

IEEE spectrum

features

61 Spectral lines: IEEE and the practicing engineer

Under a new editorial policy, IEEE Spectrum is seeking, and selectively printing, "how-to-apply-it" articles, which will necessarily be narrow in scope but highly practical

63 Some practical aspects of digital transmission

John R. Pierce

Digital transmission is ideally suited to sending signals that are inherently discrete; now, through the advance of the solid-state art, it offers many additional advantages

71 Linear integrated circuits in communication systems

Robert A. Hirschfeld

It appears that most mass-produced communication equipment in the future will be designed by engineers who will apply standardized one-chip subsystems

78 The story of Bonneville Power: 1937-1968-1987

Dams of the Columbia River Basin

Gordon D. Friedlander

The development of the full hydroelectric potential of the Columbia River Basin to meet the skyrocketing demand for electric power in the Pacific Northwest is scheduled for 1987

99 Continuing education for the engineering manager

Joseph M. Biedenbach

RCA has developed a continuing education program for its executive personnel that may be a forerunner to a generalized involvement of industry in the education of top personnel

102 Phased arrays for radars

T. C. Cheston

Through the use of electronic scanning techniques, radar equipment can perform multiple functions interlaced in time, such as tracking, searching, and illuminating targets

112 Ceramic IF filters for consumer products

Franz Sauerland, William Blum

Recent improvements in manufacturing techniques show promise of making application of ceramic filters in consumer electronic products practical, economical, and desirable

127 Spark breakdown in air at a positive point

Essam Nasser

The existence of two possible mechanisms of spark formation may help explain some of the peculiar occurrences of various ionization phenomena

62 Authors

Departments: *please turn to the next page*



THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

9 Transients and trends

10 News of the IEEE

24th NEC to feature refresher seminars and special sessions	10
Call for papers issued for Summer Power Meeting	21
IEEE awards—opportunity for nominations	22
Program set for 21st Annual EMB Conference	24
Advanced maintainability to be studied in St. Louis	24
W. R. Saunders fills new post with IEEE Publishing Services	26
Vehicular Technology Group to sponsor December meeting	26
Reliability Physics Symposium due in December	28
1969 SWIEEEO meeting scheduled for San Antonio	28
H. L. Nicol joins IEEE Educational Services	28
Nominations of candidates for IEEE Fellow grade	30
New conference is devoted to nuclear electronics	34
Mexico City conference to be held in Olympics building	34
Radiation effects meeting draws over 300 participants	36
Papers wanted for EMC Symposium	38
Aerospace meeting is planned for P.I.B. center	38

41 Calendar

44 People

49 1969 IEEE Japan Directory

135 Scanning the issues

137 Advance abstracts

Future special issues, 138

155 Translated journals

158 Special publications

160 Focal points

New above-elbow artificial arm responds to amputee's will	160
On uniformity of television color reproduction	161
Applications are invited for White House Fellowships	161
Electronic vehicle control to be tested	163
Satellite communications station is operating in Chile	163
Large negative ion offers key to ionosphere	164
Space satellites guide ships at sea	164

165 Technical correspondence

Computers aren't people, *L. Stephen Coles, J. R. Pierce*
To convert or not to convert, *A. A. Jeffery, Giovanni F. Marotta*

168 Book reviews

New Library Books, 170 Recent Books, 171

Advertising

Positions Open, Positions Wanted, 172 Advertising Index, 186

the cover

Photographic techniques are commonly employed in studies of spark breakdown phenomena, as discussed in an article beginning on page 127. This month's cover shows a typical discharge in air at a voltage of 11.5 kV.

Second-class postage paid at Easton, Pa. Printed in U.S.A.

PROCEDURE FOR IEEE MEMBERSHIP APPLICATION

In order to make the benefits of IEEE membership more readily available to the international engineering community the Board of Directors of the IEEE has recently approved an amendment to the Bylaws reducing the number of references required on applications for admission or transfer to Member and Senior Member grades.

For the Member grade — 3 references are required of either Fellow, Senior Member or Member grade.

For the Senior Member grade — 3 references are required of either Fellow or Senior Member grade.

For the grade of Associate, the number of references required continues to be only one, of either Member, Senior Member or Fellow grade.

Student membership privileges are now extended to any student enrolled in an institutional course of study related to IEEE fields of interest, provided he is carrying at least 30% of a normal full-time program.

