VOLUME 2 □ NUMBER 10 OCTOBER 1965

IEEE spectrum

features

41 Spectral lines: Role of the PROCEEDINGS

+ 42 The Land Mobile Radio Services
Kenneth A. Cox *The tremendous growth in land mobile radio has created serious congestion since less than five percent of the suitable spectrum space has been allocated for land mobile use*

+ 44 A review of radar astronomy—Part I
Duane O. Muhleman, Richard Goldstein, Roland Carpenter
A wealth of information about the moon and several planets has been obtained through the use of radar astronomy techniques

+ 56 Airborne asphyxia—an international problem
Gordon D. Friedlander *Air pollution is no longer a minor inconvenience; it is a tangible, frightening, and intolerable situation that affects virtually every major industrial city in the United States and Europe*

+ 70 Do scientists and engineers need each other?
John R. Dunning *Future technological progress depends on scientists and engineers maintaining the close relationship of the war years*

+ 73 Ultrasonic traveling-wave devices for communications
J. E. May, Jr. *A research physicist discusses the latest techniques, developments, and materials that are used in a highly specialized and experimental electronic application*

+ 86 Discounted-cash-flow analysis of the return on engineering investment
Richard Hackborn *Element of time has become one of the most important factors governing engineering investments and accounts for the increasing use of discounted-cash-flow techniques*

+ 96 Report on the Symposium on Signal Transmission and Processing
H. E. Meadows

98 Authors

100 NEREM sessions

Departments: *please turn to the next page*

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

9 Transients and trends

10 Reflections

22 Focal points

Computer to help engineering class consider problems ... 22
Microwave net links three countries ... 22
Electron bombardment engine clocks 2600 hours of operation ... 24
White House invites fellowship nominations ... 26
Women engineers will meet in Cambridge, England, in 1967 ... 26
Analog computer simulates heart functions ... 26
University of Rochester will join in two-degree program ... 28
Longest, deepest simulated dive is successfully completed ... 28
Audio Engineering Society to hold annual convention ... 29
University of Arizona offers reliability engineering program ... 29
'Human voice' system informs pilots of weather ... 29

100 News of the IEEE

Northeast Electronics Research and Engineering Meeting (NEREM) sessions ... 100
Allerton Conference planned on Circuit and System Theory, October 20–22 ... 101
Papers solicited on Design and Construction of Large Steerable Aerials ... 104
Materials Handling Conference scheduled October 18–20 in Pittsburgh ... 106
IEEE Region Six Annual Conference scheduled for April 26 to 28 in Tucson ... 108
Quantum Electronics Conference slated for April 12–15 in Phoenix, Ariz. ... 108
AIAA and IEEE are joint sponsors of Communications Satellite System Conference ... 109
Attendance of 700 is reported at 1965 Aerospace Technical Conference ... 109
"Discuss-Only" sessions planned for the 1965 FJCC ... 110
Transactions devoted to high-speed memory systems ... 110
Fortescue Fellowships for 1966–1967 are announced ... 110
IEEE Group chairmen and Transactions editors ... 112
Annual Textile Conference will take place October 21–22 ... 113
Ultrasonics Symposium slated December 1–3 in Boston ... 113
Advance call for papers, 1966 IEEE International Convention ... 113

116 Calendar

120 People

128 IEEE publications

Scanning the issues, *128* Advance abstracts, *130* Translated journals, *144*
Special publications, *148*

149 Book reviews

Organic Semiconductors, Y. Okamoto and Walter Brenner. *Reviewed by N. E. Wolff* Fundamentals of Relay Circuit Design, Alan R. Knoop. *Reviewed by W. H. Holcombe* Plasma Diagnostics with Microwaves, M. A. Heald and C. B. Wharton. *Reviewed by George Bekefi* A History of the General Radio Company, Arthur E. Thiessen. *Reviewed by Frank A. Gunther*

the cover

The grime-encrusted wall graphically illustrates one of the major problems of contemporary civilization—the corrosive and polluted atmospheres in our major cities—in which the soot-fall has been measured in quantities up to 60 tons per square mile per month. For a detailed discussion of this menace to life and health, see the article that begins on page 56.

Spectral lines

Role of the Proceedings. Now that most of the transients associated with the merger of AIEE and IRE into IEEE have passed, it is well to re-examine the roles played by various IEEE publications.

SPECTRUM has had a significant effect upon other IEEE publications, particularly PROCEEDINGS. SPECTRUM has relieved PROCEEDINGS of the rather ambiguous role it played in the IRE, where it was both the "core publication" of that Institute and also the vehicle for articles of the most advanced nature. Clearly, it was not understandable to a rather large segment of the membership and this segment of the membership was not given much assistance in keeping informed about new developments. Now the role of PROCEEDINGS is more clearly defined as the appropriate medium for articles of broad interest written at as high a technical level as the subject warrants.

In view of these changed circumstances, the Editorial Board has developed for the PROCEEDINGS a new program and a mechanism for its implementation. A major feature of the program is to expand the highly successful Special Issue feature, which, in the past, has made such a significant impact on many developing fields. These issues, devoted to a single subject with articles written by leading authorities, have frequently been the most authoritative sources of information available. They are to be found in the ready-access file of leading workers in the field and on the shelves of many research libraries. Starting with the August 1965 issue, it is planned that these Special Issues will appear more frequently—approaching an every-other-month basis.

Another feature is to invite leading specialists to write papers which review, at an advanced level, topics of current interest to a sizable fraction of the membership. The length of these articles will not be subjected to the usual restrictions. As the first article in this series, Professor E. C. Kuh of the University of California, Berkeley, and Professor Rohrer of the Polytechnic Institute of Brooklyn have written an article in the July issue of PROCEEDINGS on the application of state-variable techniques in the solution of electric network problems. The use of state variables has had a major impact on automatic control theory, and the importance of their application to electric network problems is starting to be recognized. The next article in this series is a summary of the work on vocoders by Dr. M. R. Schroeder of Bell Telephone Laboratories. Vocoders, which permit more effective transmission and storage of speech, are being increasingly used to facilitate the transmission of speech in encrypted form. Vocoder applications are expected to be of major importance to long-distance telephony, voice-readout systems, and secure voice transmission.

As a final innovation, it is planned to improve the correspondence section by decreasing the delay between the submission and publication of correspondence. This part of PROCEEDINGS—which has expanded by a factor of ten in the last ten years—will be handled on an "express" basis. It is anticipated that the average delay between submittal and publication will be less than eight weeks. It is planned to rename the section "PROCEEDINGS Letters" to highlight this new development.

In listing these innovations, it should not be assumed that the PROCEEDINGS will not continue to publish contributed papers of broad appeal in all fields of IEEE interest. The review procedures for both articles and letters are being improved so that a high standard of technical excellence will be maintained.

To implement this program, the Editorial Board has concluded that the PROCEEDINGS should be directed by a vigorous editor with a well-balanced board of editorial advisors. The Board of Directors has appointed Professor M. E. Van Valkenburg of the University of Illinois as Editor of PROCEEDINGS. Professor Van Valkenburg brings to this position a wealth of experience as a leading engineering educator, a previous editor of the IEEE TRANSACTIONS ON CIRCUIT THEORY, and the author of several widely used textbooks. He will be assisted by an editorial board that will include leading contributors in the fields of interest of the IEEE.

These developments will strengthen the PROCEEDINGS and will provide a fitting future for a publication that has served the profession so well in the past. —*F. Karl Willenbrock*

The Land Mobile Radio Services

The tremendous growth in land mobile radio has created serious congestion since less than five percent of the suitable spectrum space has been allocated for land mobile use

Kenneth A. Cox *Federal Communications Commission*

Of the myriad uses of radio, probably the least known—and in many respects the least appreciated—are those encompassed under the general heading of the Land Mobile Services. These services involve, principally, two-way base-to-mobile and mobile-to-mobile voice communication systems. Every part of our industrial, social, and economic structure has found it possible to operate more efficiently and more effectively by adapting radio to its present-day needs. For the most part, these radio systems are now indispensable and are an integral part of the entities concerned, whether they be industrial, public-safety oriented, or any other type. But public knowledge of the extent of the organization of these activities around radio is extremely limited even though the public uses radio-equipped taxicabs and watches television enough to know that the police use two-way radio.

The average person probably cannot list half a dozen of the 40-odd Land Mobile Services. He would be astonished to learn just how deeply they affect his everyday life. The rapid dispatching of highway, telephone, and public-utility emergency vehicles to trouble spots; the continuous transmission of messages to trucks engaged in pickup and delivery; the detection of hot boxes and front-to-end train communications in the railroad service; remote control of machinery and industrial processes; and the paging of key personnel in large industrial complexes are but a few of the uses of radio about which the average member of the public knows little or nothing. There are no programs, no daily schedules in the press, and no syndicated columnists extolling the virtues of such operations. Yet the quality and lower costs of the goods the public receives, the speedier service, and the public's personal safety are all dependent to a very large degree on two-way radio.

Top management also is often only vaguely aware of the extent of a company's dependence on radio systems. Key executives are not fully aware of radio's present and potential value. To them it represents another tool, a somewhat intangible asset that is but a small percentage of their capital investment as compared with physical plant or rolling stock. An executive may equate, understandably, relative intrinsic value or importance of radio in terms of its relationship to the company's total assets.

Yet many of the organizations which these executives direct depend substantially on two-way radio for efficient management and control—and some highly competitive businesses would encounter severe difficulties without it. To these concerns, mobile radio may represent the difference between profit and loss. The increasingly rapid pace of modern life in turn accelerates the demand for new tools, new techniques, and new methods. In our dynamic free-enterprise society, modern communications are vital to progress and must keep pace.

Had there been a fuller realization in the thirties of the uses to which two-way radio could be put, a better climate for its development might have been provided. The truth is that two-way radio did not come into its own until the end of World War II. Its advent could hardly have been more inauspicious. The atomic age was upon us and television was beginning its meteoric rise to a place of central importance in American life. In the middle fifties, when the problems of mobile radio were beginning to be felt seriously, the public's attention was again diverted, this time by the marvels of the space age. But land mobile radio continued to grow quietly at an ever-increasing pace, until now there are more than 2 200 000 transmitters licensed in these services. Figure 1 indicates this rapid and consistent growth over the past 18 years—and these figures are for the Land Mobile Radio Services only; they do not include the Aviation, Marine, Amateur, Citizens, and Disaster Services. The addition of the latter would bring the total number of transmitters to more than 5 million. The ratio of transmitters to licenses has been increasing steadily, but this trend is being arrested as channel loading becomes excessive in the various services.

The tremendous growth in Land Mobile Radio has created a variety of serious problems. Foremost among these is congestion of the air waves in large cities and the difficulty in finding sufficient interference-free spectrum space to accommodate the expansion. Equitable distribution of the limited spectrum is imperative, but despite the ever-increasing progress in the technology of communications, improvements in efficiency of frequency utilization have not kept pace. Frequency management today constitutes a major governmental problem.

Small wonder, then, that the Land Mobile Radio Services find themselves in their present predicament. Of that portion of the spectrum generally regarded as suitable for land mobile radio, less than 5 percent has been allocated to these services, as can be seen from Fig. 2. The total exceeds 100 percent because 9.1 percent of the government's figure (cross-hatched area) is shared with other services. One of the guidelines the Federal Communications Commission must observe in the allocation of frequencies is "the total number of people who would probably receive benefits from a particular service." Because of the lack of recognition of the merits of mobile radio, not necessarily on the part of the Commission alone but among the public generally, these services have had to get along with less than adequate space. Congress

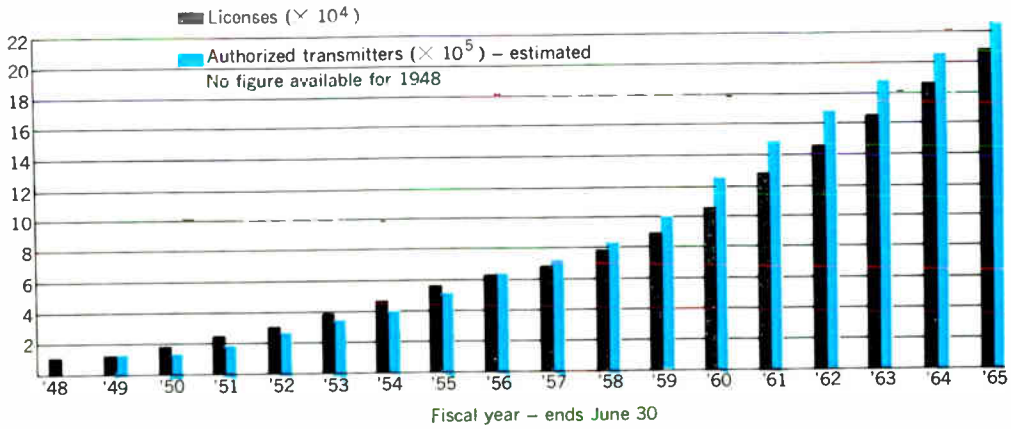


Fig. 1. Chart showing the rapid growth in Land Mobile Radio Services from 1948 through June of 1965.

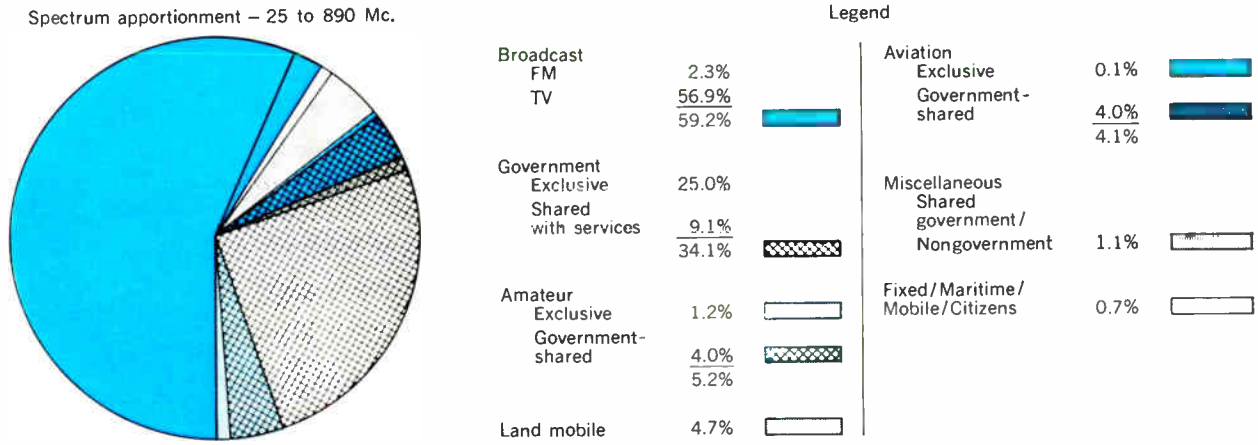


Fig. 2. Apportionment of the frequency spectrum in the range from 25 to 890 Mc/s.

has begun to recognize the growing problems in this area, but as yet there have been no heated debates concerning land mobile needs.

The Commission is now well aware of the growing crisis facing the Land Mobile Radio Services. It has become sensitive to the problems and will continue its efforts to alleviate them. A very important part of that effort is exemplified by the Advisory Committee for the Land Mobile Radio Services, established by the Commission in March 1964. While establishment of the Committee does not in itself guarantee a prompt and complete solution to all problems, it means that some progress is being made. Hopefully, the Committee will provide suggestions and recommendations for Commission action in the near future. The Committee is made up of more than 175 of the most talented, knowledgeable, and public-spirited people in the field of land mobile communications. Although many diverse interests are represented, there is excellent cooperation. Service on the Committee has brought the realization that mutual problems can be solved by working together, and prospects for at least partial solutions to some of these problems are good.

I am confident that a great deal of value will come out of this group. It is our hope that the Committee will objectively consider the various problems selected for study and will bring forth new concepts and procedures for dealing with them—not just a summation of the individual problems of the various services represented. New ideas, and perhaps even substantial breakthroughs, are required. This may mean the obsolescence of various systems, but systems become obsolete rather quickly no matter how one tries to deter progress. We must face up to the fact that our present systems will be as obsolete tomorrow as the old spark transmitter is today. We are fully aware that this will be costly—all new systems are. But cost is the one problem that is most easily solved by the right idea at the right time.

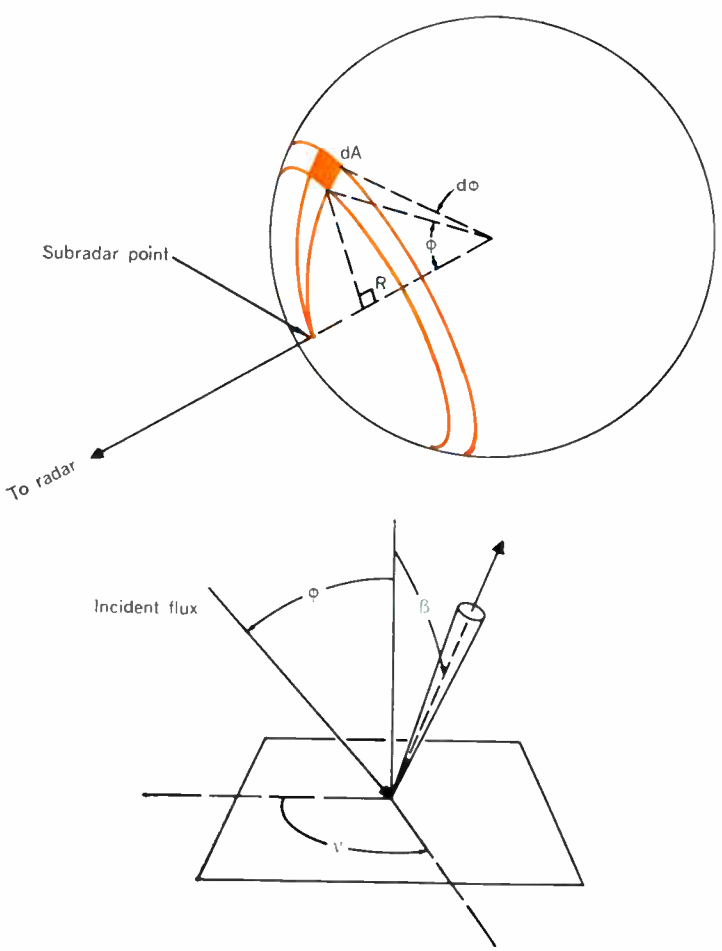
The Advisory Committee is an instrument of the Commission and can only make recommendations to the latter. However, its findings and recommendations will bear considerable weight, and its proposals are assured of an interested and sympathetic reception. The degree to which these proposals will be implemented is dependent, to a large extent, on the thoroughness and objectivity with which the problems are attacked and on the logic and completeness of their analysis and presentation to the Commission. The Commission did not establish the Advisory Committee merely to survey and present a rehash of already well-known and well-documented land mobile problems. Instead it hoped that meaningful answers to many of the problems the Committee is considering are within its competence. I believe they are. There is a real challenge here for all concerned, as participation on the Committee represents a tremendous opportunity for constructive work. If I were a user concerned with these problems I would not want to see this chance lost or frittered away. The Committee must develop objective and meaningful recommendations, and the Commission must see that each of them is fully explored and implemented to the fullest extent possible. And this, of course, may require both the land mobile users and the Commission to change some of their established concepts.

A review of radar astronomy— Part I

The past few years have seen spectacular advances in the science of radar astronomy. This two-part article develops fundamental concepts and equations, describes new measuring techniques, and summarizes significant results from planetary and lunar radar experiments

Duane O. Muhleman
Richard Goldstein
Roland Carpenter

Jet Propulsion Laboratory,
California Institute of Technology



Radar astronomy evolved shortly after World War II when groups in Hungary and the United States independently detected radar echoes from the moon by utilizing radar equipment developed during the war. During the following ten years radar research was motivated primarily by the desire to use the moon as a passive reflector in communication systems. It was not until the middle 1950s that lunar radar experiments were conducted as a pure science.[1] It was realized at that time that the distortions on the returned signal caused by the target and the medium could be analyzed in such a manner as to yield fundamental information on the nature of the lunar surface and cislunar space.

Radar astronomy has made spectacular advances with the successful detection of echoes from Venus near the time of its 1961 inferior conjunction. These first planetary observations were obtained by groups at the California Institute of Technology's Jet Propulsion Laboratory (JPL) and M.I.T.'s Lincoln Laboratory. These events were quickly followed by experiments in the U.S.S.R. and in England. A remarkable wealth of information concerning the moon and Venus has been obtained since 1961. Progress in this field has been closely tied to the developments in masers, high-gain antennas, low-noise feeds, and, perhaps most importantly of all, in digital-detection and data-processing procedures. The overall effect of these developments has been to increase the radar detectability by about an order of magnitude per year since 1961 and there are good reasons to believe that this rate of improvement will be maintained for many years.

We shall attempt here to review the new methods and techniques, as well as to present a summary of results, with particular attention to progress toward understanding the characteristics of Venus. Because of the limited length of a review article, it has been necessary to present most equations and techniques without complete derivations, most of which are available in the literature of the past few years. The phenomena considered involve the following:

1. The measurement of the total power reflectivity of the moon, Venus, Mercury, Mars, and Jupiter, and its interpretation in terms of the planetary surface characteristics.
2. The calculation of the radar backscatter properties of the planetary topography for the moon, Venus, and Mercury.
3. The effects of the planetary atmospheres and interplanetary space at radar wavelengths.

Fig. 1. Geometry for the return of a narrow pulse from points situated on a planetary surface.

4. The interpretation of the measurements with regard to the motion of the earth, Venus, and Mercury in their orbits as well as the determination of the astronomical unit (AU).

The radar equation

Most of the radar targets of interest in radar astronomy are essentially spherical in shape—e.g., the moon and the planets. Actually, these bodies deviate slightly from the spherical, but since these deviations would only be detectable with extremely sensitive radar experiments they can be neglected for the purposes of this discussion. It is well known from optics that a plane electromagnetic wave will be scattered isotropically (in the far field) from a smooth sphere whose material has an infinite electrical conductivity. If such a sphere is illuminated by a continuous-wave radar whose beam is wide relative to the diameter of the sphere, then the power received back at the radar is given by the so-called radar equation

$$P_{r_{iso}} = \frac{P_t}{(4\pi)^3} \frac{G^2\lambda^2}{r^4} \pi R^2 \quad \text{watts} \qquad (1)$$

where

P_t = power transmitted
G = gain of the antenna over isotropic
λ = wavelength of the transmitted power
r = distance between the radar and the sphere
R = radius of the sphere

The power returned from a real planet will differ from that given by (1) for several reasons. The surface material of the planet will certainly not possess infinite conductivity; rather, it will exhibit a finite complex dielectric constant with an associated impedance and ohmic loss. This fact is most conveniently expressed as a power reflectivity η, which is equal to or less than unity. Furthermore, the planet in general will not scatter the energy isotropically (primarily due to surface roughness). Instead, there will be some effective surface "gain" pattern whose value in the direction back to the radar is called the backscatter directivity g, the value of which can be greater than unity. If the radar equation is augmented by these two multiplicative factors, the ratio of the power actually received to the power received from the equivalent isotropic scatterer is just

$$\frac{P_r}{P_{r_{iso}}} = g\eta \qquad (2)$$

It is the nontrivial task of the experimenter to separate these factors if he wishes to interpret his observations in terms of the surface roughness on the one hand and the electrical parameters on the other. It has been customary in the past to obtain η from a measurement of the total power by dividing the measured ratio, Eq. (2), by a value of g estimated from the detailed scattering response of the target. The value of η so obtained has been set equal to the expression

$$\eta = \left(\frac{\sqrt{\epsilon_c} - 1}{\sqrt{\epsilon_c} + 1}\right)\left(\frac{\sqrt{\bar{\epsilon}_c} - 1}{\sqrt{\bar{\epsilon}_c} + 1}\right) \qquad (3)$$

where ϵ_c is the complex specific dielectric constant and $\bar{\epsilon}_c$ its complex conjugate. If η were found to be a function of the wavelength it would be possible to solve (3) for several of the electrical parameters, such as the real part of ϵ_c and the electrical conductivity. However, according to Evans and Pettengill,[2] η is independent of wavelength over the range of wavelengths utilized for lunar experiments, which implies that the electrical conductivity (which would normally be expected to be wavelength dependent) is very small. In such a case, (3) becomes

$$\eta = \left[\frac{1 - \sqrt{\epsilon}}{1 + \sqrt{\epsilon}}\right]^2 \qquad (4)$$

where ϵ is the real part of ϵ_c. Equation (4) is rigorously true for a smooth, homogenous dielectric sphere that is large compared with the wavelength. The use of this equation for surfaces that deviate significantly from this ideal model is suspect.

Backscatter laws

In general, the received echo signal from a planet will be spread in both time and frequency due to the spherical nature of the target. The spreading in time arises from the "depth" of the target; for example, if an infinitely narrow pulse is transmitted, the returned energy will vary as a function of time since the individual reflecting areas on the planet are located at various distances from the radar. The spreading in frequency is caused by the rotation of the planet with respect to the radar. If a monochromatic wave is transmitted, for example, the Doppler frequency shifts arising from the individual scattering areas will vary depending on the radial velocity of the reflecting area relative to the radar. By virtue of these phenomena it is possible to decompose the echo signal according to well-defined regions on the planet's surface.

Let us first consider the time spreading of an infinitely narrow pulse. We will select the time of arrival of the front edge of the *returned* pulse as the time origin. Clearly, this initial energy has been returned from the point on the surface nearest to the radar, which is the subradar point for a sphere. The energy arriving at time τ is then the return from points on the surface with

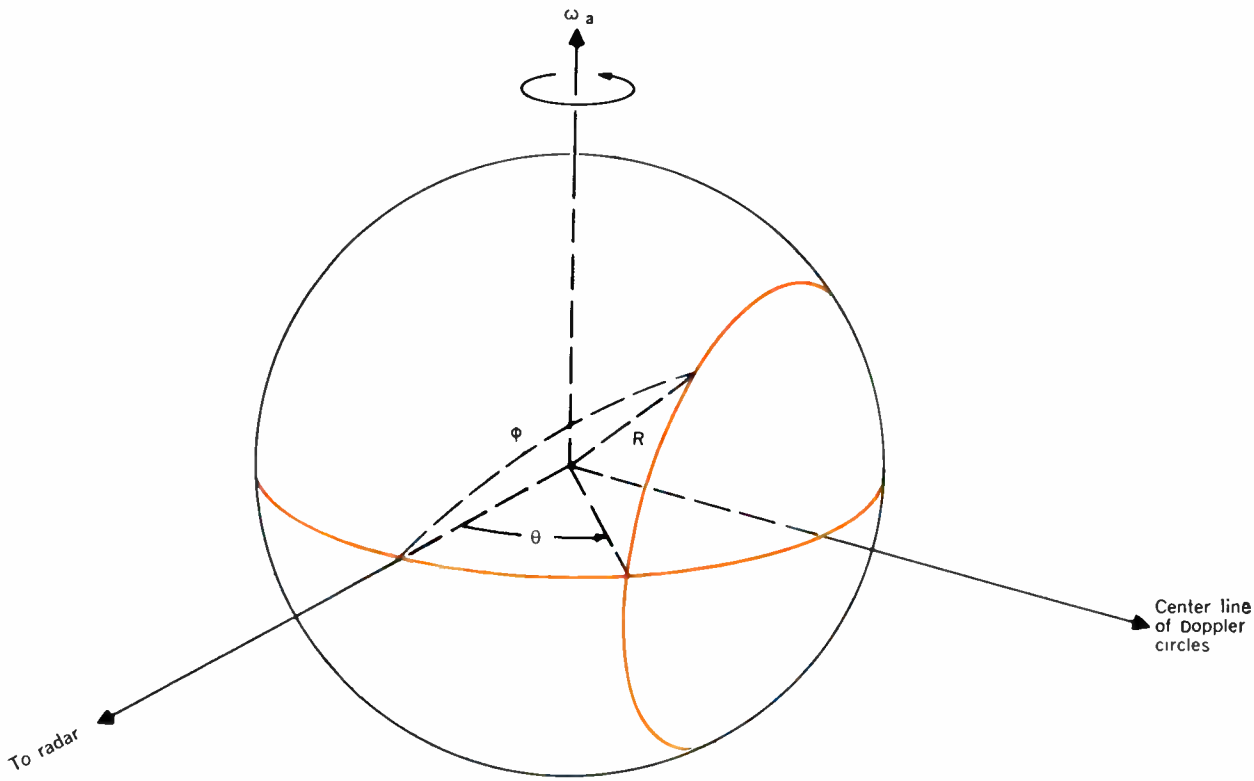


Fig. 2. Geometry for the return of a CW radar signal from the planetary surface.

distance $\tau c/2$ (where c is the speed of light) back from the subradar point and these points lie in a circle centered about this point. The radius of this circle subtends an angle ϕ at the center of the sphere as shown in Fig. 1. It can be shown from the geometry that this angle is related to τ by the simple relationship

$$\tau = \frac{2R}{c}(1 - \cos\phi) \tag{5}$$

where R is the radius of the planet, which allows us to map the received energy as a function of time with successive rings on the surface of the planet. In particular, the energy received in a time interval $d\tau$ at time τ is the contribution of area of an annulus width $R\,d\phi$ centered at the angle ϕ, where from (5),

$$d\tau = \frac{2R}{c}\sin\phi\,d\phi \tag{6}$$

Assuming that the incident power per unit area is W, the power intercepted by the annulus is

$$dP_i = 2\pi W R^2 \sin\phi\cos\phi\,d\phi \tag{7}$$

It is convenient to introduce the concept of the phase function $F(\beta, \nu)$ which determines the relative intensity of the scattered energy in an element of solid angle $\sin\beta\,d\beta\,d\nu$, where β is the polar angle from the *local* surface normal and ν is an azimuth angle (see Fig. 1). Then the power scattered into the direction (β, ν) from an area element of the annulus dA is

$$dP_s = W\cos\phi\,F(\beta, \nu)\sin\beta\,d\beta\,d\nu\,dA \tag{8}$$

Since all of the power incident on dA must be scattered in 2π steradians above the surface, we obtain the normalization for $F(\beta, \nu)$

$$\int_{\beta=0}^{\pi/2}\int_{\nu=0}^{2\pi} F(\beta, \nu)\sin\beta\,d\beta\,d\nu = 1 \tag{9}$$

The phase function, of course, is the mathematical expression for the surface scattering law. However, in planetary radar astronomy we are practically limited to monostatic radar observations, and consequently we are limited to measuring the phase function in the backscatter direction only; that is, $\beta = \phi$ and $\nu = 0$. Now the power returned from a given annulus may be computed from Eqs. (7) and (8), which, when normalized by the *total* power returned from an equivalent isotropic scatterer, Eq. (1), is

$$dP_R(\phi) = 8\pi F(\phi, 0)\cos\phi\sin\phi\,d\phi \tag{10}$$

Clearly, the integral over ϕ from 0 to $\pi/2$ yields the backscatter directivity

$$g = 8\pi\int_0^{\pi/2} F(\phi, 0)\cos\phi\sin\phi\,d\phi \tag{11}$$

Since the phase function always appears with the $\cos\phi$ as a multiplicative factor we will define the *backscatter function* $S(\phi)$

$$S(\phi) = F(\phi, 0)\cos\phi \tag{12}$$

The pulse response as a function of time can be obtained from (10) and (12) by the use of (5) and (6); that is,

$$\Phi(\tau)d\tau = 4\pi\frac{c}{R}S[\phi(\tau)]d\tau \tag{13}$$

and

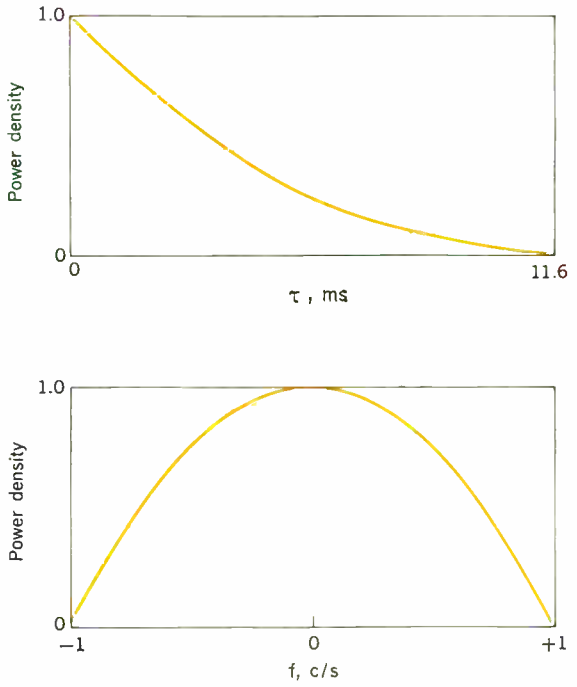


Fig. 3. Narrow-pulse response (A) and Doppler spectrum (B) from a rotating Lambert sphere. (Peaks normalized to unity.)

$$g = 4\pi \frac{c}{R} \int_0^{2R/c} S[\phi(\tau)]\,d\tau \tag{14}$$

The reader should realize that in the derivation of these results we have assumed a unity surface reflectivity and, consequently, in a practical application the integral of the pulse given by (14) yields $g\eta$.

As an example, we will consider a Lambert sphere for which $S(\phi) = (1/\pi)\cos^2\phi$. [The normalization is obtained from Eq. (9).]

Such a scattering law yields

$$\Phi(\tau) = 4\left(\frac{c}{R}\right)\left(1 - \frac{c\tau}{2R}\right)^2$$

and, from (14),

$$g = 8/3$$

We will next consider the Doppler spreading of CW illumination arising from a rotating planet characterized by a backscatter function $S(\phi)$. It can be shown that the locus of points, on the surface of a rotating sphere, that are moving with the same radial velocity with respect to the radar is a circle on the sphere centered about the line perpendicular to the plane formed by the radar line of sight and the apparent rotational axis of the sphere. Consequently, these points of constant Doppler shifts appear from the radar as straight lines parallel to the rotational axis. Furthermore, the Doppler shift relative to the frequency shift of the subradar point associated with the circle shown in Fig. 2 is

$$f_d = \frac{2R\omega_a}{\lambda} \sin\theta \tag{15}$$

and the maximum shift is

$$f_m = \frac{2R\omega_a}{\lambda} \tag{16}$$

The total spectral width, $2f_m$, depends on the apparent rotation rate ω_a, which is the component of the total rotational rate that is perpendicular to the radar line of sight. The total rotational rate arises from the vector sum of the intrinsic rotational velocity of the planet and fictitious rotational velocity caused by the translational motion of the planet relative to the radar. The apparent rotation rate changes with time, since the aspect of the planet with respect to the radar varies with time. It can be shown that power returned from a strip of angular width $d\theta$ at an angle θ from the subradar point is, after normalization with Eq. (1),

$$\Psi(\theta)\,d\theta = 8\cos\theta\,d\theta \int_\theta^{\pi/2} \frac{S(\phi)\sin\phi\,d\phi}{\sqrt{\cos^2\theta - \cos^2\phi}} \tag{17}$$

Furthermore, we can define the power spectral density of the echo signal, $\Theta(f_d)$, to be the power in a frequency interval df_d at a frequency f_d from the center of the spectrum. From (15) and (16) we see that

$$df_d = f_m \cos\theta\,d\theta \tag{18}$$

and, consequently, from (16),

$$\Theta(f_d)df_d = 8f_m \int_{\sin^{-1}f_d/f_m}^{\pi/2} \frac{S(\phi)\sin\phi\,d\phi}{\sqrt{\sin^2\phi - (f_d/f_m)^2}} \tag{19}$$

Thus, a measurement of $\Theta(f)$ (the power spectral density of the echo signal) contains the backscatter law in the integral of (19). The total power in the spectrum yields $g\eta$ by virtue of the normalization procedures used in the derivation.

