake when complex waves are used to represent continuum dynamics.[2]

Ten years ago, Sturrock[3] drew attention to these issues through a series of simple examples, which were represented by the same mathematical models we use here. Thoughtfully used, the examples should raise the right questions for further study.

Ordinary waves

The "string" is an elastic model for many physical situations. One of these is shown in Fig. 1, in which a

* These films, along with other films and demonstrations, are used to vitalize an undergraduate course in electromechanics at M.I.T. They are intended for an audience that, prior to viewing, has given some thought to the simple mathematics involved. The text developed for the course is *Electromechanical Dynamics*, by H. H. Woodson and J. R. Melcher, John Wiley & Sons, Inc., New York, 1968, published in three separately bound parts. The fundamental points underlying the examples presented here are treated in Chapters 9 and 10 of *Part II: Fields, Forces, and Motion*.

† These plots were computer-animated by William H. Huggins and William T. Pullin of Johns Hopkins University.

helical spring about a meter long and a centimeter in diameter is placed under a longitudinal tension. The transverse deflections $\delta(x, t)$ are predicted by Eq. (3), which has the physical significance of requiring that the transverse inertial force per unit length on the string be equal to an elastic restoring force per unit length caused by the longitudinal tension. (We have divided by the mass/unit length, so that c^2 is the tension divided by the mass/unit length.)

An experiment as simple as the one shown in Fig. 1

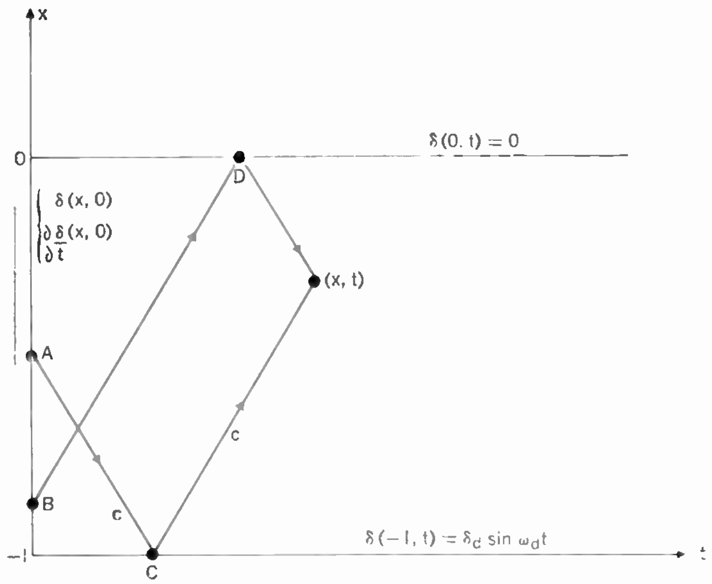
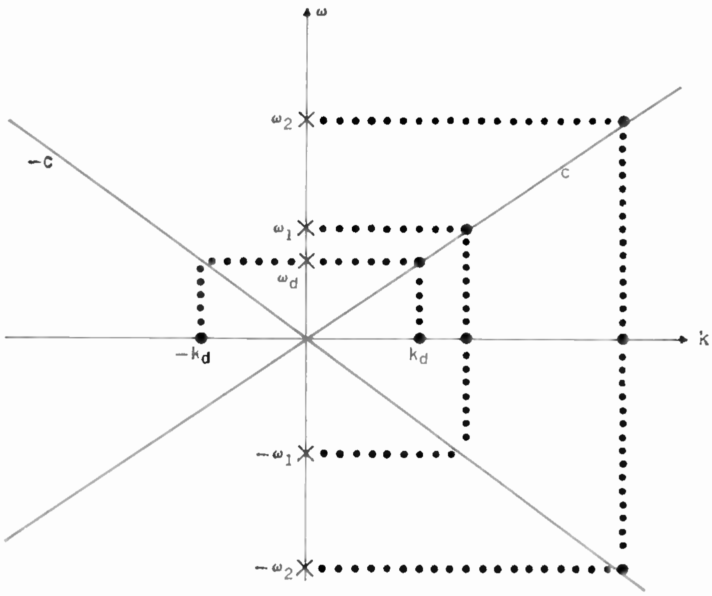


FIGURE 2. Typical boundary and initial conditions in the x-t plane for a stationary medium. For ordinary waves, deflection δ at space-time coordinate (x, t) is fully determined by initial conditions at A and B, and boundary conditions at C and D.

FIGURE 3. Dispersion plot for ordinary waves showing relation of driven and natural frequencies ω_d and $\omega_1, \omega_2, \ldots$ to driven and normal mode wave numbers $k_d, \pi/l, 2\pi/l, \ldots$



demonstrates traveling waves. In mathematical terms, we recognize that solutions to Eq. (3) are of the general form

$$\delta = \delta_+(x - ct) + \delta_-(x + ct) \qquad (5)$$

and these traveling waves are easily seen if the string is plucked. A wave traveling in one direction is reflected by the termination, and thus travels in the opposite direction. To take the simple case of Fig. 1, the string is given one boundary condition at each end. In addition, to determine completely the motions initiated from time $t = 0$, two initial conditions must be given. Thus, the boundary and initial conditions can be pictured as shown in Fig. 2, where the problem reduces formally to finding the solution at some coordinate (x, t).

For ordinary waves, the deflection at (x, t) is determined by the initial conditions at A and B, and by the boundary conditions at C and D. The characteristic lines along which the waves propagate have the slopes $\pm c$. A formalization of this approach to finding the dynamics in terms of the x-t coordinates is analogous to using convolution techniques in lumped-parameter systems. For the continuum cases, we would be concerned with Green's functions and the theory of characteristics.[4] As with circuit theory, we can take the alternative approach of using Fourier series and transforms with the time and space dependence respectively represented by ω and k. It is this latter approach that is most often used in *linear* dispersive systems.[2]

The key to picturing the dynamics in terms of (ω, k) is the dispersion equation, which for ordinary waves is Eq. (4), graphically represented in Fig. 3. For illustrative purposes, consider the response of the string to a driving deflection at $x = -l$ (the right end). This drive is turned on when $t = 0$, and is sinusoidal with frequency $\omega = \omega_d$. The familiar notion of dividing the response into two parts—one due to the drive and one due to the initial conditions—leads to a solution of the form

$$\delta(x, t) = -\delta_d \frac{\sin k_d x}{\sin k_d l} \sin \omega_d t$$
$$+ \sum_{n=1}^{\infty} (A_n^+ e^{j\omega_n t} + A_n^- e^{-j\omega_n t}) \sin k_n x \qquad (6)$$

The first term is the driven response, a superposition of two waves having the form of Eq. (2), with Eq. (3) satisfied because ω and k are related by the dispersion equation (4). Thus,

$$k_d = \frac{\omega_d}{c} \qquad (7)$$

Figure 3 illustrates this relationship in graphical form.

The additional term of Eq. (6), which takes the form of a Fourier series, is a superposition of the natural modes. These modes must also satisfy the dispersion equation so that the wave equation is satisfied everywhere on the string; in addition (in order not to disturb the boundary conditions at the ends of the string, which are satisfied by the driven response), they must be zero at $x = -l$ and $x = 0$. Hence, the "transient response" (response to the initial conditions) is a superposition of waves, of the form of Eq. (2), having discrete wave numbers

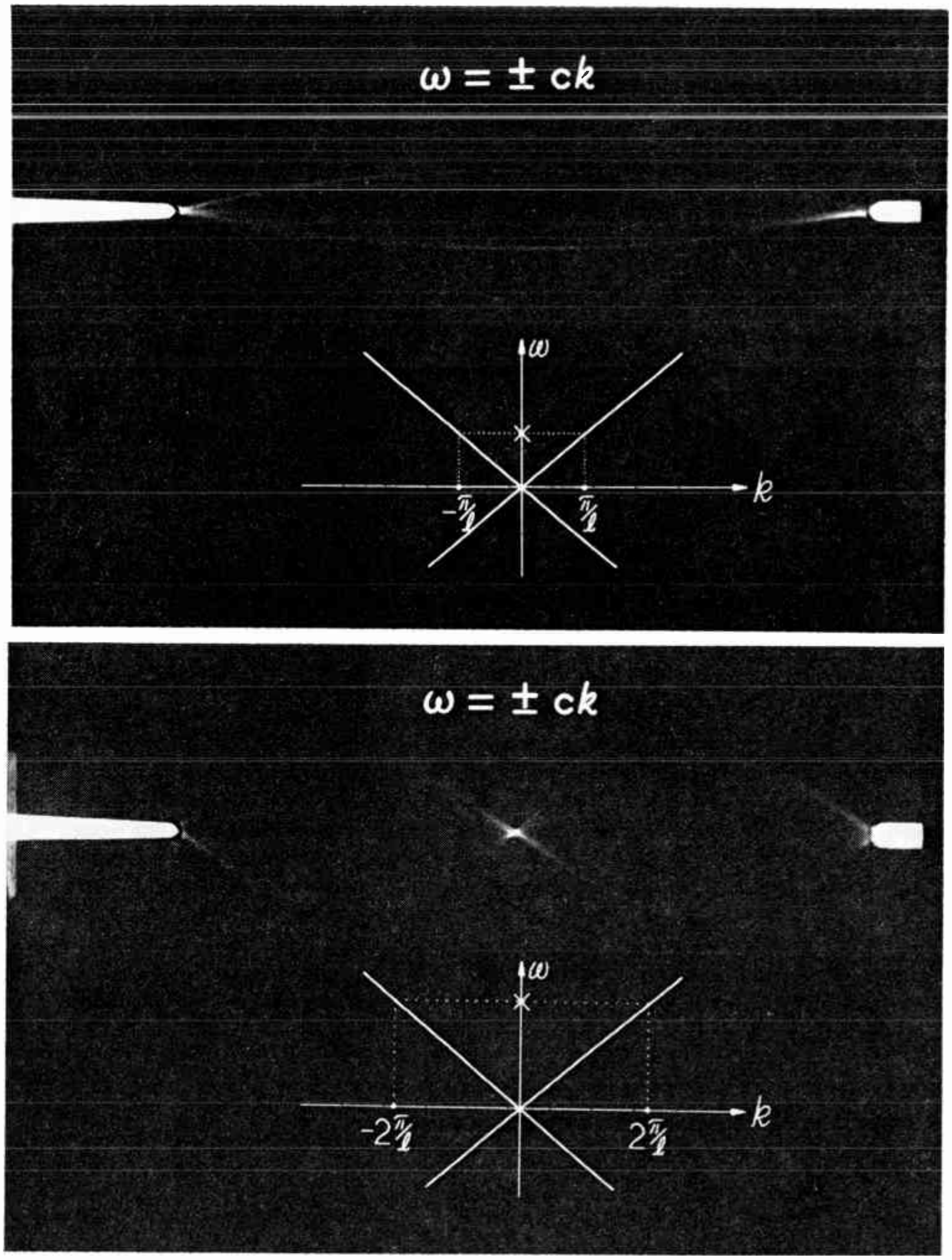


FIGURE 4. Ordinary standing waves resulting as the driving frequency is tuned to the neighborhood of the two lowest natural frequencies and resonances appear in the response.

$$k = k_n = \frac{n\pi}{l} \tag{8}$$

and, therefore, to satisfy the dispersion equation, having real discrete natural frequencies

$$\omega = \pm\omega_n \qquad \omega_n = ck_n \tag{9}$$

The Fourier coefficients A_n^+ and A_n^- are determined by the particular initial conditions, and if we had included damping in the model, the consequences of these conditions would die out with time (the natural frequencies would have positive imaginary parts). The remaining driven response still reflects these natural frequencies, however, because as ω_d approaches one of the natural frequencies ω_n, the denominator of Eq. (6) becomes small and hence the deflection becomes large. Thus, as we tune the frequency ω_d upward along the ω axis, resonances appear in the response in the neighborhoods of the natural frequencies. (Of course, once again the damping that actually exists in the experiment prevents an infinite response as $\omega_d = \omega_n$. In terms of lumped parameters, the s-plane poles are just to the left of the imaginary axis.)

The relationship between these "standing wave" resonances and the ω-k plots is shown in Fig. 4. These pictures, showing the wave envelope, characterize the case of ordinary waves, in which k is real, a periodic drive leads to a spatially periodic wave, the natural frequencies and hence resonance frequencies appear as harmonics, and the wavelength $\lambda = 2\pi/k_r$ is inversely proportional to the frequency.