THE BENEFITS OF MEMBERSHIP

- Monthly receipt of IEEE SPECTRUM as a part of dues. (Students have the option of receiving either SPECTRUM or the STUDENT JOURNAL.)
- Eligibility to join one or more of the 31 IEEE Groups and receive their TRANSACTIONS and JOURNALS.
- Conference Records, Standards, Translated Journals and Membership Directory at special rates.
- Attendance at conferences and technical meetings in many countries.
- Participation in local IEEE Section meetings and activities.
- For Students, participation in Student Branch activities and student paper prize contests.

For the convenience of engineers outside the United States and Canada, arrangements have been made for payment of the annual IEEE dues in local currency in the following countries: France, Germany, Italy, Japan and the United Kingdom. For membership application and payment details write to your local Section Secretaries:

EUROPE AND NORTH AFRICA

BENELUX

Frederik Valster
c/o N. V. Philips
Gloeilampenfabrieken
BibliotheekCentrale
Eindhoven, Netherlands

EGYPT

Abdel-Latif I. Ahmed
Electric & Electronics Industries Co.
Research & Planning
26 Adly Street 7th Floor
Cairo, Egypt

FRANCE

Alain J. Profit
C. N. E. T.
3, Avenue De La Republique
92 Issy Les Moulineaux, France

GERMANY (WEST)

Hans H. Burghoff
Stresemann Allee 21 VDE-HAUS
6 Frankfurt M 70 FR, Germany

ISRAEL

Mordechai Weissenstern
Elta Electronics
Lod, Israel

MIDDLE & SOUTH ITALY

Valario R. Cimaglia
Via Andrea Busiri
Vici 42
Rome, Italy

NORTH ITALY

Gianfranco Cariolaro
Istituto Di Electrotecnica
Via Marzolo 9
Padua, Venetia, Italy

NORWAY

Per A. Bergstad
Norwegian Def. Res. Est.
Postboks 25
Kjeller, Norway

SPAIN

Rogelio T. Segovia
Arturo Soria 98
Madrid, Spain

SWEDEN

Bo Stjernberg
Telefonaktiebolaget
L. M. Ericsson Box S 100
Postfack
Molndal, Sweden

SWITZERLAND

Eric Bernhardt
c/o Brown Boveri Co.
Sales Directorate
Baden, Switzerland

UNITED KINGDOM

& REPUBLIC OF IRELAND
Robert C. Winton
Mullard House
Torrington Place
London, W. C. 1, England

LATIN AMERICA

BUENOS AIRES

Henry Setaro
Julian Alvarez 2838
Buenos Aires, Argentina

CHILE

Victor G. Gutierrez
Casilla 16 D
Santiago, Chile

COLOMBIA

Leonardo Gomez
Apartado Nacional 5149
Bogota, Colombia, S. A.

MEXICO

Carlos A. Maigler
Av. Cuauhtemoc 1203
Department 1
Mexico City 13, Mexico

PERU

Guillermo Castillo
Laboratorio De Electricidad
Apartado 1301-UNI
Lima, Peru

PUERTO RICO & VIRGIN ISLANDS

Carlos R. Rivera
1799 San Alejandro
URB San Ignacio
San Juan, Puerto Rico

RIO DE JANEIRO

Joao A. Wiltgen
Rua Barao De Jugaribe 145
Ipanema
Rio de Janeiro 20-37, Brazil

SAO PAULO

J. R. Costa De Lacerda
I. E. F. Controles Automaticos SA
Caixa Postal 1044
Sao Paulo, Brazil

ASIA

KANPUR

Prof. H. K. Kesavan (Chairman)
Indian Institute of Technology
Dept. of Electrical Engineering
Kanpur, Uttar Pradesh, India

NEW ZEALAND

Robert Adams (Chairman)
20 Coronation Road
Mangere Bridge
Auckland, New Zealand

TOKYO

Yukimatsu Takeda
840 Seijo-Machi
Setagayaku
Tokyo, Japan

or to: Miss Emily Sirjane
IEEE, 345 E. 47 Street
New York, N.Y. 10017, U.S.A.



THE INSTITUTE OF
ELECTRICAL AND
ELECTRONICS
ENGINEERS, INC.

Nominations of candidates for IEEE Fellow grade

April 30, 1969, has been established as the deadline date for completed nominations of candidates for Fellow grade of the IEEE.