As an example, the Lambert sphere for which $S(\phi) = (1/\pi)\cos^2\phi$ yields the power spectrum

$$\Theta(f_d) = 2(1 - f_d^2) \tag{20}$$

where we have taken $f_m = 1$. Also, integration of (19) over f_d from -1 to $+1$ yields $g = 8/3$. The pulse response and power spectrum for the Lambert sphere are shown in Fig. 3.

It is clear from these results that the backscatter function of any planet, which contains the information available concerning the surface roughness, can be obtained directly by measuring the pulse response of the planet; see (13). The same information is contained in a measurement of the power spectrum in the form of an integral equation; see (18). This integral equation can be solved for $S(\phi)$ if the spectrum is symmetric. Under such circumstances, (18) becomes the Abel integral equation, whose well-known solution yields the Abel transform pair (with $f_m = 1$):

$$\Theta(f) = 8 \int_{\sin^{-1}f}^{\pi/2} \frac{S(\phi)\sin\phi}{\sqrt{\sin^2\phi - f^2}}\,d\phi \tag{21}$$

and

$$S(\phi) = \frac{1}{4\pi}\cos\phi \int_1^{\sin\theta} \frac{\Theta'(f)}{\sqrt{f^2 - \sin^2\phi}}\,df \tag{22}$$

where the prime in (22) denotes differentiation with respect to f and $\Theta(f)$ is assumed to be an even, differentiable function of f.

Mapping

If a planet is illuminated with the equivalent of pulses with finite time length, it is possible to simultaneously analyze the echo signal in terms of range rings and Doppler strips. In particular, if the echo signal is passed through a narrow range gate and subsequently spectrally analyzed, the power in a spectral width df_d at frequency f_d corresponds to the echo from the two regions of the planet's surface formed by the intersection of the particular Doppler strip and range-gate annulus. This technique has been utilized by Pettengill[3] for isolating the echo from specific regions of the lunar surface. Goldstein[4] has extensively developed digital correlation techniques for similar experiments with Venus. The technique has been particularly useful with Venus in that it has offered a nearly unambiguous determination of the Venus rotational rate.

Planetary motions

The measurement of the radar range and range rate of Venus and Mercury relative to the earth has added powerful new observational material for the study of the motion of these planets in the solar system.[5,6] The general approach to this task is to compare the observed values of range and range rate with the corresponding observables computed from planetary theory with its associated system of fundamental constants. These constants include the astronomical unit, the orbital elements of the earth and the observed planet, and the masses of the earth and moon, etc. The discordance between the observations and the computed observables is then used to redetermine new values of the constants. The power of these techniques arises from two separate facts: (1) measurements of range and range rate are highly complementary to the traditional measurements of angular positions and timing of periodic events in the solar system; and (2) the measurements of range and range rates have been accomplished for Venus with an accuracy of about 10 km and 0.005 m/s, respectively. These results will be summarized in the following sections.

Total power measurements

This section deals with the several techniques of measuring total echo power that are currently in use in radar astronomy for determining the radar cross section of the target planet. The relationships between radar cross section and physical characteristics of the planet were discussed at the beginning of this article.

Special techniques are required for measurement of echo power because the signal is always immersed in background noise, which is often much stronger than the signal itself. The measurement of signal power is fundamental to radar astronomy. Methods of dividing the power up into its frequency distribution or into its time distribution, or both, will be discussed subsequently. However, even after such partitioning, the actual measurement made is always the quantity of signal power that remains in the given "cell."

The most straightforward way to measure the signal power is to determine the signal-to-noise ratio and to compute the signal strength from a determination of the signal noise. The noise power is given by the product of the Boltzmann constant, the system temperature, and the system bandwidth or kT_sB. The system temperature may be accurately measured for each data run. We examine two methods and their associated accuracies in this section.

Because power is defined as the average of the square of the signal voltage, the natural method of measurement is that shown in Fig. 4. Here $u(t)$ is the voltage input to the square-law device and $v(t)$ is its output voltage, which, after averaging, becomes the signal power estimate w.

A filter is needed to limit the amount of noise presented to the square-law device, but otherwise the block diagram follows the definition closely.

The integrator output, on the average, equals the sum of the signal power and the noise power

$$\overline{w} = w_s + w_n$$

However, because of the random nature of both the signal and noise, the output will fluctuate about this average.

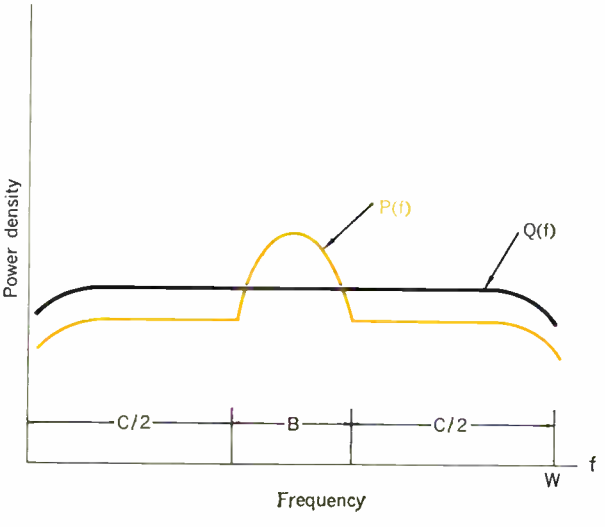
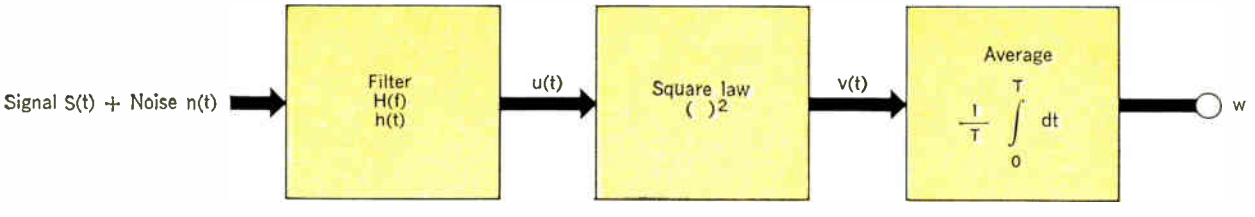


Fig. 5. Measurement of total echo power by use of the spectral analysis method.

Fig. 4. Measurement of echo power by use of a square-law device.



Let us denote the variance of the power estimate by σ_w^2. Then the performance index of the power measurement, SNR (or postdetection signal-to-noise ratio), is

$$\text{SNR} = \frac{w_s}{\sigma_w} \tag{23}$$

The experiment should be designed to maximize SNR. The following equations hold:

$$u(t) = \int_{-\infty}^{\infty} h(x)[s(t-x) + n(t-x)]dx$$

where $h(x)$ is the filter time response, $s(t)$ the signal voltage, and $n(t)$ the noise voltage. Also,

$$v(t) = u^2(t)$$

The power estimate is

$$w = \frac{1}{T}\int_0^T v(t)\,dt$$

whose ensemble average is

$$\bar{w} = \frac{1}{T}\int_0^T \bar{v}(t)\,dt$$

with variance

$$\sigma_w^2 = \frac{1}{T^2}\int_0^T\int_0^T v(t)\,\overline{v(x)}\,dt\,dx - \bar{w}^2$$

It can be shown that, under the assumptions of white Gaussian noise whose power is much greater than the signal power, SNR is maximized by the following choice of $H(f)$:

$$|H(f)|^2 = \Phi_s(f) \tag{24}$$

where $\Phi_s(f)$ is the power spectral density of the signal. When such an $H(f)$ is used, the resulting performance is

$$\text{SNR} = \frac{1}{N}\sqrt{T\int_0^{\infty}\Phi_s^2(f)\,df} \tag{25}$$

where N is the (one-sided) noise density (watts per cycle per second) and T is the integration time. Since the performance factor SNR depends on the integral of the *square* of the power spectral density, we see that, for a given signal power, the narrower the spectrum the more easily it is detected—one of the implacable laws of radar astronomy. Since the signal half-bandwidth is given by $f_m = 2R\,\omega_a/\lambda$, Eq. (16), we see that the planets possessing smaller $R\omega_a$ (smaller radius and/or angular rotational rate) for a given radar cross section are most easily detected. For example, Mercury is more easily detected than Mars due to their respective ω_a's.

There are two difficulties in applying this idealized formulation to a practical situation. First, a practical system has troublesome gain variations, and it is difficult to separate the average output into its signal and noise components. This is often accounted for by switching the signal off and on (easily done at the transmitter) and comparing the received signal-plus-noise power to the power for noise only. The measurement is then independent (to first order) of varying system gain.

The second difficulty is the fact that the power spectrum of the signal $\Phi_s(f)$ is often unknown. Indeed, this may well be the quantity to be measured. In practice, it is usually assumed that $\Phi_s(f)$ approximates a rectangular function of some bandwidth B. Then Eq. (3) becomes

$$\text{SNR} = \frac{w_s}{N}\sqrt{\frac{T}{B}} \tag{26}$$

where w_s is the signal power.

When switching is applied as described above, this equation is modified to

$$\text{SNR} = \frac{w_s}{2N}\sqrt{\frac{T}{B}} \tag{27}$$

which is known as the switched radiometer formula. The factor of ½ results from the switching process and is the cost of that practical consideration.

Equations (24) and (27) imply that a search problem exists in the case of a very weak signal in the background noise. One must find the correct bandwidth to use in order to detect the signal at all.

A second method of estimating the total power makes use of the power spectral density measurement. The raw material for this method is two spectrograms, one of signal-plus-noise (P) and one of noise only (Q). These spectra are obtained by a method to be described in a later section and have been normalized to unit power in order to eliminate the effect of varying system gain; an example of the spectra is given in Fig. 5. $P(f)$ and $Q(f)$ are given by the normalized spectral densities

$$P(f) = \frac{\Phi_s(f) + \Phi_n(f)}{w_s + w_n} \tag{28}$$

$$Q(f) = \frac{\Phi_n(f)}{w_n} \tag{29}$$

where $\Phi_n(f)$ is the noise spectral density.

Assume, again, that the maximum bandwidth B of the signal is known, and c is that fraction of the total bandwidth of Fig. 5 which contains no signal. A good estimate $M(f)$ of the signal spectrum is

$$M(f) = P(f) - aQ(f) \tag{30}$$

where a is a constant to be determined. If a is chosen properly, the average of $M(f)$ over the interval c will be zero. The value of a is defined by the following relation:

$$a = \frac{\int_c P(f)df}{\int_c Q(f)df} = \frac{w_n}{w_s + w_n} \text{ (on the average)} \tag{31a}$$

since the integral of $\Phi_s(f)$ over the region c is zero. Then it can be shown that

$$\frac{w_s}{w_n} = \frac{1-a}{a} \tag{31b}$$

and that the performance index of this method is[7]

$$\text{SNR} = \frac{w_s}{N}\sqrt{\frac{cT}{2(1-c)F}} \tag{32}$$

When $c = \frac{1}{2}$, Eq. (32) reduces to (27), the switched radiometer formula: and when c approaches unity (the spectrometer bandwidth F must become large relative to the signal width), Eq. (32) approaches (26). Thus the square-law method and the spectral method produce power measurements of approximately the same quality.

In a practical case where we have available a spectrogram of the echo signal plus noise the ratio of signal power to noise power is obtained utilizing (31a) and (31b).

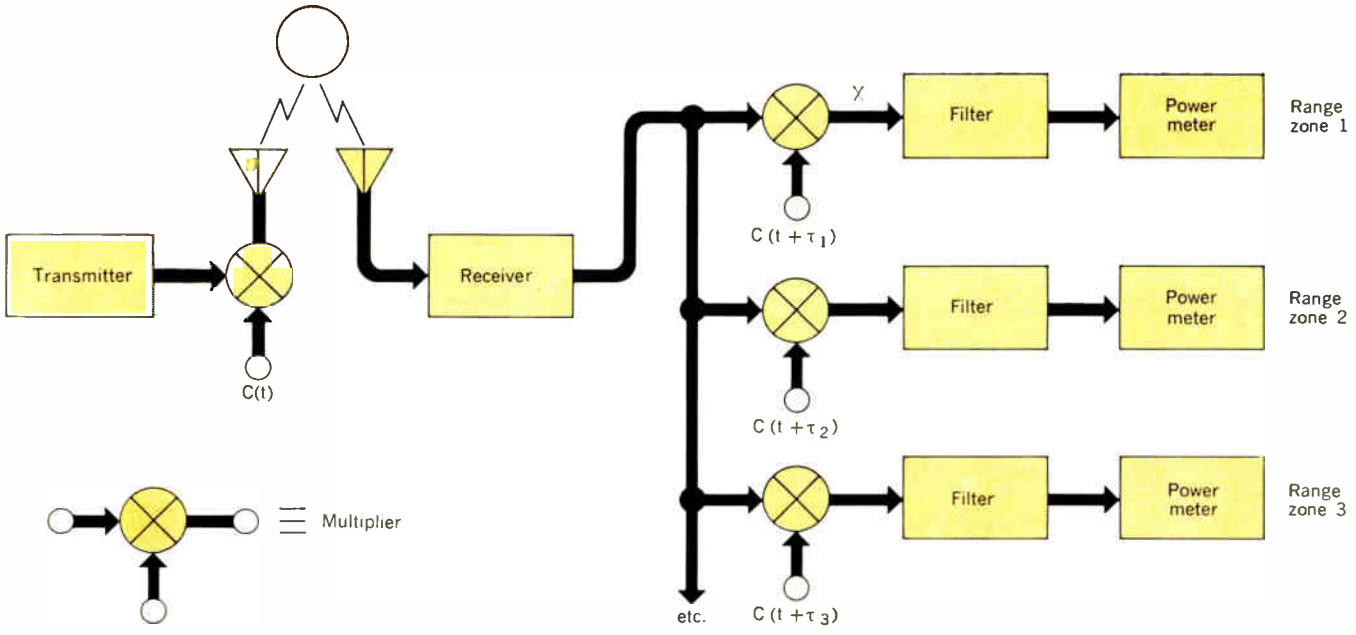


Fig. 6. Block diagram for an AM range-gate system.

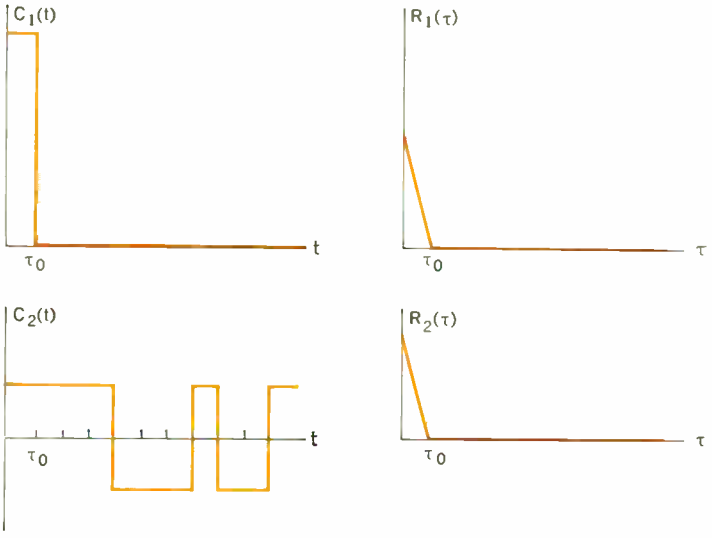


Fig. 7. Two modulating waveforms and their corresponding autocorrelation functions.

The signal power is then obtained by computing the noise power from a measurement of the system noise temperature and the known bandwidth. The *quality* of the determination is given by (32).

Range

The measurement of range is essentially a problem of estimating time delay. Here the transmitter is modulated by some waveform $C(t)$, and the signal returns after it has been delayed by the round-trip time of flight. The signal has been distorted by reflection from the target and it has been corrupted by noise at the receiver. We must estimate the time delay of this signal relative to the transmitted waveform.

Detection theory tells us (under certain simplifying assumptions) that the best technique is to compare the received signal with the possible signals (of various time delays) that might be expected. The closest comparison then gives the best estimate of the time delay. Of course, even when the assumptions are not strictly met, comparison with the expected signals is always good strategy. An example of this technique is given in Fig. 6. Although amplitude modulation is shown, frequency, phase, single-sideband modulation, etc., could be, and have been, used just as well, with a suitable change in the block diagram. The analysis would be similar. Comparison, for this type of modulation, takes the form of correlation.

Each of the power meters records the power that returns from the neighborhood of its particular time delay, τ_i. Certain design considerations must now be formulated for the modulating signal $C(t)$. In order for each power meter to respond only to its specified delay τ_i (range zone) and not to other delays, to eliminate cross-talk between zones, $C(t)$ must have a special autocorrelation function. The autocorrelation function $R(\tau)$ is defined as

$$R(\tau) = \text{average } [C(t)\, C(t+\tau)]$$

The desirable $R(\tau)$ for ranging must have a narrow peak at $\tau = 0$ and be near zero for $\tau > \tau_0$. An example for two waveforms of $C(t)$ is given in Fig. 7.

$C_1(t)$ is simply a narrow, rectangular pulse. $C_2(t)$ is a random square wave, such that after each time interval of length τ_0, a random choice is made as to whether $C_2(t)$ will be $+1$ or -1 for the next time interval. Both of these waveforms have the same $R(\tau)$, and they both result in the same time-of-flight resolution, τ_0; but they lead to entirely different radar hardware. $C_1(t)$ corresponds to a pulse radar, where the power output is concentrated into extremely narrow and high pulses. $C_2(t)$ corresponds to a CW radar, where the power remains uniform throughout the cycle.

The spectra of the two systems at x in Fig. 6 also differ markedly. The spectra are modified by the characteristics

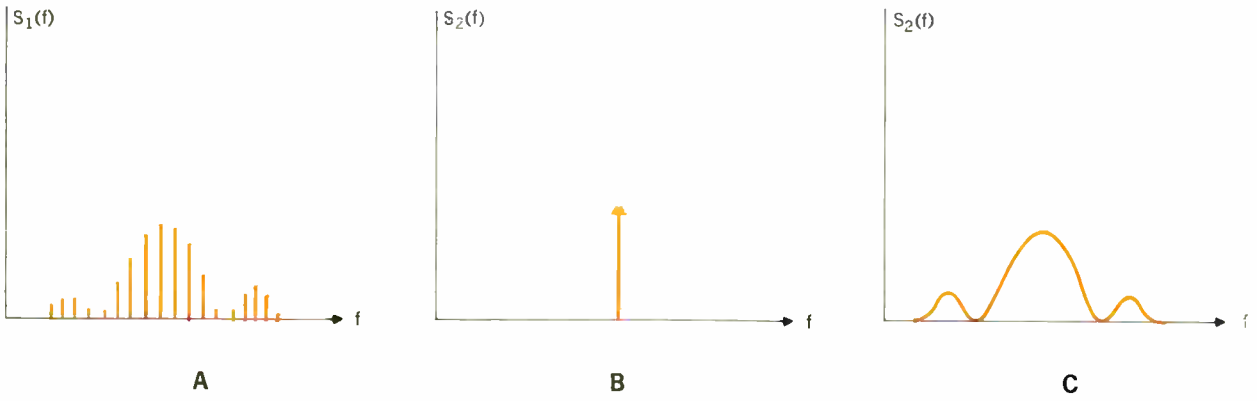


Fig. 8. Spectra of the received signal from a point reflector for (A) pulse radar with the signal in the correct range gate, (B) CW radar with the signal in the correct range gate, and (C) CW radar using an incorrect range gate.

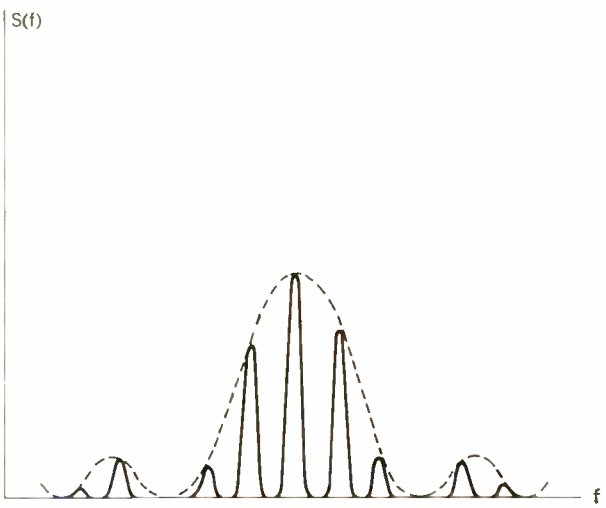


Fig. 9. Ideal Doppler-spread spectrum for a pulse radar (range zone through subearth point).

of the target, but for now we will consider the ideal case of a perfect mirror for the target. More will be said of the general case subsequently.

The pulse radar will produce a line spectrum at point x, for the range zone which contains the signal. The envelope of the spectrum depends on the shape of the transmitted pulse, and the line spacing depends upon the repetition rate. This is illustrated in Fig. 8(A). Of course, range zones that do not contain the signal have no spectrum except for the system noise, which appears in all of the zones. It can be seen that the optimum filter for the pulse radar is a "comb" filter, passing each line of the line spectrum, but eliminating all of the noise between lines.

We shall mention here that the block diagram of Fig. 6 is more complicated than necessary, since all the range channels can be mechanized by one set consisting of a filter, power meter, etc., which are time-shared. We have shown it this way to make the discussion of the "mapping" technique easier to follow.

The spectrum for the CW radar, for the correct range zone, is narrow band. The receiver multiplier serves to cancel out exactly the modulation impressed upon the signal at the transmitter. The incorrect range zones have a wide-band signal, similar to the spectrum that is transmitted. Both of these spectra are also illustrated in Fig. 8.

The optimum filter for the CW method is a narrow bandpass filter. Very little of the wide-band power from the wrong zones gets through the bandpass filter; and that amount, in practice, is negligible compared to the thermal noise, which is always present.

One other similarity between pulse and CW radar is important. For both of these systems, it is the average transmitter power, not the peak power, which is the important design parameter determining the quality of the range measurements.

Range and velocity mapping

The range gates and the spectrometers described previously can be combined to produce two-dimensional (range–velocity) maps of the surface of the target. The mechanization is simply to replace each power meter of Fig. 6 by a spectrometer. Each spectrometer, then, accepts power from only one range zone and subdivides this power into its frequency content.

The target can no longer be thought of as a perfect mirror for this experiment; its depth and rotation must be taken into account. Indeed, those are the properties that allow a map to be made by these techniques. All of the range zones that are placed on the target will contain some power. Because of the rotation, this power will no longer be narrow band, but will be spread into a spectrum. The filters in the range gates must be widened to allow the power to pass through.

In the case of the pulse radar, each line of the line spectrum is spread by the target. An example is given in Fig. 9 for the case of the range zone passing through the front cap, or subearth point. A design restriction appears in Fig. 9. The pulse repetition rate must be high enough that there is no overlapping of the individual spectra of Fig. 8, but low enough that echoes from the full depth of the planet do not interfere with those from the selected range zone. There is no corresponding restriction for the case of the CW radar.

An example of actual data for the CW radar case is given in Fig. 10. The target was the planet Venus at a distance of 30 million miles from earth. Eight range zones are shown, with a spacing between them of 11 miles. Zone 1 is centered about the front cap of Venus and each

successive zone contains energy from 11 miles farther back on Venus. The slight asymmetry of the spectra of zones 2 and 3 can be related to a topographical feature on the Venusian surface near the subearth point.

Resolution in the range dimension ΔR is set by the speed of the modulating waveform $C_2(t)$ and is roughly

$$\Delta R = \frac{c}{2\tau_0}$$

where c is the speed of light and τ_0 is the switching interval of $C_2(t)$. Resolution in the frequency dimension is limited by the length of time the transmission lasts.

Theoretically, these resolutions can be increased simultaneously to the limit of the stability of the oscillators of the system. In practice another limit, the signal-to-noise ratio, is often reached first. As the power from the target is divided into finer and finer "cells," the echo from each must become weaker until they are no longer distinguishable from the ever-present background noise.

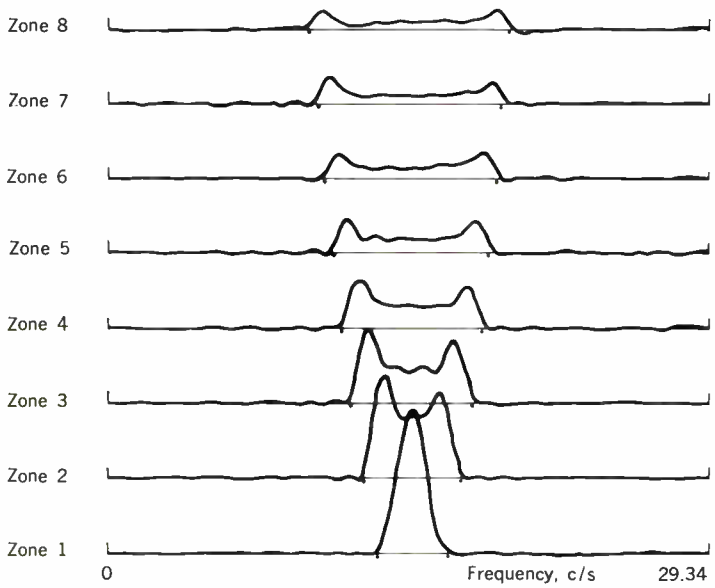
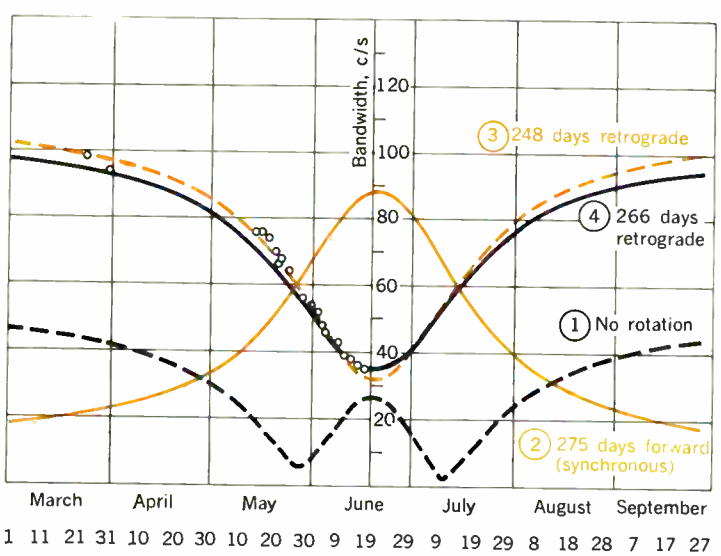


Fig. 10. An example of Venus range-gated spectra taken with the Goldstone radar system (June 6, 1964, eight 11-mile range zones).

Fig. 11. Change in total spectral width with time for several possible rotational vectors of Venus.



Spectral analysis—CW systems

The study of the planets by the use of CW radar techniques is, in principle, relatively simple compared with the pulse and modulation methods. A pure RF signal is transmitted to the planet and the spectrum of the resulting echo is derived. The echo spectrum may be thought of as a one-dimensional map of the surface, assuming, of course, that the reflection occurs from the true surface and not the planet's ionosphere. This interpretation of the spectrum is a consequence of the fact that for any distant rotating object the regions of constant radial velocity (and hence Doppler frequency) consist of straight lines or zones parallel to the axis of apparent rotation. The pertinent geometry for the problem is shown in Fig. 2, and the necessary equations are (15) through (19). If the spectrum of the echo signal has a resolution of df_d c/s at a frequency f_d measured from the center of the spectrum, then the power in this frequency band is given by the integrated reflected power from the strip on the planetary surface of angular width $d\theta$, see (18), as indicated in (19). The center zone, along the planet's axis, has zero radial velocity and usually corresponds to the peak power of the spectrum. Zones farther removed from the center have greater radial velocities (of either sign) until the limb of the planet is reached. The relative power from zone to zone is equivalent to a one-dimensional map measured parallel to the planet's rotational equator.

The rotation of Venus has long been a mystery, and its determination is an important objective of radar astronomy. By observing Venus radarmetrically at different points in its orbit, the true rotation period and axis can be found with the aid of spectral analysis techniques. This is possible because the limb-to-limb bandwidth of Venus, see (16), changes as Venus passes the earth in space. Figure 11 shows the limb-to-limb bandwidth that would arise for three different periods of rotation. Curve 1 is for no rotation at all—that is, for Venus not rotating relative to the stars. In this case the change in the bandwidth is due to an apparent rotation as Venus passes the earth-based radar. Before conjunction we see one side of the planet and as it passes us we see more and more of the other side. Curve 2 is for the case in which Venus always keeps the same face toward the sun. This is usually referred to as synchronous rotation because its rotation period is synchronized or equal to its period of revolution around the sun. Note that in this case the bandwidth reaches a maximum at conjunction—that is, when Venus passes nearest to the earth. Curve 3 shows the bandwidth assuming a retrograde rotation period of 248 days. Note that here the bandwidth is a minimum at conjunction. For all three cases it was assumed that the rotation axis of Venus was perpendicular to its orbit. For other axis orientations, the curves have the same general shape but become more distorted and asymmetrical. By comparing the observed bandwidths with different theoretical curves, both the period and axis can be found. Some preliminary values from the 1964 JPL measurements are also shown in Fig. 11.

As described earlier, the CW spectrum is a one-dimensional map of the surface. By examining how these maps

change with time, we can determine the location of surface features that scatter back more (or less) power than surrounding regions. These features will appear as statistically significant "bumps" in the observed spectra. As the planet rotates, different parts of the surface will pass through the zones of constant Doppler shift. Hence, if a "bright" surface feature is present, it will move from zone to zone and show up on the spectrum as a bump or feature that slowly moves from the high-frequency to the low-frequency side of the spectrum. If the rotation period and axis of the planet are known, then the Doppler history of a feature can give its location on the planet.

An example of a spectral feature is shown in Fig. 12. This is one of the best CW spectra obtained during the 1964 conjunction of Venus. The feature is at −10 c/s. It is, of course, small in comparison with the central peak; however, its amplitude is some 18 standard deviations above the background noise. This particular feature persisted for several weeks, showing up on each of the daily spectra that were taken. In addition, it slowly moved down the side of the spectrum and eventually disappeared.

A remarkable demonstration of spectral features is exhibited by the depolarized CW spectra. When a circularly polarized wave is reflected from a smooth surface, the sense of polarization is reversed; consequently, the receiving antenna must be matched to this polarization in order not to degrade the received power. However, reflections from a rough surface will partially depolarize the signal so that both right and left circularly polarized components are present. If the receiving antenna is in its mismatched polarization mode, a depolarized spectrum is obtained. Figure 13 shows the depolarized spectrum for June 13, 1964. The received power is much less than in the matched polarization case and, therefore, the relative background noise is quite large.

Two large peaks, which are statistically significant, can be seen on the spectrogram of Fig. 13. These peaks correlate strongly with spectral features observed on the same day with matched polarization. Furthermore, they were found to move across the spectrum from day to day in correspondence with the planet's rotation.

The true nature of these features can only be surmised. Nevertheless, it appears likely that they represent actual physiographic entities on the surface of Venus and that their motion is the result of the planet's rotation.

The measurement of CW spectra

The measurement of the CW spectrum is accomplished by the JPL group in two ways: One is to heterodyne the received signal down to 1 kc/s, filter it with a bandpass filter, and compute the spectrum with a small general-purpose digital computer located at the radar site. The output of the bandpass filter is sampled at twice the folding or Nyquist rate corresponding to the bandwidth of the filter used. The effect of sampling is to generate a series of spectra each symmetrically placed about multiples of the sampling frequency, including zero frequency. If we now compute the spectrum of the sampled signal from zero frequency up to the lowest subharmonic of the folding frequency—that is, half the sampling frequency—we obtain the spectrum of the signal in the bandpass filter. Because of this effect of sampling, the signal does not have to be heterodyned down to the desired low-pass frequency range before the spectrum is computed.

The other method is to heterodyne the signal down to *near* 455 kc/s and filter it with a 500-c/s bandwidth IF mechanical filter and then heterodyne it still further to 100 c/s. The signal is positioned at approximately 100 c/s from the skirt of the IF filter so that when it is heterodyned to 100 c/s, the image of the audio spectrum is sup-

Fig. 12. Observed spectrum of Venus, with enhanced return in left shoulder of spectrum (June 16–17, 1964).

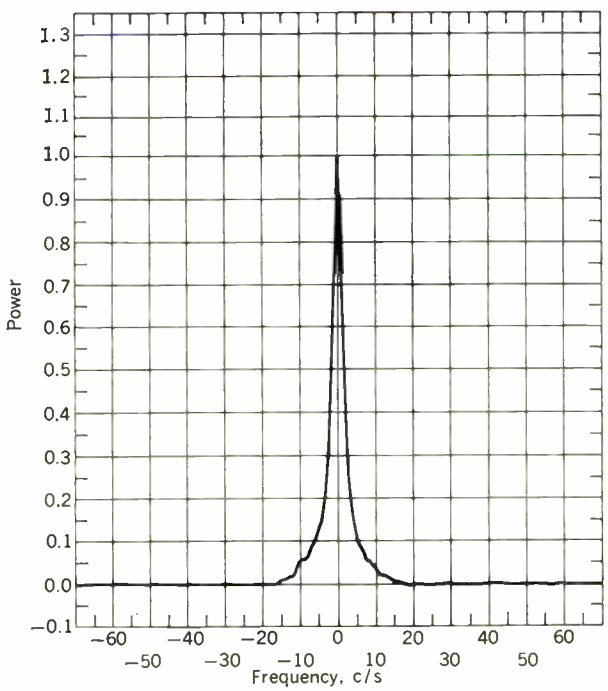
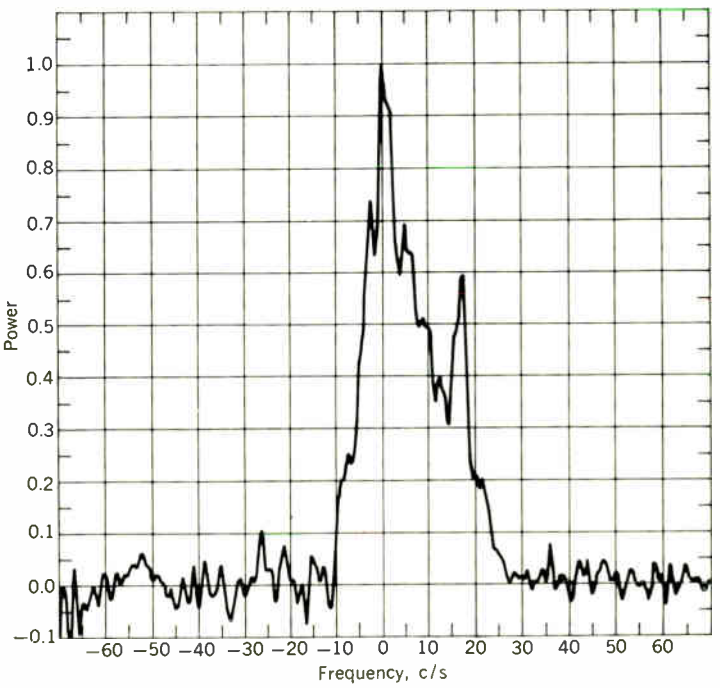


Fig. 13. An observed "depolarized" spectrum of Venus (June 12–13, 1964).



pressed; that is, there is very little noise power folded around zero frequency onto the positive frequency side of the spectrum. A sharp-cutoff low-pass filter is then used to eliminate all frequencies above 200 c/s. This procedure is possible in the cases of Venus, Mercury, and the moon due to their slow rotation. The audio signal is then sampled, digitized, and recorded on magnetic tape in the correct format so that the spectrum can be computed on the IBM 7094 computer located at JPL.