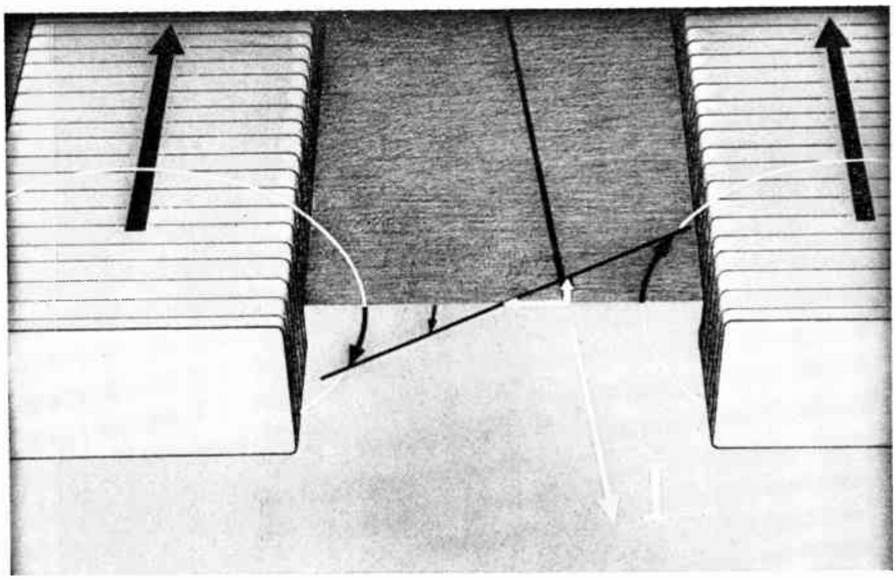


FIGURE 5. Magnet coils, shown in cross section to left and right, carry currents in the same direction, as represented by left and right arrows. Thus there is a line of symmetry between the coils, where magnetic fields induced by these currents cancel, as shown. String, in the form of a conducting spring, has an equilibrium position along this line of symmetry and carries current I. Thus, when deflected from the line of symmetry in plane of coils, it assumes appearance shown in cross section. The deflection leads to a magnetic restoring force per unit length tending to return the string to its equilibrium position. At low frequencies, waves on the string are evanescent.

Evanescent waves

The phenomenon of spatial decay without dissipation, or wave evanescence, is illustrated by making the string carry a longitudinal current I and placing it in a magnetic field, as shown in Fig. 5. The field coils are arranged so that as the string is deflected from its equilibrium position, it moves into a region of magnetic flux density **B** that depends on the direction and magnitude of the deflection δ. Thus, there is an $\mathbf{I} \times \mathbf{B}$ restoring force per unit length on each element of the string (similar to that of a distributed spring) tending to restore the string to its equilibrium position.

The electromechanical coupling requires that a force per unit length, which is proportional to the string current I and deflection amplitude δ as well as to the gradient in magnetic flux density transverse to the equilibrium axis of the string, be added to the equation of motion.

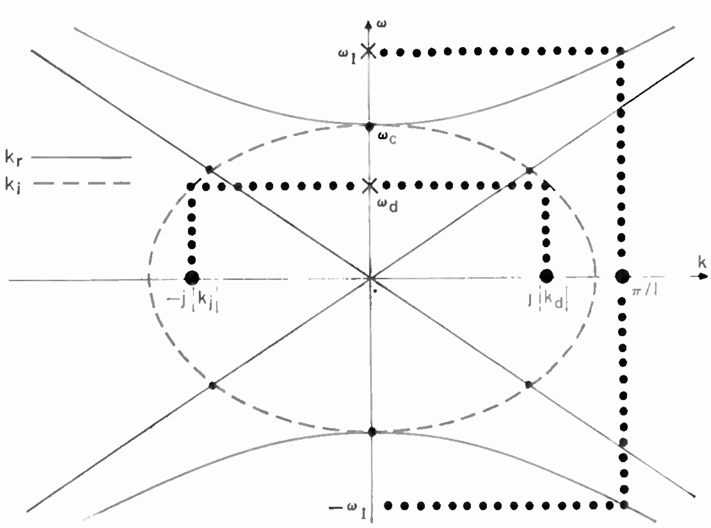


FIGURE 6. Dispersion equation for evanescent or cutoff waves, showing complex k for real ω.

The current and the flux-density gradient are absorbed in the parameter ω_c^2, and Eq. (3) becomes

$$\frac{\partial^2 \delta}{\partial t^2} = c^2 \frac{\partial^2 \delta}{\partial x^2} - \omega_c^2 \delta \qquad (10)$$

The dispersion equation, which follows from a substitution of Eq. (2) into Eq. (10), is

$$\omega^2 = c^2 k^2 + \omega_c^2 \qquad (11)$$

All of the ω–k relations discussed here will take the geometric form of quadratic functions. Thus, the hyperbolas shown in Fig. 6 represent the values of ω given by Eq. (11) for real values of k.

The response once again takes the form of Eq. (6). Thus, the transient response is again a superposition of waves having real discrete wave numbers $k = n\pi/l$. The natural frequencies ω_n are found from Eq. (11) using these discrete k's and can be pictured on the ω–k plot as shown in Fig. 6. Remember that raising the string current increases ω_c. As the current is raised the hyperbola ω-axis intercepts move away from the origin, and the ω_n's are increased. That is, increasing the current (ω_c^2) has the effect of increasing the natural frequencies. Note that they are no longer harmonics, but have the same spatial appearance as ordinary waves.

For the driven part of the response, we are interested in the values of k implied by a given real value of $\omega = \omega_d$. For $\omega_d < \omega_c$, Eq. (11) gives imaginary values of k which have the geometric form of an ellipse, also shown in Fig. 6. In this range of frequencies, the waves are evanescent (cut off) and the driven response of Eq. (6) assumes the form

$$\delta = -\delta_d \frac{\sinh |k_i| x}{\sinh |k_i| l} \sin \omega_d t \qquad (12)$$

At each point, the string deflections have the same temporal phase but bow inward rather than outward.

The relationship of the ω–k plot to the experiment is shown in Fig. 7. In the sequence of pictures the driving frequency remains fixed near the lowest resonance fre-

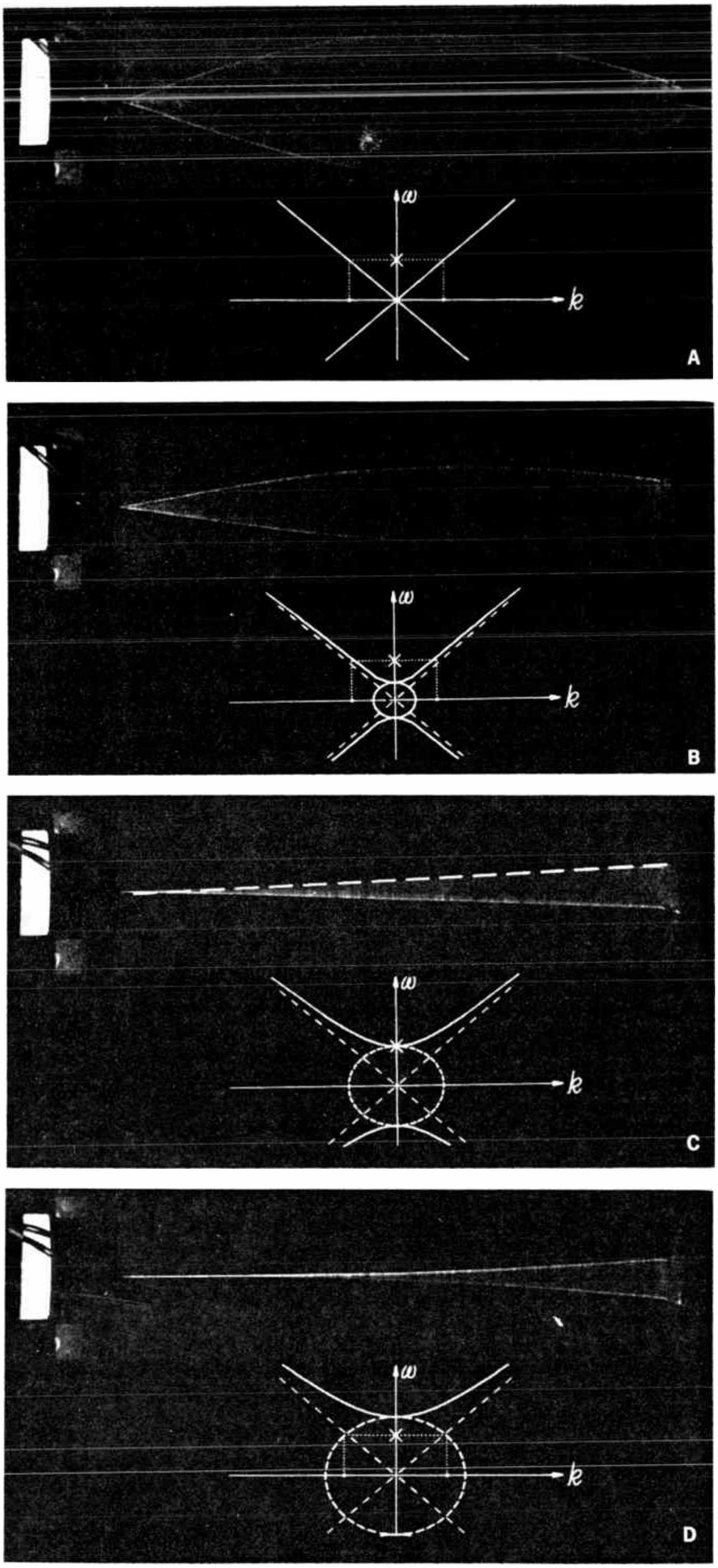


FIGURE 7. Demonstration of evanescent waves. With the frequency and amplitude of the drive at the right fixed, the current is raised. A: I = 0, to give ordinary waves nearly in resonance. B: I > 0, but ordinary waves still propagate. C: Current I adjusted to give cutoff. D: Current I greater than required for cutoff, hence an evanescent wave.

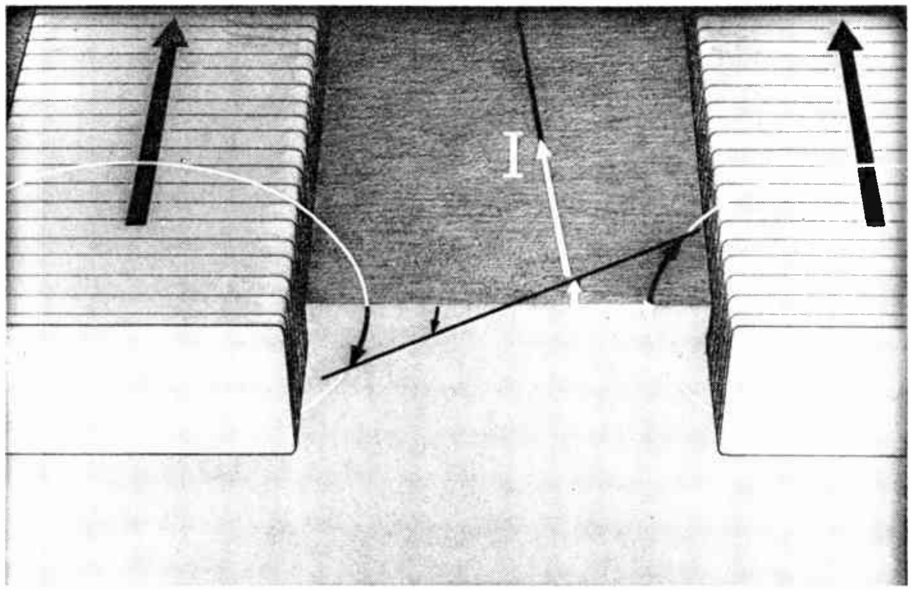
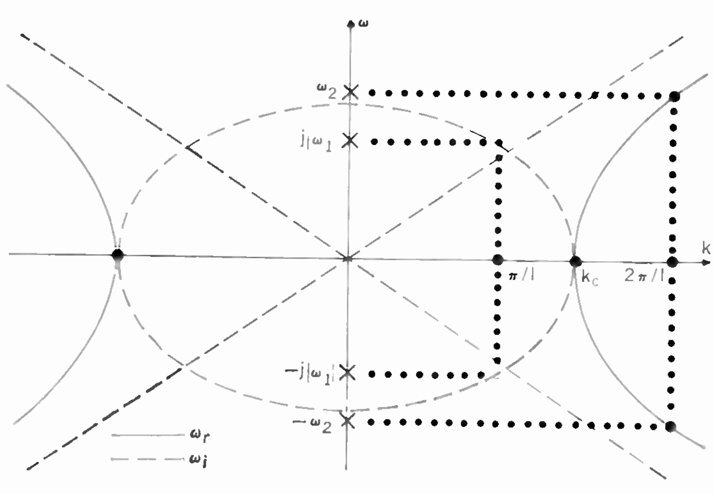


FIGURE 8. String in magnetic field as shown in Fig. 5, except current I is reversed.

quency in the absence of a current ($\omega_c = 0$). Thus, in Fig. 7(A) the ω–k plot and the deflections have the familiar appearance of ordinary standing waves: k is real. As the current is raised, the ellipse and hyperbola expand outward, k approaches zero, and the deflections of Fig. 7(C) have a linear dependence on x. At this cutoff point the inertial effects are exactly canceled by the magnetic restoring force, and the string behaves as it would with ordinary waves at nearly zero frequency. With a further increase in current [Fig. 7(D)] the string bows inward, k is purely imaginary, and the waves decay away from the point of excitation.