The IEEE Bylaws define the qualifications for this grade as follows: "The grade of Fellow is one of unusual professional distinction and shall be conferred only by invitation of the Board of Directors upon a person of outstanding and extraordinary qualifications and experience in the fields of electrical engineering, electronics, radio, allied branches of engineering or the related arts and sciences, who meets the requirements for Senior Member as stated in these Bylaws and who has been a member in any grade for a period of seven years preceding the year of nomination, except that the seven-year provision in any individual case may be waived for cause by the Board of Directors."

A Fellow grade nomination form and a minimum of five confidential Fellow grade reference forms must be submitted by the deadline date of April 30, 1969. Appraisal of candidates to determine their order of qualification is based solely upon the information submitted. This makes *careful and complete* preparation of all forms essential if each candidate is to receive the highest appraisal to which he is entitled. The duty of the Fellow Committee is to recommend to the Board of Directors, from those presented to it, the most deserving candidates (up to a maximum number allowed under the IEEE Bylaws). To assure objectivity, the accomplishments of each candidate are appraised in a number of reasonably well-defined categories (more fully described in the next paragraph).

It is of prime importance to read the leaflet "IEEE Fellow Grade Nominations" before serious consideration of possible nominees. This brief publication has just been revised in an effort to assist in the effective preparation of nominations. Following are the criteria upon which nominations are appraised—all are important; those of primary emphasis are listed first: technical contributions, theoretical or practical; personal distinction earned, as evidenced by positions held, operational as well as technical; output of technical papers and patents; evaluations by confidential references; duration of particularly distinguished professional activity in positions of major responsibility; duration of active participation in the profession; service to IEEE and its predecessor societies; additional support of the nomination by an IEEE organizational unit, such as Section, Group, Committee, etc.; and service to the engineering community, broadly interpreted, other than through IEEE.

To insure receipt of the required number of references by the April 30 deadline, some sponsors have found it desirable to solicit as many as seven or eight references;

others have followed the excellent plan of monitoring the responses by requesting those named as references to indicate, on a postal card supplied by the sponsor, the date when the form was completed and mailed to the Fellow Committee. All qualified reference forms submitted on behalf of a Fellow grade candidate are considered by the Fellow Committee.

When the nomination has the formal support of a Section, Group, committee, or organization of equivalent standing (or more than one of these), the candidate's competitive position will be enhanced *only* if such additional support is documented by a *separate written statement* from each entity to the Fellow Committee, signed by an appropriate officer. The statement (which is in addition to the five reference forms) should be concise and yet provide all relevant facts in support of the nominee; it must be received by the Fellow Committee by the April 30 deadline.

Should a sponsor of a candidate outside the United States and Canada (i.e., Region 8, 9, or 10) certify that he is unable to furnish the names of five Fellow grade references on the nomination, the Fellow Committee may, in its discretion, accept references from other IEEE members, preferably Senior Members, who are engineers or scientists of recognized standing.

The number of nominations received each year far exceeds the number of new Fellows permitted by the IEEE Bylaws. Inevitably, favorable action on a large number of nominations is impossible. In these cases the proposer (not the nominee) is notified. Such candidates are not reconsidered in a subsequent year *unless* a new, updated nomination form is submitted by the regular deadline date. Likewise, new written statements of support by a Section, Group, committee, etc., must be filed by the regular deadline date. Reference forms submitted on behalf of a candidate are acceptable to the Fellow Committee for a period of three consecutive years, including the year of original submission. However, new reference forms may be submitted. All qualified reference forms submitted on behalf of a candidate are considered. Members of the Board of Directors, the IEEE staff, or the Fellow Committee are not eligible to serve as references on a Fellow grade nomination.

All information furnished on behalf of a candidate must be typewritten so that legible copies can be duplicated and furnished to all members of the Fellow Committee.

Fellow grade nomination kits are available from Miss Emily Sirjane, IEEE Staff Secretary of the Fellow Committee. These kits contain an FC-I Form, seven reference forms, seven reference envelopes, and a copy of "IEEE Fellow Grade Nominations."



Spectral lines

IEEE and the practicing engineer. In the office next to mine there sits a member of the Institute of Electrical and Electronics Engineers. As an orphan, he was forced by economics to drop out of high school, but he resumed it at night; now in his early forties, he expects a bachelor's degree within a few years. I am fortunate in having him work with me, because no matter what electronic problem you take to him, he will design for you a little assembly, and the little assembly will work. No doubt his name is Legion.