The frequency of the received signal, of course, changes continuously due to the relative motion between the radar and Venus. To remove this effect, a programmed local oscillator (PLO) is changed continuously in accordance with a precomputed estimate of the Doppler frequency derived from the ephemerides of the earth and Venus. The PLO keeps the signal's frequency constant to within $\pm \frac{1}{2}$ c/s. There is an offset control on the PLO that shifts the frequency by any desired amount and allows the positioning of the signal anywhere in the bandpass of the IF amplifiers. This offset control is used to place the signal near one side of the IF bandpass filter as described earlier.

The amount of signal integration necessary to produce a useful CW spectrum is extensive and it is essential that the spectral computations be as rapid as possible. This is true whether the on-site computer or the IBM 7094 computer is used. One technique makes use of the fact that the spectrum of a clipped or hard limited Gaussian signal is in one-to-one correspondence with the unclipped spectrum. This remarkable mathematical fact was first demonstrated by J. H. van Vleck in a study he made of the spectrum of limited noise.[8] He showed that if a Gaussian-type signal has an autocorrelation function of $\rho(\tau)$ then the autocorrelation function of the limited signal is

$$R(\tau) = \frac{2}{\pi} \sin^{-1} \rho(\tau)$$

Therefore, if we limit the signal, compute its autocorrelation function, solve the above equation for $\rho(\tau)$, and then take its Fourier transform, we have the spectrum of the original signal. The advantage of using the limited signal is that it is considerably faster to compute the autocorrelation function of a sequence of plus ones and minus ones obtained by clipping than to utilize the full measurement of the amplitude of the signal at each sampled point.

When using the on-site computer for spectral analysis a limiter is used as an analog-to-digital (A/D) converter, quantizing to only two levels. A specially constructed digital correlator can compute 45 points on the autocorrelation function at a maximum sampling rate of $66\frac{2}{3}$ kc/s and can integrate at this rate for weeks without overflow. The Fourier transformation is done on a small general-purpose computer (PB 250) and the spectrum is plotted whenever desired. This on-site system has the important advantage of being able to compute and plot the spectrum in real time. It has the disadvantage of being able to compute only 45 points on the autocorrelation function, which means that it is not possible to compute a spectrum with more than 45 resolvable frequency elements.

When using the off-site IBM 7094 computer, there is usually a lag of two or three days between the recording of the signal and computation of the spectrum. The signal is first recorded on magnetic tape by means of an A/D recorder, which is capable of sampling the signal from 375 samples per second to 33 300 samples per second. The resulting magnetic tapes can be directly read into the IBM 7094 computer. A special program was constructed to compute the spectrum utilizing either the full amplitude resolution of the A/D recorder (16 bits), eight levels of amplitude resolution (4 bits), or two levels of amplitude resolution (binary). Most of the spectra are computed with the binary mode.

As an example of the speed with which a spectrum may be computed, consider a typical signal recording session and the spectral computation. A CW signal is transmitted to the planet for the duration of the signal's round-trip time. For Venus near conjunction, this is about 5 minutes. The radar is then switched to its receive mode for the same duration to receive the reflected signal.

After the signal run, a noise run is recorded with no signal present. It is used to determine the spectral characteristics of the entire radar system and to correct the signal spectrum accordingly. A total of perhaps 15 signal-noise run pairs, which represents $2\frac{1}{2}$ hours of actual station time, usually constitutes a recording session. The computation of a spectrum with one-cycle resolution over a bandwidth of 185 cycles requires the computation of 370 points on the autocorrelation function. It takes approximately 30 minutes on the IBM 7094 to do the calculations and plot the spectrum when the binary mode is used. To perform the same spectral analysis using the full 16-bit samples of the recorded signal would require about 9 hours of computer time. The 4-bit (eight-level) mode would require about 3 hours.

In actual practice, the computer time required for the 4- and 16-bit modes is not as different from that of the two-level spectrum as it appears. When the binary mode is used, approximately $2\frac{1}{2}$ times more data must be recorded in order to obtain as smooth a spectrum as when the 16-bit mode is used. This requirement arises because the noise on the spectrum is increased by a factor of $\pi/2$ when the limited signal is used and, therefore, $(\pi/2)^2$ or 2.46 times more signal time is needed to reduce the noise to the same level.

Power spectral analysis

The calculation of a power spectrum can be done in two ways: we may compute the spectrum directly by taking the Fourier transform of the data itself or we may do it indirectly by first computing the signal's autocovariance function (the unnormalized autocorrelation function) and then taking its Fourier transform. The power spectrum $P(f)$ is found from the signal $X(t)$ by

$$P(f) = \lim_{T \to \infty} \frac{1}{T} \left| \int_{-T/2}^{+T/2} X(t)\, e^{-j2\pi ft} dt \right|^2 \qquad (33)$$

This equation gives the power spectrum directly. The indirect approach entails first the computation of the autocovariance function $C(\tau)$ of the signal:

$$C(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{+T/2} X(t)\, X(t + \tau)\, dt \qquad (34)$$

The power spectrum is then

$$P(f) = \int_{-\infty}^{\infty} C(\tau)\, e^{-j2\pi f\tau} d\tau \qquad (35)$$

$P(f)$ is the "two-sided" power spectrum and exists for

both positive and negative frequencies. For the power spectrum defined for positive frequencies only, one must use $2P(f)$ for $0 \leq f < \infty$.

Now, in any real experiment the signal cannot be of infinite duration as the definition of $C(\tau)$ appears to require. Thus, the spectrum of a signal of finite duration is an approximation to the "true" spectrum. Since we cannot integrate for an infinitely long time, the spectrum will be corrupted by noise, which will show up as random fluctuations across it. Also, we cannot compute the spectrum with infinite resolution, and hence it will be blurred. This brings us to consider two aspects of spectral analysis that are necessary before spectral computations are made. These are the stability of the computed spectral estimates (the points on the spectrum) and the frequency resolution. This problem has been analyzed in detail by Blackman and Tukey.[9]

Stability is a measure of the smallness of the variability or noise associated with the spectral estimates. The longer the signal is integrated, the smaller the resulting noise. Care must be taken to distinguish between the noise on the spectrum and the spectrum of the noise. Consider the calculation of the spectrum of a noise signal that is known to be flat. If we could integrate for an infinite length of time, the computed spectrum would be perfectly flat and smooth. However, in the practical case where the integration time is finite, the spectrum will not be smooth, but will show random fluctuations. These fluctuations are commonly referred to as the noise on the spectrum.

If σ is the rms deviation of the fluctuations on a spectrum having a "true" value M, the stability S is

$$S = \frac{M}{\sigma} = C\sqrt{WT}$$

where C is of the order of unity, W is the frequency resolution in cycles per second, and T is the duration of the signal in seconds. We have found both theoretically and experimentally that when the binary mode of spectral analysis is used, $C \approx 2/\pi$. For the 4-bit mode, it has been found experimentally that $C \approx 0.8$. For the 16-bit mode, $C \approx 1.0$.

The blurring of the spectrum is determined by the frequency resolution W. It is related to the number of points or lags used in computing the autocovariance function $C(\tau)$ and the sampling rate of the signal. For sampled data, $C(\tau)$ becomes a sum called the mean lagged product:

$$C(\tau) = \frac{1}{n-\tau}\sum_{q=0}^{q=n-\tau} X_q X_{q+\tau} \qquad \tau = 1, 2, \ldots m$$

where n is the total number of sampled points and m is the maximum number of lags. If the sampling rate is Ω samples per second, then the duration T_m corresponding to the maximum lag m is

$$T_m = \frac{m}{\Omega}$$

Since the frequency resolution is approximately equal to the reciprocal of T_m, then

$$W \approx \frac{\Omega}{m}$$

Any very narrow spectral feature will be broadened and have a half-power width of W cycles per second because of the limited number of lags that can be computed. This limit is imposed by such factors as the stability of the spectrum needed and the amount of computer time available for spectral analysis.

There is another consequence of limiting the number of lags. This is the effect of truncating the autocovariance function. When $C(\tau)$ is computed for only m lags we are in effect saying that $C(\tau) = 0$ for $\tau > T_m$. If $C_t(\tau)$ is the "true" autocovariance function and $D(\tau)$, called the lag window, is equal to one for $\tau \leq T_m$ and zero for $\tau > T_m$, then the computed autocovariance function is

$$C(\tau) = C_t(\tau)\, D(\tau)$$

This means that the computed spectrum $P(f)$ is the convolution of the "true" spectrum $P_t(f)$ and the so-called spectral window $Q(f)$:

$$P(f) = \int_{-\infty}^{\infty} P_t(g)\, Q(f-g)\, dg$$

The spectral window $Q(f)$ is the Fourier transform of the lag window $D(f)$. The convolution means that the value of each computed spectral estimate is an average of a region on the true spectrum. If the region includes a steep part of the true spectrum, then the spectral estimate can have a large error. The overall effect is to give the spectrum an oscillatory appearance. These oscillations are caused by the "side lobes" of the spectral window. They can be markedly reduced by judiciously choosing the form of the lag window $D(\tau)$. The idea is to choose a lag window so that the autocovariance function gently goes to zero as τ approaches its maximum value. A commonly used lag window, referred to as the hanning window,[9] is

$$D(\tau) = \frac{1}{2}\left(1 + \cos\frac{\pi\tau}{T_m}\right)$$

It contains a spectral window (the region of the true spectrum that contributes to a spectral estimate) that dies out rapidly and thus is not too influenced by large peaks some distance away on the spectrum.

(This article will be concluded in the November issue.)

This two-part article presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract NAS7-100, sponsored by the National Aeronautics and Space Administration.

REFERENCES

1. Browne, I. C., Evans, J. V., Hargreaves, J. K., and Murray, W. A. S., "Radar echoes from the moon," *Proc. Phys. Soc. (London)*, sec. A, vol. 69, p. 901; 1956.

2. Evans, J. V., and Pettengill, G. H., "The radar cross section of the moon," *J. Geophys. Res.*, vol. 68, p. 5098; 1963.

3. Pettengill, G. H., Unpublished lecture notes, Massachusetts Institute of Technology, 1960.

4. Goldstein, R. M., "Venus characteristics of earth-based radar," *Astron. J.*, vol. 69, p. 12; 1964.

5. Muhleman, D. O., Holdridge, D. B., and Block, N., "The astronomical unit determined by radar reflections from Venus," *Astron. J.*, vol. 67, p. 191; 1962.

6. Muhleman, D. O., "The relationship between the system of astronomical constants and the radar determinations of the astronomical unit," presented at IAU Symp. No. 21, Paris, France, 1963; to be published in *Bull. Astron. (France)*.

7. Goldstein, R. M., "The measurement of signal power," Space Programs Summary 37–28, vol. IV, Jet Propulsion Laboratory.

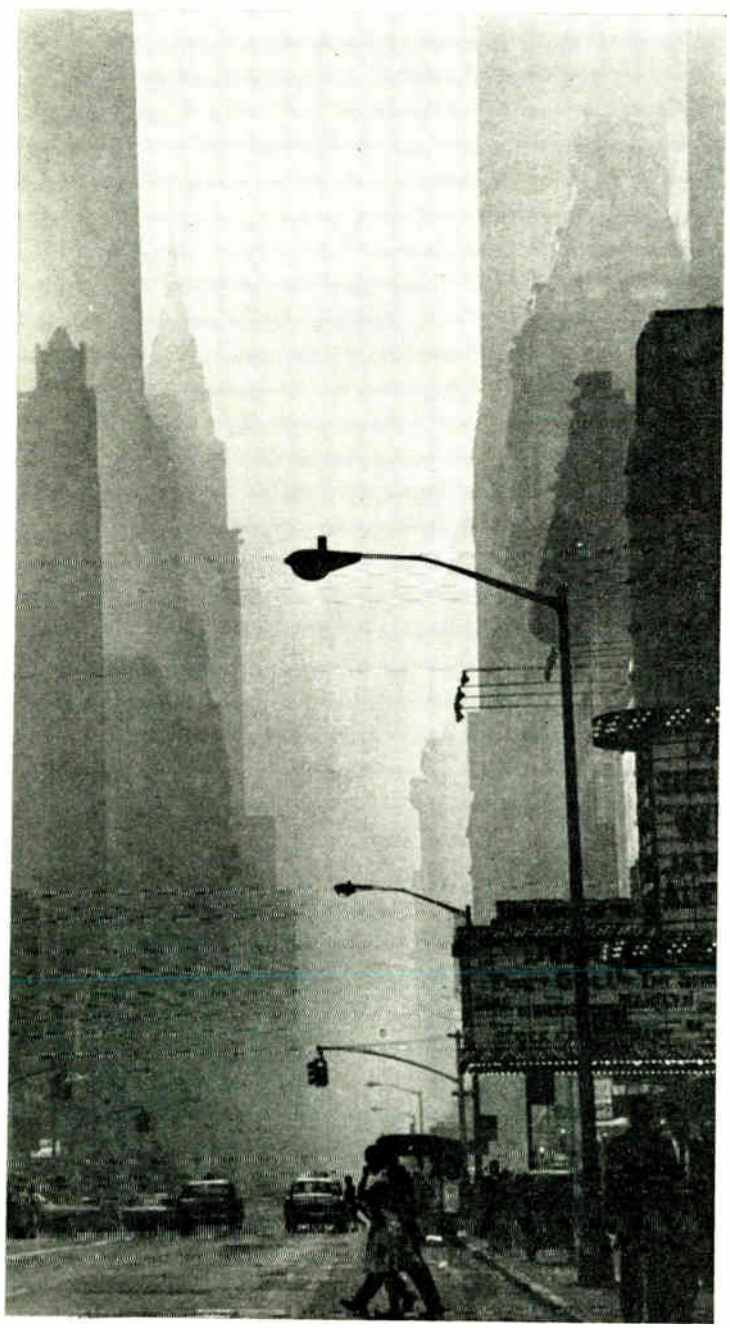
8. Lawson, G. E., and Uhlenbeck, J. L., eds., *Threshold Signals*. New York: McGraw-Hill, 1950.

9. Blackman, R. B., and Tukey, J. W., *The Measurement of Power Spectra*. New York: Dover, 1958.

Airborne asphyxia—an international problem

Air pollution is no longer a minor inconvenience; it is a tangible, frightening, and intolerable situation that affects virtually every major industrial city in the United States and Europe

Gordon D. Friedlander *Staff Writer*



Until quite recently, the smog in Los Angeles was the butt of every radio and television comedian. It was subjected to ridicule as the world's worst example of an out-of-control condition that was apparently without remedy.

But today the joke has gone sour. A recent—and disturbing—survey made by a special committee under the aegis of the New York City Council reveals that the type of air pollution—a dangerously high level of sulfur dioxide—in this huge metropolis, and particularly in the Borough of Manhattan, may be more hazardous to human life and health than that of Los Angeles.

It is estimated that the air is so foul over the island of Manhattan (see Fig. 1) that a monthly sootfall of more than 60 tons per square mile is commonplace during the winter. The City Council report alleges that most of the pollutants are belched out of the chimneys of industrial plants, electric utilities (see Figs. 2 and 3), and some 14 000 apartment house incinerators. Additional noxious fumes are spewed into the air from the tailpipes of thousands of motor vehicles that clog almost every square inch of city streets.

The commission's report asserts that breathing city air for one day would have the same effect on human lungs as inhaling two packs of cigarettes within the same time interval. So, as one wag put it: if one happens to be a habitual smoker in the two-pack-a-day category, and cannot shake the habit, it may be more healthful for one to give up breathing!

The sad story of air pollution, however, does not end with New York; Houston, Atlanta, Louisville, Chicago, St. Louis, Birmingham, and a host of other American and foreign cities are suffering the same atmospheric miseries (Fig. 4) in varying degrees and from various sources. In Germany, England, Belgium, France, and Luxembourg, where coal is the principal industrial fuel, the air pollution and smog problem is equally intense. Cities such as Essen, Dusseldorf, and other industrial centers in the Ruhr Valley are plagued by air

Fig. 1. "Darkness at high noon"—a view of 42nd Street in New York City under a blanket of heavy smog caused by a temperature inversion, on October 20, 1963.

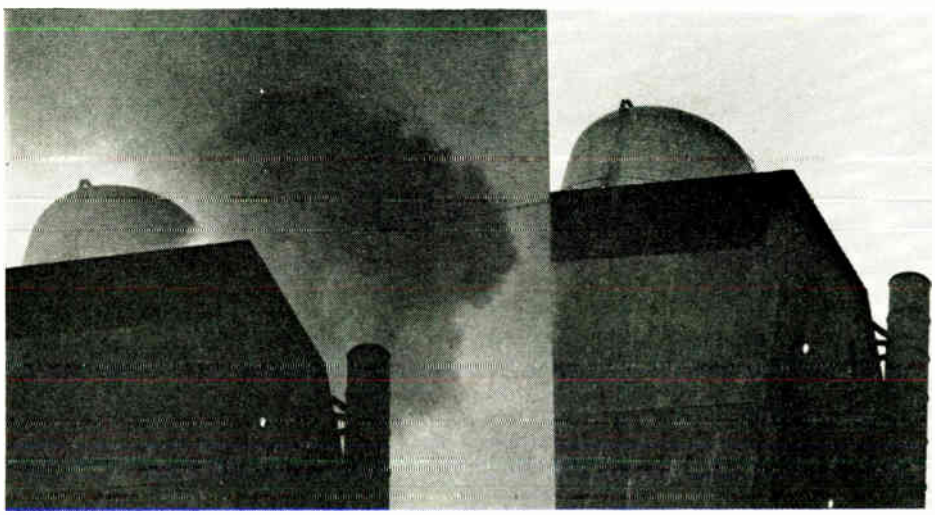


Fig. 2 (left). Photo showing stack operation of industrial plant before the installation of air pollution control equipment. (Right) Same stack after installation of control equipment. Fig. 3 (below). Aerial view of St. Louis showing a several-square-mile area of the city under a pall of industrial air pollution.

Fig. 4. Smoke belching from an electric utility company stack blankets an entire area.

contamination. In addition to London, the industrial midland cities of Birmingham, Manchester, Sheffield, and Liverpool suffer from the same malady. With the advent of the megalopolis and vast urban industrial sprawls, air pollution is no longer the exclusive property of any one country—it is rapidly becoming a worldwide menace. Later in this article we will discuss in detail the individual air pollution problems of five American cities and what efforts—if any—are being made to remedy the conditions.

The Fig. 5 bar graph dramatically indicates the major air contaminants and quantities emitted per year in the United States, while Fig. 6 shows the amount of suspended dusts and other particulates in relation to urban population classifications. Note that there is a steady rise in the number of particulates with every increase in population classification.

Mining, meteorology, and medicine

An entire session at the recent American Power Conference in Chicago clearly indicated that air pollution has become a prime interdisciplinary concern to both the electric utilities and industry. The participants in the panel discussion[1] on this subject included Dr. Bertram D. Dinman, Associate Professor of Preventive Medicine at Ohio State University; Francis E. Gartrell, Assistant Director of Health, Tennessee Valley Authority; Abraham Gerber, secretary of the System Development Committee, American Electric Power Service Corporation; James R. Jones, combustion engineer, Peabody Coal Company, Inc.; Harry Perry, Director of Coal Research, Bureau of Mines, U.S. Department of the Interior; and Dr. L. A. Ripperton, Associate Professor of Air Hygiene, University of North Carolina. This diverse group of panelists was moderated by T. T. Frankenberg, consulting mechanical engineer, American Electric Power Service Corporation, and R. L. Ireland of the Consolidation Coal Company, Inc., Cleveland, Ohio. Many of the following facts and figures were presented at this symposium.

Coal, CO, SO_2, and temperature inversion. At the present time, coal is used for the production of about 55 percent of the electric power in the United States, and electric generation represents almost half of the total domestic market for coal. And, although nuclear energy will be used increasingly for power production, estimates projected to the year 2000 indicate that the electric utility industry will require about 40 percent of the total available energy—the equivalent of 1.6 million tons of bituminous coal. Of this amount, coal is expected to account for at least 600 000 tons, even at the optimum rate of nuclear power introduction. This level of coal use would be 25 percent more than the total amount of coal mined in the United States in 1964. These statistics indicate the continuing importance of coal in providing a significant portion of the nation's total energy requirements.

Realistically speaking, any restrictions on the use of coal for electric generation would have a serious economic impact on the welfare of the coal-producing regions of the United States, and it would impose cost penalties for present and future power generation.

Air pollution, however, presents one of the most serious threats to the coal industry in the fulfillment of its predicted future as a source of fuel for electric generation. The 2000-MW conventional steam-electric power plants, presently in the design stages, will consume about 20 000 tons of coal per day. This could mean that, in addition to a large volume of carbon monoxide being produced through incomplete combustion, between 700 and 800 tons of sulfur would be burned to produce an intolerable level of sulfur dioxide.

In the vicious cycle wherein urban population densities are increasing, the quantities of energy required in our burgeoning economy will also spiral upward. At some predictable time it will be necessary to decrease the quality of coal used for electric generation as the reserves of high-grade fuel dwindle. And, since the lower-grade bituminous coals contain a higher sulfur content, the problem of air pollution control will be intensified.

Temperature inversion is a meteorological phenomenon which, when occurring over large cities, can have very serious consequences. Essentially, temperature inversion is a perverse atmospheric condition in which the air temperature increases with height above the earth's surface. Normally, temperature decreases with height in the lower atmosphere up to the troposphere, and then the temperature increases in the stratosphere. The rate of decrement, or *lapse rate*, is about 3.3°F per 1000 feet of altitude.

Inversions are caused by radiative cooling of a lower air layer, subsidence heating of an upper layer, or the advection of warm air over cooler air or of cool air under warmer air. Radiative exchange between the earth's sur-

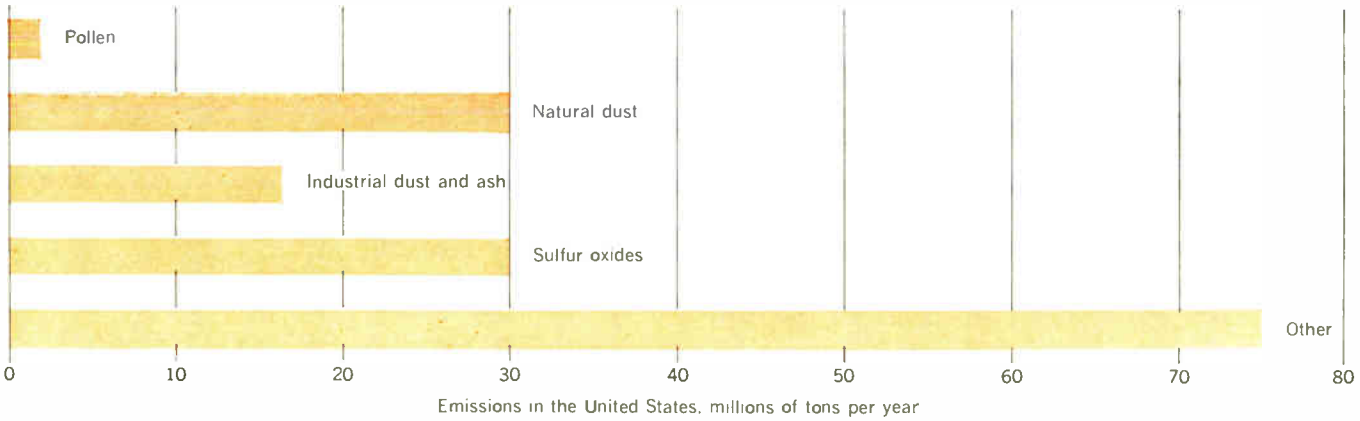
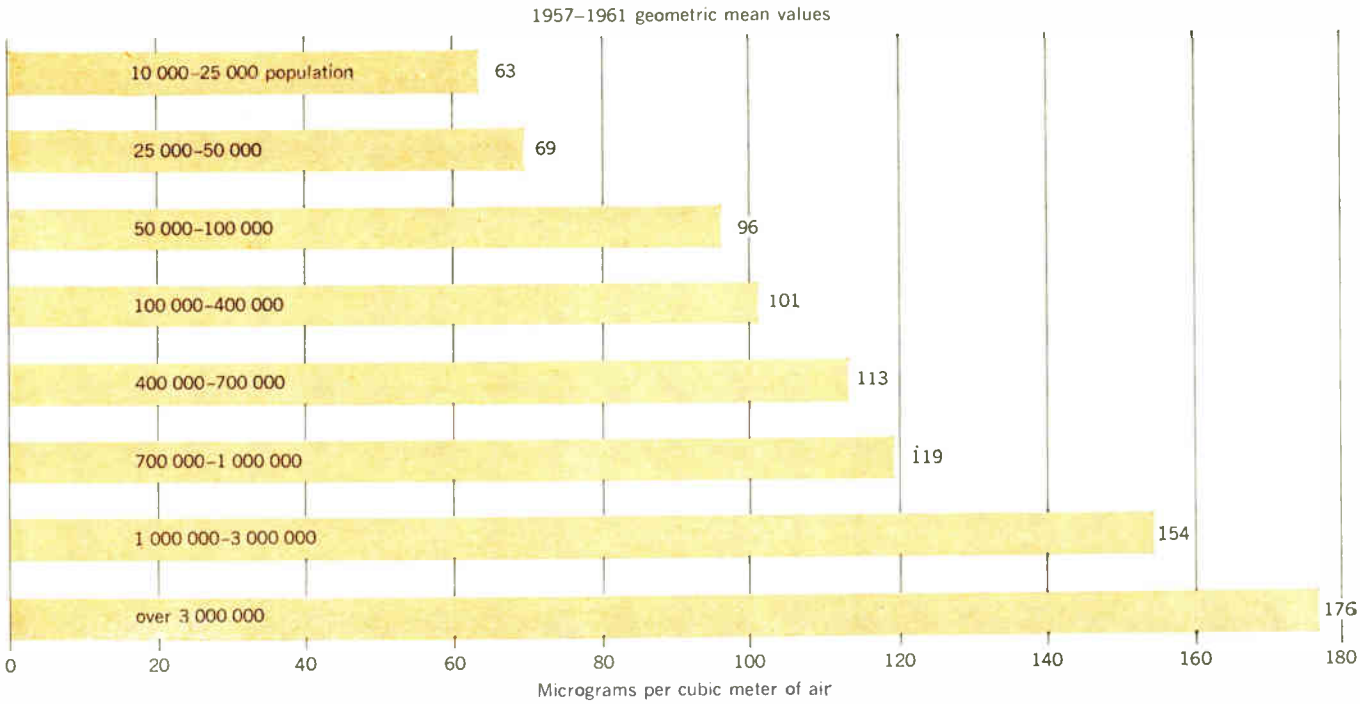


Fig. 5. Horizontal bar graph showing principal major air contaminants that are emitted per year in the United States.

Fig. 6. Graph showing the quantities of suspended particulates and their relationship to urban population concentrations.



face and the atmosphere on clear nights cools the ground and the adjacent layer of air. This makes the adjacent layer colder than the layers immediately above, and thereby creates a ground inversion that can vary from a few feet to a few thousand feet in thickness.

Radiative cooling of the top of a cloud bank or dust layer can also create an inversion. In this case, the sinking air warms at the adiabatic lapse rate of 5.5°F per 1000 feet, and this sinking air can produce a layer warmer than the layer of air that is immediately adjacent to the earth's surface.

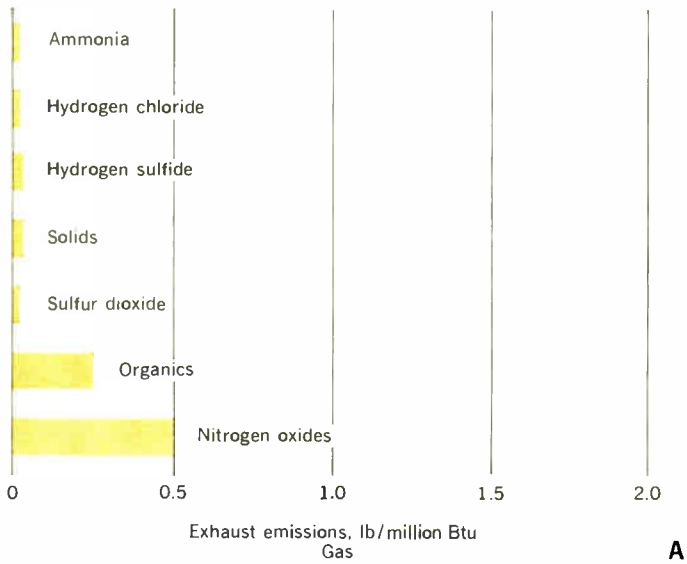
Cool air that displaces warmer air, such as air that blows from a cool ocean onto a warmer land, can cause a pronounced inversion that persists as long as the flow continues. Similarly, warm air may flow over a cold surface layer, especially one trapped in a valley, and this may cause an inversion. The episodes of acute air pollution in the Meuse River Valley of Belgium, in 1930; in Donora, Pa., in 1948; and in London, England, in 1952 were caused by this latter phenomenon.

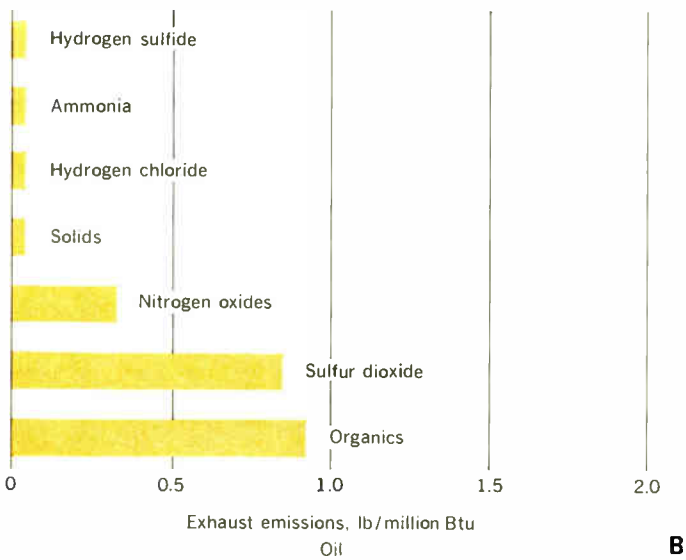
Inversions effectively suppress vertical air movement and cause an atmospheric stagnation in which smoke and other volatile contaminants cannot rise from the earth's surface. Persistent inversions have been experienced in Los Angeles, New York, London, and other industrial metropolises. Under such conditions, lethal layers of sulfur dioxide, soot, carbon monoxide, ozone, and nitrous oxide can become statically entrapped for days at a time.

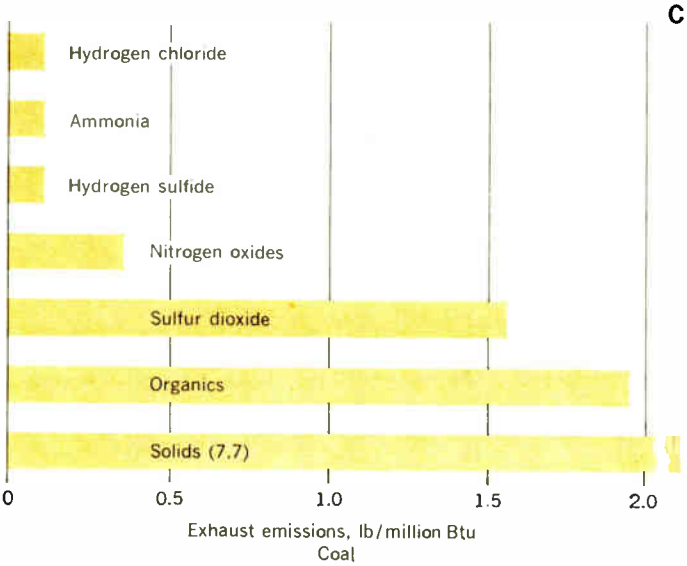
Climatological factors. There are meteorological and climatological factors that influence the action of airborne, volatile chemicals. The most common of these variables are temperature, wind velocity and turbulence, humidity, atmospheric pressure, and intensity and duration of sunlight.

Thermal reactions, involving the corrosion of materials,

Fig. 7. A—Chart of exhaust emissions from gas-fired furnaces. B—Similar chart for oil-fired furnaces. C—Chart for coal-fired equipment indicates highest contaminant emissions.







will approximately double in rate for each 18°F rise in temperature. While temperature is not usually considered to have an effect on photochemical reactions, recent research has indicated that photochemical oxidant production and the rate of photooxidation of hydrocarbons are accelerated by increased temperature in synthetic smog. And there is evidence to indicate that eye irritants could also be increased in the atmosphere by the elevation of air temperature.

It is known that when the air temperature is raised, the respiratory rates of humans and animals are increased. And the toxic effects of many pollutants are affected by temperature changes. For example, in one experiment, the lethal dose of ozone for rats in a temperature environment of 90°F was 2.6 ppm for a 4-hour exposure; while at 75°F, the lethal dosage for a 6-hour exposure was 6–8 ppm. This would seem to indicate that the permissible levels of air pollution should be revised in accordance with seasonal mean temperatures.

Humidity can influence the effects of air pollution. Its presence often causes more rapid corrosion of metals by certain chemical substances. Many acidic gases, such as sulfur dioxide, nitrous oxide, hydrogen sulfide, and chlorine, are much more corrosive in atmospheres that contain high humidity than they are in the presence of drier air. And since humidity directly affects the heat transfer between humans and their environment, it will, in turn, influence the effect of exposure on humans.

High wind velocity and air turbulence are generally beneficial because pollutants are dispersed and diluted more rapidly.

It is known that sunlight is an important factor in the effects of air pollution since eye irritants, plant toxins, and ozone are formed in the air by photochemical reactions. Air pollution experts are aware, for instance, that some types of soot deposited on motor vehicles will damage the lacquer in the presence of sunlight. If the soot is removed, however, before sunlight can touch the finish, no damage will result.

Atmospheric pressure has a relevant influence in air pollution. The oxygen pressure in the air decreases as the height above sea level is increased. The immediate physiological effect of increased altitude is a more rapid blood flow rate; then, the involuntary rate of respiration increases. As the body adapts to the new atmospheric environment, the concentration of blood hemoglobin rises. The ambient pressure in populated areas of the United States varies somewhat more than 0.2 atmosphere (assuming that 1.0 atmosphere equals about 15 psi). For relative comparison, the instantaneous rate for a given concentration of air contaminant may be expressed as $dx/dt = K$. At a higher elevation, such as Denver, Colo., with a pressure approximately 0.8 of that at sea level, the instantaneous rate would be $dx/dt = 0.64K$, or 36 percent slower. Therefore, pressure considerations are important in establishing standards that are designed to prevent the formation of secondary pollutants that are synthesized by either photochemistry or oxidation from primary contaminants.

Medically, a dim view. Sulfur dioxide is increasingly emerging as a prime villain in the air pollution drama. This contaminant is a major by-product of fossil-fuel combustion from the lower grade fuel oils and coal.

London's smog, a true smoke suspension in fog, has for many centuries been a prime example of traditional air

Fig. 8. Paint damage caused by air pollution. In this case, the pollutant was hydrogen sulfide.

pollution. Because England imports limited quantities of high-grade fossil fuels, the typical London "pea souper" consists mainly of sulfur compounds, particularly SO_2, which are produced by the combustion of bituminous coal, low-grade heating oil, acid manufacture, ore smelting, and other industrial manufacturing processes.

In the United States, New York and Chicago have record quantities of SO_2 in their atmospheres that are second only to London's. In all, about 60 percent of the American population is exposed to continuous peril from atmospheric contaminants (see Fig. 7). And it does not require a medical opinion to suggest that pollutants capable of corroding metal, darkening white paint (Fig. 8), disintegrating stone, dissolving nylon hose, and cracking rubber are somewhat less than beneficial to human lung tissue. There is ample circumstantial evidence to link air pollution with asthma, pneumonia, tuberculosis, pulmonary emphysema, lung cancer, and even the common cold. In 1962, the chairman of a panel of medical experts at the National Conference on Air Pollution Control stated: "The evidence that air pollution contributes to the pathogenesis of chronic respiratory disease is overwhelming."