It is important to recognize that the positive imaginary wave number given by the dispersion equation as the waves are cut off does not represent spatial growth, but rather represents a wave reflected by the fixed end of the string to decay in the $-x$ direction. This is in contrast to a spatially amplifying wave, in which complex values of k are also found for real values of ω. Before we can present a case of this type, it is necessary to appreciate those classes of continuum instability in which the growth is temporal.

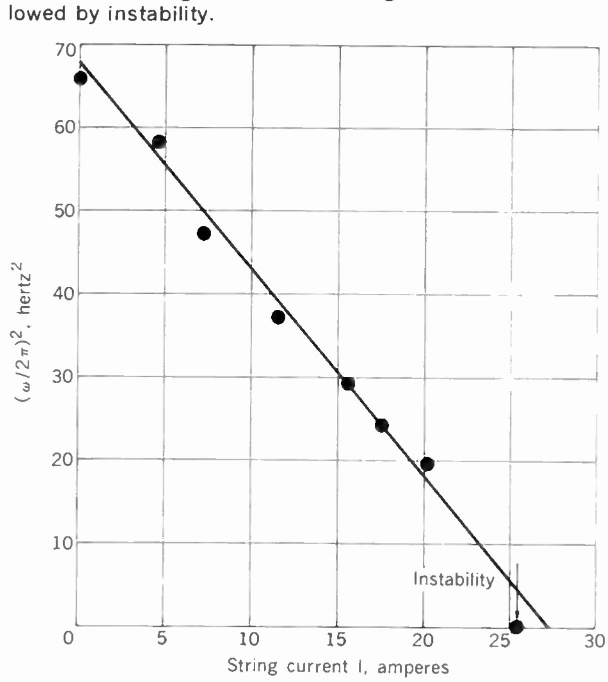
FIGURE 9. Dispersion equation for absolute instability, showing complex ω for real k.



Absolute instability

If the string current I is reversed in the experiment of Fig. 5, the magnetic force per unit length is still proportional to the deflection but it tends to carry the string away from its equilibrium position, as illustrated in Fig. 8. The magnetic interaction has the effect of a distributed spring, but now the spring constant is negative. We expect that, physically, as the current is raised the

FIGURE 10. Lowest resonance frequency squared as a function of string current I, showing downward shift followed by instability.



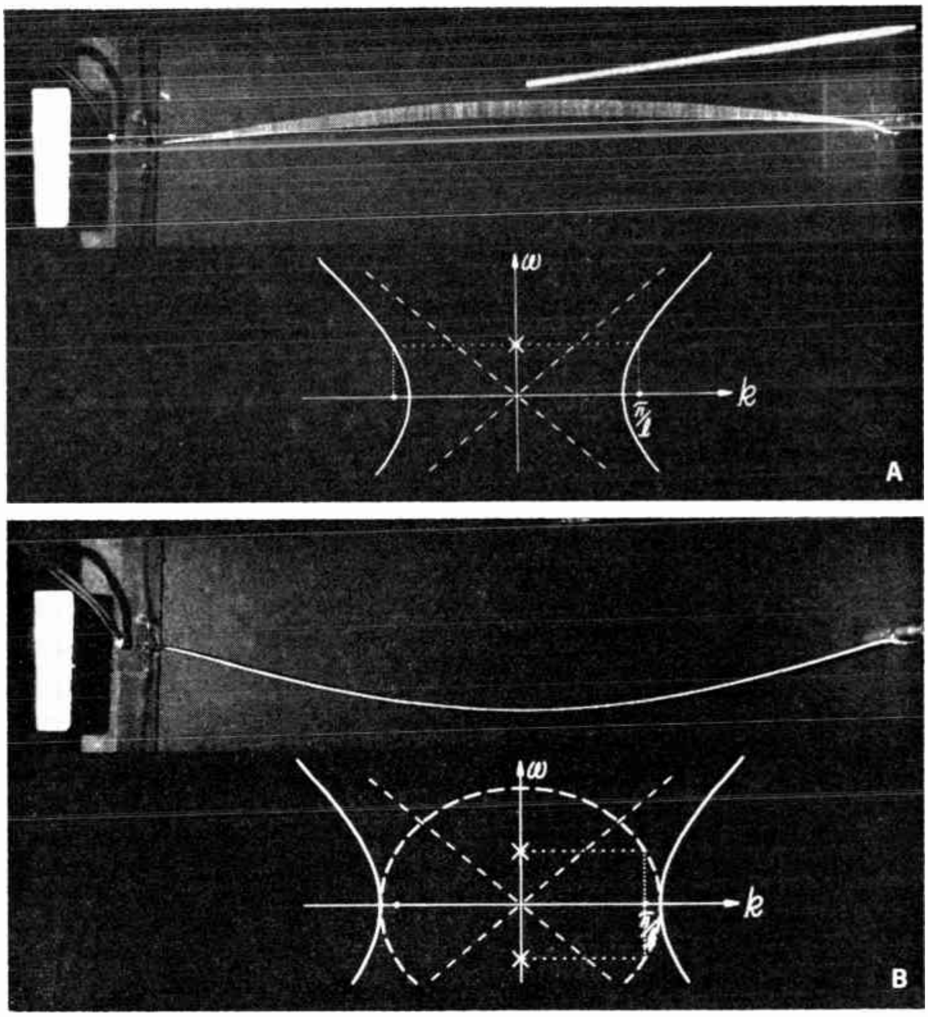


FIGURE 11. Experiment of Fig. 5, but with current as in Fig. 8. A—A small current I reduces the natural frequency observed as the string is plucked. B—Current greater than required for instability; no matter how carefully the string is centered, as it is released it bows to one side or the other.

point will be reached at which equilibrium of the string is unstable. The simple model and experiment allow us to form a physical and mathematical picture of how this happens.

The equation of motion, with the current reversed, is conveniently written as

$$\frac{\partial^2\delta}{\partial t^2} = c^2\left[\frac{\partial^2\delta}{\partial x^2} + k_c^2\delta\right] \tag{13}$$

where k_c^2 is proportional to the string current and to the gradient in magnetic flux density. (We have replaced $\omega_c^2 \rightarrow -c^2k_c^2$.) The dispersion equation follows as

$$\omega^2 = c^2(k^2 - k_c^2) \tag{14}$$

so that now the ω–k plot is as shown in Fig. 9. By contrast with Fig. 6, there are now only real values of k for real values of ω, the hyperbola passing through the k axis at $\pm k_c$.

Once again, we consider the significance of the dispersion equation for the driven and transient parts of the response [Eq. (6)]. First, note that regardless of the driving frequency, the driven response has the same general appearance as for ordinary waves: the string bows outward and the wave numbers are real. However, the resonance frequencies, which reflect the natural frequencies, are now shifted downward by increasing the string current (k_c^2).

We can conduct an experiment in which the lowest resonance frequency is measured as a function of string current. According to Eq. (14), the resulting plot of frequency squared as a function of I has the form of the solid line shown in Fig. 10. The data are from the actual experiment. In terms of Fig. 9, as the current is raised the hyperbolas expand outward (that is, the intercepts $\pm k_c$ move outward). Eventually, the point is reached at which $k_c = \pi/l$ and the lowest resonance frequency has reached zero. As far as the driven response is concerned, we can consider still greater values of k_c, but it is essential to recognize that the lowest natural frequencies measured with our frequency shift experiment are now imaginary.

According to Eq. (14), imaginary values of ω are represented by the ellipse of Fig. 9. For the case shown in Fig. 9, where k_c exceeds π/l, the transient solution will eventually dominate the driven response, because initial noise establishes the component, having a negative imaginary frequency, which grows exponentially with time.

The experiment illustrates this point in Fig. 11. With the current below the threshold for instability, the plucked string vibrates in its lowest natural mode, though with a natural frequency reduced by increasing the current. With the current beyond the threshold, no

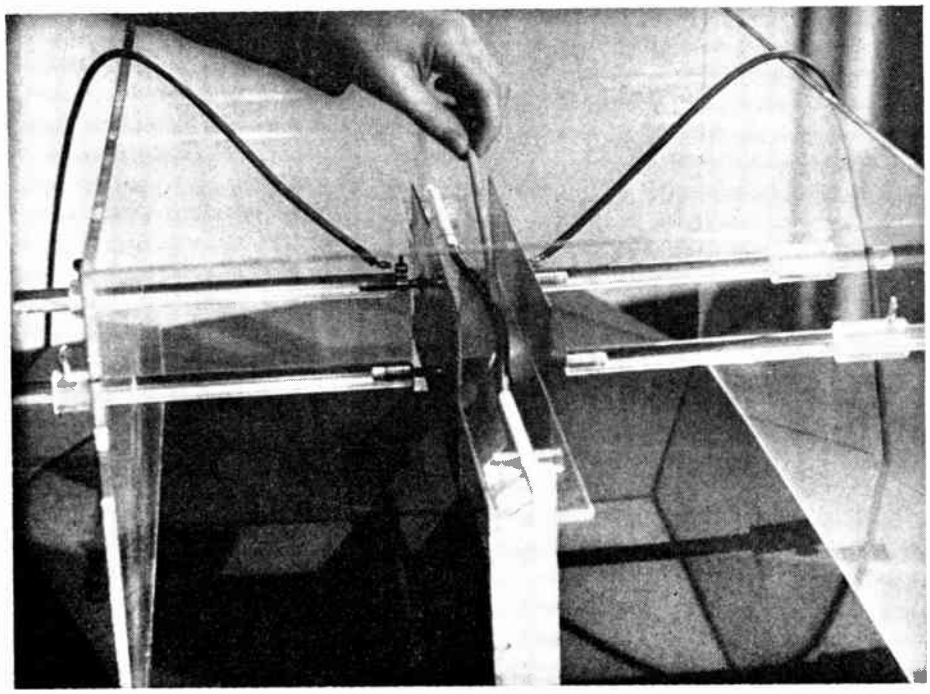


FIGURE 12. Conducting spring, constrained to execute horizontal vibrations between plane, parallel electrodes.