During the symposium on the clean air problem at the recent American Power Conference, Dr. Dinman, in his opening statement, gave a concise description of the pathological effects of sulfur dioxide:

"To understand the effects of SO_2 on health, it is necessary to delineate those mechanisms whereby sulfur oxides alter human function. On a mechanistic basis, we may conceive of air conduction tubes (the tracheobronchial tree) as a series of interconnecting ducts. These ducts have the unusual capacity of changing their cross-sectional area. This is accomplished by contraction of circumferentially aligned muscles. Thus, given a proper stimulus at certain receptions in the wall of this air conduction system, input from these receivers arrives at the brain. A flow of impulses, in turn, is transmitted to these surrounding muscles, which leads to their contraction with a decrease in cross-sectional area.

"The consequences of such decrease in cross-sectional area are apparent. Consider that a fixed volume of air per unit time must be available for oxygen extraction by the blood. Therefore, an increase in velocity is the only method whereby this fixed volume may be moved through this attenuated system. Obviously, the energy required per unit time to obtain this work function is increased. In individuals who have cardiac disease, these increased demands are met with difficulty and subsequent deterioration. Another complication stems from one other consequence of SO_2 or SO_3 (sulfur trioxide) or H_2SO_4 (sulfuric acid) impingement on the lining of the gas–blood exchange surfaces. Impingement of these irritant polar compounds stimulates the release of a diluent at such affected surfaces. While this dilution increases pH toward normal levels, at the same time it imposes a *thickened barrier* to gas transfer across the membrane. Since this barrier is but a few hundred microns thick, this imposes no significant load on gas transfer in the normal person.

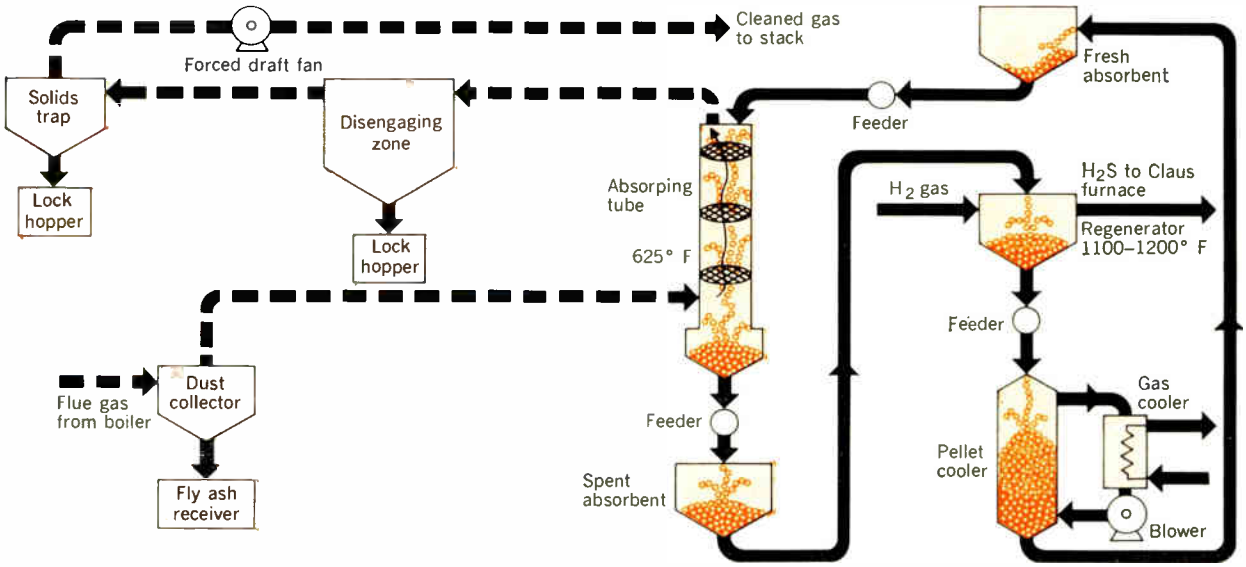


Fig. 9. Block diagram of the Bureau of Mines' alkalized alumina process.

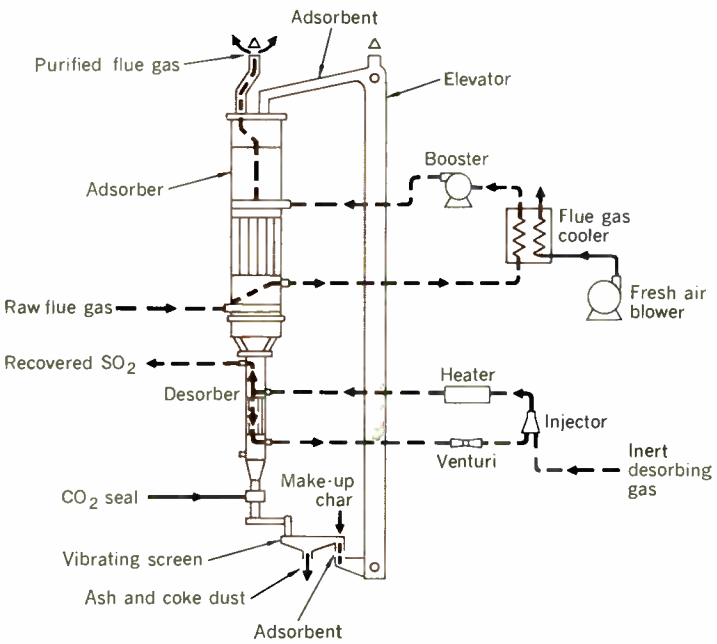


Fig. 10. Diagram showing the mechanical configuration of a Reinluft process pilot plant.

However, in those persons with already thickened gas exchange membranes due to chronic lung disease, this increased thickness essentially causes a barrier to oxygen diffusion with resultant asphyxia, and further deterioration as a result of decreased oxygen availability to the vital organs.

". . . While there is much data from animal experimentation, there is relatively little human data even in normal persons. *In such persons, 4–6 ppm of SO_2 produces consistently reproducible changes in airway resistance within 10 seconds to 4 minutes . . .*"

Nitrous oxide is no laughing matter. Some medical experts regard nitrous oxide (N_2O), the common "laughing gas" administered as an anesthetic by dentists for tooth extraction, in the same insidious category as SO_2. Recent evidence indicates that N_2O has teratogenic—and possibly carcinogenic—effects on animal and human receptors. In air pollution, this gas is emitted as a by-product of hydrocarbon combustion.

In therapeutic medicine, bone marrow depressant effects have been observed as a result of the protracted administration of N_2O to control convulsions in tetanus victims. And, although any conceivable concentration of the gas as an air pollutant would be very low, the cumulative effects of continued exposure to this gas may be deleterious to human receptors.

Catalysts and buffers. Dinman further observed that with the addition of any suspended particulates, of less than 5 microns in size, the response of the human receptor to SO_2 may be accentuated. The reasons for this physiological reaction are twofold, and may be ascribed to

1. The interposition of particulate surfaces of relatively large areas for irritant gas adsorption, which would increase the gas concentration per unit volume.
2. The increased probability of impingement, which is a result of the different kinetic behavior of particulate vs. gas phase.

Although many particulates tend to aggravate the biological damage potential of SO_2, it is apparent from actual case histories that the biopotency of these particulates is a function of their chemical properties. It is known that manganese catalyzes the conversion of SO_2 to H_2SO_4, but it is also possible for some particulates to buffer the physiological reaction of SO_2.

In the phenomenon of buffering it is known that inhaled hydrogen sulfide (H_2S) and other sulfhydryl compounds can protect mice from otherwise lethal exposures to ozone. Inhaled formaldehyde and SO_2 produced much more resistance to air flow in guinea pigs when applied with a physiologically inert aerosol than did the same concentrations of the gases alone.

Knowing the concentration of a pollutant does not necessarily indicate its physiological effect upon a receptor. The presence of other contaminants may either inhibit or increase the expected effect. Yet, almost no research has been conducted to determine the effects on re-

ceptors of long-term exposure to known admixtures of pollutants. Nevertheless, many proposed air standards for urban areas are predicated on consideration of the exposure effects of a single pollutant at a time.

The final medical factor is individual sensitivity or allergic reaction. Heretofore this has been a rather vague and nebulous concept, but recent evidence is accumulating to verify this phenomenon.

Methods of reducing SO_2 and SO_3 emissions from coal

Based upon statistics available for the year 1962, and by combining a knowledge of the sulfur content of coal seams being mined in various states, Table I indicates the range of sulfur contents and the corresponding percentages consumed in the United States.

In view of the very large estimated demand for electric power generation in geographic areas where coal is the preferred fuel, the quantities of SO_2 released to the atmosphere will rise alarmingly unless means are developed either to remove the sulfur before combustion or to remove the SO_2 from the stack gases. One of the prime difficulties in achieving the former objective is that a portion of the sulfur content (20–60 percent) is chemically bound as organic sulfur, and this can only be removed by very complex and expensive chemical processes.

At present, three high-temperature processes[1] are being operationally tested for electric utility applications. These are: the alkalized alumina, the Reinluft, and the Pennsylvania Electric.

Bureau of Mines' alkalized alumina process. In this process, flue gas containing SO_2 and SO_3 are absorbed by alkalized alumina—$Al(OH)_3$—in a vessel at a temperature of 625°F (see Fig. 9). The alkalized alumina is regenerated in a second vessel at 1200°F, by using producer gas or re-formed natural gas. The product gas from the regenerator is then introduced to a sulfur recovery plant in which elemental sulfur is produced.

The flue gas used in the Bureau's pilot plant is made from the combustion of powdered coal and it contains all the impurities that might affect the absorption and regeneration cycles. Tests have been conducted to establish the optimum conditions of temperature and time for both the absorption and regeneration, and various procedures for preparing the alkalized alumina have been tried because, in the repeated cycles, physical and chemical changes occur that may degrade or poison the absorbent.

The experimentation with this process has led to the construction of a larger pilot plant in which variables can be studied more efficiently. The advantages of the process include a low pressure drop of the flue gas during absorption, operation over a wide temperature range of 250°–650°F, and the ability to obtain elemental sulfur as the end by-product.

Reinluft process. In this method, flue gas at 300°F is forced upward through a filter bed of activated charcoal (see Fig. 10) that is slowly descending through the adsorber. The SO_3 is adsorbed directly and the flue gas is then cooled to 220°F. At this temperature, SO_2 is oxidized to SO_3, which is then adsorbed on the activated charcoal. The SO_3 combines with the adsorbed water from the flue gas to form dilute H_2SO_4.

The activated charcoal, with the adsorbed dilute H_2SO_4, is next regenerated in a separate vessel by the recirculation of product gas heated to 700°F. The dissociated H_2SO_4 products react chemically with a portion of the carbon to form a gas that contains a high concentration of C_2O and SO_2. The latter gas is converted to H_2SO_4 in a contact acid plant. After cleansing, the regenerated char is recycled to the adsorber.

At present, two commercial plants are under construction in Germany to use this method of SO_2 removal. One of these plants will service flue gases produced from low-grade fuel oil, and the other will be used in connection with a coal-fired installation.

The Reinluft process is particularly feasible if there is a nearby industrial requirement for sulfuric acid.

The Pennsylvania Electric process. The Pennsylvania Electric Company has constructed a pilot plant at its Sewart generating station to remove SO_2 by the catalytic conversion of SO_2 to SO_3. Sulfuric acid is formed and collected on the cooling water stream that contains the SO_3. The objectives of the pilot plant are fourfold:

1. To determine the effect of actual flue gas, with time, on catalyst activity.
2. To establish the degree to which the flue gases must be precleaned to prevent catalyst fouling.
3. To calibrate the rate of catalytic oxidation of SO_2 and flue gas pressure drop so that large-scale plants can be sized.
4. To determine removal methods for the acid and the quality of the acid produced.

Reports indicate that the pilot plant has been operated successfully and that adequate data have been gathered for the design and construction of a full-scale plant. Many variables, however, still must be investigated, such as the life expectancy of the equipment, required construction materials, and the character of the maintenance problems that will be experienced.

At the present time there is insufficient information

I. Quality of coal used by electric utility industry in the United States—1962

Sulfur Content, percent	Production, percent
0–1.5	32.0
1.5–2.0	9.0
2.0–2.5	6.6
2.5–3.0	19.7
3.0–3.5	23.0
3.5 plus	9.7

"Despite the evidence of our senses, only in relatively recent years have we recognized that pollutants in the air have a direct bearing on health. And there are still doubting Thomases who maintain that pollution cannot be linked to disease because no specific etiologic agent in the air has been identified as responsible for a specific disease."

—James E. Perkins, M.D.,
Managing Director,
National Tuberculosis Association

available to make accurate cost estimates for the comparison of the three methods. Each of the described processes has its unique advantages, but each requires additional development to verify critical process variables that have an important bearing on the eventual economics. Also, market studies are needed to evaluate the industrial requirements of the manufactured by-products.

Magnetic separation of sulfur. An interesting new process, called magnetic separation,[2] has been reported by the Russians. Essentially, this method involves a flash coking step in which the iron pyrites containing the sulfur fraction in the coal are modified into a magnetic form to simplify the separation procedure.

In this technique, pulverized coal is heated to about 640°F for a period of 2–5 minutes, during which time a jet of steam and air is blown through the heated coal. This treatment produces a surface layer of magnetized material on each pyrite crystal, and separation is made in a magnetic field of 10 000 gauss. By discarding this magnetic fraction, the sulfur content of the coal can be reduced by about 0.5 percent and the ash content by 50 percent. The U.S. Bureau of Mines is investigating this process to determine the feasibility of its application to the treatment of coal in the United States.

SO_2 monitoring and control—TVA experience

During the past 15 years, the Tennessee Valley Authority has added 52 coal-fired, steam-electric generating units, located in eight plants and ranging in size from 125 to 650 MW, to its power production facilities. And a single-unit plant of 900-MW capacity (Bull Run) is scheduled to go on the line in 1966.

Beginning with the first large steam-electric station, TVA conducted extensive air pollution studies at each plant site. The experience and data obtained by these studies have been applied in planning air pollution control at subsequent plants and for additional units in existing plants.

Long-term records of meteorological data and SO_2 concentrations from permanent monitoring stations have

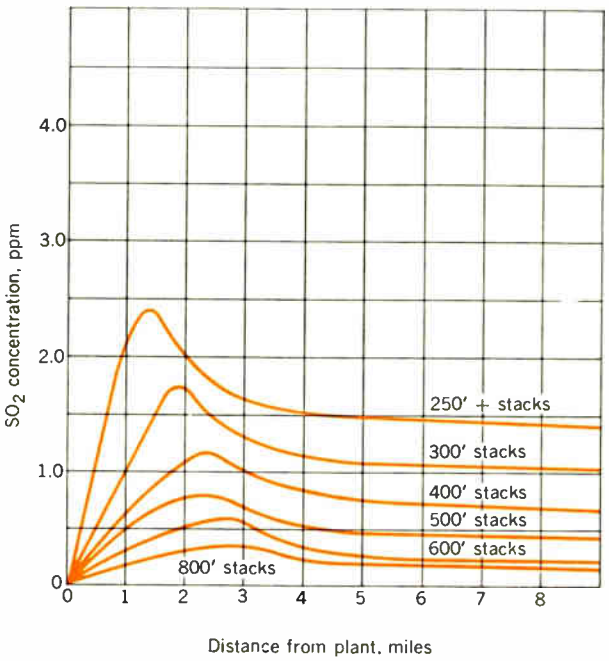


Fig. 12. Calculated profiles of SO_2 ground level concentrations, with various stack heights, for power plant with two stacks and an SO_2 emission rate of 810 tons per day.

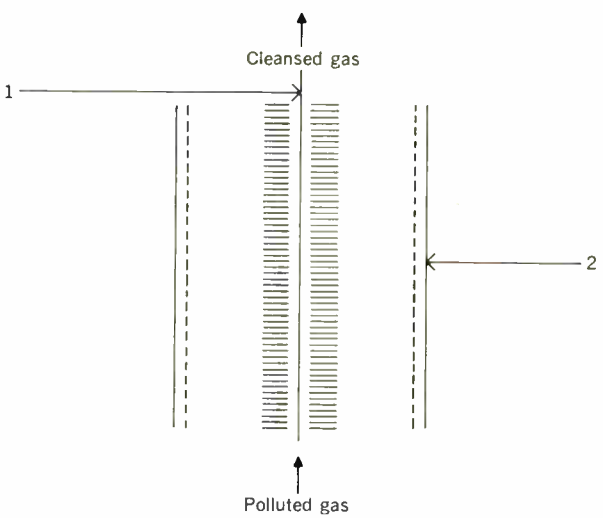


Fig. 13. Diagram of a simple electrostatic precipitator. "1" indicates corona wire; "2," grounded tube.

Fig. 14. Vertical-view diagram of a two-stage electrostatic precipitator. "1" indicates grounded cylinders; "2," corona wires; "3," grounded collector plates; "4," charged plates are similar to "2" in polarity.

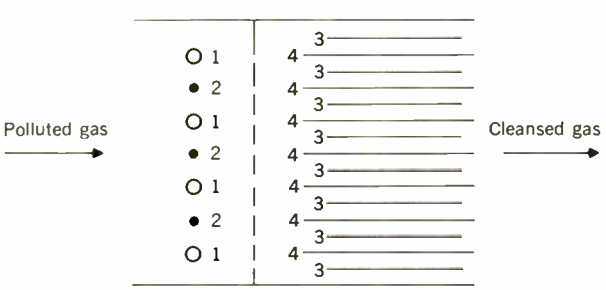
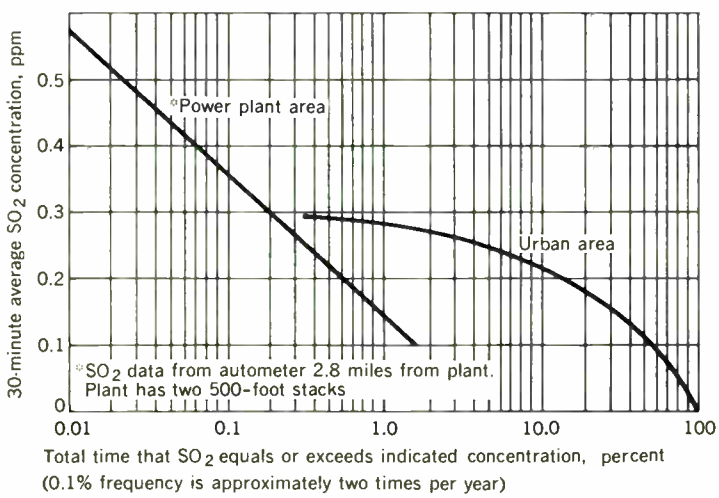


Fig. 11. Graph showing comparison-frequency distribution of 30-minute average SO_2 concentrations in an urban area and an area in the vicinity of a steam-electric power plant.



been augmented by extensive mobile sampling by car and helicopter equipment and by full-scale atmospheric dispersion studies. Since most of the steam-electric plants in the TVA system are geographically located in areas remote from other significant sources of SO_2, it is felt that the test results are most closely representative of the flue gas distribution patterns of modern coal-fired power plants. The SO_2 concentrations in these tests refer to 30-minute average concentrations.

Frequency distribution of SO_2 concentrations. A logarithmic plot of frequency of SO_2 concentrations at fixed monitoring stations has consistently indicated a fairly straight line. Figure 11 shows this distribution as measured by an autometer situated where maximum concentrations occurred at ground level in the vicinity of a 4-unit plant, with two 500-foot-high stacks, and with a total generating capacity of 1050 MW. For the sampling period—approximately 19 months—the highest recorded concentration was about 0.6 ppm for three 30-minute periods. And SO_2 concentrations were 0.2 ppm or above for only eighty-four 30-minute periods, or approximately 0.40 percent of the time.

Similar data obtained from Public Health Service studies in Nashville, Tenn., are also plotted in Fig. 11. Although the maximum SO_2 concentration recorded was only about 0.3 ppm, the concentrations of this gas were 0.2 ppm or more 14.1 percent of the time. And the estimated SO_2 emissions in the urban area were less than half of that recorded at the power plant.

Although higher concentrations of pollution in urban areas tend to occur during periods of low wind velocity and temperature inversion, *the higher levels of pollution in the vicinity of large power plants tend to occur during moderate to high wind speeds and neutral atmospheric stability conditions.* Since none of the TVA plants are located in large urban areas, the data do not provide a direct quantitative measurement of the contribution of a large power plant to an urban pollution problem. But data analysis from an autometer located in a small town near one of the large plants indicated that SO_2 in detectable amounts was present 14 percent of the time.

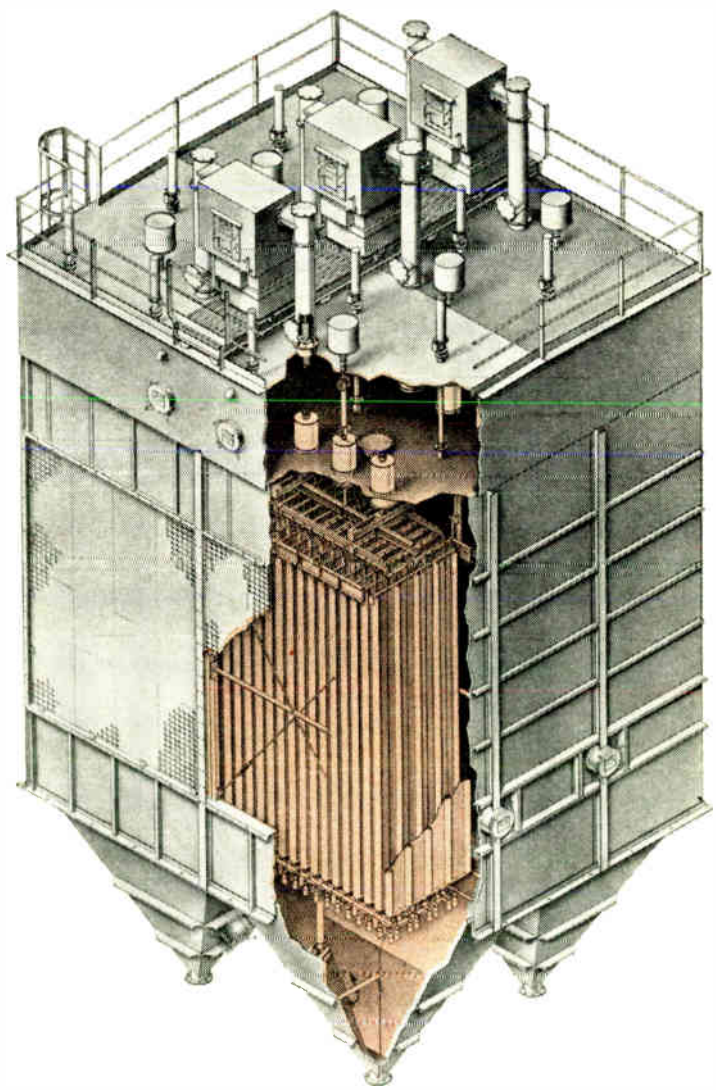
Effect and influence stack height. Monitoring data and data obtained from full-scale dispersion studies have been used in estimating stack height requirements for TVA plants. An example of estimates made by empirically derived formulas, based upon monitoring data, is shown in Fig. 12. The subject of this graphic plot was a two-unit, two-stack plant, with a generating capacity of 1800 MW, and an estimated SO_2 emission rate of 810 tons per day. This emission rate is calculated empirically from SO_2 monitoring data at plants with 250-, 300-, and 500-foot-high stacks. The curve for a 400-foot-high stack is interpolated, while the curves for the 600- and 800-foot-high stacks were extrapolated.

The electrostatic precipitator—TVA experience. The practical application of the electrostatic precipitator was first demonstrated by F. G. Cottrell in 1906. Essentially, it is a device that is used to remove liquid chemical mists or solid particulates from a gas in which they are suspended. Electrostatic precipitation is a two-stage process. In the first step, the gas containing the suspended particulates is passed through an electric, or corona, discharge area in which ionization of the gas occurs. The ions produced collide with the suspended particles and impart an electric charge to them. These charged particles then drift toward an electrode of opposite polarity, and they are deposited upon this electrode, where their electric charge is neutralized.

In its most elementary form, the precipitator configuration may consist merely of a vertical tube that contains an insulated concentric wire (see Fig. 13). When a dc potential of 10–100 kV is applied to the central wire, a corona discharge occurs in a small area surrounding the wire. The suspended particulates are ionized in the corona discharge and migrate to the tube wall. If the suspension is liquid, it will accumulate on the wall and coalesce into droplets that can be drained from the base of the tube. Suspended solid particulates can be removed from the tube wall by mechanical vibrators or scrapers, and then discharged into a cyclone or dust collector at the bottom of the apparatus.

In more complex configurations, the ionization may occur in one vessel and the deposition and precipitation in another. Figure 14 shows the plan view of a simplified two-chamber apparatus. In the first chamber, the particles become charged but are prevented from depositing

Fig. 15. Cutaway isometric view of an Opzel Plate Precipitator, a type manufactured by Research Cottrell, Inc.



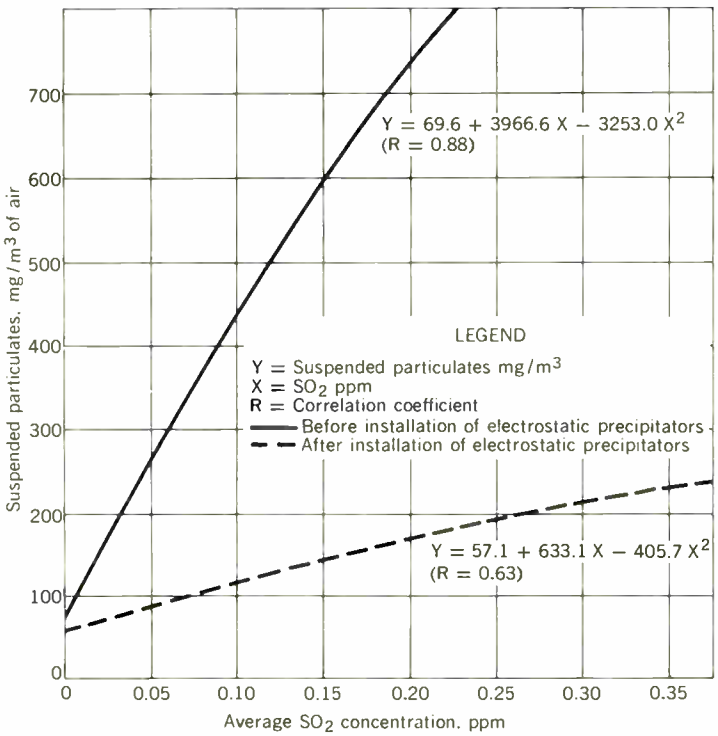
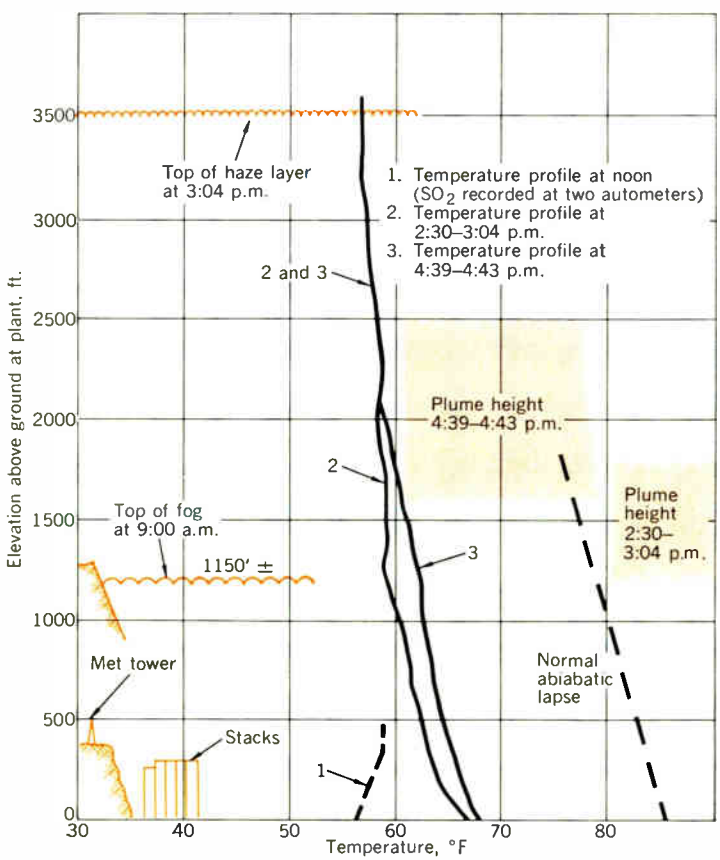


Fig. 16. Graph showing relationship between suspended particulates and SO_2 concentration before and after installation of electrostatic precipitators.

Fig. 17. Profile plot from aerial observations and temperature soundings at Kingston steam-electric plant (TVA system) on November 10, 1964.



on the grounded cylinders by the appropriate adjustment of the rate of gas flow. In the second chamber—consisting of alternately charged, loosely packed and parallel plates—satisfactory precipitation can be attained by applying a lower potential than that required in the charging chamber, since there is no need for a corona discharge.

The corona discharge is usually produced by making the center wire the cathode, because precipitation efficiency is greater under such an operating condition. Less ozone is produced, however, by reversing the polarity, and a positively charged wire is employed in the cleaning of air if the presence of ozone is objectionable.

The high-voltage direct current is generally produced by either mercury vapor or vacuum tube rectifiers. The power requirements vary from 2–5 kWh per million cubic feet of polluted gas being treated, and the variation will be a function of the quantity, size, and physical properties of the particulates that are being removed.

Figure 15 is a cutaway view of an Opzel plate precipitator, manufactured by Research-Cottrell, Inc.

Effect of precipitators on suspended particulates in an ambient atmosphere. Electrostatic precipitators were installed to supplement mechanical ash collectors at one of the TVA plants after a special investigation was conducted to determine the relationships between SO_2 concentrations and suspended particulates in the vicinity of the plant where maximum ground level concentrations of stack emissions occurred. Data were analyzed with the results shown in Fig. 16. From the two derived equations indicated on the graph, TVA estimated that the electrostatic precipitators reduced the suspended particulates by 85 percent in the ambient atmosphere at ground level during those periods when SO_2 was present.

The TVA believes that an additional—although unproved—possible benefit from the electrostatic precipitators is a reduction in the maximum ground level SO_2 concentrations in the vicinity of steam plant emissions. This is predicated on data indicating that the maximum recorded SO_2 concentration during the four years of precipitator operation is about 25 percent less than that recorded prior to precipitator operation.

Power plant pollution potential under air stagnation (temperature inversion) conditions. Air pollution control plans developed for the Kingston plant, until recently the largest steam-electric station in the TVA system, gave special attention to the SO_2 problem associated with periods of atmospheric stagnation.

This plant is located on the floor of an Appalachian valley. The local terrain has parallel ridge features that vary from 400 to 1000 feet above the plant grade level.

During the period from November 9–11, 1964, a temperature inversion occurred in the Kingston plant area. The U.S. Weather Bureau, by prearrangement, alerted the TVA beforehand, and precautionary air pollution control measures and monitoring were initiated.

Sulfur dioxide autometers were checked at regular intervals, and mobile sampling was conducted during the 3-day period by specially equipped helicopters and cars.

The frequency and concentration of SO_2 recorded at ground level were no higher than during normal atmospheric conditions. The absence of an SO_2 buildup was attributed to penetration of the low-level inversion by the hot, high-velocity stack gases and dispersion of the smoke plume from the area by light and steady winds. Figure 17 indicates the time and temperature conditions aloft,

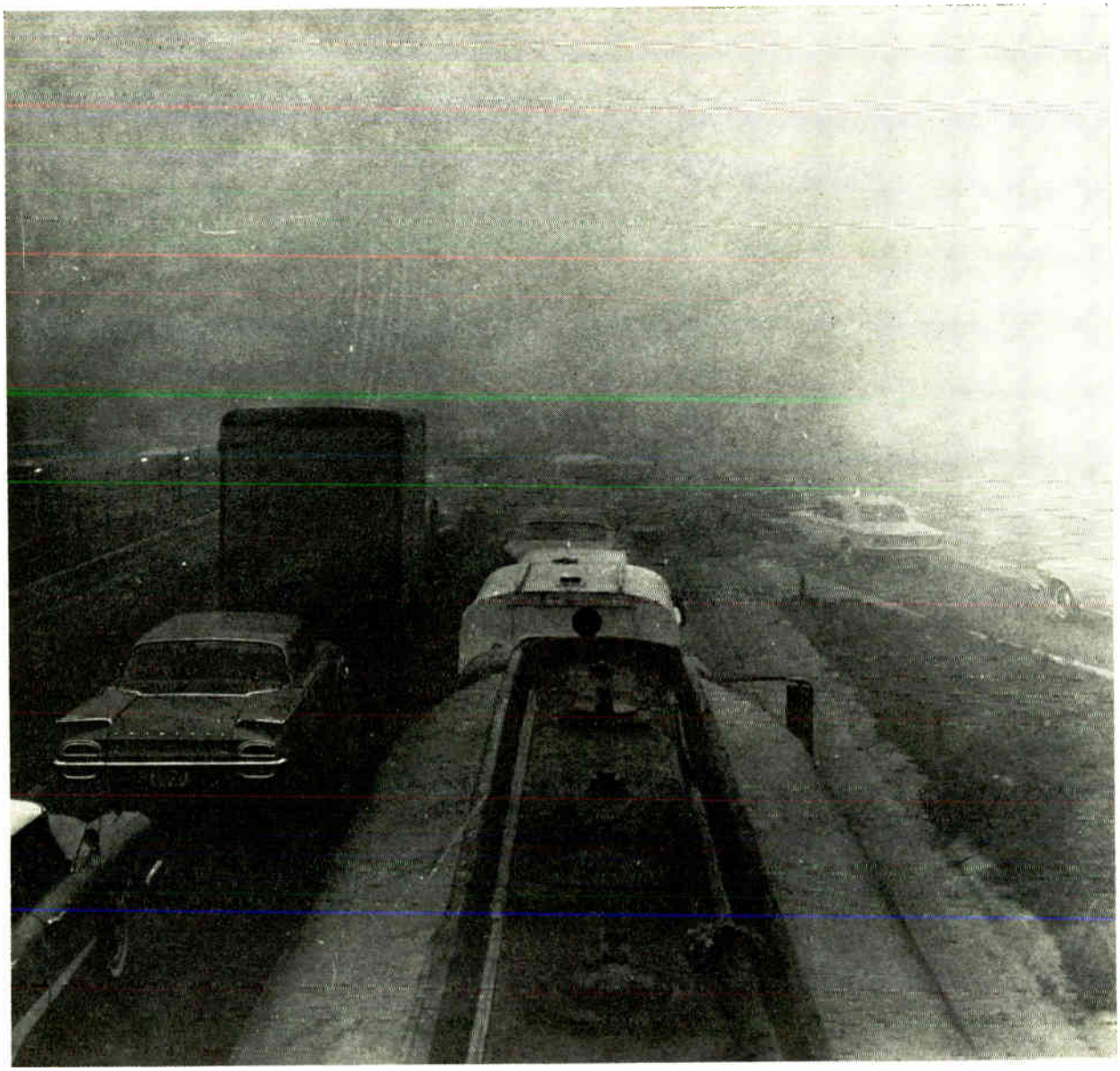


Fig. 18. View of a heavily trafficked Los Angeles freeway shrouded in smog.

and the plume elevation during a timed sequence on November 10.