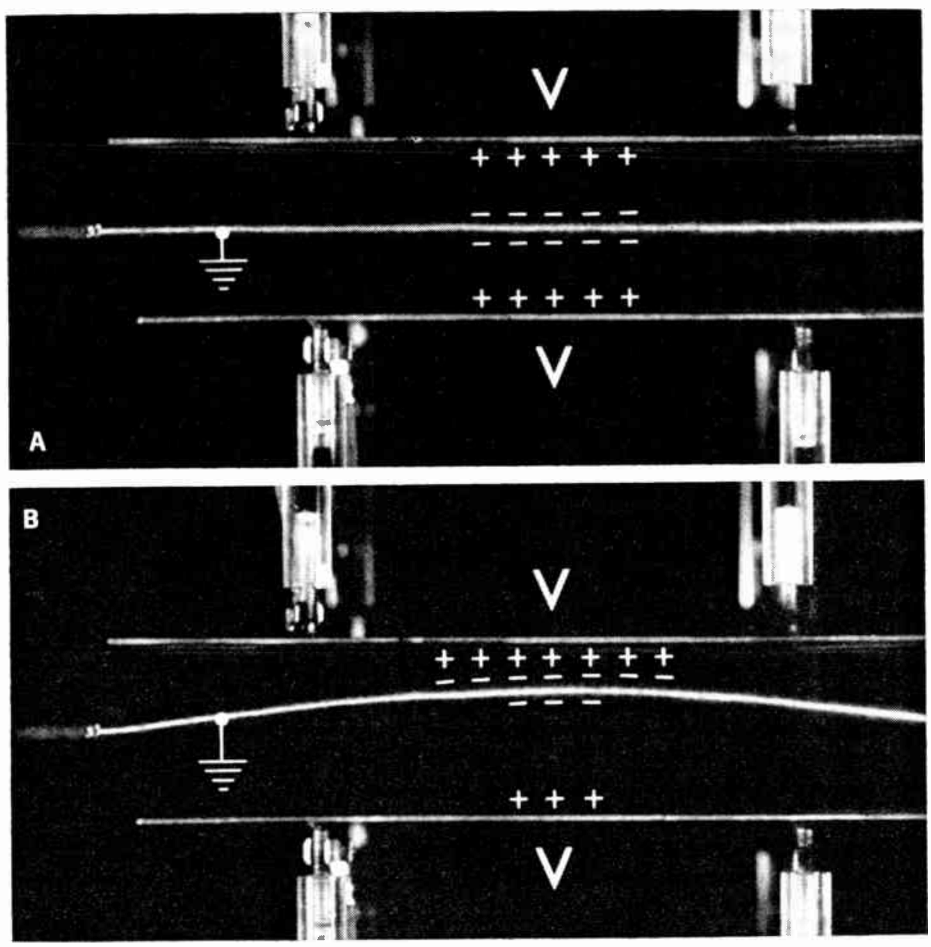


FIGURE 13. Top view of Fig. 12, showing potential constraints. A—Static equilibrium with net force due to charges zero. B—Charges induced by deflection.

matter how carefully the string is released from a centered position it deflects to one side or the other. Note that the wave number of the deflection is $k = \pi/l$, as predicted by the theory. Of course, the amplitude of the unstable deflection is limited in reality by nonlinear effects not accounted for by the simple model.

We can use electric, rather than magnetic, forces on the string to create a dynamical situation identical to the one just considered. In Fig. 12 the string is shown centered between plane, parallel electrodes. These electrodes each have the constant potential V with respect to the string, which is grounded. Thus, with the string centered, surface charges are induced as shown in Fig. 13(A), and the forces of electrical attraction between unlike charges cancel; there is no net transverse force on the string. However, because the string-electrode potential difference remains constant, a slight deflection leads to an accumulation of charge, hence an attendent transverse force which tends to further increase the deflection as illustrated in Fig. 13(B). For small deflections, the electric force has the same linear dependence on the deflection amplitude δ as for the case of the destabilizing magnetic force.

The equation of motion and dispersion equation again take the form of Eqs. (13) and (14): now k_c is proportional to V. As the voltage V is raised, the ellipse and hyperbola of the ω–k plot expand outward and the static equilibrium becomes unstable as k_c passes through the critical value π/l. At each position x, the string deflection increases until the field intensity increases to the point where electrical breakdown occurs, as shown in Fig. 14.

The temporal nature of this instability and its relationship to the driven response are similar to familiar situations in lumped-parameter systems. Now, however, we deal with an infinite number of natural modes, each having its own spatial dependence.

The next two sections will illustrate that, in cases involving a moving continuum, lumped-parameter concepts are not so immediately applicable. First, we study the effect of motion on ordinary waves. Then we combine the effect of the motion with the electrically induced instability of this section to create an instability that grows spatially, but not temporally.

Ordinary waves on moving media

We now replace the spring of Figs. 12–14 with a jet of water, as shown in Fig. 15. Transverse or kinking motions of the jet can be modeled as those of a string having a longitudinal velocity U in the x direction.[5] The string tension in the jet is supplied by surface tension; the water, like the metal spring of the previous section, can be considered as perfectly conducting. The parallel-plate electrodes will be used in the next section.

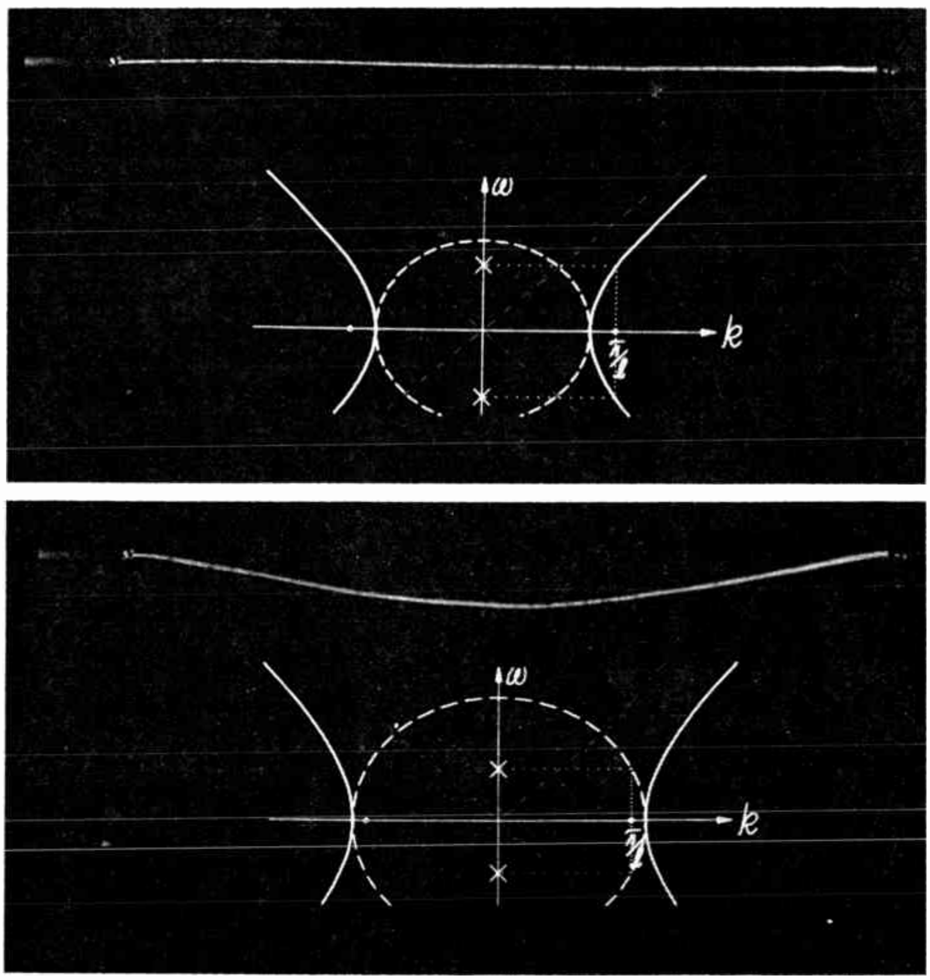


FIGURE 14. As the voltage V is raised in the experiment of Figs. 12 and 13, the static equilibrium becomes unstable just as the ellipse intercepts the k axis at $k = \pm\pi/l$.

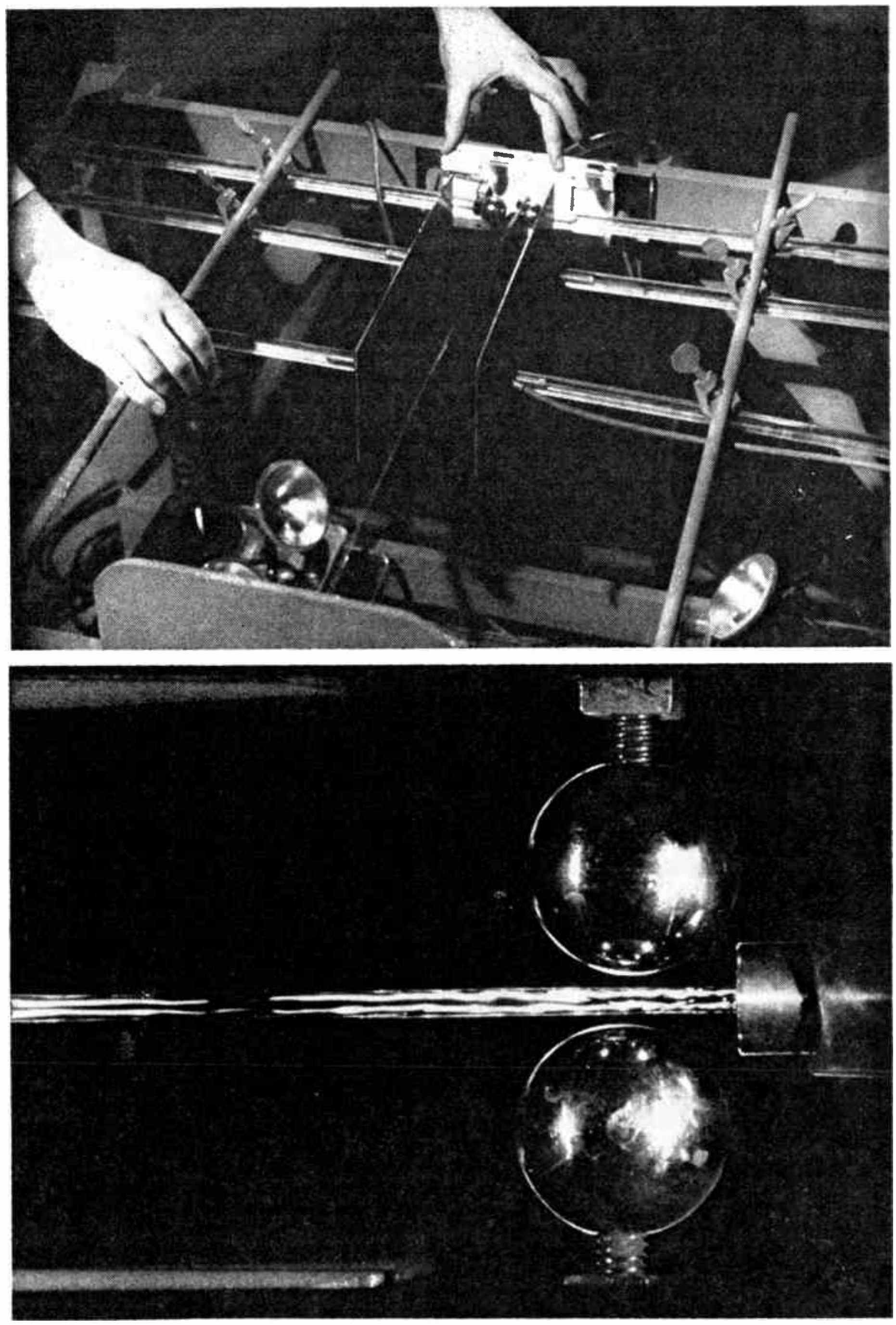


FIGURE 15. The string of Fig. 12 is replaced by a jet or water issuing from a nozzle adjacent to the exciter spheres. Kink motions of the jet are modeled as those of a conducting moving string with tension supplied by surface tension. A transverse sinusoidal driving force is produced by the spheres, which do not touch the stream.

much as they were in Figs. 12–14. The spheres, shown in Fig. 15, are driven in "push-pull" by biased sinusoidal driving potentials having the frequency ω_d, to produce a transverse sinusoidal excitation at the upstream end of the jet. First consider the effect of longitudinal motion on ordinary waves (the electrodes of Fig. 15 are grounded). In a frame of reference moving with the same longitudinal velocity U as the jet, transverse deflections would be described by the wave equation (3):

$$\frac{\partial^2\delta}{\partial t'^2} = c^2\frac{\partial^2\delta}{\partial x'^2} \qquad (15)$$

Here, (x', t') are the position and time in the moving reference frame. The spatial derivative is computed holding t' constant, but since $t' = t$, we can replace

$$\frac{\partial}{\partial x'} \rightarrow \frac{\partial}{\partial x}$$

By contrast, the time rate of change is computed holding x' constant, and because $x' = x - Ut$, we must write $\partial\delta/\partial t'$ in the fixed coordinates as

$$\left[\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right]\delta$$

Thus, in the fixed-coordinate system, Eq. (15) becomes

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)^2 \delta = c^2\frac{\partial^2\delta}{\partial x^2} \qquad (16)$$

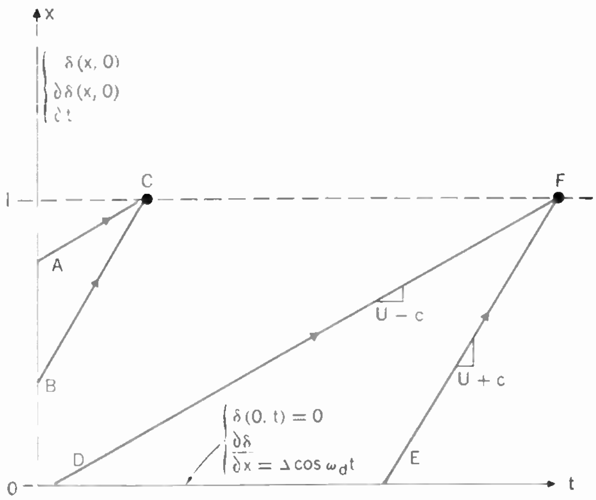
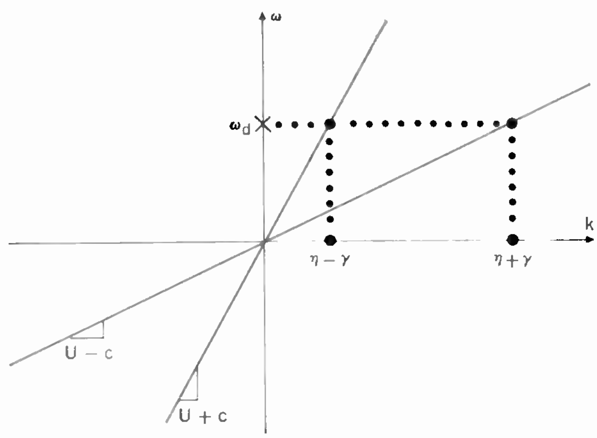


FIGURE 16. Typical boundary and initial conditions in the x–t plane for "supersonic" moving medium (U > c). Upstream driving conditions fully determine response at F.