From the data obtained during this monitoring, it is possible that air pollution predictions should consider power plants as a special case for which normal meteorological criteria may not apply.

The other half of the foggy, smoggy brew

If the witch's brew concocted in the smokestacks of American industry seems appalling, cheer up; for it is only half of the dirty story. Most of the other half may be found in Fig. 18. The U.S. Public Health Service estimates that the 88 million motor vehicles on the nation's highways belch into the atmosphere about 350 000 tons of carbon monoxide, volatile hydrocarbons, and nitrogen dioxide daily. The automobile population in the United States is increasing at more than double the rate of the human population! Urban traffic jams are providing civilization with one of its most diabolical methods of multiplying the problems of air pollution; and, unless something is done to correct this man-made plague, our metropolises will soon become obsolete and deadly places for human habitation.

Pollution control devices for cars. In addition to assuring more complete combustion of hydrocarbon fuels in internal combustion engines by simple carburetor and ignition system adjustments, there are two basic devices that can effectively reduce the amount of air contaminants released from motor vehicle exhausts. For an investment of about $5, crankcase blowby units, which return unburned gases to the combustion chambers, can be installed. Properly maintained blowbys can reduce an automobile's hydrocarbon emission by at least 25 percent. There is also an "afterburner" device, which, when attached to the tailpipe of a motor vehicle, will complete the combustion of carbon monoxide and other unburned gases. When used in conjunction with a crankcase blowby, the afterburner would eliminate the major portion of pollutants now poisoning the air from this source.

It is no secret that the automobile industry has been dragging its feet for a long time in resisting the factory installation of these control accessories. The weight of public opinion, and public officials who have been aroused by the menace of the motor vehicle situation, has forced the hand of the manufacturers. These devices will be required by law for 1966 model cars to be sold in California.

A pollution survey, city by city

The Pittsburgh story. Since the 1940s, a generation of young Pittsburghers has grown up convinced that the sky above them is really blue, and not murky gray as they were assured by their parents. And it is possible to put on a white shirt in the morning, wear it all day, and still find it reasonably clean. Housewives, too, are able to hang out a line of wash without having to relaunder it the same day.

Before the 1940s, Pittsburgh was the classic American example of continual air pollution. Since the middle period of the Industrial Revolution, the "Smoky City" blamed its railroads and concentrated industry for its plight. The Pittsburgh Chamber of Commerce appointed its first committee on smoke abatement in 1899, and a smoke control ordinance was passed in 1906. For the next 35 years, the Chamber of Commerce and other civic groups hoped that education in matters of proper fuel combustion and pollution control devices would help to solve the problem. But the results of self-regulation were negligible; the situation deteriorated, and enforced regulation seemed to be the only answer.

The City of Pittsburgh Smoke Control Ordinance was passed in 1941, and more stringent amendments were added in 1943, 1946, and 1951. Dual enforcement by Allegheny County and the city is in effect today.

Under the provisions of the ordinance for smoke and air pollution control, limitations are established on the emission of smoke, fly ash, soot, cinders, toxic or radioactive substances, noxious acids, fumes, oxides, gases, and odors, or any other matter that may create atmospheric pollution. These regulations further stipulate that the home owner should do one of several things:

1. Burn highly volatile bituminous coal by installing a stoker.
2. Burn briquettes or other forms of prepared coal, coke, anthracite, or less volatile bituminous coal.
3. Use gas or oil.

Emissions from furnaces of the steel and allied industries are generally restricted to two parts of solids in every 3000 parts of gas, but these limits can vary somewhat dependent upon the types of furnaces employed.

The installation and repair of furnaces, boilers, and other fuel-burning equipment must be done under a permit issued by the terms of the Rules and Regulations of the Pittsburgh Smoke Control Ordinance. With the exception of installations in private residences, furnaces or other fuel-burning equipment are subject to periodic inspection by the Bureau of Air Pollution Control.

To date, the result has been gratifying. Pittsburgh has an average sootfall of only 30 tons per square mile per month as compared with 60 tons for a similar area and time period in New York.

Chicago—soot, smoke, and steel. The nation's second largest metropolis—population 4 million—is gravely endangered by both air and water pollution from the Calumet region of neighboring Indiana. The concentration of steel mills at Hammond and Gary, a scant 15 miles from the Chicago loop, belch a continuous barrage of acrid fumes, and the smoke plumes from these emissions are an ever-present facet of the seascape over Lake Michigan. The legal aspects of this situation are complex. As one leading Chicago spokesman put it: "The cities and towns along the lakefront are not doing a good job of pollution control. And Illinois state agencies are unable to cope satisfactorily with the issues since interstate problems are involved." In view of this, the general sentiment among Chicago officials is that the Federal Government should enforce antipollution controls under Congressional authority.

Air pollution control is being handled by a special city agency. In the three years since its formation, this agency has held 2800 hearings, handled 5000 smoke complaints, inspected 50 000 furnaces, and has filed almost 600 suits for the prosecution of chronic violators. There are also 20 rooftop monitoring stations in the city that were established to trap pollutant particulates in the air, and another 35 specially equipped stations measure the dustfall and SO_2 content.

A three-year Federal grant of $357 000 has assisted Chicago in putting teeth into its enforcement of pollution control. It is hoped that, by 1968, all improper burning of combustible refuse will be banned. This would put the damper on about 30 000 Chicago apartment houses that foul the atmosphere with soot from coal-fired boilers.

Like all major cities, Chicago has its smog—a noxious blend of automobile fumes, industrial pollutants, and soot. The present sootfall is about 43 tons per square mile per month.

Recently, the four major steel producers in the Calumet region agreed to appropriate $50 million over the next seven years for pollution control equipment. It is believed, hopefully, that this equipment will cut the 60 000 tons of pollutants released annually to half that amount by next year.

Chicago has no illusions about quick and easy solutions to its overall pollution problems, but at least the city is doing something about them.

Houston—an industrial complex of chemicals and contaminants. Houston, Tex., is probably the fastest growing city in the United States. And concurrent with this phenomenal growth is the tremendous increase in the number of petrochemical plants, oil refineries, paper mills, and other heavy industries that gird the city's perimeter. Although pollutants from coal combustion are no problem here (natural gas is the most common fuel for both industrial and domestic use), the melange of chemical fumes released into the air is causing considerable apprehension. When the wind is from the southeast quadrant, the sickening fumes from "old smoky"—a sulfite process paper mill in suburban Pasadena—are very apparent in downtown Houston.

Carbon monoxide and the nitrogen dioxides, caused primarily by the heavy automobile traffic, are found in the same ranges of concentration as those reported in other large cities.

However, the general feeling in Houston is that the air pollution problem, at least, is under fairly good control. Private industry has shown an inclination to comply with local city ordinances, primarily because it knows that the

"The problem of sulphur pollution has grown (in New York) until now the measured concentrations of sulphur dioxide are higher than in any other major city in which such measurements are taken."

—V. G. MacKenzie,
Assistant Surgeon General,
Chief of the Air Pollution Division,
U.S. Public Health Service

city is quite willing to initiate court action to curb flagrant violations.

Los Angeles and New York. Reams of copy have been written over the past 20 years about the lethal smogs, principally caused by the fantastic motor vehicle population and domestic trash burners, in Los Angeles; and any further emphasis on the situation in this West Coast metropolis would seem boring and repetitious to the reader. Suffice to say that, in 1960, California passed a law requiring all new motor vehicles sold in the state to be equipped by the owners with the crankcase and tailpipe devices previously described. State officials estimate that more than 4 million California vehicles are now equipped with these devices and they are removing about 200 000 gallons of unburned gasoline from the atmosphere each day.

The air pollution situation in New York, however, is so far out of control that this writer could be panicked into leaving the city after reviewing his own copy. New York channels most of its air contamination problems into the municipal Department of Air Pollution Control, whose commissioner, Arthur J. Benline, frankly confesses that far too little money is appropriated either by the city or by the Federal Government to do an adequate job. During a recent television interview, Benline summed up the New York situation in two terse sentences: "The air over our city is helping our citizens to shuffle off this planet at a much higher rate than they would ordinarily go. Yet there has not been any overall demand from the public to clean up our dirty air." So, in the meanwhile, the city muddles through, and everyone hopes that, by some miracle, adequate corrective measures will be taken before any future prolonged temperature inversion produces the disaster that many health authorities fear.

The air pollution situation throughout most urban areas in the United States closely parallels the samplings just described in five major cities.

A utility answers its critics

At a recent hearing called by the Special Committee to Investigate Air Pollution (authorized by the New York City Council), the Consolidated Edison Company of New York replied to the Committee's accusation that Con Edison is one of the principal contributors to the air pollution problem. Otto W. Manz, Jr., the company's executive vice president, contended that the utility has done a conscientious job of providing adequate air pollution control equipment and devices in its city generating plants to minimize the emission of contaminants. In outlining future plans on the subject of air pollution control, Manz listed the following activities, presently under way, that will either decentralize electric generation to remote areas or significantly reduce objectionable stack emissions in city-located generating stations:

1. The construction of the Cornwall pumped-storage station, which will produce about 2000 MW of power, and will permit the retirement, or placement on cold standby, of about 750 MW of generating capacity that is presently being supplied by city steam-electric stations.

2. The possible substitution of more natural gas for coal as an alternative fuel for generating purposes.

3. The establishment of firm interconnections, by EHV transmission, with the CONVEX group of New England utilities and with the Pennsylvania–New Jersey–Maryland (PJM) complex.

Manz believes that the ultimate answer to the problem of air pollution in metropolitan areas, insofar as power generation is concerned, is the use of nuclear energy. He cited the Indian Point nuclear generating station in suburban Westchester County that has been in operation for almost three years and is a firm and reliable producer of power for the Con Edison system.

The Con Edison spokesman sounded a warning note, however, on the subject of who will eventually pick up the tab for air pollution control measures. To quote from Manz' 8-page prepared statement:

"Our estimate is that on and after October 1, 1969, when the maximum sulfur content in these [coal] fuels will be 2.2 percent by weight, the additional annual cost of the fuels used by Con Edison and which will be passed on to the consumer could be as much as $20 million."

Where do we go from here?

As one public health official bluntly put it: "Perhaps we are worrying about the wrong menace. Our urban civilization stands a much better chance of being suffocated by air pollution than being annihilated by atomic weapons."

From numerous surveys, it is obvious that the sources of air pollution are numerous and complex. There is no single group or interest toward whom we can point an accusing finger as being the sole villain of the drama. Until heavy industry, the utilities, car manufacturers, and the general public become totally aware of the truth about air pollution and the consequences of irresponsibility, the already intolerable situation will become impossible.

Piecemeal solutions are not the answer. A city committee cannot correct a situation—such as in Chicago and to some extent in New York—whose origins are interstate. Total and effective remedial action must be undertaken by joint commissions that represent municipal, state, Federal, and private interests.

Life depends upon the air we breathe.

The author wishes to acknowledge the following picture credits: Fig. 1, *N.Y. Journal-American;* Figs. 2 through 8, U.S. Public Health Service, Division of Air Pollution; Figs. 9 and 10, *Mechanical Engineering* magazine; Fig. 13, Research-Cottrell, Inc.; Fig. 18, *Los Angeles Times.*

REFERENCES

1. Dinman, B. D., Frankenberg, T. T., Gartrell, F. E., Gerber, A., Ireland, R. L., Jones, J. R., Perry, H., and Ripperton, L. A., "Panel discussion of the clean air problem," presented at the American Power Conf., Chicago, Ill., Apr. 27–29, 1965.

2. Frankenberg, T. T., "Sulfur removal: for air pollution control," *Mech. Eng.*, vol. 87, no. 8, pp. 36–41; Aug. 1965.

Do scientists and engineers need each other?

The close links between scientists and engineers formed during the war must be maintained if development of new technologies for man's welfare is to continue at a pace consistent with the rate of fundamental discoveries. Scientists and engineers are urged to seek elective posts both to enlighten the public and to strengthen governmental processes

John R. Dunning *Columbia University*

Historically, man has been a curious animal, and this has spurred his progress. He has searched for the invisible and has probed for the unknown; he has groped for the inaccessible and has struggled for the unattainable. He has disobeyed every injunction that he not eat the fruit of the tree of knowledge. But he has accepted the great challenge that if you know the truth, "the truth shall make you free." The fruit of the tree of knowledge has been sweet, and it has been bitter. If knowledge and truth shattered the dream of some placid Eden, man has been content to labor outside the gates. And today, even though the prospect may sometimes be awesome, he is using all the keys of his genius to unlock the doors that hide Nature's secrets.

We are all familiar with the unprecedented burgeoning of scientific programs that have been supported by the government, the universities, and industry since the beginning of the Second World War. And it is only natural, of course, that the attention of the general public has been focused upon the more spectacular material results of these programs. The events leading up to and culminating in the first nuclear weapons, the advent and perfection of high-speed jet aircraft, the economic effects, both adverse and favorable, of the rapidly extending realm of automation, and the inspiring spectacle of American astronauts orbiting beyond the limits of the earth's atmosphere—all of these things have been indelibly impressed upon the public consciousness as images of science.

The invisible organism

What the general public does not properly recognize, and what is not so evident to the majority of our citizens, is the great invisible organism of science that forms the basis of our technological achievements. It is difficult for society to visualize this aspect of science, and even more difficult for it to realize that much of what we have seen materialize in recent years has behind it a profound history of growth in human knowledge that stretches into the past for several centuries.

But let us consider just a few of the more recent discoveries and go back 80 years. In electronics. Hertz, by finding concrete evidence of electromagnetic waves, gave experimental proof of Maxwell's electromagnetic theories, the basis for all of today's radio, television, and microwave transmission. In physics, Balmer formulated his equation for the spectral hydrogen lines series; in chemistry, Moissan isolated fluorine and Winkler discovered the chemical element germanium, a basic ingredient in today's transistors; in medicine, Merck synthesized cocaine and Halsted used it in nerve block anesthesia, Ramon y Cajal contributed to the physiology of the brain, and Pasteur perfected his rabies treatment; in astronomy, Pickering investigated the planet Neptune. In other fields of technology, Hall, of Oberlin College, perfected the inexpensive reduction of aluminum oxide to produce aluminum metal, until then a rare and expensive commodity; Burroughs invented the adding machine; Dunlop the pneumatic tire; Gassner the first practical dry cell; and Eastman began his pioneering photography with transparent film.

The list of scientific and technological discoveries continued to increase in the years immediately following. In the last decade of the 19th century, Acheson discovered carborundum; Behring and Ehrlich discovered the diphtheria antitoxin; Dewar and Redwood pioneered the cracking of petroleum; Diesel made valuable contributions to the internal combustion engine; Dubois discovered the fossil of the ape man in Java; Duryea perfected the gasoline automobile; Edison's kinetoscope emerged as the first motion picture; Freud pioneered the analysis of dreams and Galton the systematics of fingerprints; Koch announced his work on tuberculin; Steinmetz announced his pioneering theory of transformers and the theory of hysteresis; and Tesla developed the high-frequency transformer.

To show that there were even then the glimmerings of today's space developments, Wilson and Gray were doing pioneering work on solar temperatures, and Bernard, looking beyond the moon, discovered the fifth satellite of

Jupiter. The list of fundamental discoveries and inventions is by no means exhausted; however, a long enumeration of those since 1900 is not needed to illustrate the theme of this discussion: the relationship between science and engineering.

Beginning before 1900 and lasting through the years until World War I, we find a casual and distant relationship between science and technology. Science developed slowly and the technological advances or economic creativity, as some have labeled it, were not wedded to pure scientific activities. It was only recently that the United States became a country of key importance in the world of science. Our technology, nevertheless, was excellent.

During this same period, several of our universities had reorganized their graduate instruction along the lines of the best European models. In scientific and educational circles greater emphasis was being placed upon the prestige value of becoming a pure scientist. At the same time the academic world began to underscore pure research. Relatively little support or encouragement was given the bright student interested in applied science.

Meanwhile, earlier graduates of our educational institutions, attracted to our industrial organizations and semiprivate research groups and stimulated by the demands of our society for "a better mousetrap," attacked engineering problems that promised immediate application and benefits, thereby contributing immensely to the continued growth of the vast industrial complex that provides the foundation for the American economy.

The events leading to and immediately following World War II, however, presented a radical new challenge to both the engineer and the scientist. Because of the stress on the practice of engineering rather than on the fundamentals of why and how a device or principle worked, World War II found a dearth of engineers with strong scientific backgrounds and advanced training. Scientists were pressed into service and found themselves working on the development of new processes and new devices, and, in fact, developing new industries that normally require engineers.

In almost every instance engineers and scientists were being asked to work more closely with each other than they had in the past. For example, engineers working side by side with chemists were engaged in translating the discoveries of the tracer experiments at Berkeley and the microscope-scaled experiments in Chicago into huge plutonium production facilities. And this had to be done without going through most of the stages of development that usually fall between the basic research studies and eventual industrial production.

Changing attitudes following World War II

Immediately after World War II a revolutionary change in attitudes toward science and technology took place. In technology, the precipitating cause of the change was clearly our military experiences, confirmed and extended by the military, nuclear, and space programs that have developed since then. The results of these large and complex undertakings that grew out of the dramatic successes of World War II, lent encouragement to the planning and public support of science ever since. Forced development became and still is an established concept for military purposes. Public support has been exceptionally spontaneous—only because the public believes that the results will give it what it wants and needs. Our society has embraced a multibillion dollar support program for research and development and expects technological achievements far greater than before.

In the United States, the federal support for basic research in physics has continued to increase to the point where the next rounds of high-energy accelerators will cost a significant fraction of the national budget. Naturally we begin to meet congressional questions and resistance. We have reached the point in the support of research for the sake of pure knowledge where society is beginning to ask why. And is high-energy physics likely to produce tangible benefits to society in the foreseeable future?

What is happening now?

We know that past applications of new knowledge radically changed the conditions of life even though there was a tremendous time lag between discovery and application; and that the time lag between discovery and application was really not an important factor in our society as it advanced with the industrial revolution. We also know that World War II brought about a drastic change that compressed the time lag and forced rapid application of new knowledge into new devices, systems, and processes, and, in addition, forced scientists and engineers into an affiliation that never existed before.

It is also apparent today that the urgency of the war years is not quite so intense nor is the close scientist–engineer affiliation quite so novel. Nevertheless, prompted by popular feeling and by the successes of the scientific breakthroughs during the war years, the government has greatly increased its support of science and technology for peaceful applications, as well as for military ones.

For the first time, the government is taking a major role in the *development of new technologies for man's welfare*. The price the United States public is paying for this government support today is more than $14 billion, which amounts to an annual tax tab of about $400 for a family of five.

With governmental support, it is now possible to pursue many lines of research leading to the application of increased knowledge at a pace consistent with the rate of new fundamental discoveries and at a pace which will satisfy the general public.

However, even though a few scientists have contributed to the formulation of a national policy that has made this support possible, scientists and engineers, as a group, have not been sufficiently sensitive to the implications of two broad facts of the new post-World War II situation: (1) the pre-1939 separation of the government, the universities, and private industry into neat little categories no longer holds true; (2) science and technology are no longer tools to be picked up when needed and then put aside—they are now an extremely important and integrated part of the basic structure of our economy and of our government process.

Conclusions and some general proposals

Using broad brush strokes I have shown:

- That throughout history, science and technology have never really gone hand in hand, in a balanced sense, except upon rare occasions.
- That deliberate applications of science and technology on a broad scale were strongly highlighted by the war and immediate postwar years, and that since then we are,

for all practical purposes, coasting with those successes.

- That substantial public support has been given to scientific research, and that society, right or wrong, accepts the idea that science is applicable to technology and looks to such application as a mainspring for progress and that it has the power to cut off its source of support.
- That too many scientists are continuing to cling, or are retreating, since World War II, to the images and conditions of a world that puts a premium on the separation of government, the universities, and private industry into neat little categories.
- That in our educational institutions we have directly and indirectly contributed to the isolation and compartmentalization of basic research, when instead we should be emphasizing a continuous flow from pure mathematics into basic scientific research, applied research, and, finally, into production and on into the economic stream.
- That today, even though science and technology are an extremely important and an integrated part of the basic structure of our economy and our government processes, scientists and engineers are not as sensitive to the implications of the integration of science into society as they really should be.

These six points suggest some significant areas that contribute to inefficiencies in the free flow of technological advances demanded by our society. How can the situation be improved? Let us consider some steps that can be taken by the government, by our educational institutions, by our engineering and scientific societies, and by individual scientists and engineers.

While the scientist and engineer have become valued advisors and policy makers in the circles of government, particularly since the advent of the Soviet Sputnik in 1957, the roles being pursued are generally those that are compartmentalized or departmentalized. For the most part, agencies of the government support mission-oriented research, with the exception of the National Science Foundation. We find clearly defined goals and objectives in each of the agencies but we do not find clearly defined scientific and engineering goals and objectives that encompass or coordinate the support for the whole spectrum of scientific endeavor. Is this enough? Have we done everything necessary to give science and technology their proper places in government? I do not think so. It is time to give careful consideration to the next steps. One such possible step is the elevation to Cabinet rank of the present Office of Science and Technology in the Executive Branch of the government. It would have the prime responsibility of clearly delineating both the boundaries of our public-supported scientific endeavor and the whole spectrum of the scientific competence of the scientists involved in these endeavors.

McGeorge Bundy, in a speech three years ago before the American Association for the Advancement of Science, said, "The first and most important proposition is that the scientist needs to be there—there in the councils of government and in the policy-making processes." The role of the scientist and engineer in government needs to be more pervasive than is generally assumed.

In our educational institutions we need to place the same emphasis on applied science as we have placed so dramatically on pure science. While faculties of universities may not easily accept applied work as similar in value to the old academic disciplines, we must now demonstrate the greatest possible initiative and imagination in revitalizing and updating our educational goals and objectives. Our emphasis should be broadened and excessive specialization curbed. While specialization is inevitable in many branches of pure science and also in engineering developments, it would have, as it has in the past, a decidedly harmful effect in applied science.

Perhaps the most important recommendation I can make—inasmuch as our society is one in which the really important problems and needs are not reducible to analytic formulations—is that our engineering faculties in particular should be composed of "men for all seasons": scientists and engineers endowed with good scientific and technical judgment, who understand the context in which their work occurs.

In addition, the status symbols for faculty and students who are oriented toward both new knowledge and useful applications are long overdue for elevation. The mantle of prestige in scientific circles now covering the pure scientists and abstract research should be enlarged to include the applied scientist and engineers who have a genuine interest both in new knowledge and in its usefulness. There should be equal pride and respect in whatever role one chooses for furthering a dynamic society.

Today there is no really effective mechanism for putting the new knowledge gained in our basic research activities into deliberate use. I suggest that our engineering and scientific societies could serve effectively together in bridging this semi-no man's land. A complete overhaul in our communications effort is indicated. President Johnson recently said, "One of the major opportunities for enhancing the effectiveness of our national scientific and technical effort and the efficiency of government management of research and development lies in the improvement of our ability to communicate information about current research efforts and the results of past efforts."

I suggest that a joint study group with representation from each of the scientific and engineering societies be formed to explore this problem, to formulate specific recommendations, and to design a constructive program to improve our ability to communicate.

Earlier it was indicated that the majority of our citizens tend to forget or do not thoroughly understand that one of the crucial features of technological progress is built upon scientific discoveries that were not motivated originally by practical aims or planned for specific purposes. A much broader program for enlightenment among the public concerning science is needed. Such public enlightenment can be fostered if, as individuals, scientists and engineers participate much more in the affairs of government than they have in the past.

It is indeed incredible that there are only one or two scientists and engineers in the Congress. In view of the increasing importance of science and technology in national life, there is no reason why more scientists and engineers should not seek elective positions. At the same time, understanding of the humanities and social sciences should not be neglected. The scientist and engineer, as the real humanists of the future, should be able to identify, adjust to, and comprehend the economic, historical, and social factors that influence our civilization.

This article is based on an address given at the Banquet of the American Physical Society, Summer Meeting, New York, N. Y., June 23–25, 1965.

Ultrasonic traveling-wave devices for communications

A comprehensive discussion of the latest techniques and materials in the fabrication of ultrasonic delay lines and transducers—plus a description of delay media and delay times, and elastic wave interactions

J. E. May, Jr. *Bell Telephone Laboratories*

As a result of advances in transducer materials, wave propagation technology, and device inventions, ultrasonic delay lines are now available with a large variety of delay characteristics and for operation over a wide range of frequencies that extend from the kilocycle to the microwave range. In this paper the principles of operation and characteristics of presently available designs will be reviewed, together with a description of some of the delay and amplification devices that are currently being developed and which should become available shortly.

Types of transducers

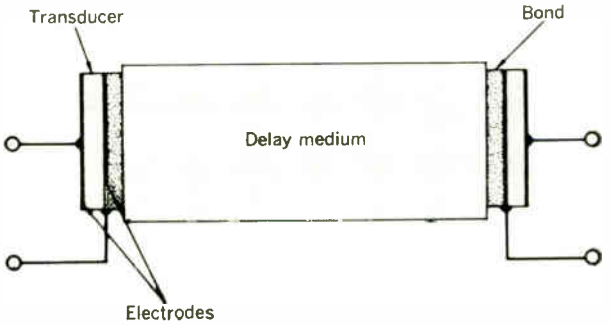
The basic structure of an ultrasonic delay device, as shown in Fig. 1, consists of a delay medium and one or more transducers for converting electric into mechanical energy, and vice versa. The transducers may be either magnetostrictive or piezoelectric. Magnetostrictive transducers,[1] consisting of a coil of wire encircling a magnetostrictive wire delay medium, have been used extensively on relatively low-frequency wire delay lines of both the longitudinal mode and the torsional mode types. These transducers have the distinct advantages of allowing continuously variable delay and of providing taps economically. The upper frequency of operation of magnetostrictive transducers appears to be limited to about 5 Mc/s because of fringing field effects. Below this frequency, however, the transducers are characterized by large fractional bandwidths and insertion loss of the order of 30 to 40 dB for a pair of transducers. Another type of magnetostrictive transducer has been used to some extent at frequencies as high as several thousand megacycles. This is a transducer formed by evaporating a thin film of magnetic material,[2] such as nickel, and exciting it under conditions suitable for ferromagnetic resonance. Although these transducers can be used to generate either longitudinal or shear modes at high frequencies, the efficiency of the transduction process is low because of the difficulty in arranging the driving field to provide high coupling to the film.

Piezoelectric transducers, fabricated from either single-crystal materials or from one of the polarized ferroelectric ceramic materials, are commonly manufactured to a thickness of approximately a half wavelength at the operating frequency (this being the thickness for maximum efficiency) and are bonded to a delay medium. An upper frequency limit to the fabrication and bonding of transducers is established by practical considerations at around 100 Mc/s and the half-wavelength thickness is about 25 microns. Also, at frequencies above 100 Mc/s, it becomes difficult to make bonds thin enough (they must be about $0.01\ \lambda$)[3] in comparison with the acoustic wavelength that the bond effects are negligible. All thicker bonds should be designed as transmission line sections, a fact that requires extremely tight tolerance in maintaining the bond thickness.

Coupling factors

Transducers may be characterized by the electromechanical coupling factor, which may be defined for certain special geometries in terms of the piezoelectric, dielectric, and elastic parameters of the material. The electromechanical coupling factor of a delay-line trans-

Fig. 1. Basic delay-line structure.



ducer is a rough measure of the fractional bandwidth obtained under matched electrical terminations. For the ceramic materials, such as potassium sodium niobate and lead titanate zirconate, coupling factors in shear mode operation as high as 70 percent have been realized and, therefore, it is possible to construct low-loss,* wideband delay lines with fractional bandwidths of approximately 70 percent. These transducers have been operated at frequencies in excess of 100 Mc/s in sections of small area. For operation of large-area transducers, however, at frequencies much above 30 Mc/s, the large dielectric constant of these materials (from about 400 to 1000) results in an unmanageable, low value of capacitive reactance. Hence, for these applications above 30 Mc/s, quartz transducers are commonly used. Because of the low inherent coupling factor of quartz (~10 percent), these transducers are usually operated with mismatched electrical terminations to achieve large bandwidths and, as a result, transducer losses of 20 to 40 dB are incurred.

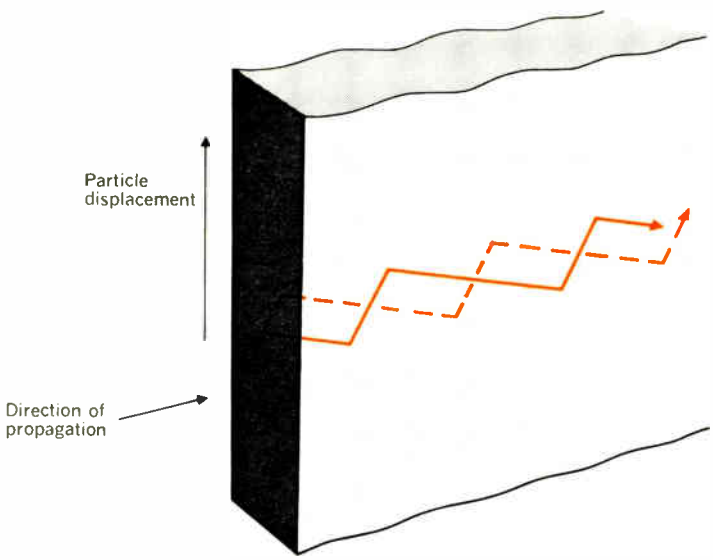
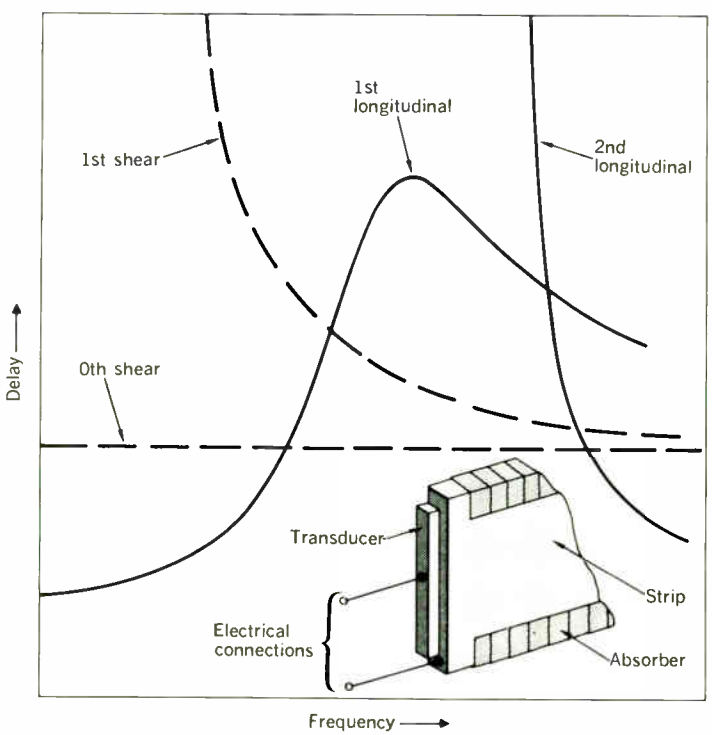


Fig. 2. Isometric view of a shear mode in a flat plate.

Fig. 3. Delay vs. frequency characteristics for low-order longitudinal and shear modes in a rectangular strip.



Delay media and delay times

For the delay medium, fused quartz is the lowest-loss material commonly employed. Other materials include polycrystalline aluminum, which is actually the lowest-loss metal, and some glasses and nickel–iron alloys that have a low temperature coefficient of delay time. The delay times realized are of the order of 7 to 10 μs per inch. The delay medium may be either a simple block or rod of a suitable material or, in the cases where longer delays are required, it is common to use one of the multiple reflection patterns, such as the familiar polygon reflection pattern originally described by D. L. Arenberg.[4] Delay lines in this form can be fabricated to achieve total delay times in one piece of material up to about 3300 μs and, by bonding three blocks together with suitable mode conversion sections, total delay times as high as 10 000 μs have been realized. Polygon delay lines have been operated at center frequencies as high as 60 Mc/s with large fractional bandwidth and with suppression of spurious responses to about 60 dB.

In the multiple reflection type of delay line the elastic wave propagates essentially as if it is in an infinite medium in which there is no interaction (at least to the first order) with the bounding surface. Therefore the delay is nondispersive. If the wave motion, however, is caused to interact strongly with the bounding surfaces, as in the case of a plate or a wire, propagation of an infinite number of dispersive modes is now possible. Consider, for example, the case of a shear wave propagating in an infinite plate with particle displacement lying in the plane of the plate as in Fig. 2. There exists an infinite set of dispersive modes, each having a cutoff frequency: the lowest, plus one zero-order mode which is nondispersive and is of prime importance for obtaining high-quality pulse transmission, free of other interfering modes. A. H. Meitzler[5] showed that the infinite-plate modes could be propagated in a rectangular strip many wavelengths in width if absorbing material is added at the edges to minimize interaction with the minor surfaces. Meitzler has developed this type of line for high-quality transmission of pulses in aluminum strip materials at frequencies as high as 5 Mc/s, and for delay times as long as 10 000 μs.

If a compressional mode transducer is bonded to the rectangular strip structure, an infinite set of longitudinal modes may be propagated. The most important of these is the first longitudinal mode as shown in Fig. 3. At the lowest inflection point there exists a region of approximately linear delay vs. frequency variation, which is use-

* The insertion loss of a delay line may be divided between the loss in the delay medium, which may be dissipative, geometric, or both, and transducer loss, which may be defined as the ratio of the electric signal input to the acoustic signal output or vice versa. Transducer loss includes the effects of electrical or mechanical impedance mismatch, which, by proper design, may be made very small, and dissipation in the transducer material is usually negligible. For the case cited, transducer losses of 2 or 3 dB are common.

ful for achieving a dispersive delay line. Meeker[6] has developed a longitudinal mode strip delay line operating in this region to perform as a dispersive network for achieving the pulse compression function in chirp radar systems.[7]

Similar families of propagating modes exist in a round wire. Analogous to the shear modes in a strip are the torsional modes, the lowest of which is nondispersive and has been utilized extensively in delay lines that are driven magnetostrictively and, more recently, piezoelectrically. The lowest-order longitudinal mode has a delay vs. frequency characteristic similar to that shown for the strip in Fig. 3. The quasi-nondispersive region of low frequency–diameter product has been employed in many magnetostrictive delay lines. Because the transducer coupling to the wire is through the magnetic field rather than by mechanical contact, the longitudinal mode lends itself well to providing continuously variable delay and multiple taps. The linear region of the delay curve surrounding the lowest inflection point can be used to provide a linear delay vs. frequency characteristic,[8] and is particularly useful at low frequencies where a strip many wavelengths in width becomes prohibitively large.

In the uniform thickness strip, the fractional bandwidth of the linear region is limited to about 15 percent by the shape of the delay characteristic. A. H. Fitch[9] devised a method of increasing the fractional bandwidth by tapering the thickness of the strip. The principle of operation may be seen in Fig. 4, in which the delay characteristics of several uniform strips of different thickness have been added to achieve a composite delay characteristic that has a much larger fractional bandwidth for the linear delay vs. frequency region. Fractional bandwidths as large as 40 percent have been achieved with this type of line.

Nondispersive and dispersive delay lines

Nondispersive delay lines are used for storage of both analog and digital information; the latter application is illustrated by the recirculating memory loop of Fig. 5. Total storage capacity is best illustrated by considering the digital capacity as shown in Table I. In this type of operation, the important quantities are the bit rate, the storage time, and the product of these, which gives the total number of bits that a given delay line can store. Table I lists the maximum values of bit rate and delay time which result in maximum storage capacity for a given type of delay line. Although either the bit rate or the delay time may be increased from the values shown, a decrease in the total bit capacity will result. This circumstance arises because the losses in the delay medium increase with frequency. Thus either the total attenuation or the increase in attenuation over the bandwidth of the delay line sets a limit on the maximum bit-storage capability for a given delay line.