FIGURE 17. Dispersion equation for ordinary waves on a string having longitudinal velocity U.



The convective second derivative on the left side of this equation is the transverse acceleration of a segment of the string that is traveling in the x direction with a velocity U.

The traveling waves, which in the absence of convection propagate in both directions with a velocity c, must now buck the streaming velocity U. Substitution into Eq. (16) shows that solutions now take the form

$$\delta = \delta_+[x - (U + c)t] + \delta_-[x - (U - c)t] \quad (17)$$

The most dramatic effect of the convection is evident if $U > c$, for then both waves propagate in the $+x$ direction—that is, downstream. In the x–t plane of Fig. 16, at the downstream position $x = l$, the response is fully determined by upstream conditions. Thus, for a point such as C, when $t < l/(U + c)$, the response is fully determined by initial conditions at upstream positions A and B. At a point such as F [when $t > l/(U - c)$ and the slow wave, which tries to buck the stream, has arrived from the origin], the response is determined by the driving conditions.

We emphasize the ω–k approach to describing the response. Yet the manner in which the response is established, as portrayed in the x–t plane, is essential to making appropriate use of the dispersion equation. We could again write the solution as the superposition of a transient and a driven response that satisfies boundary conditions to the left and right of a point of interest. In terms of the x–t plane, it is clear that this procedure leads to nonsense, because upstream conditions fully determine the response, which would be overspecified by a further downstream boundary condition. Implicit to this reasoning is the condition that the past affects the future, and not the reverse. Whereas with $U < c$ we could impose only one boundary condition at $x = 0$, we now can impose both a zero deflection and a sinusoidally varying slope, as illustrated in Fig. 16.

Consider the response at the downstream position $x = l$ to this sinusoidal excitation, which is turned on when $t = 0$. If the stream is initially undisturbed (zero initial conditions), there is no response at $x = l$ until the fast wave arrives. More important, regardless of the initial conditions, the sinusoidal steady state is fully established when the slow wave arrives from the origin as $t = l/(U - c)$. This behavior is in sharp contrast to the $U = 0$ case of Fig. 2, where the damping is the only means by which the transient response, as a remnant of the initial conditions, can disappear. Most important are the implications of these observations for the case in which the string is unstable, for with $U < c$ the transient response dominates the driven response, and the amplitude becomes unbounded with time at any fixed position x. On the supersonic string, $U > c$, this is not possible because after $t = l/(U - c)$, there is no effect at $x = l$ from the initial conditions. The string responds only to the drive at $x = 0$, and if that drive is in the sinusoidal steady state, the response also will be.

In summary, with $U > c$, the transient response is of limited duration; there is no danger that the driven response will be dominated by the transient as time passes, and hence the driven response is of primary interest. Instability in the sense of the previous section is not possible.

Substitution of Eq. (2) into Eq. (16) gives the dispersion equation

$$(\omega - kU)^2 = c^2k^2 \quad (18)$$

which has the graphical form of two straight lines, as shown in Fig. 17. The response to the real frequency ω_d is a superposition of spatially periodic waves with wavelengths $2\pi/(\eta - \gamma)$ and $2\pi/(\eta + \gamma)$, where

$$\gamma = \frac{\omega_d c}{U^2 - c^2} \qquad \eta = \frac{U\gamma}{c} \quad (19)$$

The particular linear combination of these waves that satisfies the given driving conditions can be written in the form

$$\delta(x, t) = \frac{\Delta}{\gamma} \cos(\omega_d t - \eta x) \sin \gamma x \quad (20)$$

Thus the waves interfere spatially to form a stationary sinusoidal envelope or "beats."

The effect of raising the frequency is shown in Fig. 18, where the jet is illuminated both stroboscopically, to see the instantaneous deflection, and with continuous light, to show the envelope; on the jet U is about 20

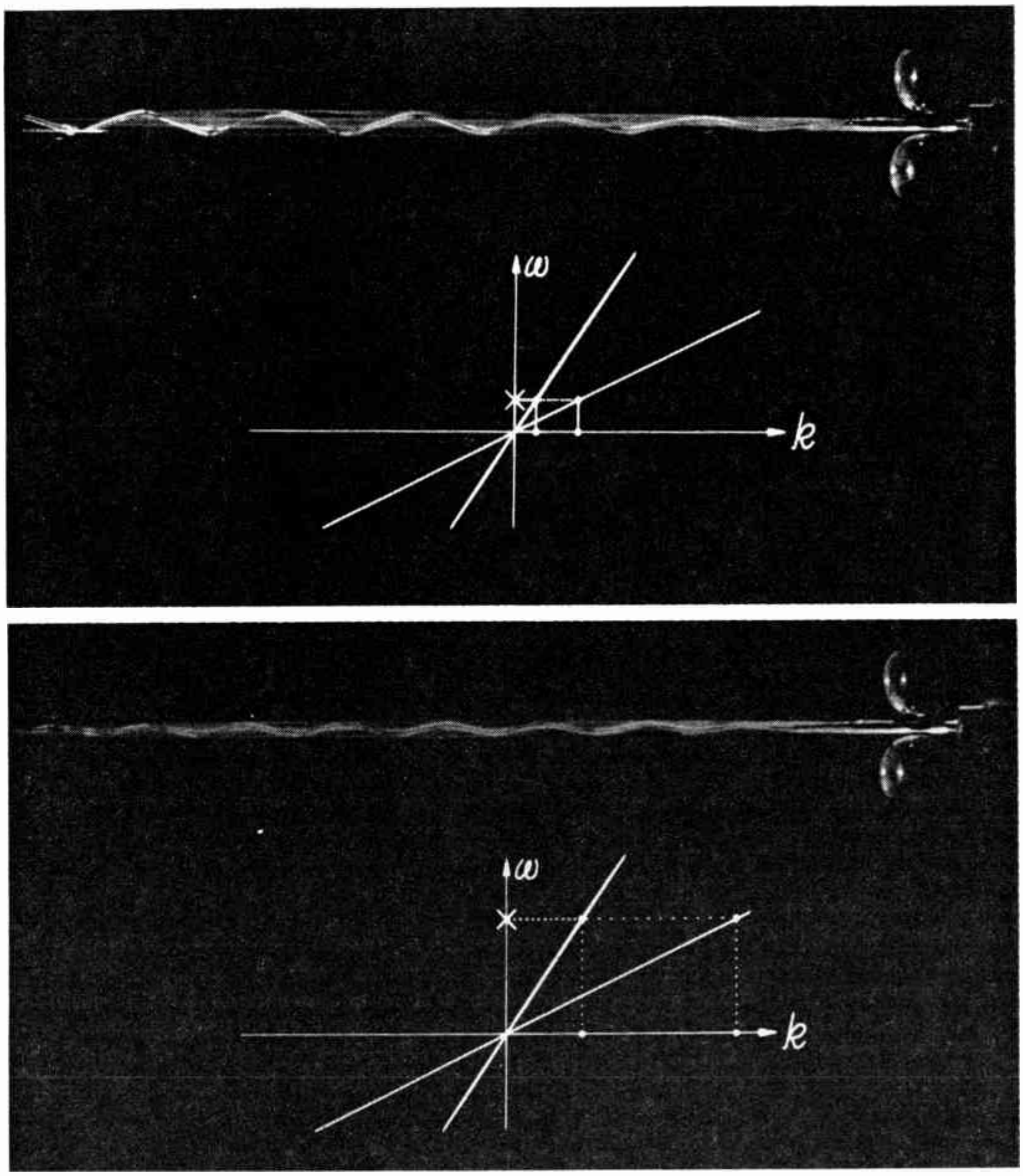


FIGURE 18. Moving string in form of a water jet with U ≈ 20c. Top—Sinusoidal steady-state response is excited electrically by the spheres at the right, and illumination is combined steady and stroboscopic light, with the latter synchronized to the excitation. Bottom—Here the frequency is sufficiently high that the envelope of the first beat can be seen.

times c. As indicated by Eqs. (19) and (20), raising the frequency shortens the distance between points of zero phase and between beat nulls. In the first picture the deflections appear to grow because only the front edge of the beat is visible. As the frequency is raised, the first beat null comes on screen at the left.

Wave amplification

We are now in a position to bring together the effects of material convection and instability by adding the electric force demonstrated in the experiment of Figs. 12–14 to the string in motion. This is done by considering the consequences of a potential V applied to the parallel-plate electrodes of Fig. 15. The mathematical model is then Eq. (16) with the addition of the electric-force term (again k_c is proportional to V):

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)^2 \delta = c^2\left(\frac{\partial^2\delta}{\partial x^2} + k_c^2\delta\right) \qquad (21)$$

It follows that the dispersion equation is

$$(\omega - kU)^2 = c^2(k^2 - k_c^2) \qquad (22)$$

This expression is quadratic in both ω and k, so it can be easily solved either for ω as a function of k or vice versa. Solution for ω shows that even with $U > c$, if k is real and less than k_c, there will be frequencies having negative imaginary parts. Still, this result does not imply that the jet deflections are unstable in the sense that they grow with time at a given longitudinal position. On the contrary, the sinusoidal steady state is all that remains at a position $x = l$ after $t = l/(U - c)$. We are primarily interested in the driven response to the sinus-

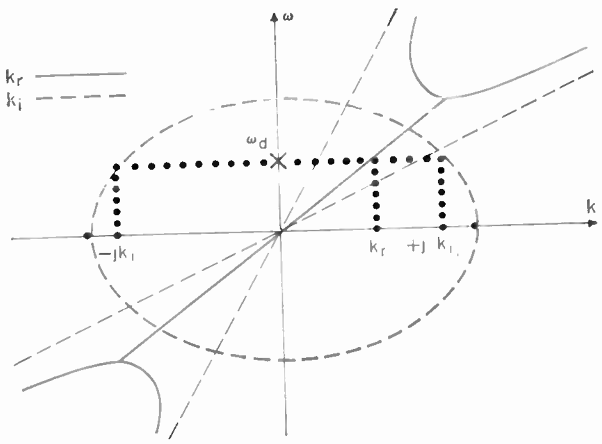


FIGURE 19. Dispersion equation for amplifying wave showing complex k for real ω.

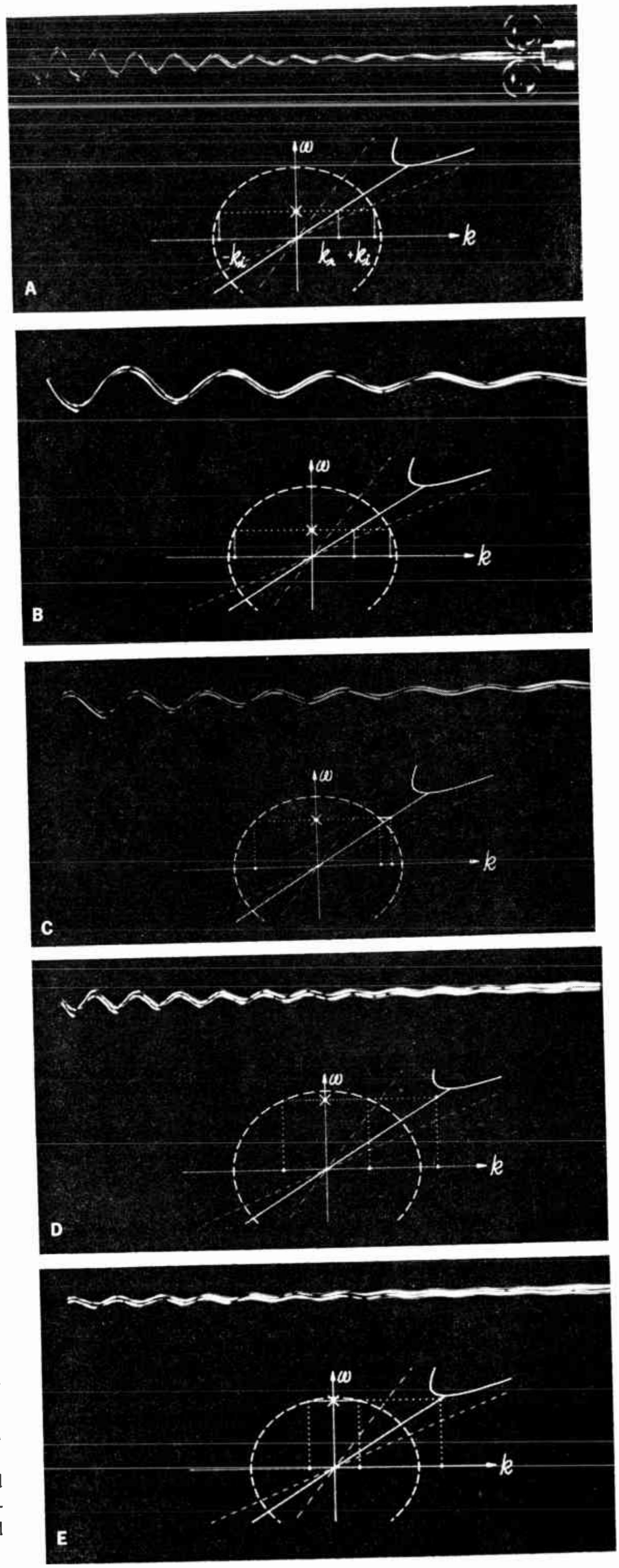


FIGURE 20 (right). Amplifying wave on the moving string excited in the sinusoidal steady state. A—Overall appearance of the experiment of Fig. 15, as viewed vertically. B–E—Closeup of stream showing decrease in wavelength and spatial rate of growth as frequency is raised.

oidal driving function—that is, in the wave numbers given by Eq. (22) for the real frequency ω_d.

The appropriate plot for finding the driven response shows complex values of k for real ω as sketched in Fig. 19. For the indicated driving frequency ω_d, there are two complex conjugate values of $k = k_r \pm j|k_i|$, where

$$k_r = \frac{\omega_d U}{U^2 - c^2}$$

$$|k_i| = c\frac{\sqrt{(U^2 - c^2)k_c^2 - \omega_d^2}}{U^2 - c^2} \qquad (23)$$

An appropriate combination of these waves to satisfy the two driving conditions of Fig. 16 can be written as

$$\delta(x, t) = \frac{\Delta}{|k_i|} \sinh |k_i| x \cos(\omega_d t - k_r x) \qquad (24)$$

We are now in a position to make observations concerning these amplifying waves, either in terms of this analytical expression or in terms of the experiment shown in Fig. 20. With the jet illuminated by a stroboscopic light synchronized with the sinusoidal excitation, the instantaneous deflections appear to stand still. As the driving frequency ω_d is raised, the dispersion equation and pictures show that the wavelength $2\pi/k_r$ becomes shorter. At the same time the spatial growth rate, which is proportional to k_i, becomes smaller. As the frequency approaches $(\sqrt{U^2 - c^2})k_c$, the imaginary parts of the wave numbers approach zero and the waves no longer exhibit an exponential spatial growth; see Fig. 20(E). A further increase in frequency would lead to the spatial beats of the preceding section.

The case of the amplifying wave should be compared to that of the evanescent wave, since both have dispersion equations that exhibit complex values of k for real values of ω.

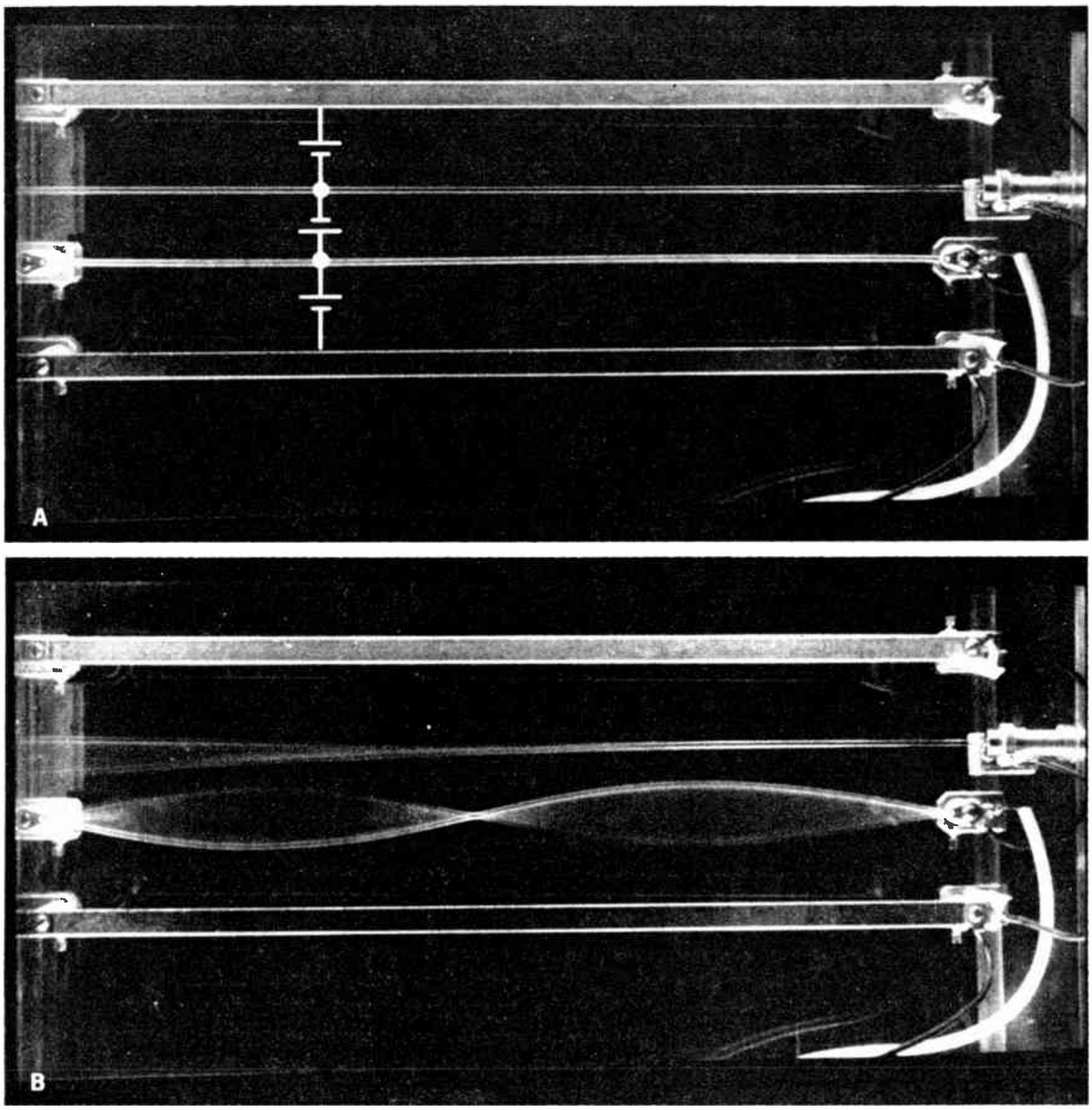


FIGURE 21. The stream interacting with a fixed string. A—From top to bottom, a rigid electrode, the stream issuing from nozzle at right, a string fixed at the right and left, and a second fixed electrode. Throughout the motions these four elements retain the potentials shown. B—Spontaneous oscillatory instability similar to that found in an improperly terminated traveling-wave tube, with the stream playing the role of the electron beam and the string playing the role of the slow-wave structure.

Applications

The simple string, interacting with magnetic and electric fields, has been presented as a basic physical model for a case study concerned with complex waves. The format of the films from which the pictures were taken parallels that used here. The series of examples makes possible a close tie between the ω–k diagram, as it represents the analytical model, and easily visualized and appreciated physical situations. But our objectives are not met unless we make the "generalized string" a nucleation point (like the *RLC* network), from which an understanding of more practical situations can grow. Hence, it is important that we connect these examples with practical applications from the various areas indicated in the introduction. Here, we can give only a small indication of the wealth of possible connections that can be made.

After reviewing pertinent aspects of ordinary waves, we considered the consequences of adding a magnetic transverse restoring force, which effectively "squashed" the driven response to excitations below the cutoff frequency. The resulting evanescent or cutoff waves are an example of one of the most common types of complex waves. Two- and three-dimensional structures used to guide elastic, acoustic, and electromagnetic waves commonly exhibit cutoff modes.[6-8] Other examples that, like our case study, involve waves in only one dimension are found in plasmas[9] and magnetized materials.[10]

Probably the most widely publicized cases involving continuum instabilities are the fusion machines, where the instability is a primary limitation on achieving hydromagnetic equilibria.[11] Our case study has many of the attributes of Rayleigh–Taylor types of instability found in plasmas[12] and fluids[13,14] as they interact with fields. As a parameter is varied to make the system unstable, instability impends at zero frequency. This is sometimes referred to as "static instability." Electrical engineers have the habit of making use of instabilities, especially those that are oscillatory in nature. These dynamic instabilities or overstabilities are the basis for making oscillators from electron beams, as will be illustrated shortly.

The electron beam confined by a magnetic field is analogous in its longitudinal space charge oscillations to our moving string.[15] To make the analogy complete, we require that the string tension be zero, and that there be a restoring force such as that demonstrated in the section on evanescent waves. However, the spatial beats described in the section on ordinary waves with convection are found not only on electron beams, but in supersonic flow of a gas[16] or supercritical flow of rivers,[17] and in fibers, paper, etc. There are examples from almost every area of engineering that require the understanding of the interrelationships exemplified here between material motion, wave propagation, boundary conditions consistent with causality, and the driven and transient response.

One of the most dramatic effects of material motion is its influence on an instability, a point emphasized here by first considering a case in which the instability grew in time at a fixed point in space, and then adding the effect of material motion in such a way that the growth appeared in space. The resulting amplifying wave can be driven, and is similar to responses found in crossed-field sheet electron-beam interactions (diochotron, magnetron, etc.).[18,19] Of course, amplifying waves do not exist entirely in the electrical engineering discipline; they are found as well in streaming fluids,[20] aeroelastic structures,[21] and semiconductors.[22]

Complex waves are of great variety. We have begun here with the simplest systems, modeled by second-order partial differential equations and quadratic ω–k relations. Our simple string model also allows us to demonstrate damped (and, in particular, diffusion) waves and even to show simple forms of resistive wall instability, where in the absence of motion waves are damped, but with motion they amplify. The art of electron-beam engineering, with its traveling-wave tubes and backward-wave oscillators, is a reminder that electrical engineers have made practical use of coupled systems represented by dispersion equations of higher order than quadratic. In the traveling-wave tube, space-charge waves on the streaming electron beam interact with the fields traveling on a fixed electromagnetic transmission line.[23] We can put together situations we have considered here to give a graphic impression of how such an interaction can be used to make an oscillator. Our moving string, in the form of the water jet, is coupled to a fixed string (a spring) by means of an electric field, as shown in Fig. 21. The fixed and moving strings taken individually have essentially the same physical nature as discussed in the sections on instability and wave amplification. Now, however, there is an electric field between the moving and the fixed strings (which play roles analogous to those of the electron beam and the transmission line), and a transverse motion of either results in an incremental force on the other. The result is the spontaneous oscillation shown in Fig. 21(B). This oscillation derives its energy from the kinetic energy of the water jet, and in slow motion gives a lucid physical demonstration of traveling-wave interactions.[24]

These examples should raise fundamental questions concerned with complex waves.[2] For instance, how can the dispersion equation be used to discriminate between amplifying and evanescent waves when both involve complex values of k for real values of ω?