The characteristics of presently available dispersive

I. Memory and storage capacity

	Maximum Bit Frequency, Mc/s	Maximum Delay, μs	Total Bits
Magnetostrictive wire	2	10 000	20 000
Shear wave strip	2	10 000	20 000
Fused quartz polygon	20	3 000	60 000

II. Dispersive delay

	Center Frequency, Mc/s	Fractional Bandwidth, μm	Delay Variation, μs	Compression Ratio
Metal strips	1 (min)	0.15	1000	150
	40 (max)	0.10	40	160
Tapered strip, metal	5	0.40	32	65
Tapered strip, glass	25	0.40	20	200

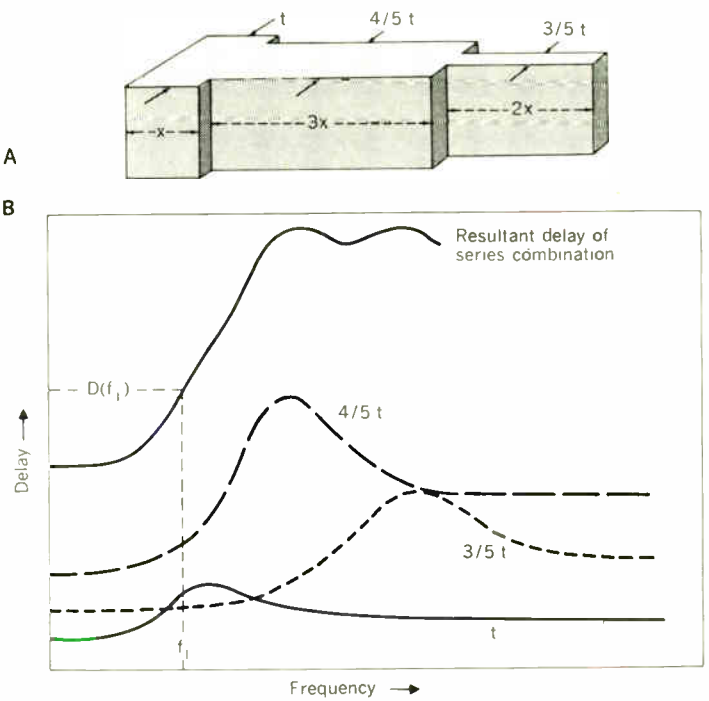
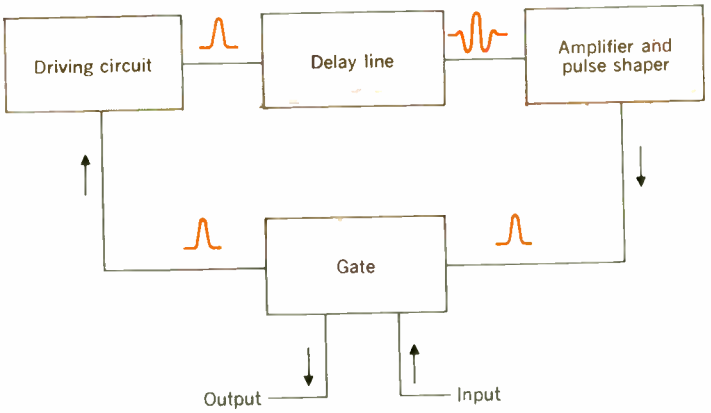
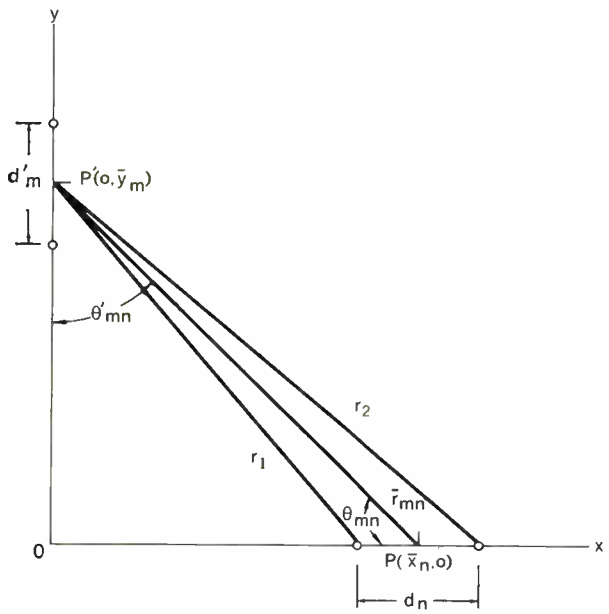
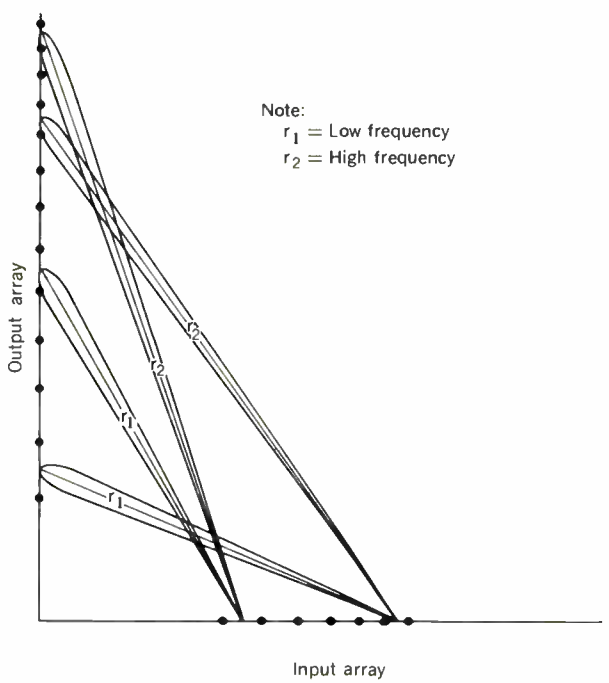


Fig. 4. Tapered strip delay line. A—Geometric configuration. B—Graphic plot of composite delay vs. frequency characteristic (after Fitch).

Fig. 5. Block diagram of a recirculating memory loop.







delay lines are summarized in Table II. Here the quantity useful for comparison purposes is the product of the bandwidth and the delay variation. This is known as the pulse compression ratio, a quantity of importance in characterizing delay lines for pulse compression radar applications. The metal-strip delay lines can be operated either at relatively low frequencies, with large delay variation, or at moderately high frequencies, with much smaller delay variations. It can be seen that in the tapered-strip case, although larger fractional bandwidths are obtained, the compression ratios are comparable with the uniform metal-strip delay line.

Grating-type delay lines

The strip delay lines just described operate at a thickness of approximately one-half wavelength (this thickness at 30 Mc/s is about 75 microns), and it thus becomes increasingly more impractical to obtain large bandwidth by raising the frequency of operation. A completely different approach to the dispersive delay line problem is taken in the family of delay lines known as grating delay lines, early versions of which were proposed by R. S. Duncan and M. R. Parker,[10] and by W. S. Mortley.[11]

In the Duncan and Parker version, called the perpendicular diffraction delay line (PDDL),[10,12] the principle of operation may be seen by considering Fig. 6, in which two sets of nonuniformly spaced transducer elements are placed at points on the x and y axes that represent two adjacent perpendicular edges of a block of suitable delay medium. In each set, all elements are connected together. For the two pairs of elements shown, maximum output will occur at P' [P] on the y [x] axis when the radiation from two adjacent elements on the x [y] axis meets the condition for constructive interference

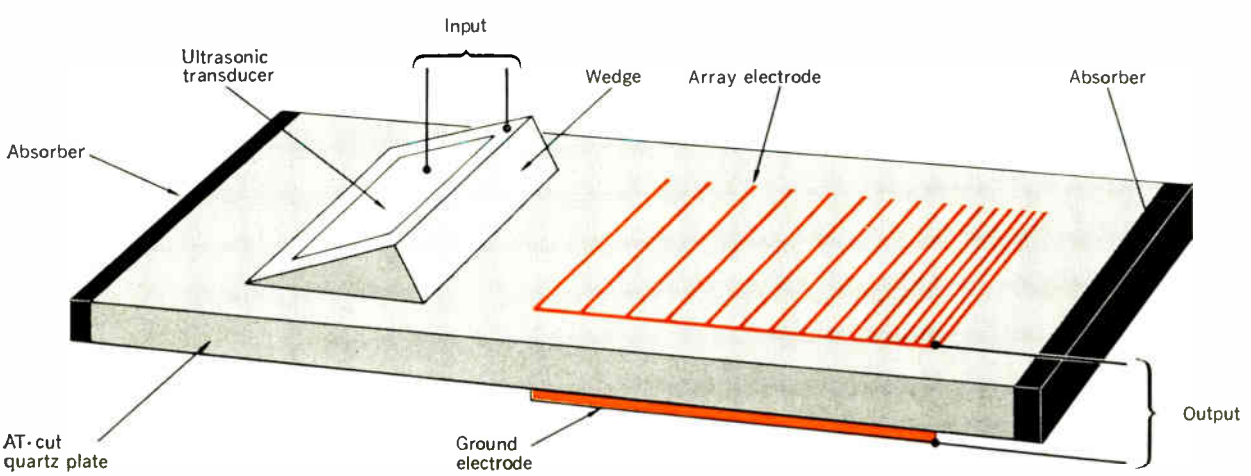
$$\frac{\bar{x}_n d_n}{r_{mn}} = \frac{\bar{y}_m d_m'}{\bar{r}_{mn}} = p\lambda$$

where $\bar{x}_n$ and $\bar{y}_m$ are mean element coordinates, d_n and

Fig. 6 (top). Ray diagram for perpendicular diffraction line (after Coquin and Tsu).

Fig. 7 (left). Perpendicular diffraction line (after Rowen).

Fig. 8 (below). Surface-wave delay-line structure showing a launching wedge for exciting surface waves with a conventional transducer and grating transducer (after Rowen).



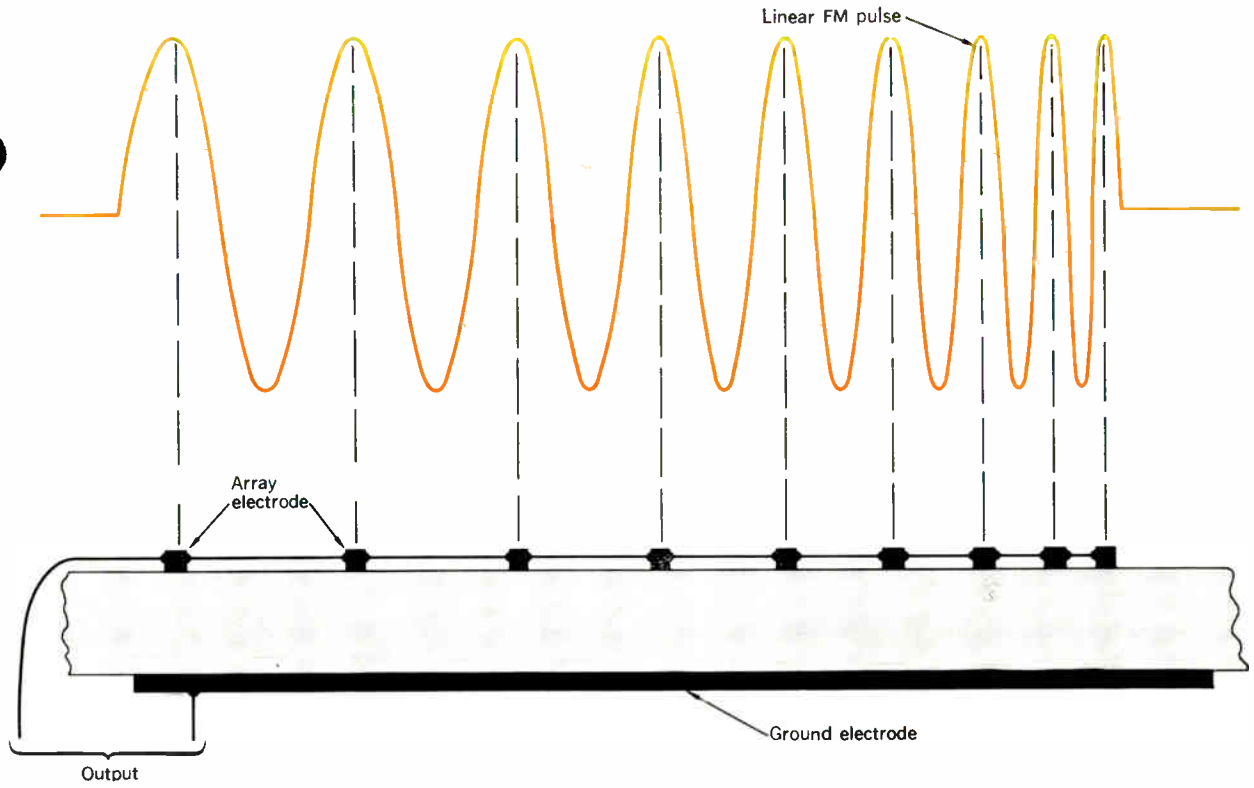


Fig. 9. Pulse compression with a grating transducer on a surface wave line (after Rowen).

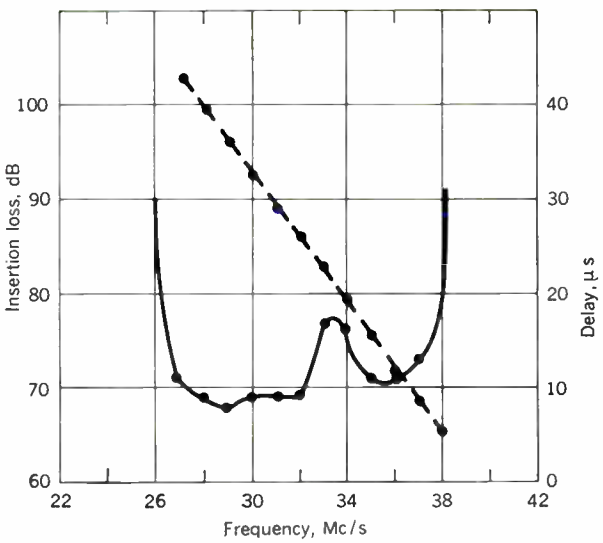


Fig. 10. Loss and delay characteristic of surface wave line (after Sittig).

d_m' are element spacings, r_{mn} is the distance from P to P', P is an integer, and λ is the wavelength. If the grating spacings d_n [d_m'] are chosen so that at a given frequency f, a large number of element pairs satisfy the condition

$$\bar{x}_n\, d_n = y_m\, d_m' = \text{Constant} = p\lambda\, r = pr\, v_s/f$$

where v_s is the sound velocity, then the delay time will be a linear function of frequency given by

$$T = \frac{r}{v_s}\left(\frac{\text{Constant}}{P v_s^2}\right) f$$

This is the delay vs. frequency relation required to produce compression of a linearly frequency-modulated pulse signal applied at the input grating into a signal at the output grating that was a (sin T)/T amplitude distribution as required for chirp radar.[7] The PDDL is shown diagrammatically in Fig. 7 where, at a low frequency, the two rays r_1 have equal lengths and, at a higher frequency, the two rays r_2 paradoxically have equal but greater lengths.

A third form of grating-type delay line is the surface wave delay line invented by J. H. Rowen[13] and shown in Fig. 8. This delay line is composed of a piezoelectric plate that serves as the propagating medium and as a transducer because an array of electrodes placed on the surface causes surface wave displacements to be generated in the region immediately beneath the electrode elements. Operation of this type of delay line can be illustrated by considering the linearly frequency-modulated pulse to be imposed by a distant transducer such as that on the wedge of Fig. 8, and to be propagated by a nondispersive surface wave. On arrival of the strain pulse beneath the transducer elements, as shown in Fig. 9, again the coincidence of strain maxima and array elements causes maximum output and the expected sin T/T distribution results. It may be seen that this type of line could also be constructed using two array-type transducers and that, when the spacing of the transducer elements decreases in the direction of wave propagation for both transducers, nondispersive operation will result. If the orientation of one of the arrays is reversed, dispersive operation results. Further, the slope of the resulting delay characteristic can be reversed merely by changing the orientation of the arrays. As an example of the operation of the surface-wave array delay line, the data shown in Fig. 10 are

presented. These data are taken from a two-array delay line developed by E. K. Sittig[14] by using a plate of crystalline quartz. The delay characteristic approximates a straight line very closely and the loss characteristic follows the shape expected from an array-type transducer with the exception of the loss peak at about 34 Mc/s. This peak could be shifted out of the band of interest by proper design of the plate thickness. Because fabrication of the transducers for the array-type delay lines lends itself to printed circuit technology, it appears quite feasible to construct these delay lines for operation at frequencies in excess of 100 Mc/s. Fractional bandwidths up to 50 percent can be achieved.

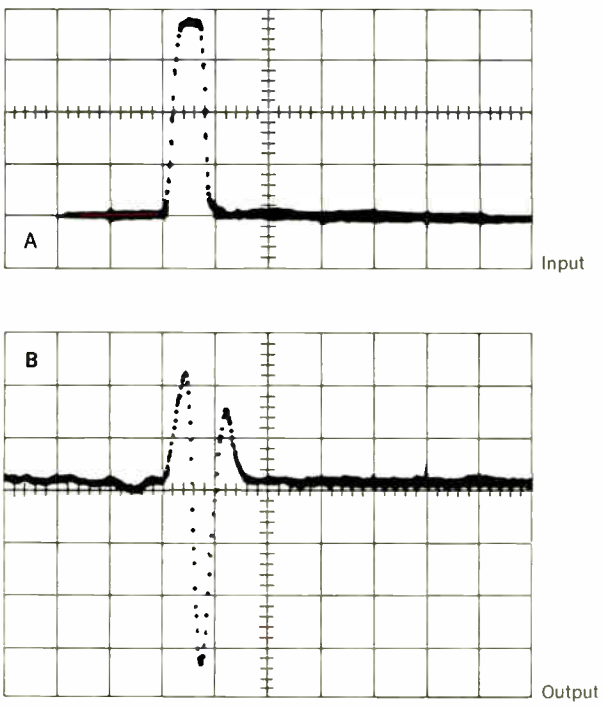
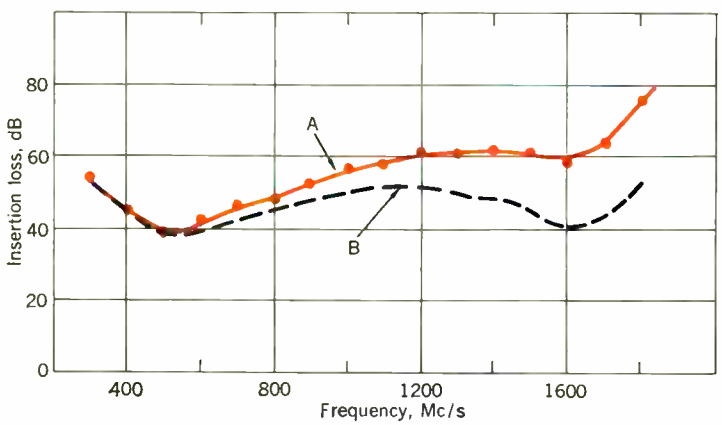


Fig. 11. Transmission of a 3-ns pulse through a 5-μs quartz line with evaporated layer CdS transducers. A—Input. B—Output; time scales, 5 ns/cm (after Foster).

Fig. 12. Plot of acoustic bandwidth for a CdS transducer on a BC-cut quartz bar. Curve A represents total loss; curve B represents residual loss after subtracting loss in BC-cut quartz bar (after Foster).



Evaporated layer transducers and UHF delay lines

The nondispersive delay lines also have similar fabrication limitations at frequencies above about 100 Mc/s because of the required thinness of the transducer. Several different schemes have been investigated for efficient high-frequency transducer action that extends into the microwave frequency range. A number of very useful new transducer structures have been developed which take advantage of the existence of materials such as ZnO and CdS. Both materials are semiconducting and strongly piezoelectric.[15] The characteristics of elastic wave propagation in such media have been extensively studied by Hutson and White.[16] Among these are the depletion layer and diffusion layer transducers invented by White.[17] The depletion layer transducer appears to be most suitable for operation in frequencies in excess of 3 Gc/s because of limitations on the thickness of the depletion layer that can be produced in commonly available materials.

The diffusion layer transducer, formed by diffusing a compensating material into a highly conductive piezoelectric semiconductor material, has been developed by Foster[18] for efficient operation over the frequency range from about 50 Mc/s to about 5 Gc/s. More recently, Foster[19] invented a method for evaporating cadmium sulfide (CdS) to form an oriented polycrystalline layer that produces transducer action with the lowest loss yet observed for high-frequency transducers. Losses as low as 10 dB can be achieved for a pair of transducers at frequencies up to 400 Mc/s, 20 dB at 1 Gc/s, and 30 dB at 2 Gc/s. These transducers are formed by evaporating CdS on a metallic layer under conditions in which a highly oriented aggregate of crystallites is produced. When a second electrode is applied over the CdS layer, it becomes possible to confine the driving field to the piezoelectric medium and efficient operation is the result. By controlling the direction of evaporation,[20] the films can be oriented to produce either longitudinal mode generation or shear mode generation, with high suppression of the other mode components. An advantage over the diffusion layer transducer is that the evaporated layer can be applied to any selected delay medium without the use of an intervening bond. An example of the operation of 300-Mc/s shear mode, evaporated layer transducers on a 25-millimeter-long fused quartz bar, is shown in the pulse photographs of Fig. 11. Intended for digital pulse delay, the input to the line is a dc pulse of about 3-ns duration, and the output pulse shows that operation at a 200-Mc/s bit rate is clearly feasible. For a transducer with a much higher fundamental frequency,* the acoustic bandwidth curve (obtained by tuning out the reactive component of transducer admittance at each frequency) for a pair of shear mode transducers deposited on a BC-cut quartz bar is shown in Fig. 12. The dotted curve indicates the transducer bandwidth and it is obtained by subtracting the losses in the BC-cut quartz.

No upper frequency limit has been established as yet for operation of these transducers; however, J. F. de Klerk,[21] using an evaporated film of CdS without metallic electrode layers, has excited these films by employing cavity excitation[22] at frequencies as high as 70 Gc/s. In the latter type of structure, however, the transduction

* The term "fundamental frequency" is used here to define the frequency at which the transducer is one-half wavelength in thickness.

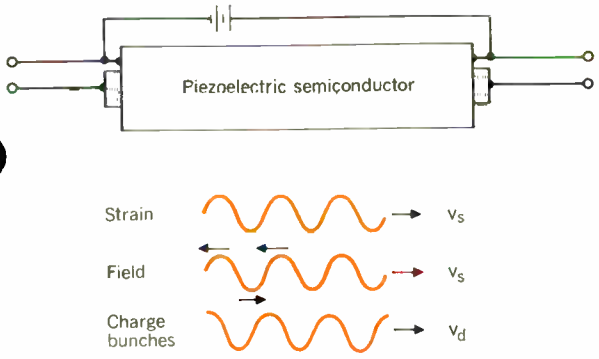


Fig. 13. Ultrasonic traveling-wave amplifier structure.

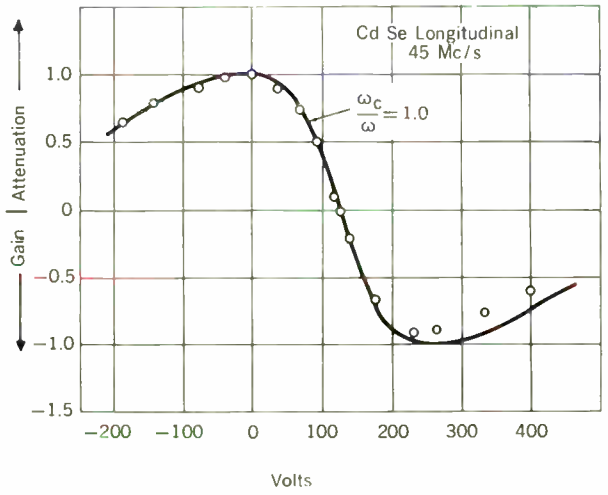


Fig. 14. Amplifier gain vs. applied field.

process is less efficient because the field is not confined to the piezoelectric medium and the bandwidth is limited by the necessity of using a high-Q cavity to achieve sufficient field intensity.

With evaporated layer transducers capable of producing transducer loss of the order of 20 dB at 1 Gc/s, one can consider a delay line constructed of single-crystal material such as BC quartz, for example, which has a loss at room temperature of about 1 dB per centimeter. Assuming that a total delay line loss of 60 dB could be tolerated, a total delay of about 120 μs could be attained. When these transducers are operated under matched conditions, bandwidths of the order of 20 percent are achieved; higher fractional bandwidths are obtainable by electrical mismatch with consequent increase in losses. Thus in the example cited, when matched, a bit rate of 200 Mc/s could be achieved, which would result in a total storage capacity of 24 000 bits. In practice, however, because of the increase in loss with frequency, the bandwidth would be limited so that a capacity of only 10 000 bits could be realized.

Ultrasonic traveling-wave amplifiers

In addition to the devices already described, which are capable of introducing various types of delay characteristics with accompanying attenuation, there also exist devices for achieving amplification of ultrasonic waves, such as the ultrasonic traveling-wave amplifier suggested by White and realized in the experiments of Hutson, McFee, and White.[23] The principle of operation of this device may be seen by referring to Fig. 13, which is basically the diagram of the delay line, except that the delay medium is now a piezoelectric semiconductor material and a dc field is applied. When the strain wave is propagated through the material, a field is produced in the direction shown; which direction causes bunching of the free carriers within the semiconductor. If the carriers are given a drift velocity slightly in excess of the velocity of sound, the interaction will transfer energy from the drifting carriers to the elastic wave and amplification results.

One of the major problems in the development of the ultrasonic amplifier is the availability of materials that have the required uniformity and resistivity. Possible piezoelectric semiconductor materials that are available are gallium arsenide, cadmium selenide, cadmium sulfide, and zinc oxide. Gallium arsenide has a very small electromechanical coupling factor and, therefore, the interaction with the free carriers is considerably less than in the other materials, and the principal importance appears to be for applications at frequencies below 100 Mc/s. Compared with the remaining materials, CdS is more readily available in crystals of reasonable size and quality and, therefore, most of the work done has been with this material.

The shape of the gain vs. voltage characteristic may be seen from Fig. 14. When the applied voltage is lower than

III. Calculated characteristics of a 1000-Mc/s CdS amplifier

Insertion Loss (two transducers), dB	Net Gain, dB	Input Impedance, ohms	Fractional Bandwidth, microns	Efficiency, percent	Noise Figure, dB	Equivalent Noise Temperature, °K
0	40	375	0.03	4.2	4.3	956
4	40	110	0.10	1.3	9.7	2410
20	40	38	0.31	0.15	15.0	8850

Dimensions and properties of CdS (L) evaporated layer transducers:

Area = 0.36 mm²

Gain section, length = 1.28 mm
ρ = 378 ohm·cm
β = 1.4
dc power input, 3780 mW

Isolator section, length = 1.6 mm
ρ = 292 ohm·cm
β = 1.08
dc power input, 3600 mW

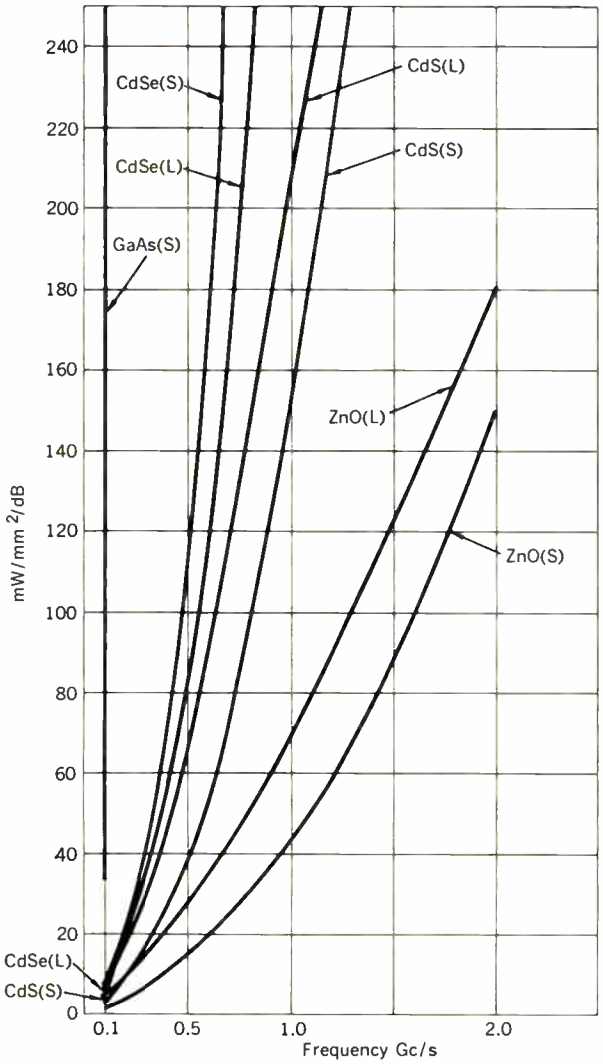
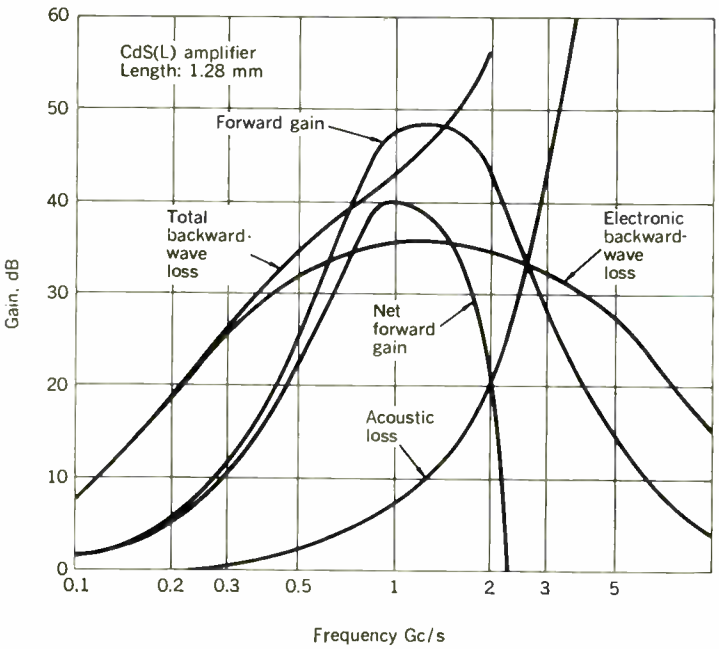


Fig. 15. Amplifier "figure of merit" for various materials.

Fig. 16. Gain vs. frequency for a longitudinal-wave, 1-Gc/s CdS amplifier, 1.3 mm in length.



that required to produce a drift velocity equal to the sound velocity, attenuation results. When the voltage is increased in excess of this value, gain is produced. With further increase in voltage, the gain passes through a maximum. Also shown in Fig. 14 are experimental points obtained by White on a cadmium selenide sample at 45 Mc/s. These show good agreement with the theoretical calculations.

In the early experiments in this field, a transducer was bonded to one end of a quartz buffer rod and the piezoelectric semiconductor was mounted between a pair of buffer rods. This structure provided isolated terminals for application of the drift field. In these earlier experiments, attenuation in the buffer rods and associated bonds was so large that they overrode the gain produced in the piezoelectric semiconductor sample. Recently, net electronic gain as high as 60 dB, by using quartz transducers bonded directly to the amplifier, has been obtained by Hickernell and Sakiotis,[24] and by Wanuga.[25] Net gain has been achieved at 60 Mc/s in gallium arsenide (GaAs) by Hickernell.[26] White and his co-workers[27] have achieved overall gains as high as 30 to 40 dB by using the buffer rod structure and CdS evaporated layer transducers at frequencies from 100 to 1000 Mc/s.

It is clear that the ultrasonic amplifier will be useful for amplification of ultrasonic waves, and it may also be useful for amplification of electronic signals. This latter application requires consideration of the loss and bandwidth introduced by the transducers. In an effort to evaluate the usefulness of such an amplifier when suitable crystals are developed, the author[28] has calculated, using White's theory, the characteristics to be expected when operating with evaporated layer transducers having various amounts of insertion loss. These are summarized in Table III. Noise figure calculations are based on a theoretical treatment by Hanlon and White,[29] which includes the effects of transducer loss and of shot noise on the drifting carriers as calculated by Blotekjaer and Quate.[30] Thus the effect of the transducer loss is an appreciable increase in the noise figure.

One of the major problems in operating the amplifier at increasingly high frequencies is the increase in dc power dissipation. The writer[28] has developed a criterion for determining an optimum operating point which minimizes the dc power dissipation. If, at these operating points, the power dissipation per unit volume per unit gain is plotted as a function of frequency, as in Fig. 15, a criterion then exists for comparing the various materials for amplification purposes. Although at 100 Mc/s there is a wide choice of materials, at higher frequencies the advantage of zinc oxide (ZnO) becomes increasingly apparent.

Another problem in the amplification of electronic signals is assuring that signals reflected at the transducers cannot propagate in a round trip through the semiconductor material with a net gain greater than unity. The acoustic attenuation in CdS is quite high and has an appreciable effect at one gigacycle per second on the overall shape of the forward gain characteristic, as can be seen in the gain vs. frequency curves of Fig. 16. The reflected wave experiences an attenuation which varies with frequency in accordance with the electronic backward wave loss curve shown. When the effect of acoustic attenuation is added to this curve, the total backward wave loss exceeds the net forward gain by about 2 dB,

thus assuring stable operation. Bandwidth of the electronic amplifier is determined primarily by the transducer characteristics. Thus the ultrasonic amplifier shows promise as an amplifier, not only of ultrasonic waves, but also of electronic signals in the frequency range from 100 Mc/s to well above 1000 Mc/s.

Other elastic wave interactions

The ultrasonic traveling-wave amplifier is based on the interaction of elastic waves with electrons. Now that high-efficiency transducers exist in the range extending up to several gigacycles per second, it is appropriate to consider what other interactions may prove to be of value for ultrasonic devices. Listed among these should be the interaction of elastic waves with electrons through deformation potential and with spins in both ferrimagnetic and paramagnetic media. Piezoelectric amplification decreases because of carrier diffusion at frequencies where the acoustic wavelength approaches the electron mean free path in magnitude. At these same frequencies the interaction of elastic waves with electrons through the deformation potential—an effect which increases in strength with frequency—becomes significant in semiconductor materials. In germanium, cooled to liquid helium temperatures to reduce the acoustic loss, Pomerantz[31] has observed the amplification of phonons at 9 Gc/s by using the deformation potential.

In ferrimagnetic materials, the interaction of elastic waves with spin waves is called magnon–phonon interactions. Early experiments of Spencer, Denton, and Chambers[32] demonstrated an extremely simple ferrimagnetic transducer action and low acoustic loss in YIG. Strauss and Eggers[33] have investigated both dispersive and nondispersive YIG delay lines at frequencies up to 4 Gc/s. Nonreciprocal rotation of the plane of polarization of elastic waves in YIG rods has been observed by Matthews and LeCraw.[34] Parametric amplification of elastic waves in YIG has been achieved by Matthews.[35]

Tucker[36] has investigated the interaction of 10-Gc/s phonons with the paramagnetic spins in ruby. When the spin resonance frequency coincided with the phonon frequency, phonon attenuation was observed. When the population of the spin levels was inverted, amplification of the ultrasonic waves by stimulated emission from the spin system was noticed; thus the first phonon maser was demonstrated. Shiren[37] has demonstrated a double quantum detector which uses the three-level spin transitions of Fe^{++} in the magnesium oxide (MgO) lattice. By supplying photons and phonons of proper frequencies so that the sum of the energies of the two equal the energy difference between the first and third level, the presence of phonons of a particular energy is detected by the spin-resonance spectrometer. In another experiment, Shiren investigated three phonon processes by studying the interaction of two microwave-frequency sound beams of different but prescribed frequencies, and of sufficient intensity to produce nonlinear effects. Both the parametric amplification of phonons and the up-conversion of the sound frequency were observed.