Few are aware of the kind of support required in a project of the type summarized here. Those who are include the members of the Advisory Committee: Profs. John G. Brainerd, James H. Mulligan, Jr., Robert M. Saunders, Louis D. Smullin, and Herbert H. Woodson. Technical assistance and advice were given by Profs. Hermann A. Haus, Frederick D. Ketterer, and Charles D. Hendricks. The computer animation assumes an essential role in the presentation, and was carried out by Prof. William H. Huggins and Mr. William T. Pullin.

REFERENCES

1. *Complex Waves I: Propagation, Evanescence and Instability* (26 minutes), and *Complex Waves II: Instability, Convection and Amplification* (23 minutes), films obtainable from the Education Development Center, 39 Chapel Street, Newton, Mass. 02160.

2. Briggs, R. J., *Electron-Stream Interaction with Plasmas.* Cambridge, Mass.: M.I.T. Press, 1964, p. 33.

3. Sturrock, P. A., "Kinematics of growing waves," *Phys. Rev.*, vol. 112, no. 5, pp. 1488–1503, 1958.

4. Courant, R., and Hilbert, D., *Methods of Mathematical Physics.* New York: Interscience, 1962, pp. 449–461.

5. Melcher, J. R., *Field-Coupled Surface Waves.* Cambridge, Mass.: M.I.T. Press, 1963, pp. 123–135.

6. Mason, W. P., *Physical Acoustics*, vol. I, pt. A. New York: Academic Press, 1963, pp. 111–164.

7. Rayleigh, J. W. S., *The Theory of Sound*, vol. II. New York: Dover, 1945, pp. 73–74.

8. Adler, R. B., Chu, L. J., and Fano, R. M., *Electromagnetic Energy Transmission and Radiation.* New York: Wiley, 1960, pp. 369–378.

9. Stix, T. H., *The Theory of Plasma Waves.* New York: McGraw-Hill, 1962, p. 13.

10. Johnson, C. C., *Field and Wave Electrodynamics.* New York: McGraw-Hill, 1965, pp. 422–424.

11. Glasston, S., and Lovberg, R. H., *Controlled Thermonuclear Reactions.* Princeton, N.J.: Van Nostrand, 1960, pp. 484–486.

12. Rose, D. J., and Clark, M., Jr., *Plasmas and Controlled Fusion.* New York: Wiley, 1961, p. 258.

13. Chandrasekhar, S., *Hydrodynamic and Hydromagnetic Stability.* New York: Oxford, 1961, p. 428.

14. Melcher, J. R., and Hurwitz, M., "Gradient stabilization of electrohydrodynamically oriented liquids," *J. Spacecraft Rockets*, vol. 4, pp. 864–881, July 1967.

15. Johnson, C. C., *Field and Wave Electrodynamics.* New York: McGraw-Hill, 1965, p. 277.

16. Shapiro, A. H., *The Dynamics and Thermodynamics of Compressible Fluid Flow*, vol. I. New York: Ronald, 1953, pp. 596–597.

17. Stoker, J. J., *Water Waves.* New York: Interscience, 1956, pp. 25–26.

18. Clauser, F. H., *Symposium of Plasma Dynamics.* Reading, Mass.; Addison-Wesley, 1960, pp. 99–101.

19. Neil, V. K., and Heckrotte, W., "Relation between diochotron and negative mass instabilities," *J. Appl. Phys.*, vol. 36, pp. 2761–2766, Sept. 1965.

20. An impressive example of fluid dynamic amplifying waves is demonstrated in the film *Flow Visualization*, by S. J. Kline, produced by the Education Development Center, Newton, Mass., for the National Committee for Fluid Mechanics Films.

21. Dugundji, J., and Ghareeb, N., "Pure bending flutter of a swept wing in a high-density low-speed flow," *AIAA J.*, vol. 3, pp. 1126–1133, June 1965.

22. For a series of articles devoted to the semiconductor instabilities, see *IEEE Trans. Electron Devices*, vol. ED-13, Jan. 1966.

23. Pierce, J. R., *Traveling Wave Tubes.* Princeton, N.J.: Van Nostrand, 1950, chap. 4.

24. Ketterer, F. D., and Melcher, J. R., "Stream-structure instabilities," Rept. CSR-TR-66-16, M.I.T. Center for Space Research, Aug. 1966; also *Phys. Fluids*, in publication.

Scanning the issues

That Gulf Again. Do people say about you, when they are apparently complimenting you, that in your technical work you are conscientious, brutally hard-working, very serious, scrupulous? To other people, these same observers may be saying that you are a hard man, difficult to deal with, stubborn, rigid, a loner, a nit-picker. And work evaluation reports about you may contain such phrases as introvert, does it all himself, never takes a vacation, has no deputy, never trained a replacement.

These pretty categories of description, though they are applied to people in all walks of life, may, in the case of the technologist, stem in part from the character of his professional training. A technologist who is extremely competent in his technical work but who is discovering that he is having problems in his dealings with other people may be making the tacit assumption that the sophisticated technological models and scientific logic he has acquired during his education are continuous into sociological situations. The assumption may be completely unwarranted. This question of the disparity between the technologist's "magnificent ability to function on the technical plane and his amateur efforts in the second world of people" is discussed with some wisdom and a generous spirit in an article by John J. Dougherty.

Dougherty notes that the mental model that the technologist has of the physical world is more modern than was Einstein's. As a "systems analyst" he optimizes *theoretical* systems using the fantastic resources of the computer. His work with the basic elements of the atom or the construction of the basic protein molecule uncovers the most fundamental secrets of life itself.

Yet this same man, five years later, may use a "system" for supervising, administering, or managing his own organization, that is often *archaic*, *outmoded*, *barbaric*, *benighted*, or, as he might say, suboptimized against the boundary of his limited insight.

Although the author rightly does not pretend to support his conclusion on a "scientific basis," the conclusion he argues seems inescapable: The outlook (the models, constructs, language, and analytic techniques) given to the technologist in his extensive technical education, although thoroughly useful in his world of science and technology, are completely inadequate substitutes for the models and semantics of another world—namely, the world of people.

For technologists who are experiencing such "people" troubles, Dougherty's prescription (which is by no means as simplistic as we shall baldly state here) is the undertaking of a *second* education, an education dealing with the limitations of Aristotelian logic and "reality," the findings of modern behavioral science, and dealing with the nature of man. Such an education is an endeavor to complement with more sophisticated insights what may prove to be a "limited personal maturity." Such a prescription of human growth, it should be clear, is no less than a lifelong endeavor. Its relevance, considering our society's present stage of technological development, goes far beyond the personal level. For there are many symptoms, some of which Dougherty cites, that the relation of technologists with people and with society at large is defective. In short, technology has made troubles for people, and technologists must find a new initiative and sense of responsibility in dealing with those troubles.

In this context, Dougherty quotes from a source, exhorting technology to assert itself, that is worth requoting here: "Objective leadership means, in part, that engineering renew the posture it has traditionally and successfully assumed in the capitalistic age. It must interpret for society, in the language of society, the objectives (and accomplishments) of science.

"But if engineering is to take on this task—and for the sake of social stability it seems imperative that it do so—then engineering and its organizations must be more responsive. Engineering must adopt a macroscopic rather than a microscopic view of technology—a view of the whole rather than its parts—the long-range dream instead of the short-range objective. Engineers must look at life rather than matter. They must evolve, without loss of objectivity, a better understanding of the social, political, and economic forces that mold the lives of men, if they are to bridge successfully the gulf between science and humanity."

We have all heard a great deal about that gulf over the years, but it bears repeating in view of the special kinds of troubles that technology has unleashed upon men. However it is described, the dissociation in one's self of one's knowledge and analytical skills from one's physical, emotional ground can mean troubles far beyond one's personal perimeter.

In the light of what technology has wrought upon society, the goals of technology are changing in radical ways. These new goals, which take into their purview the reality of the nature of people, demand a new growth and new characteristics on the part of engineers, not just an intensification or "more of the same." Engineers must become a different kind of people, both as individuals and as a group. One might call it an even deeper professionalism. The prospect is worth pondering. (John J. Dougherty, "The Two Worlds of the Technologist," *IEEE Trans. on Engineering Management*, March 1968.)

Innovating Innovation. The most recent issue of IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT also deserves attention. It is a special issue—based on a seminar—on management of new developments. Its contents may give many readers who are seeking a broader perspective some pause.

In "The Innovation of Innovation," J. A. Morton discusses how management's role has evolved from that of a programmer of an inanimate, already existing system to that of a system engineer concerned with the process of innovation. This process, he says, is the perception, creation, and transformation of relevant science into new and improved products and services. He thus suggests the view that a manager is the selective agent of change. Morton treats, among other things, the deep changes going on in the whole electronics industry stemming from the development of integrated circuits.

In "Venture Activities in the Large Corporation," J. Hillier discusses the problems involved in getting a product from research into production in a large company where everyone is busy in his own short-range activity.

In "Patents and the Stimulation of Innovation," which contains a sharp criticism of the recent report of the Presidential Commission on U.S. Patent Laws by R. H. Rines, we hear strong reverberations of the argument that technologists must broaden their sights. As to problems relating to the patent system, Rines states: Our Congress currently wants help. It wants to know what you, the technologists, need in terms of law or changes in laws that are going to stimulate you, aid you in your problems of innovation, and assist all of the myriad of individual inventors, and small and large innovative companies.

The message I bring to you, Rines goes on, is appalling. It is that Congress is met by substantial silence from you. There is silence on the part of the engineering, the scientific, the university, and the business societies and organizations, who are in the center of this field, but who decline to study the problem or participate in the related debate. Yet these are the groups best qualified to assist in the formulating of constructive concepts and ideas. This silence exists, Rines declares, despite the fact that the laws that will be formed to govern innovation, the protection of industrial property, and the relationship between you and the United States Government or others with whom you contract, will underlie, plague, and control your own daily activities.

Accordingly, Rines issues a call to the technological community to wake up, and to broaden its sights and its knowledge in the realm of the economic, social, and political world. This community, he states, must assume the degree of leadership in the nation that the technological era demands.

In "Xerography—A Single Idea Transforms a Company," J. H. Dessauer tells the fascinating story of how that now-universal process was invented and brought to its present great success. In the end, one major reason for its success, Dessauer declares, is that the significant people involved in the product development process had the faith that they were right, and they persevered through the scientific, technical, and financial problems until they proved the validity of their convictions.

In "A Product Line for a Government-Oriented R & D Company," K. J. Germeshausen discusses a question that many government-contract supported companies face, namely, how to develop a commercial product base to supplement or supplant their governmental activities.

For those who may at first tend to pooh-pooh the notion of studying the innovation of innovation, the chairman of the seminar, R. C. Levine, offers a classic example of how difficult it is to identify at an early stage the really good innovative ideas. He cites how industrialist Chauncey Depew, president of the Telegraph Company, was offered the purchase of the telephone patent by Hubbard and Bell in 1876 for a nominal sum. Depew consulted his technical experts who rejected the "utterly unreasonable" device in a document that is priceless from our present vantage. Among other points, the technical experts then said: "Messrs. Hubbard and Bell want to install one of their 'Telephone' devices in virtually *every* home and business establishment in the city. This idea is idiotic on the face of it. Furthermore, why would any person want to use this ungainly and impractical device when he can send a messenger to the local telegraph office and have a clear written message sent to any large city in the United States?"

Moreover, the experts advised, "Hubbard's fanciful predictions, while they sound very rosy, are based upon wild-eyed imagination and a lack of understanding of the technical and economic facts of the situation, and a posture of ignoring the obvious technical limitations of his device, which is hardly more than a toy, or a laboratory curiosity."

But the point is, as Levine brings out, actually putting a wire into virtually every home in America *is* an idiotic idea on the face of it. And who knows which idiotic idea going begging today may not be the foundation of an empire in the years to come?

These and other such points make this special issue an unusually provocative one. (*IEEE Trans. on Engineering Management*, July 1968.)

The Illiac IV. A major piece of work in the computer field and one that is being looked at by everyone in the computer industry is described in a solid paper in the current issue of the IEEE Transactions on Computers. The new computer design, called Illiac IV, is significant because it is a parallel-type processor, thus promising much higher computing speeds than strictly conventional sequential processors and the capacity to handle huge computational problems. The paper summarizing the structure of the entire Illiac IV system is by members of departments at the University of Illinois and the Burroughs Corporation. A companion paper provides an overview of the software aspects of the system.