Although this listing of recent phonon interaction experiments is necessarily brief and incomplete, it is indicative of the wide range of interaction possibilities which exist. These phenomena are, at present, in the research experiment stage, but it may be expected that some of these principles will form the basis for the sonic devices of the future.

REFERENCES

1. Bradfield, G., *Electron. Eng.*, vol. 20, p. 74; 1948.
2. Bommel, H. E., and Dransfeld, K., *Phys. Rev. Letters*, vol. 3, p. 83; 1959.
3. Konig, W. F., Lambert, L. B., and Schilling, D. L., *1961 WESCON Convention Record*, paper 37/3.
4. Arenberg, D. L., *J. Acoust. Soc. Am.*, vol. 20, p. 1; 1948.
5. Meitzler, A. H., "Ultrasonic delay lines using shear modes in strips," *IRE Trans. on Ultrasonics Eng.*, vol. UE-7, no. 2, pp. 35–43; June 1960.
6. Meeker, T. R., "Dispersive ultrasonic delay lines using the first longitudinal mode in a strip," *IRE Trans. on Ultrasonics Eng.*, vol. UE-7, no. 2, pp. 53–58; June 1960.
7. Klauder, J. R., Price, A. C., Darlington, S. J., and Albersheim, W. J., *Bell System Tech. J.*, vol. 39, p. 745; 1960.
8. May, J. E., Jr., "Wire-type dispersive ultrasonic delay lines," *IRE Trans. on Ultrasonics Eng.*, vol. UE-7, no. 2, pp. 44–53; June 1960.
9. Fitch, A. H., *J. Acoust. Soc. Am.*, vol. 33, p. 1658; *Ibid.*, vol. 35, p. 709; 1961.
10. Duncan, R. S., and Parker, M. R., "The perpendicular diffraction delay line: a new kind of ultrasonic dispersive device," *Proc. IEEE*, vol. 53, no. 4, pp. 413–414; Apr. 1965.
11. Mortley, "Dispersers for pule compression systems," AGARD Sym. on Radar Tech. for Detection, Tracking and Navigation, Apr. 7, 1965.
12. Coquin, G. A., and Tsu, R., "Theory and performance of perpendicular diffraction delay lines," *Proc. IEEE*, vol. 53, no. 6, pp. 581–591; June 1965.
13. Rowen, J. H., 1964 Ultrasonics Sym., paper J6, Santa Monica, Calif., Oct. 14–16.
14. Sittig, E. K., Private communication.
15. Hutson, A. R., *Phys. Rev. Letters*, vol. 4, p. 505; 1960.
16. Hutson, A. R., and White, D. L., *J. Appl. Phys.*, vol. 33, p. 40; 1962.
17. White, D. L., "The depletion layer transducer," *IRE Trans. on Ultrasonics Eng.*, vol. UE-9, no. 1, pp. 21–27; July 1962.
18. Foster, N. F., *J. Appl. Phys.*, vol. 34, p. 990; 1963. "The performance of dilational mode calcium sulphide diffusion layer transducers," *IEEE Trans. on Ultrasonics Eng.*, vol. UE-10, no. 1, pp. 39–44; July 1963.
19. Foster, N. F., "Ultra-high frequency cadmium-sulphide transducers," *IEEE Trans. on Sonics and Ultrasonics*, vol. SU-11, no. 2, pp. 63–68; Nov. 1964.
20. Foster, N. F., 1964 Ultrasonics Sym., paper K2, Santa Monica, Calif., Oct. 14–16.
21. deKlerk, J. F., and Kelly, E. F., *Appl. Phys. Letters*, vol. 5, p. 2; 1964.
22. Bommel, H. E., and Dransfeld, K., *Phys. Rev. Letters*, vol. 2, p. 298; 1959.
23. Hutson, A. R., McFee, J. H., and White, D. L., *Phys. Rev. Letters*, vol. 7, p. 237; 1961.
24. Hickernell, L. F., and Sakiotis, N. G., "An electroacoustic amplifier with net electrical gain," *Proc. IEEE*, vol. 52, no. 2, pp. 194–195; Feb. 1964.
25. Wanuga, S., "CW electroacoustic amplifier," *Proc. IEEE*, vol. 53, no. 5, p. 555; May 1965.
26. Hickernell, L. F., 1964 Ultrasonics Sym., paper K6, Santa Monica, Calif., Oct. 14–16.
27. White, D. L., Hanlon, J. T., and Handelman, E. T., Private communication.
28. May, J. E., Jr., 1964 Ultrasonics Sym., paper K3, Santa Monica, Calif., Oct. (to be published in *Proc. IEEE*).
29. Hanlon, J. T., and White, D. L., Private communication.
30. Blotekjaer, K., Quate, C. F., and Hanson, W. W., Microwave Lab. Contract Rept., Stanford University, July 1963.
31. Pomerantz, M., *Phys. Rev. Letters*, vol. 13, p. 308; 1964.
32. Spencer, E. G., Denton, R. T., and Chambers, R. P., *Phys. Rev.*, vol. 125, p. 1950; 1960.
33. Strauss, W., and Eggers, F. G., *J. Appl. Phys.*, vol. 34, p. 1180; 1963.
34. Matthews, H., and LeCraw, R. C., *Phys. Rev. Letters*, vol. 8, p. 937; 1960.
35. Matthews, H., *Phys. Rev. Letters*.
36. Tucker, E., *Phys. Rev. Letters*, vol. 6, p. 547; 1961.
37. Shiren, N., *Phys. Rev. Letters*, vol. 13, p. 308; 1964.

Discounted-cash-flow analysis of the return on engineering investment

Technical and business factors governing engineering investments in new product development programs have grown more complex. Time has become an especially crucial factor, accounting for the increasing use of discounted-cash-flow techniques, which include the effect of the time value of money in financial studies

Richard Hackborn *Hewlett-Packard Company*

In recent years, manufacturing firms have shown growing concern over the problem of achieving their profit goals. The main causes of this concern are (1) the rapid increase of competition in the market place; and (2) the accelerating rate of obsolescence of newly introduced products. Firms faced with these problems or wishing to improve on their profit objectives are placing greater emphasis on management decision making, which facilitates more efficient use of the firm's resources.

This emphasis has focused attention on the effect of investment decisions on the selection of new product proposals. Capital budgeting has played a major role in this area. Crucial to the whole problem of making effective capital budgeting decisions is the method used to measure the *worth* of an investment. The use of an incorrect or meaningless method will greatly reduce the value of the other areas of capital budgeting. What makes for a method that can correctly and effectively measure an investment's worth bears further elaboration.

From a survey of the literature on investment worth methods, it appears that many of the authors place these methods into one of two categories: those that have "serious weaknesses" and those that are "theoretically sound."[1,2] A generally accepted breakdown by category of the more regularly used methods follows:

Serious weaknesses
1. Ranking by inspection
2. Payback period
3. Proceeds per dollar of outlay
4. Average annual proceeds per dollar of outlay
5. Average income on the book value of the investment

Theoretically sound
1. Yield of investment (also called discounted rate of return, internal rate of return, return on investment, or marginal efficiency of capital).
2. Net present value (also called excess present worth and present value).

There is common agreement among the authors of these classifications that a method must use discounted-cash-flow techniques in order to be considered "theoretically sound." (Both the yield and net present value methods use discounted-cash-flow techniques.) The reason for this agreement is that the most objectionable fault of all of the methods listed as having serious weaknesses is their failure to take into consideration the *time value of money*. This fault, in many instances, not only leads to incorrect measurements of the individual worth of a group of investments (ranking) but will also cause the serious-weakness methods to contradict each other directly for an identical set of investments.

Time value of money

The discounting, or compounding, approach is based on the fundamental assumption that the value of money depends not only on its amount but also on the time at

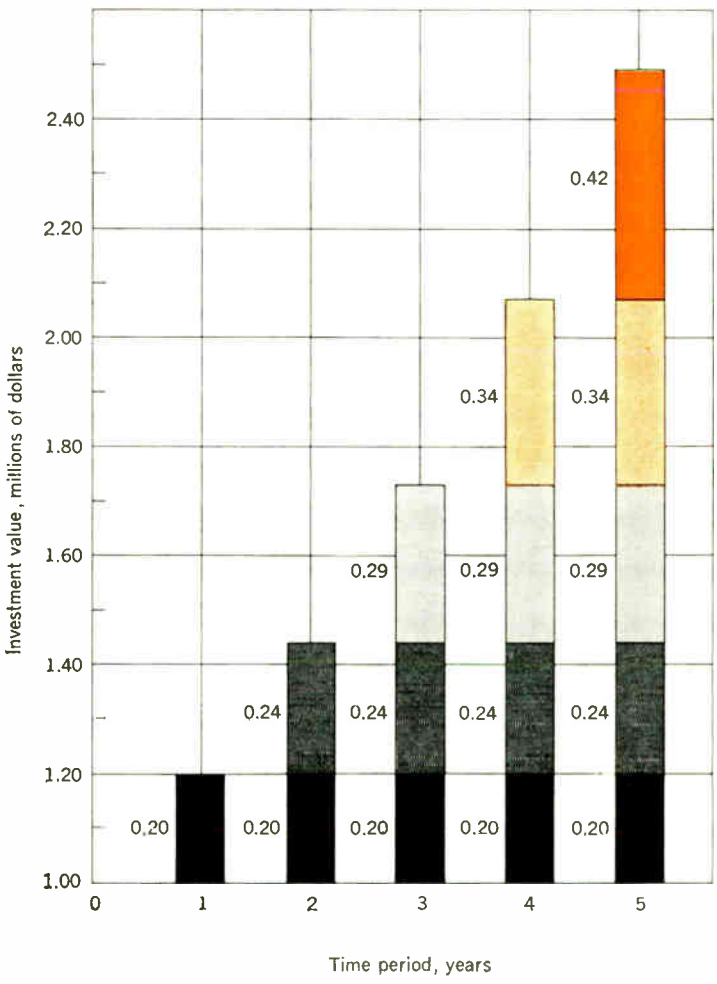
which it is available. This is the basic philosophy of the discounting approach: money has a time value. The time value of money exists as a business fact. Money that is available today can be invested in new projects that are expected to yield profits to the firm at some time in the future. On the other hand, money that is not available until some future date can not be invested profitably until it is received. Therefore, future funds have no profit-generating potential in the intervening time between the present and the date they are received. The discounted-cash-flow concept is simply a precise algebraic statement of the above business fact. A time value is placed on money by using the compound interest technique. (Compound interest should not be confused with simple interest, which is the interest rate on some fixed amount of money over time.) The compound interest concept is used in discounting techniques in the same manner as it is used in the familiar savings account procedures. In order to give a detailed description of the time value of money and compound interest concepts, the following investment example will be used.

Let us assume that we have a choice of having either $1 million today or five years hence. In addition, we will assume that we can invest the money annually and receive 20 percent increase on its value at the end of each year. This 20 percent per year increase could be due to the following possibility: The original $1 million would be invested in producing and selling $1.2 million of finished goods during the first year. The $1.2 million would then be invested in goods that sell for $1.44 million by the end of the second year (20 percent increase above $1.2 million). This process would then continue for the next three years.

Our $1 million investment is now growing at 20 percent per year. Let us now look at how this growth rate is obtained. Figure 1 shows the contributions made each year in order to achieve the 20 percent per year increase in the value of the money. At the end of the first year, the value has increased $200 000, reflecting the 20 percent increase in the original $1 million. (For reasons given later, it will be assumed that cash flows representing one time period occur entirely at the end of that period.) The second year shows that there are two contributions to the total increase of $440 000. The $200 000 earned at the end of the first year has been reinvested and is worth $240 000 at the end of the second year. Also, the original $1 million has been reinvested again, resulting in another $200 000. This mechanism of compound growth continues in succeeding years, and by the end of the fifth year the original $1 million invested has increased to $2.49 million. Our example, using a growth rate of 20 percent per year, shows that $1 million today will be worth at the end of five years $1.49 million more than $1 million received at that time. Thus, it is clearly evident that money is worth more today than an equal amount at some future time—money has a time value. Another relationship seen from Fig. 1 is that the increase in value of the original investment over time is generated by a compounding of growth. This relationship is based on the concept that the growth obtained in one time period is the basis for the growth generated in the next time period. Figure 1 demonstrates this compound growth concept by showing that the profits in any given year are a combination of the investment of profits from the previous years as well as from the present year. For example, during the third year the profits of the previous year ($440 000) grew to about $530 000 in addition to the present $200 000 contribution (obtained from the reinvestment of $1 million). Thus, the increase in value of money over time is a direct result of the compound growth of the original investment, and this is equivalent to the compound interest rate. In fact, the higher the compound interest rate (or growth rate) per year and the farther into the future the comparison is made, the greater will be the difference in the time value of money between these two different periods. These relationships are shown in Fig. 2.

The upper curve in Fig. 2 (compound interest rate equals 20 percent per year) is obtained directly from the bar graph in Fig. 1. Therefore, the curves in Fig. 2 only have meaning for zero and positive integer values (0, 1, 2, 3, etc.) on the time period axis. This happens because we have assumed that growth over one time period, or year, is only realized at the end of that year. In other words, we have assumed *discrete cash flows*. The reason for this assumption is that cash flows occurring at discrete times are the most common experience in the business world (e.g., weekly or semiweekly payrolls, quarterly tax payments, etc.). Also, by using this approach, the advantage is gained of having a time base that is consistent with most of the discount tables used for determining the

Fig. 1. Contributions to the increase in the value of a $1 million investment when the growth rate of the investment is 20 percent per year.

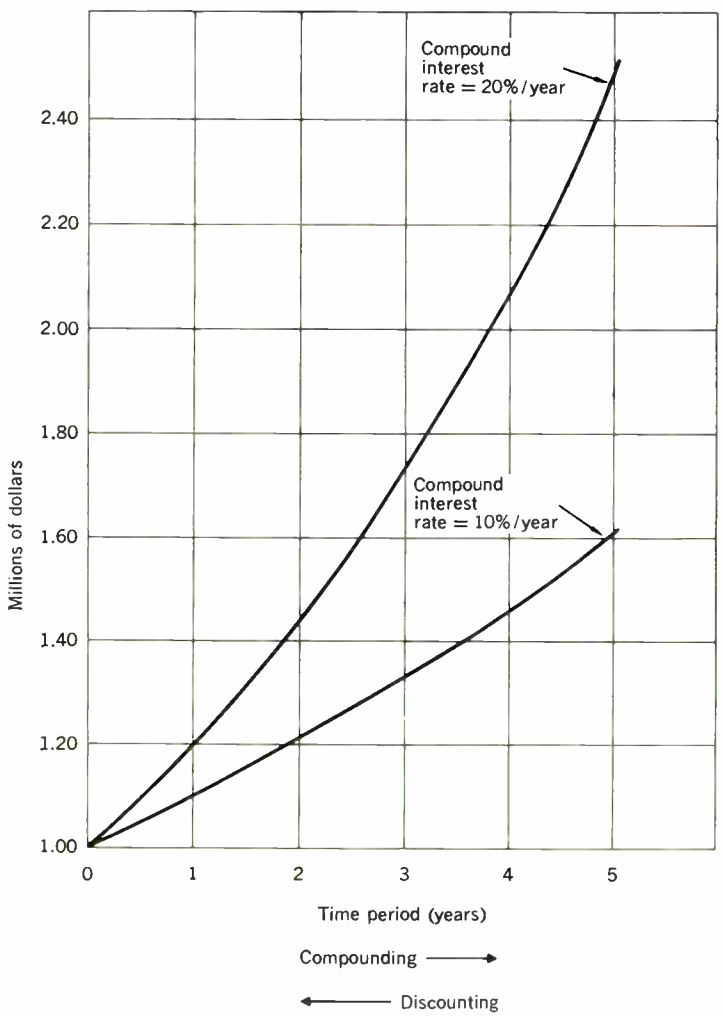


time value of money.[2-4] Even though the number of time periods would increase if smaller time intervals are desired, this would present no new theoretical problems but only an increased computational burden.

One further aspect of these growth-rate methods must be pointed out. We have made the implicit assumption that there are investment opportunities available to the firm that will allow it to realize a certain growth rate—in other words, it is economically a "going concern." If this is not the case, the compound interest rate methods have no meaning and will not work. Therefore, we must assume that we are applying these methods under conditions relevant to a "going concern."

Figure 2 also shows the relationship between funds available today and those available in the future. If the initial investment of $1 million grows to about $2.49 million in five years, then $1 million at that time is only worth about $402 000 today (1/2.49). This devaluation of future funds is called *discounting* and is the inverse of compounding. It can be seen (Fig. 2) that funds carried forward in time are compounded in value while future funds brought back in time to some equivalent present value are discounted. Thus, the time value of money can be determined by two exactly inverse, or reciprocal, relationships that are based on which direction in time we take. It is clear that the reason future funds have a lower equivalent present value is that they represent foregone investment opportunities. Using Fig. 2, we would say that $1 million five years from now is equivalent to $402 000 today based on a compound interest (growth) rate of 20 percent per year, because of the opportunity to invest the $402 000 for the next five years.

Fig. 2. Graph depicting future value of $1 million available today, as a function of time for two different compound interest rates.



Before we proceed to the actual discounted-cash-flow approach, it will be helpful to develop the fundamental mathematics used in discounting. Figure 1 will be the basis for the derivation of the following equations. We will recall that the profits realized after our first year of investment were $200 000. Thus the funds available for the following year consisted of the $200 000 of profit, in addition to the $1 million of original outlays. Using this approach, the compound amount (CA) of the original investment at the end of the first year equals (all values in millions of dollars):

$$\text{CA}\,[1] = \text{original outlay} + \text{profits} = 1 + r = 1.2$$

at the end of the second year,

$$\text{CA}\,[2] = 1.44 = \{\text{CA}\,[1]\}^2 = (1 + r)^2$$

This trend can be directly translated into the compound amount equation

$$\text{CA}\,[n] = (1 + r)^n$$

where

r = compound interest rate per year
n = number of years

The rate of increase in the amount of the compounded funds over time is obtained by taking the logarithm of the compound amount factor

$$\log \text{CA}\,[n] = n \log (1 + r)$$

Hence the funds increase at a rate that is equal to the logarithm of $(1 + r)$. This relationship is shown in Fig. 3 for the same data that are used in Fig. 2.

As was discussed previously, the discounting method is the exact inverse, or reciprocal, of the compounding method. It is conventional to call the discount amount factor the present value factor (PV $[n,r]$). The reason for this convention is that there is a discount method of investment analysis that devalues all future funds to a common present value. This method will be presented shortly. Mathematically, the present value factor is

$$\text{PV}\,[n,r] = \frac{1}{(1 + r)^n}$$

where

r = compound interest rate per year
n = number of years

Tabulated values of this factor can be found in the literature.[2,3]

There are two different variations of the discounted-cash-flow approach used to measure the worth of an investment. One is called the net present value (NPV) method and the other is called the yield of investment method. The latter method can lead to erroneous results

under conditions that are not uncommon to engineering-type investments. (These conditions are realized when a comparison, or ranking, of investment alternatives is required, and these alternatives have different lifetimes with individual yields that are considerably different than the average reinvestment rate. This problem is described and analyzed elsewhere.[4]) For this reason, we will only discuss the net present value method.

Net present value method

The net present value method is a direct result of applying the principle of "equating" future funds to present values. It will be remembered that this concept equates a sum of money available at some future time to an equivalent amount of money available today, for a given discount rate. This equivalence is, of course, a function of the compound interest rate used and the number of time periods over which the money is discounted (see Fig. 3). The net present value method uses this principle to reduce all future cash flows of an investment to an equivalent present dollar value for a given compound interest rate. As an example (using Fig. 3), an investment that costs $1 million today and generates $1.2 million in the first year and an additional $720 000 in the second year has an equivalent net present value of $500 000 for a compound interest rate of 20 percent per year:

Cash flow	Cash flow occurs at end of	Present value factor	Present value
$1 200 000	1st year	$\frac{1}{1.2}$	$\frac{1\,200\,000}{1.2} = \$1\,000\,000$
$720 000	2nd year	$\frac{1}{1.44}$	$\frac{720\,000}{1.44} = \$500\,000$

NPV = $1 000 000 + $ 500 000 − $1 000 000 = $500 000

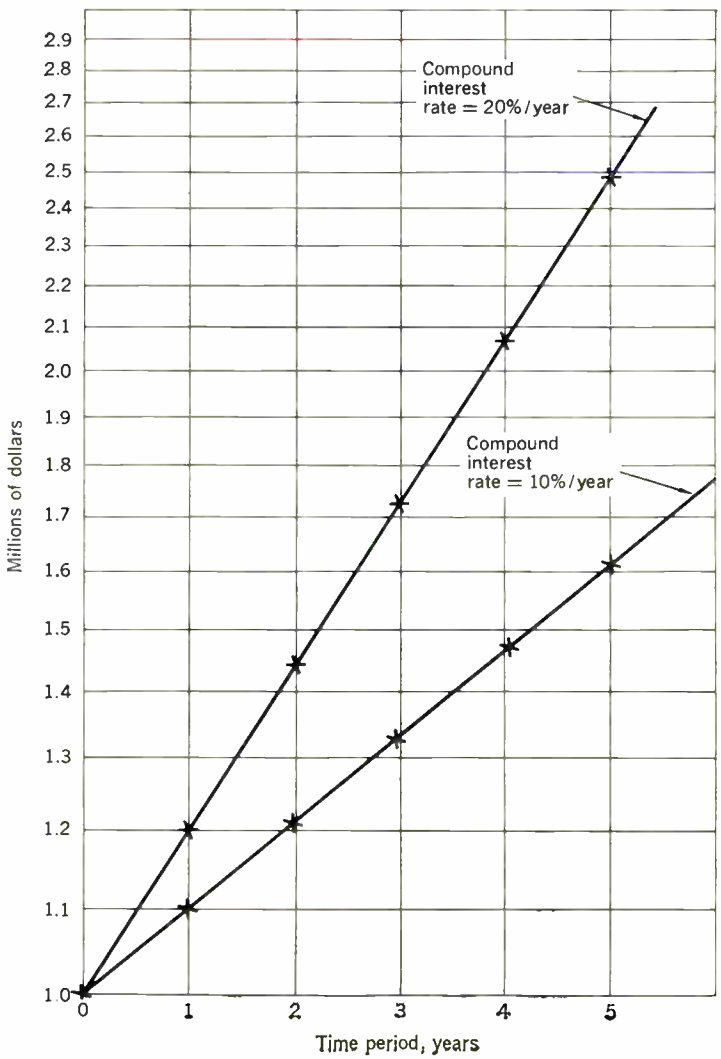
The net present value of this investment proposal as well as other investment alternatives would provide a "theoretically sound" comparison basis for management decision making. While this approach involves a straightforward use of the discounting concepts, there is one aspect that is open to a considerable amount of controversy. The area of disagreement is: What should be the criterion for picking the compound interest rate? Many different answers to this question can be found in the pertinent literature. They involve such criteria as "cost of capital," "minimum attractive rate of return," and others.[1–4] At this time, no single criterion appears to be the best under all conditions of capital budgeting. The following compound interest criterion is presented since it is particularly useful for describing discounted-cash-flow analysis of engineering investment.

It will be recalled that the "going concern" assumption is the main basis for placing a time value on money. Because a firm is not thinking of declaring bankruptcy but instead is making engineering investments with expected future returns, its money today has more value than an equivalent amount at some future time. In fact, the firm's time value of money for engineering investment, through the effect of the compound interest rate, is proportional to the overall profit growth rate that the firm has and/or expects. This profit growth rate can be thought of as representing an "average" engineering investment. Thus, if the firm invests money in a particular engineering project, an *opportunity cost* is incurred. This cost is a result of foregone investment opportunities in which the funds could have been invested. For our purposes, the foregone opportunity is the alternative of getting a return from the "average" engineering investment previously mentioned. Therefore, the compound interest rate represents the cost to the firm of not being able to invest in an "average" engineering project. It is assumed that this foregone investment opportunity is always available to a "going concern." From now on, compound interest rate will be called reinvestment rate.

Based on the above discussion, the criterion for picking the compound interest rate would be determined in the following manner:

1. Place the proposed engineering investment into one of several broad categories. These categories classify the proposed investment according to risk and other factors.
2. The return from an "average" engineering investment in the category that best describes the proposed investment is used to determine the reinvestment rate. This approximation should be done very carefully.

Fig. 3. Money value as a function of time for two different compound interest rates. Data from Fig. 2 are plotted on semilog paper.



3. The reinvestment rate corresponding to the selected category is used to determine the net present value of the proposed investment.

The following are some examples that illustrate the NPV technique.

Example I. This example is the first possibility of three different engineering investment alternatives for a new product development program. These alternatives all have identical expected profits ($300 000 per year for three years) with progressively shorter development times (going from alternative one to three) at increasingly higher total costs. This analysis is being made at the end of 1964.

Alternative One

End of Year	Cash Flows,* dollars
1964	0 (initial outlay)
1965	−100 000
1966	−100 000
1967	−100 000
1968	+300 000
1969	+300 000
1970	+300 000

Example II. The second investment alternative is to accelerate the new product development program by one year.

Alternative Two

End of Year	Cash Flows, dollars
1964	0 (initial outlay)
1965	−160 000
1966	−160 000
1967	+300 000
1968	+300 000
1969	+300 000

Example III. The third alternative would be to consider a highly accelerated development program for the new product.

Alternative Three

End of Year	Cash Flows, dollars
1964	0 (initial outlay)
1965	−450 000
1966	+300 000
1967	+300 000
1968	+300 000

We will now use these examples to calculate the net present value for each of the engineering investment alternatives.

It will be recalled that the net present value method was nothing more than discounting all future cash flows, of a given investment, back to a present value at some reinvestment rate. Let us assume that the reinvestment rate for the category describing the proposed investments is 30 percent per year.

The following equations are a direct result of the previous discussion:

Let

$$\mathrm{CF}[n] = \text{Cash flows at end of period } n$$

$$\mathrm{PV}[n, \mathrm{RI}] = \text{Present value factor} = \frac{1}{(1 + \mathrm{RI})^n}$$

$$\mathrm{RI} = \text{Reinvestment rate}$$

$$\text{Net present value (NPV)} = \sum_{n=0}^{N} \mathrm{PV}[n, \mathrm{RI}] \times \mathrm{CF}[n]$$

For Example I:

$$\mathrm{NPV} = \frac{0}{1} - \frac{100\,000}{1.3} - \frac{100\,000}{(1.3)^2} - \frac{100\,000}{(1.3)^3} + \frac{300\,000}{(1.3)^4} + \frac{300\,000}{(1.3)^5} + \frac{300\,000}{(1.3)^6}$$

$$\mathrm{NPV} = \$66\,379 \text{ (Example I)}$$

By similar calculations for Examples II and III:

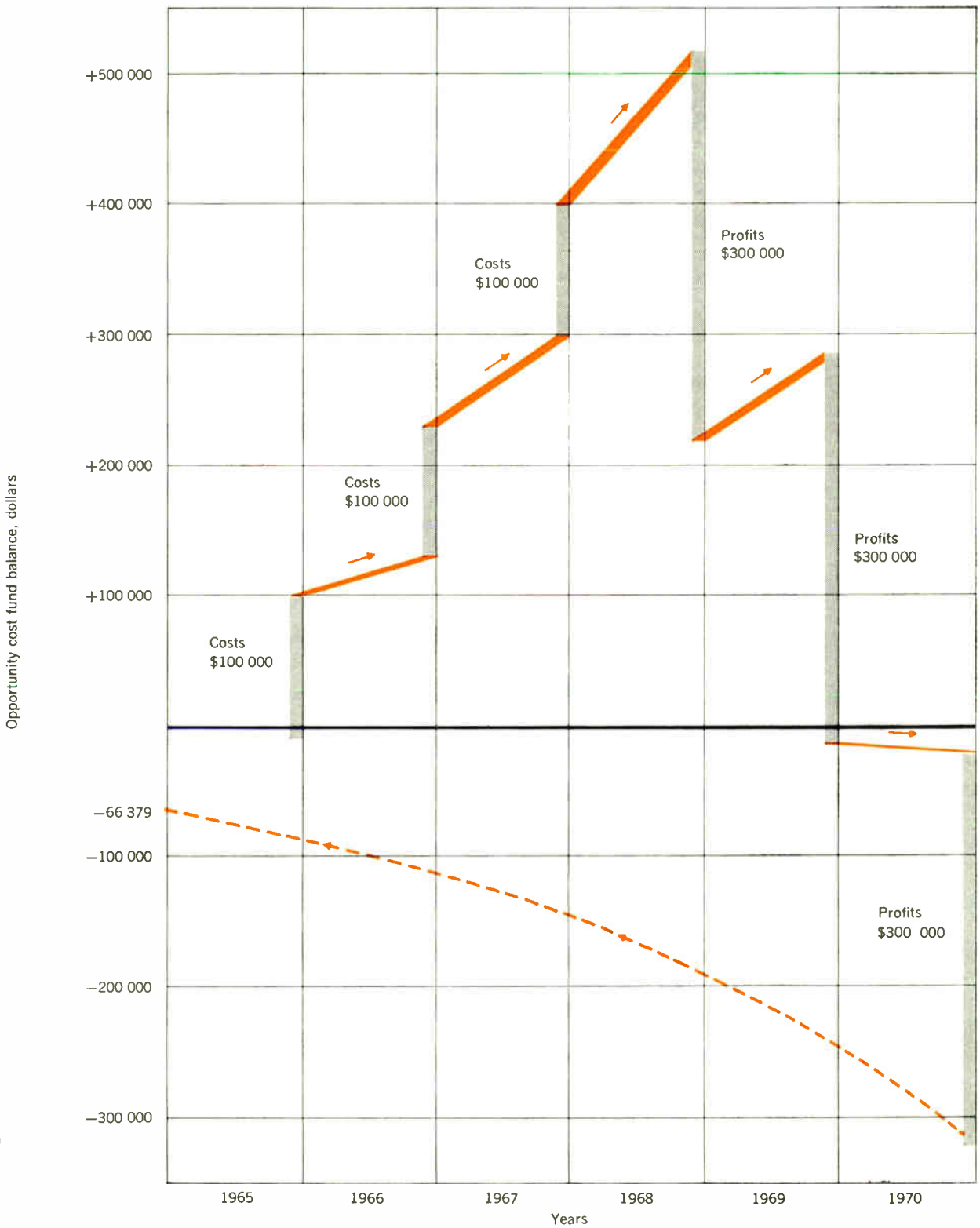
$$\mathrm{NPV} = \$104\,635 \text{ (Example II)}$$

$$\mathrm{NPV} = \$72\,949 \text{ (Example III)}$$

Thus we see that each of the three engineering investment alternatives is more profitable than the investment opportunity foregone ("average" investment in the group). Also, Alternative Two (Example II) is the best investment of the three. Figure 4 gives a graphic example of what is happening in the net present value method. This graph is determined by the following approach. Let us assume that instead of investing funds in Alternative One (Example I), we invest an equal amount of funds, at identical times, in the engineering investment opportunity foregone ("average" engineering investment in the group). Thus, the costs incurred in Example I would be transformed into deposits (positive additions) that are put into a fund whose balance is compounded at the reinvestment rate. Similarly, the profits are transformed into withdrawals from this fund. This fund will be called the "opportunity cost fund" (a fund whose balance is compounded at the reinvestment rate regardless of whether it is positive or negative), and its balance represents the advantage of investing in an "average" engineering investment, which grows at the reinvestment rate, over investing in the proposed engineering project. A negative balance, due to the inverse transformation, is identical to the positive NPV of a given proposed investment. In Fig. 4 we see that the first $100 000 of cash outlays at the end of 1965 are invested in the opportunity cost fund. This means that instead of putting the money into the proposed investment we are placing it in an investment that produces profits that are typical for this group of investments. The balance of this fund grows at 30 percent per year, resulting in $30 000 plus $100 000 at the end of 1966. The process continues until the end of 1968, when $300 000 are withdrawn from the balance. This represents the lost profits from the proposed engineering investment that is not undertaken. By the end of 1970 the comparison is finished and the remaining balance in the opportunity cost fund is discounted back at the growth, or reinvestment, rate to an equivalent present value. This value of −$66 379 is expected opportunity cost of not investing in the proposed engineering project. It means that if we invested in an "average" engineering project in the selected category instead of the proposed engineering investment, we are losing money (minus sign means costs). This approach makes good use of the "management by exception" principle. Simply by looking at the sign of its NPV, we can determine if the proposed engineering investment will generate more, or less, expected profits than the "average" engineering investment

* Note that the convention used in these methods represents cash outlays (costs) with a minus sign and cash proceeds (profits) with a plus sign.

Fig. 4. Graphic description of net present value method for Example I. Bars represent end-of-period cash flows. Reinvestment rate = 30 percent per year.



opportunity that is foregone. From our previous discussion and Fig. 4, it can be seen that a positive NPV indicates a proposed investment that will produce more potential future profits than an "average" investment in the selected category.

So far the net present value method has been used for determining the worth of a proposed engineering investment with respect to a given reinvestment rate. This is an external type of criterion because the proposed investment is being compared with a predetermined "average" investment. If a number of engineering investment possibilities existed, the net present value method would indicate that investment should be made in those engineering proposals that have a positive NPV. (In addition, investments in some that have a negative NPV may be made because the reinvestment rate represents an "average" engineering investment.) Thus the net present value method would rank the proposed investments according to the size of their NPV—the most desirable investment would have the largest positive NPV and the least desirable investment would have the largest negative NPV. This is correct *if* there are enough funds available for all of the desirable engineering investments. However, if funds are limited, investment proposals would be ranked on the basis of how efficiently the funds invested are used to generate future profits. A criterion that accomplishes this is called the "net present value return" method. This method gives a percentage measurement of the worth of an investment, which is defined as the ratio of the NPV of the proposed investment to the present value of its costs (cash outlays). Mathematically, the NPV return is described by the following equation:

$$\text{NPV return} = \frac{\text{NPV}}{-(\text{Present value of the costs})} \times 100\%$$

$$= \frac{\text{NPV} \times 100}{-\sum_{n} \text{PV}[n, \text{RI}] \times \text{CF}[n]}$$

where n = cash flows that represent costs.

To show how the NPV return method is used, three proposed engineering investments will be ranked according to how efficiently each investment generates future profits for each present value dollar invested. The reinvestment rate is 30 percent per year.

Example IV.

End of Year	Cash Flows, dollars
1964	−100 000 (initial outlay)
1965	+ 50 000
1966	+100 000
1967	+ 50 000

$$\text{NPV} = \frac{-100\,000}{1} + \frac{50\,000}{(1.3)} + \frac{100\,000}{(1.3)^2} + \frac{50\,000}{(1.3)^3}$$

$$= \$20\,391$$

$$\text{NPV return} = \frac{20\,391 \times 100}{-(-100\,000)} = 20.4\%$$

Example V.

End of Year	Cash Flows, dollars
1964	−20 000 (initial outlay)
1965	−40 000
1966	+70 000

$$\text{NPV} = \frac{-20\,000}{1} - \frac{40\,000}{(1.3)} + \frac{70\,000}{(1.3)^2}$$

$$= -\$9349$$

$$\text{NPV return} = \frac{-9349 \times 100}{-(-50\,769)} = -18.4\%$$

Example VI.

End of Year	Cash Flows, dollars
1964	−10 000 (initial outlay)
1965	+18 000

$$\text{NPV} = \frac{-10\,000}{1} + \frac{18\,000}{(1.3)}$$

$$= \$3846$$

$$\text{NPV return} = \frac{3846 \times 100}{-(-10\,000)} = 38.5\%$$

If the proposals are ranked according to their NPV returns, we would get

Proposed Investment	NPV Return, percent	NVP, dollars	Total Funds Required, dollars
VI	38.5	3 846	10 000
IV	20.4	20 391	100 000
V	−18.4	−9 349	60 000

We would first invest $10 000 in Example VI in the expectation of receiving $3846 in present value dollars of future profits. If there were sufficient funds available, we would next invest $100 000 in Example IV. On the other hand, if funds are insufficient, then management has to decide between obtaining additional funds or foregoing the opportunity to invest in Example IV. In other words, the constraint on the amount of investment (caused by limited funds) costs $20 391 in present value dollars of potential future profits. These examples indicate how discounted-cash-flow concepts through the use of the NPV and NPV return methods aid rational, objective, management decision making in the measurement, and selection, of proposed engineering investments.

It is true that while the increased sophistication of the discounted-cash-flow approach allows a more accurate measurement of the worth of an investment, it unfortunately requires a considerable number of additional calculations as compared with the serious-weakness methods. To reduce this computational burden, two special nomographs have been developed (using simple graphical techniques).[5] Which nomograph is used depends upon the pattern of the cash flows of the investment under consideration.

Case I. Uniform cash flows. The conditions required for uniform cash flows is that there are a given number of time periods of equal costs (cash outlays) per time period followed by a given number of time periods of equal profits (cash proceeds) per time period. Examples I, II, III, and VI have uniform cash flows.

Case II. Nonuniform cash flows. These cash flows correspond to all investments that do not meet the conditions in Case I. Examples IV and V have nonuniform cash flow patterns.

The nomograph used for uniform cash flow investments is shown in Fig. 5. To demonstrate how the nomograph is used to determine the NPV and NPV return of a proposed investment, let us use Example II.

Uniform cash flows nomograph (no initial outlay).

For Example II:

Investment lifetime = 5 years
Cash outlay duration = 2 years
Cost per year = $160 000
Profits per year = $300 000
Reinvestment rate per year = 30 percent

Find the point on the five-year curve (right-hand graph of Fig. 5) where the reinvestment rate is 30 percent per year. Make a mark on the annuity factor scale that is directly to the left of this point, as shown. Make a mark on the actual dollars per time period scale at 3, corresponding to $300 000. By drawing a straight line between the two marks, a lifetime present value (LPV) of 7.3, corresponding to $730 000, is obtained. Next, find the point on the two-year curve where the reinvestment rate is 30 percent per year. Proceeding as before, place a mark on the annuity factor scale. Add the costs per year and profits per year figures to obtain $460 000. Place a mark on the actual dollars per time period scale at 4.6.

Fig. 5. Present value nomograph for uniform cash flows.

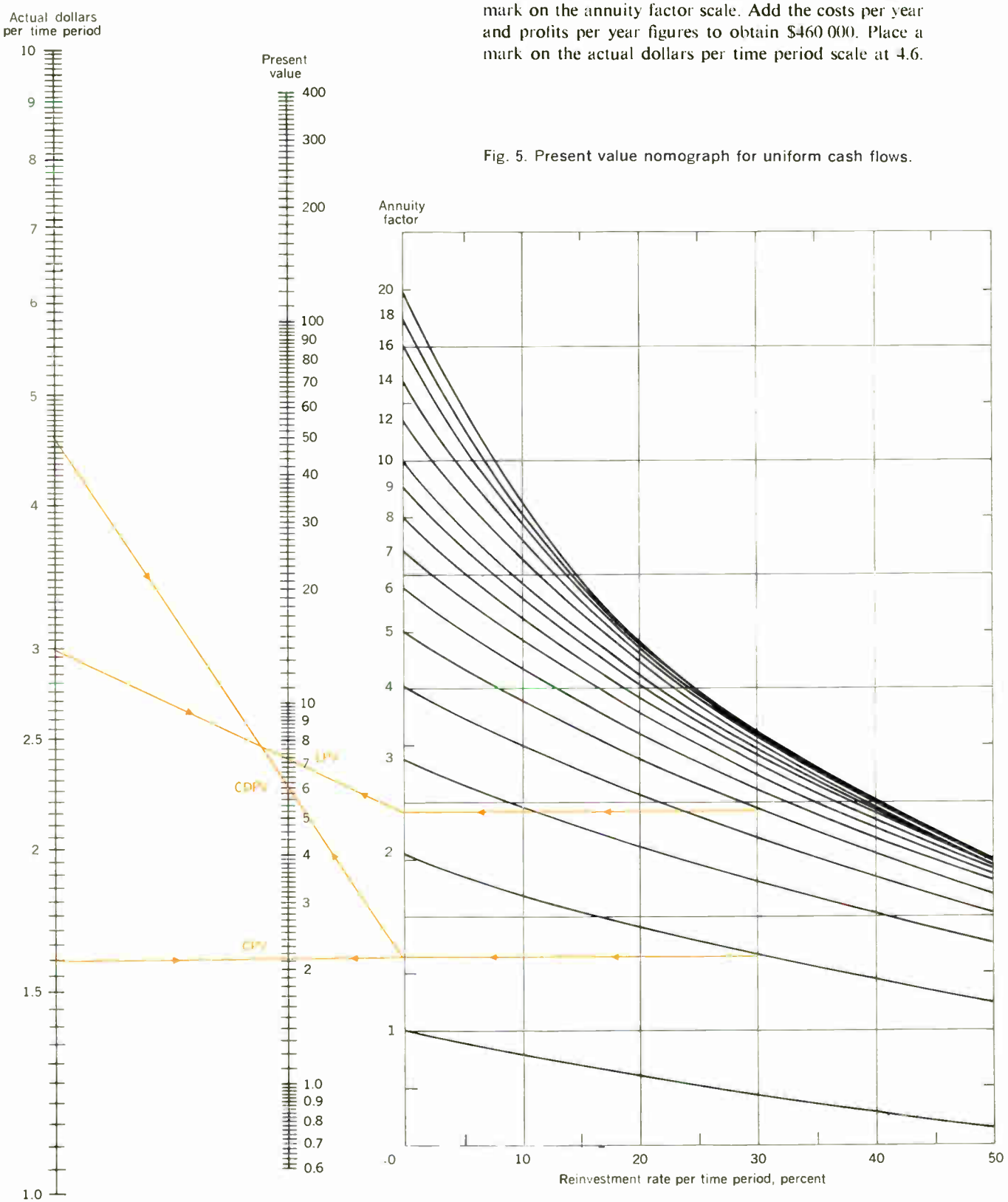
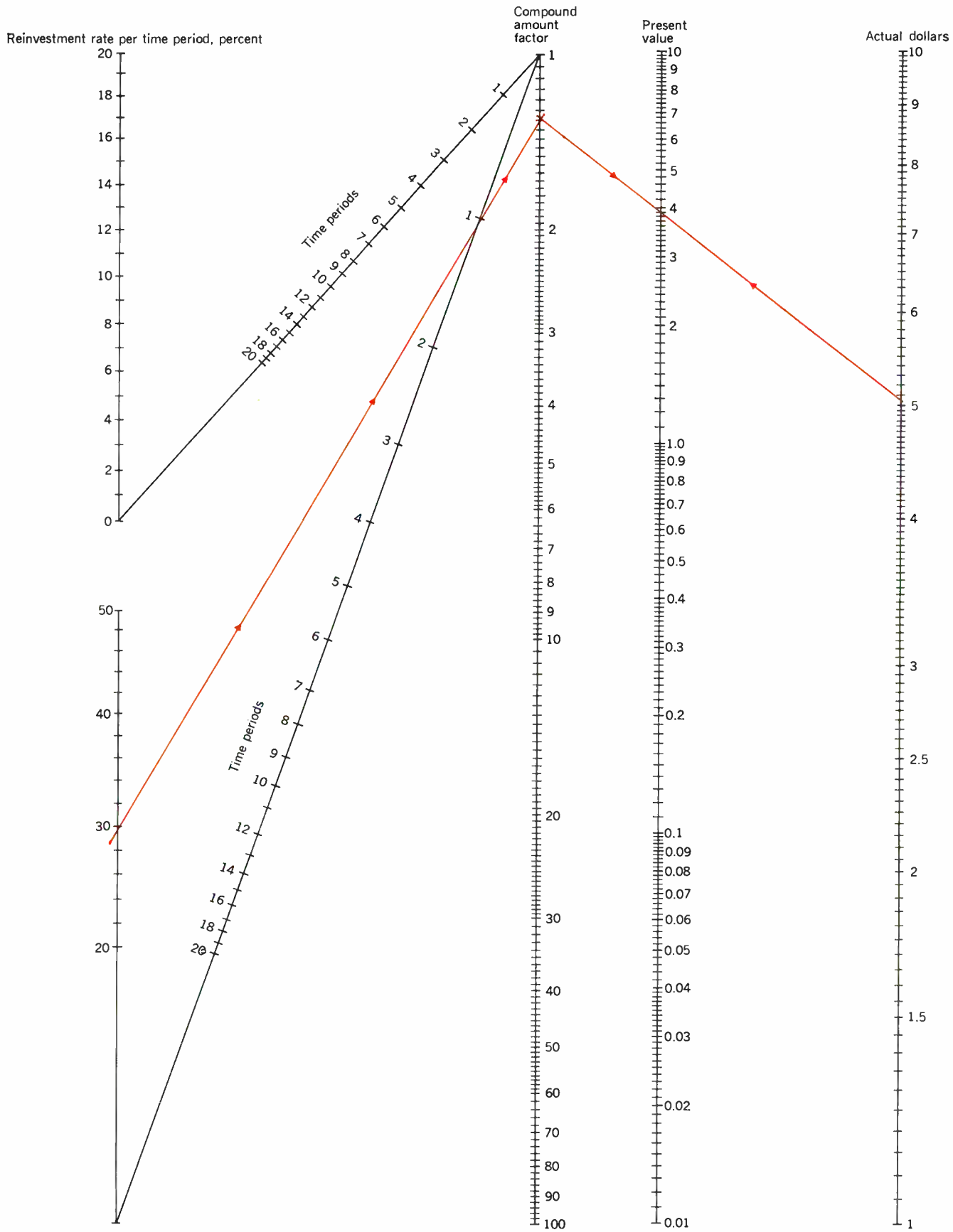


Fig. 6. Present value nomograph for nonuniform cash flows.



A cost duration present value (CDPV) of 6.3, corresponding to $630 000, is obtained as shown in Fig. 5. The NPV is equal to the LPV minus the CDPV. In this case, the NPV is $730 000 (LPV) minus $630 000 (CDPV) or $100 000. This agrees with our previous calculation to within the accuracy of the nomograph.

The NPV return calculation is very simple. Using the mark on the annuity factor scale from the two-year curve and a mark on the actual dollars per time period scale at 1.6, corresponding to the costs per year of −$160 000, we get a cost present value (CPV) of approximately 2.18 or −$218 000. From the NPV return equation, we get

$$\text{NPV return} = \frac{\text{NPV}}{-(\text{CPV})} \times 100\ \%$$

$$= \frac{100\,000}{-(-218\,000)} \times 100\ \%$$

$$\text{NPV return} = 46\%$$

Uniform cash flows nomograph (initial outlay). If the investment has an initial outlay, there is only one simple change in the above procedure—the initial outlay is added to the CDPV before it is subtracted from the LPV to obtain the net present value of the proposed investment. For example, if the previous investment had an initial outlay of $60 000, the CDPV would be $630 000 plus $60 000 or $690 000. The NPV would then be $40 000 ($730 000−$690 000).

Likewise, in the NPV return calculation the initial outlay would be added to the CPV before it is divided into the NPV. Or,

$$\text{CPV} = -(218\,000 + 60\,000) = -278\,000$$

$$\text{NPV return} = \frac{40\,000}{-(-278\,000)} = 14\%$$

It is readily seen that the nomograph in Fig. 5 reduces the calculations required for uniform cash flow investments to a few simple arithmetic operations. Unfortunately, this great a reduction in computations cannot be achieved for nonuniform cash flow investments, but the nomograph in Fig. 6 does keep the computations simple.

Nonuniform cash flows nomograph. For purposes of explanation we will use Example IV:

Time Period	Actual Dollars
0	−100 000
1	+ 50 000
2	+100 000
3	+ 50 000

Reinvestment rate per year = 30%

The initial outlay, obviously, is already on a present value basis so that in this case its value is −$100 000.

The present values of the cash flows after the initial outlay are obtained in the following manner:

1. For the first time period a straight edge is aligned to cross 30 percent on the reinvestment rate per time period scale and 1 on the time periods scale. Next, a mark is placed on the compound amount factor scale where the straight edge intersects it. Moving the straight edge so that it crosses the compound amount factor scale at the pencil mark (1,3 in this case) and $50 000 on the actual dollars scale, about $38 000 is obtained for the present value of the cash flow.

2. Using the nomograph in the same manner as above, a present value of about $59 000 is obtained for the $100 000 cash flow that occurs at the end of period 2.

3. Finally, by applying the nomograph to evaluate cash flows in period 3, a present value of $23 000 is obtained.

The NPV of the proposed investment is the algebraic sum of the present values determined above, or

Time Period	Present Values, dollars
0	−100 000
1	+ 38 000
2	+ 59 000
3	+ 23 000
	NPV = + 20 000

This value compares very favorably with the value previously obtained ($20 391). It should also be observed that the nomograph approach reduced the computational burden to only the addition of four numbers.

The calculation of the NPV return is very simple once the NPV has been determined. Just add all negative present value terms. This value is equal to the cost present value (CPV). Thus, for Example IV,

$$\text{CPV} = -\$100\,000,\ \text{NPV} = \$20\,000$$

$$\text{NPV return} = \frac{\text{NPV}}{-(\text{CPV})} \times 100\%$$

$$= \frac{20\,000}{-(-100\,000)} \times 100\%$$

$$\text{NPV return} = 20\%$$

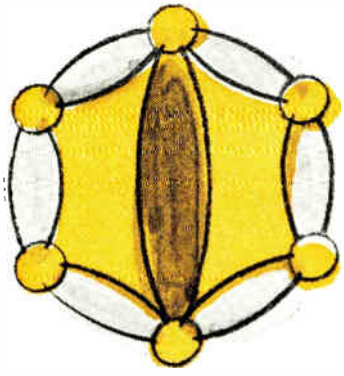
The above discussion demonstrates that the present value nomographs allow much more effective management decision making to be realized by facilitating a rapid and convenient discounted-cash-flow evaluation of many different investment alternatives.

In conclusion, it can be stated that there has been a definite trend for many industrial investments to extend over much longer periods of time and to increase greatly in size. This has been caused in large part by the tremendous rise in the complexity of technical and business factors that govern the investments undertaken. Engineering investment in new product development programs is clearly an example of the trend in this direction. A factor that has become crucial is the very important role that time plays in the analysis of proposed engineering investments. This is why discounted-cash-flow methods, which include the effect of the time value of money in financial studies, have received an increased emphasis. There is little doubt that as engineering investment proposals have become more complicated and relate to longer periods of time there have been an increasing number of engineering managers turning to discounted-cash-flow techniques.

REFERENCES

1. Canada, J. R., "Capital budgeting: Its nature, present practice and needs for the future," *J. Ind. Eng.*; Mar.–Apr. 1964.
2. Bierman, H., Jr., and Smidt, S., *The Capital Budgeting Decision*. New York: Macmillan, 1960.
3. Grant, E., and Ireson, W. G., *Principles of Engineering Economy*. (4th ed.). New York: Ronald Press, 1959.
4. Solomon, Ezra, ed., *The Management of Corporate Capital*. New York: The Free Press of Glencoe (Macmillan), 1960.
5. Baker, W. W., and Simons, H. P., "Simplified procedure for nomograph construction" (Eng. Experiment Sta. Tech. Bull. 34), W. Va. Univ. Bull.; Feb. 1957.

Special Conference Report



Symposium on signal transmission and processing

H. E. Meadows *Columbia University*

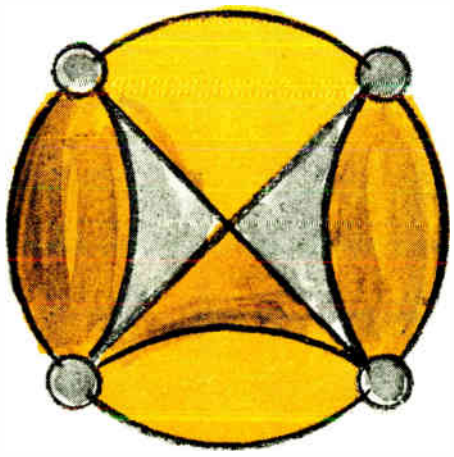
Extensive present demands and much greater projected future needs for high-speed error-free relaying of data have provided an additional stimulus to the communications engineer's interest in efficient transmission and processing of information-bearing signals. A topically well-unified view of current work motivated by this interest was evident in the technical program of the Symposium on Signal Transmission and Processing held at Columbia University in New York City, May 13 and 14, 1965. The symposium was sponsored jointly by the IEEE Circuit Theory Group and the Department of Electrical Engineering of Columbia University. Held in conjunction with the Centennial Celebration of Columbia's School of Engineering and Applied Science, the symposium was part of a year-long series of events commemorating the founding of the Columbia School of Mines in 1864.

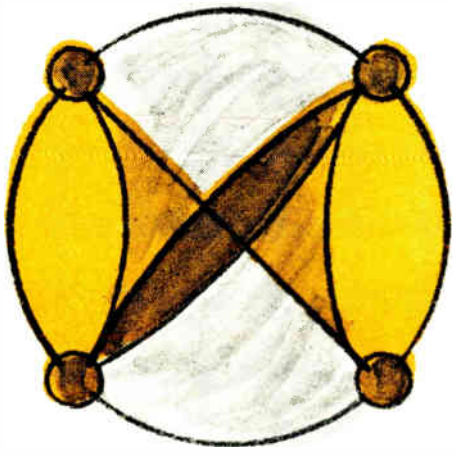
The major theme of the symposium was the need for sophisticated techniques in the design of optimal signals and signal processors to provide immunity against the effects of transmission distortions, noise, and mutual signal interference. Papers presented generally dealt with fundamental aspects of the characterization and design of signal transmission systems, with particular reference to matching of signal and channel characteristics for optimum performance.

The technical sessions of the symposium were held in Wollman Auditorium of Columbia's Ferris Booth Hall. At the opening session, a welcome to Columbia University was extended by Vice President Lawrence H. Chamberlain, who spoke briefly of society's need for meaningful communication on various levels. Dean J. R. Dunning brought greetings from the School of Engineering and Applied Science. Prof. R. J. Schwarz, Chairman of the IEEE Circuit Theory Group and of the Department of Electrical Engineering of Columbia, then greeted the participants on behalf of the cosponsors of the symposium.

The technical program comprised four sessions: "Random Signal Processes," "Signal Design," "Signal Characterization," and "Signal Detection and Processing." All of the papers were published in a conference record that was available to registrants at the symposium. The papers had been reviewed and selected for presentation in accordance with the customary policy of the IEEE Circuit Theory Group. Copies of the conference record of the symposium may be obtained from IEEE Headquarters (Publication 4C9; $3 for members, $6 for nonmembers).

The session on random signal processes began with a discussion by A. Papoulis of "Systems with Nonsta-





tionary Inputs." Since both deterministic and random systems were treated, this discussion established a desirable foundation and preparation for the rest of the session. The three papers which followed, presented by A. W. Drake, T. L. Booth, and C. T. Tsitsera, dealt generally with statistical observations of systems or signal sources in which noise or other natural fluctuations significantly influence the measurements. Prof. Mischa Schwartz of the Brooklyn Polytechnic Institute served as chairman of the session.

A group of papers dealing closely with signal design was presented in the second session. Problems of signal selection were approached from two general points of view: (1) performance optimization in the presence of given channel limitations (noise, variable delay distortion, nonzero memory, etc.); and (2) pulse shaping to approximate prescribed characteristics such as limited duration and specified energy density spectrum. The first approach was employed by H. Schwarzlander and J. C. Hancock and by M. K. Simon and L. Kurz in two papers on the optimum choice of signals to minimize error rates. A paper presented by S. H. Chang and written by four authors from Northeastern University adopted the second approach, and pointed out an interesting use of the theory of entire functions in pulse shaping. The necessity for engineering compromise in signal design was indicated by R. W. Lucky, who showed how, in a typical case, amplitude-frequency shaping of signals (meeting Nyquist's zero-crossing criterion) can provide immunity to a variable delay distortion of known form at the expense of more stringent receiver timing requirements.

The papers dealing with signal characterization and with signal detection and processing were more varied in subject and approach than those presented in the earlier sessions. Topics ranged from a class of learning receivers, described by J. C. Hancock and R. W. Chang, to symbol transition properties of codes generated by sequential machines, discussed by C. V. Ramamoorthy, and theorems in sequential detection, presented by K. B. Gray. Synthesis of oriented communication nets (the accompanying illustrations are abstracts of these nets) was treated by D. K. Sen and I. T. Frisch; their results included necessary and sufficient conditions as well as an algorithm for realization of a restricted class of terminal capacity matrices. B. K. Kinariwala analyzed the timing jitter build-up in a pulse circulating loop consisting of a regenerative repeater and a finite delay.

Nonuniformly sampled signals were the subject of two papers. Calling upon nonharmonic Fourier series methods and the theory of entire functions, K. Yao and J. B. Thomas derived results for sampling representations of band-limited signals that are valid when samples are taken at instants differing significantly from uniform spacing at the Nyquist rate. P. A. Franaszek and B. Liu determined an optimum linear filter for reconstitution of signals sampled at a periodically varying rate. The filter, chosen to minimize the mean-square error due to additive noise at the sampler input, is realizable in a periodically time-variable network of special form.

A method of quantizing spherically symmetric n-dimensional Gaussian signals was analyzed by J. G. Dunn. The quantizing technique is based on quantizing regions derived from the symmetry group of the regular simplex. Results derived relating quantization distortion and information transmission rate were compared with the known performance bound of Shannon. Several quantizers were shown to give operation close to the bound.

Those in attendance at the symposium, numbering more than 230, represented industrial, academic, and governmental organizations, and included visitors from Canada, France, and Germany. In addition to engineers engaged in industrial development and research in communications system concepts and technology, it was interesting to note the presence of a number of geophysicists and oceanographic experts. Registrants from universities included both faculty members and graduate students. Ample opportunity for informal discussions among participants was to be found at lunch and during midsession coffee breaks.

Although a number of persons contributed to the success of the symposium, particular credit should be given to the symposium chairman, Prof. Omar Wing, and to the local arrangements chairman, Prof. C. C. Halkias, both of Columbia University, for a smoothly organized meeting. The difficult task of selecting papers for presentation was the responsibility of the technical program committee, headed by L. E. Franks of Bell Telephone Laboratories. The efforts of the sponsors and the interest of the participants combined to produce a worthwhile symposium.

Authors

Kenneth A. Cox attended the University of Washington, from which he received the bachelor's degree in 1938 and the law degree in 1940. In 1941 he received the master's degree in law from the University of Michigan. He began his legal career in 1941 as a law clerk to Justice W. J. Steinert of the Washington Supreme Court. Later he worked for the State Attorney General and was briefly associated with the Seattle law firm of Weter, Roberts and Shefelman. He served in the U.S. Army from 1943 to 1946, attaining the rank of captain. He then returned to the University of Michigan, as assistant professor of law. In 1948 he joined the Seattle law firm of Little, Palmer, Scott and Slemmons, of which he became a partner in 1953. In 1955 he was named special counsel for the Senate Interstate and Foreign Commerce Committee, in which capacity he helped direct that committee's television inquiry of 1956–1957. In 1961 the Federal Communications Commission named him chief of its Broadcast Bureau, a position he held until he became commissioner in 1963.

Duane O. Muhleman received the B.S. degree in physics in 1953 from the University of Toledo. He worked for two years with the National Advisory Group for Aeronautics on dynamic stability and control problems of experimental aircraft. He then joined Jet Propulsion Laboratory, where he was concerned with guidance and communication systems analysis and radar. He attended Harvard University from 1961 to 1963, and received the Ph.D. degree in astrophysics. He then rejoined Jet Propulsion Laboratory, where he is now working in planetary and radar astronomy.

Richard Goldstein received the B.S. degree in electrical engineering from Purdue University in 1947 and the M.S. and Ph.D. degrees, also in electrical engineering, from the California Institute of Technology in 1959 and 1962 respectively.

Dr. Goldstein joined the staff of the Telecommunications Division of the Jet Propulsion Laboratory in 1958, and at the present time is manager of the Communications Research Section of that division. He has been associated with JPL's Radar Astronomy Program since the first radar detection of the planet Venus, in 1961.

Roland Carpenter (A) received the B.A. degree in psychology from Los Angeles State College in 1951. From 1952 to 1959 he was employed by Collins Radio Company, Burbank Division, as an electronics technician, engineer, and group supervisor. In 1959 he joined JPL as a communication research engineer. At present he is supervisor of the Radio Astronomy Group in the Space Science Division. He received the M.A. degree in astronomy from U.C.L.A. in 1964. Mr. Carpenter is a member of the American Astronomical Society, the Astronomical Society of the Pacific, and Sigma Xi.

John R. Dunning has been dean of Columbia University's School of Engineering and Applied Science since 1950. He received the A.B. degree from Nebraska Wesleyan University in 1929 and the Ph.D. degree from Columbia in 1934. He joined the faculty of Columbia in 1935 as assistant professor of physics. He was named associate professor in 1938 and professor in 1946. He pioneered the first U.S. neutron experiments beginning in 1932 and directed the development of Columbia's first cyclotron in 1936, and demonstrated the first fission of separated U-235 in 1940.

Dr. Dunning is the recipient of seven honorary doctorates and numerous medals and awards, including the Medal for Merit, Presidential Citation (1946) and the Medal for Distinguished Service, City of New York (1948). He has served or is serving on many U.S. Government and New York City advisory committees. He is coauthor of the book, *Matter, Energy and Radiation*. He is a Fellow of the American Association for the Advancement of Science, American Physical Society, and American Nuclear Society.

J. E. May, Jr. (SM) received the A.B. degree from Wesleyan University in 1943, the M.S. degree from Tufts College in 1949, and the Ph.D. degree from Yale University in 1953, all in physics. From 1952 to 1959 he was engaged in research and development work in ultrasonic delay lines at the Bell Telephone Laboratories; he became a supervisor in that field in 1959. In 1962 he assumed responsibility for a group at the company's Murray Hill, N.J., facility concerned with new types of transducers and amplifiers for generation and amplification of sound in the microwave frequency range. He was appointed to his present post as head of the Ultrasonic Device Department, Allentown, Pa., in July 1965.

Dr. May has been granted four patents and is the author of a number of published technical articles. He is a Fellow of the Acoustical Society of America and a member of the American Physical Society, Sigma Xi, and Sigma Pi Sigma.

Richard Hackborn (M) received the B.S. degree in electrical engineering from the University of Minnesota in 1960 and the M.B.A. degree from Santa Clara University in 1963. In 1960 he joined the Hewlett-Packard Company in Palo Alto, Calif., as a design engineer in advanced research and development. While in this capacity he helped in the design of a multichannel recorder, for which he was granted a patent. He transferred to the company's corporate operations staff in 1963. In his present position of operations research analyst he is concerned primarily with the development and implementation of laboratory management systems. Concurrent with his work at Hewlett-Packard, he is taking graduate courses in operations research and statistics at Stanford University.

News of the IEEE

calendar
people
obituaries

Northeast Electronics Research and Engineering Meeting (NEREM) announces sessions

The technical program for the Northeast Electronics Research and Engineering Meeting, which will be held November 3 to 5 in Boston's War Memorial Auditorium, has been announced.

The tentative list of sessions follows:

WEDNESDAY, NOVEMBER 3

10:00 a.m. Sessions

Microwave Antennas

Chairman: John Ruze, MIT Lincoln Lab.
Variable Time Delay Microwave Devices Electronically Scanned and Signal Processing Antennas. *R. A. Sparks, Litton Systems*
A Temperature-Stable High-Power C-Band Digital Phase Shifter. *J. A. Kempic, R. R. Jones, Westinghouse Def. & Space Center*
The Ferrigas Duplexer. *F. Jellison, N. Brown, Microwave Assoc.*
A Near Isotropic Circularly Polarized Antenna. *V. Galindo, Bell Tel. Labs.; K. Green, Dalmo Victor Co.*
Early Test Results with the Circularly Polarized Tracking Feed in the Haystack Antenna. *K. J. Keeping, MIT Lincoln Lab.*

Space Electronics Reliability

Chairman: L. R. Doyon, Raytheon Co.
Mathematical Model for Gemini Countdown. *W. H. Sellers, Raytheon Co.*
Acquisition of Surveyor Reliability Data. *W. R. Kuzmin, Hughes Aircraft*
Reliability Demonstration of a Space Digital Computer. *J. E. Anderson, L. E. Peters, IBM*
Assured Gemini Radar Reliability. *J. H. Scrivner, McDonnel Aircraft*
Voyager Program Feasibility Evaluation—A Second Look. *J. C. Dyer, AVCO*

Communication Switching Systems and Techniques

Chairman: R. C. Stiles, Sylvania
Application of Operational Amplifier Techniques to Conference Circuit Switching. *E. F. Haselton, N. S. Van Buren, Sylvania*
The Relative Effectiveness of Assigned and Shared Storage in a Switching System. *N. Saber, H. Young, Sylvania*
Stored Program Switching Systems in Use in the Bell System. *J. B. Connell, Bell Tel. Labs.*
Novel Features Arising from the Tides–Time Division System Concept of Integrated Control and Switching Logic. *A. A. Kunze, RADC, Griffin Air Force Base*
Measuring the Statistical Accuracy of Communication Network Simulation Results. *S. A. Sobel, MITRE Corp.*

High Power, High Frequency Transistors

Chairman: H. H. Loar, Bell Tel. Labs.
A High-Power VHF Overlay Transistor for Single-Sideband Applications. *R. Rosenzweig, Z. F. Chang, RCA*
Microwave Power Generation and Amplification Using Transistors. *A. J. Anderson, H. F. Cooke, B. T. Vincent, Jr., Tex. Instr.*
A 40-Watt, 175-Mc/s Transistor Family. *S. H. Barnes, G. G. Luettgenau, G. Erkan, TRW Semiconductors*
Design Considerations of Microwave Power Transistors. *G. W. Parker, D. Richardson, Fairchild Semiconductor*
Practical Performance Limits in Silicon UHF Power Transistors. *R. M. Scarlett, ITT*

2:30 p.m. Sessions

Space Instrumentation

Chairman: H. B. Merrill, Jr., Acton Labs.
A Theoretical Investigation of an MHD Gyroscope Employing an Axial Magnetic Field. *G. K. Bentley, A. D. Pierce, AVCO*
Primate Instrumentation During Orbit. *R. H. Schiffman, C. E. Alvine, Univ. of Calif.*
Space Instrumentation for Martian Exploration. *W. H. Foster, L. M. Snyder, Electro-Optical Systems*
Meteoroid Measurements with Pegasus Satellites. *E. Stuhlinger, Marshall Space Flight Center*
Pseudo-Random Scan Applications to Data Processors. *R. A. Reise, Univ. of Conn.*

Biomedical Electronics

Chairman: G. O. Barnett, Mass. General Hospital
An Interpretive Communication System for Hospitals. *J. Baruch, Bolt Beranek & Newman; G. O. Barnett, Mass. General Hospital*
A Special Purpose Computer for Use in Computing Cardiac Output. *H. S. Goldberg, Texas Instr.; H. Sherman, MIT Lincoln Lab.*
The Use of a Small Computer in the Physiology Laboratory. *W. Simon, Harvard Medical School*
A Special Purpose Hybrid Computer for Monitoring Electro-Cardiographic Abnormalities. *A. N. Pappalardo, Mass. General Hospital*

Satellite Communication Systems

Chairman: W. E. Morrow, Jr., MIT Lincoln Lab.
The Initial Defense Communications Satellite Project. *J. J. Cohen, Defense Communications Agency*
The Early Bird Project. *M. J. Votaw, Communications Satellite*
Results of Experiments with the Lincoln Experimental Terminal Using a Variety of Channel Types. *P. Rosen, B. E. Nichols, MIT Lincoln Lab.*
NASA's Planned Program in Communications Satellite. *L. Jaffe, NASA*

Novel Materials for Solid State Electronics

Chairman: Roger Newman, Sperry Rand Res. Center
Physics of the Gunn Effect and Its Relevance to Devices. *D. E. McCumber, Bell Tel. Labs.*
Materials for Solid State Optical Devices. *J. E. Geusic, Bell Tel. Labs.*
Materials for Acoustic and Magnetic Microwave Delay Lines. *R. W. Damon, Sperry Rand*
Applications of Magnetic Materials and Magnetic Fields. *K. J. Button, MIT*

Transistor Circuits—Linear

Chairman: E. G. Nielsen, General Elec. Co.
Optically-Coupled Linear Circuit Techniques. *E. L. Bonin, Tex. Instr.*
Linear Power Amplification Using Switching Techniques. *A. W. Carlson, C. A. Furcinity, R. E. Bach, Northeastern Univ.*
An Integrated Negative Impedance Converter. *M. Nagata, J. G. Linvill, Stanford Univ.*
Feedback Phase Modulator. *J. C. Otto, General Elec. Co.*
Characterization of Transformers Through the Use of h-Parameters. *A. Olsen, Jr., Bell Tel. Labs.*

8:00 p.m. Panel Discussion—Characteristics of the Creative Company

Organizer: K. C. Black, Scientific Analysis Corp.
Panelists: Donald Sinclair, pres., General Radio; Howard McMahon, pres., A. D. Little Co.; E. L. Ginzton, board chairman, acting pres., Varian Assoc.; A. Y. Taylor, pres., Jackson & Moreland

THURSDAY, NOVEMBER 4

10:00 a.m. Sessions

Plasma Astrophysics

Chairman: M. A. Levine, AFCRL
Spectroscopy in Solar Research. *J. M. Beckers, Sacramento Peak Observ.*
Shock Waves in Astrophysics. *R. S. Kushwaha, Hanscom Field*
The Magnetosphere Flow. *H. E. Petscheck, AVCO*

Biological Effects of Laser Radiation

Chairman: A. S. Ketcham, Natl. Cancer Inst.
Bio-Engineering Aspects of Laser Radiation. *P. E. McGuff, Children's Hosp. Res. Foundation*
Experiences with Treatment of Patients with High Energy Ruby and Neodymium Lasers. *L. Goldman, Children's Hospital Res. Foundation*
The Effect of the Laser on Cellular Respiration. *D. E. Rounds, R. S. Olson, Pasadena Foundation for Medical Res.; F. M. Johnson, Electro-Optical Systems*
Threshold Studies and Reversible Depigmentation in Rodent Skin. *E. Klein, Y. Loar, L. C. Simpson, Roswell Park Mem. Inst.; S. Fine, Northeastern Univ.; J. Edlow, M. Litwin, Harvard Med. School*

Broadband Analog Communications Systems

Chairman: D. D. Sagaser, Bell Tel. Labs.
Bell System Development of SSB-AM Coaxial Cable System. *F. J. Herr, Bell Tel. Labs.*
The Line Equipment of a Carrier System for 2700 Voice Circuits on Coaxial Cable Equipped with Transistor Amplifiers. *H. Keil, Siemens & Halske AG*
Features of a 12-Mc/s Coaxial Line Equipment. *F. Tropf, Telefunken AG*
Transistorized Coaxial Line Equipment for 2100 Channels. *J. F. Lansu, N.V. Philips'*