In their discussion of the system, G. H. Barnes and his colleagues point out that the study of a number of well-formulated but computationally massive problems is now limited by the computing power of available or proposed computers. Some such problems involve the manipulation of very large matrices (e.g., linear programming), others the solution of sets of partial differential equations over sizable grids (e.g., weather models), and still others need extremely fast data correlation techniques (phased array signal processing). Real progress in these problem areas, the authors continue, requires computing speeds several orders of magnitude greater than conventional computers.

At the same time, signal propagation speeds have become a serious barrier to increasing the speed of strictly sequential computers. Thus, in recent years a variety of techniques have been introduced to overlap the functions required in sequential processing; however, only incremental speed gains have been achieved at the cost of more complex and expensive hardware and greater problems in machine checkout and reliability. Explicit parallelism, on the other hand, offers speeds that increase linearly with the number of gates. Such parallelism, which was used in the Solomon computer in 1962, featured such things as: a large array of arithmetic units controlled by a single control unit so that single instruction streams sequenced the processing of many data streams; memory addresses and data common to all data processing were broadcast from the central control; and others.

The Solomon computer showed that a parallel approach was both feasible and applicable in many areas. Moreover, LSI circuitry with gate times of 2 to 5 nanoseconds suggested the possibility of a Solomon-type array of potentially 10^9 word operations per second. Furthermore, memory technology had reached the point where 10^6 words with cycle times of 200 to 500 ns could be produced at acceptable cost. These kinds of considerations led to the present Illiac IV.

The main structure of the computer consists of 256 processing elements arranged in four reconfigurable Solomon-type arrays of 64 processors each. The individual processors have a 240-ns add time and a 400-ns multiply time for 64-bit operands. Each processor is provided with 2048 words of 240-ns cycle time thin-film memory. Each processor is constructed of approximately 10 000 ECL gates. (G. H. Barnes *et al.*, "The Illiac IV Computer," *IEEE Trans. on Computers*, August 1968.)

Advance abstracts

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

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

Proceedings of the IEEE
IEEE Transactions on
- Audio and Electroacoustics
- Automatic Control
- Bio-Medical Engineering
- Broadcasting
- Communication Technology
- Computers
- Electron Devices
- Industrial Electronics and Control Instrumentation
- Industry and General Applications
- Information Theory
- Magnetics
- Microwave Theory and Techniques
- Nuclear Science
- Parts, Materials and Packaging
- Power Apparatus and Systems
- Quantum Electronics
- Vehicular Technology

Proceedings of the IEEE

Vol. 56, no. 10, October 1968
M—$1.00; L—$1.50; NM—$2.00

Sensitivity in Optimal Control Systems, *M. Sobral, Jr.*—Some recent developments in the area of sensitivity in optimal control systems are summarized. The several sensitivity definitions, analysis methods, and design techniques are placed in perspective. A selected list of references is included.

Communication in Time–Frequency Spread Media Using Adaptive Equalization, *M. J. DiToro*—In high-speed communication through a medium with time–frequency spread (such as in HF ionospheric, sonic underwater, and voice-quality telephone-line transmission), the transmission speed and system errors are determined by an overall system variance (or equivalent noise). This overall variance comprises: (1) the intersymbol interference variance from time spread (or dispersion in the unit impulse response of the medium); (2) the variance from frequency spread (or dispersion in the CW response of the medium); and (3) the variance from channel noise at the receiver. In order to mitigate intersymbol variance from time spread, after discussing some early and limited attempts, a review is made of the general synthesis of the infinite and finite Wiener optimal networks or equalizers using delay lines with feedforward and/or feedback taps (or their shift register digital network counterpart). Also considered are optimal finite feedforward delay line approximations to the Wiener networks, synthesized nonsequentially and sequentially using a steepest descent evolutionary network synthesis, resulting in a monotonically decreasing and convergent overall variance. Applications of these networks are shown in the transmission of speeded-up analog facsimile pictures over unconditioned voice-quality telephone lines. It is shown that the received smeared picture can be refocused or desmeared, or its time spread mitigated, to allow overall satisfactory high-speed picture transmission. Moreover, in over-the-horizon HF ionospheric analog and/or data transmission, time spread occurs in the form of discrete multipath reception, such as the familiar ghosts observed in television. A particularly stringent multipath comprises two equal or nearly equal magnitude paths, a situation that is actually observed in HF ionospheric communication between two ships at sea. For this case it is shown that serial data transmission with pulses thin enough to resolve the multipath achieves, with optimal equalization, an irreducible bit error ratio (BER), which is a number of orders of magnitude less than the irreducible BER of the parallel data transmission method used in all contemporary HF modems (*mod*ulators and *dem*odulators). Design curves are given showing the intersymbol interference variance from time spread as a function of the number of taps in the delay line correction networks, along with the variance arising from channel noise, and the optimal allocation of both variances for minimal overall BER with a given number of taps for the finite realizable correction networks. Because of frequency spread, the ever-changing unit or impulse response of the medium (e.g., HF ionospheric) causes the correction or equalization networks to become aged, giving rise to frequency-spread variance. This is formulated both for determinate and for random changes of the medium unit response. For transmission media having simultaneous time spread and frequency spread, the equivalent overall variance is a simple function of the time–frequency spread product of the medium. These formulas are applied to find the overall minimal BER as a function of the time–frequency product for a contemporary parallel data modem and for a new serial adaptive data transmission system (ADAPTICOM), which periodically in real time and with digital techniques rejuvenates the time spread digital correction networks. It is shown that the transmission limit of communication of the parallel data modem has a time–frequency spread product of about 1/2000, and that for the new serial adaptive data modem is about ten times larger, or 1/200. As a result, new communication channels are opened up for serial adaptive transmission, such as the HF spectrum below the maximum usable frequency.

Cryoelectric Hybrid System for Very Large Random Access Memory, *R. A. Gange*—The recent introduction of structured cells characterized by no cell-to-cell interaction, adequate sense signal, and array tolerance under repetitive worse-case disturb evaluation and coincident current operation, has made possible new systems approaches. A coincident current "hybrid" ("word-bit") system utilizing room-temperature decoders and structured loop cell arrays was designed to satisfy the criteria pertinent to a very large random access cryoelectric memory. Besides offering the advantage of an early realization of a cryoelectric memory, compared with previous systems advanced, several other benefits are realized: (1) low electronics cost, especially in large system sizes; (2) low heat load; (3) potential plane yield improvement through redundancy techniques. Although a disadvantage is a large number of interconnections, analysis shows that the overall memory cost can be minimized through proper substrate design. The technology required by the hybrid system embodies present art, and very large random access cryoelectric memories employing hybrid organization and loop cell arrays appear to be feasible.

Air-Cooled Two-Pole Generators for Gas Turbine Peaking Service, *B. H. Smith, J. E. Doescher*—Presented are considerations in the application and matching of two-pole cylindrical rotor generators to gas turbines for peaking service. Various criteria are reviewed, including generator thermal rating and limitations, gas turbine ratings, altitude factors, cyclic operating duty, rapid loading considerations, generator ventilation and air filtration, noise, steady-state, and transient performance, and special generator features.

The Linear Random Process, *I. F. Blake, J. B. Thomas*—A survey of the theory and applications of the class of linear random processes is presented. Topics discussed include the law of large numbers, covariance estimation, and the relationship of linear to normal processes. Various applications of the linear process to problems of communication theory are considered. These include prediction, signal extraction, and detection, using the linear process as a model for the signal or noise.

Radiation Models Using Discrete Radiator Ensembles, *S. Karp, R. M. Gagliardi, I. S. Reed*—In recent literature radiation emitted or reflected from a body or surface has been modeled as equivalent radiation from a collection of individual discrete radiators. This model has application to the analysis of forward- and backscatter from rough surfaces, clutter and chaff models for radar, and cavity emissions at optical frequencies. Radiation using a generalization of the above radiator ensemble is investigated. Some degree of coupling is allowed to exist between individual radiators, and each radiator is assumed to emit random energy bursts as a Poisson process. Analysis is confined to power spectral densities and first-order statistics of the resulting scalar radiation fold. The results indicate that the pulsed radiator model can account for many properties of radiation. A byproduct of the analysis is a filter model for dispersive channels, applicable to radar design problems. Theoretical results are compared with previously reported experimental results wherever possible. Some consideration is given to a relativistic interpretation of the radiator ensemble.

Proceedings Letters

Because letters are published in PROCEEDINGS as soon as possible after receipt, necessitat-

The day Andy Califano's in-depth ASW thinking surfaced some unusual dividends.

It took awhile. For Andy had been working as part of Electronics division's ASW team for some time. In close partnership with data acquisition and processing specialists, mathematicians, acoustics experts, human factors engineers, physicists and a multitude of others.

Creating state-of-the-art high density/dynamic range digital data systems, adaptive spatial and temporal filters, computerized performance prediction models and advanced transducer designs. Plus innovations in other areas. All in an environment with design tools and test facilities among the best in the ASW community.

Then the day came. The day his contributions to the computer math model, the new tracking system, the prebeamformer... and his occasional frustrations ... all paid off.

The team won the biggest yet addition to Electronics division's already sizeable list of long term ASW projects. Prime contractor for the AUTEC Sonar Range, the U.S. Navy's first in-situ fixed sonar calibration facility. Including precision underwater tracking and calibration arrays, shipboard data acquisition systems, shore-based digital data analysis and reduction systems and a whole spectrum of underwater range problems.

If you have background and interests in ASW (or, for that matter, AGE, navigation aids, data equipment or communications), why not surface some excitement of your own in one of our labs? We can't promise your boss will buy cigars. But, celebrations here get more often all the time. Start your first one by writing J. P. O'Reilly, Dept. 194, for more details.

GENERAL DYNAMICS

Electronics Division

1400 North Goodman Street, Rochester, New York 14601

An Equal Opportunity Employer (M&F)

the greatest admiration for these developments. They are wonderfully interesting, they are great advances, and they certainly have an impact on electrical engineering and on our economy. But they are a part of engineering; they are not, as some seem to think, nearly all of it.

Sehn's final comment really does call for us to get our feet under the same table. Lacking that, I can only do my best here. In saying that engineers have a tribal, cultural life, I meant it, and say it again. In no sense is Sehn justified in thinking that I want engineers to withdraw from society, to set up our own rude culture. Like everyone else, I want engineers to be normal human beings first—fully integrated with our general culture—and engineers second. Is that clear? Then let it also be clear that every professional group has its own tribal, or sub-culture. Within it there are its heroes; there is its history; there is unity of purpose; there is justified pride of accomplishment. When a professional group loses such aspects of its subculture, it disintegrates, and something else takes its place. If in engineering we were guilty of widespread dishonesty, we would fall apart. There is no such risk. But if, say, we permit our teaching to become too diluted by what is otherwise immensely admirable in itself—the math-theory-science complex—we could be in danger of losing our identity. How far have we already gone in that direction?

I wish also to welcome the communication from William L. Clements (Sept., pp. 134–135) and his agreement with some of my points. I especially like his term *academic incest*. Yes, we are loaded with it. Universities used to tend to avoid inbreeding in building up their faculties. "He is just the man for us, but is one of our own graduates; we had better look elsewhere." Ever hear that? But it could be that there is an even greater danger from Ph.D. inbreeding, or academic incest, but it seems far from being widely recognized.

It so happens that, though retired, I have not left the scene. When in town, I am in my laboratory seven days a week. A rewarding feature of this is the many contacts I have with faculty members, struggling doctoral candidates, and undergraduates. Some, nicely enough, bring their problems to me. From all this, I get samplings of how some of our most brilliant youngsters are unprepared to cope with real engineering. And not only the youngsters. Think of the faculty member I know of (he is elsewhere) who committed the same crime twice—by assigning, to graduate students, theoretical problems that were indeterminate. The problems didn't exist, even in theory. How far out yonder can you get?

A. D. Moore
Professor Emeritus
The University of Michigan
Ann Arbor, Mich.

Additional reference

"Men, Machines, and Languages," by John R. Pierce in the July IEEE SPECTRUM (pp. 44–49), is a fine review of this subject, a difficult one, to be sure, in view of the complex problems involved in language translation by electromechanical means.

Perhaps Dr. Pierce would not object to one addition to his list of references: *Sorrows of a Super Soul: The Memoirs of Marie Mushenough*, translated by machinery out of the original Russian, from the fifth edition of Stephen Leacock's "nonsense novels" (London: Lane, 1913).

Benjamin F. Miessner
Miessner Inventions, Inc.
Miami Shores, Fla.