Not long ago Mr. Legion, noticing me examining the manuscripts that arrived in the mail, said: "I don't know why I keep paying dues to IEEE; nothing they do is of any use to me." I'm not sure that he expected an answer, but his comment certainly deserves one.

There are at least four reasons why the publications of the IEEE do not include more articles that are of use to Legion:

1. A manuscript submitted for publication is reviewed by the three most sapient experts in the field it deals with, and two of them say, "This is already well known," which may mean that a month ago they learned through the grapevine of this advance in technology, or else that it follows, by implication, from a general theory published in Iceland in 1906. The manuscript is rejected.

2. An engineer in the industry makes a fresh advance in the state of the art. He writes it up for publication and submits it to the company management for release. He gets called into the front office and asked if he realizes what would happen if this information were disseminated. He departs chastened, knowing that if he does succeed in making an innovation, the last thing in the world that his management wants to allow is for him to tell their competitors how to go and do likewise. His counterparts in the competing firms are, almost simultaneously, having the same experience.

3. A bold management, realizing that it can hardly be very many months ahead of the competitors, gives permission for one of its engineers to write up a new development. Full of pride in his association with this excellent company, and carefully kept ignorant by competing managements of the advances made by their engineers, he writes an article that does not conceal his enthusiasm for his firm and its products; the referees unanimously condemn his manuscript as merely a commercial that publicizes his own organization.

4. The practicing engineer likes to read about new developments. The prospect of writing reduces him to a

state near collapse. Therefore, situations 1, 2, and 3 hardly ever arise.

A great part of the problem has nothing to do with IEEE policy or practice. The two major factors limiting IEEE publication that interest men like Legion are company secrecy and the unwillingness of practicing engineers to write. Legion keeps up his membership because the IEEE plays a strong role in promoting growth of the field from which he draws his income. But he wonders whether there couldn't be some benefits of a more direct and personal sort.

There can be, and there will be. A new editorial policy of IEEE SPECTRUM will seek out, and selectively print, articles of a "how-to-apply-it" sort. Articles for this category will not have to describe ideas completely new to the profession, and they will not have to avoid all mention of specific products. They will necessarily have narrow scope, but they will be highly practical.

Also, there is an important move in the area of continuing education. The new Educational Activities Board—which generated the list of short courses that appeared in the July issue—is producing sets of preserved (what I mean is canned, but in a nice way) lectures for use by Sections and by Student Branches. The first offerings will be packaged versions of the two special courses given at the 1968 Convention: "Computer-Aided Circuit Design" and "Integrated Circuits and Their Incorporation Into Equipment." The presentations employ slides, audio tapes, and course notes. Guides to precourse reading enable all comers to acquire the necessary background before the first lecture. Other courses will become available; the first of these will go out on video tape. If experience shows that the program fills a need, it can become an extensive one. The methods and logistics will have to be worked out by trial. Regardless of what storage medium proves best, the means seem to be at hand for enabling the Sections to increase the usefulness of IEEE to the men who see that the industry's work gets done.

The new editorial policy of SPECTRUM is manifested in this issue by the article of Hirschfeld on application of integrated circuits and by that of Sauerland and Blum on ceramic IF filters. These two articles are representative of what we have in mind, but we expect a good deal of diversity. Give us three months for trying out this new policy, and then let us have a letter saying whether you are pleased or displeased with the first examples. Even better: overcome your natural bashfulness and the fears of your company's patent attorney, and send us a manuscript of the how-to-apply-it sort. —J. J. G. McCue

Authors

Some practical aspects of digital transmission (page 63)

John R. Pierce (F). A biographical sketch of Dr. Pierce appears in the July issue, page 42.



Linear integrated circuits in communication systems (page 71)

Robert A. Hirschfeld is head of the Communications Circuits Section at the National Semiconductor Corporation, Santa Clara, Calif. He received the B.S.E.E. degree from the Massachusetts Institute of Technology in 1963. Prior to joining National Semiconductor he was associated with Motorola Inc. and the Amelco Semiconductor Division of Teledyne Inc. Mr. Hirschfeld has been responsible for the design of National's LM170, LM171, and LM172 communication circuits. He is also the author of a number of articles in various electronic periodicals.

Continuing education for the engineering manager (page 99)

Joseph M. Biedenbach received the B.S. degree in engineering in 1950 and the M.S. degree in education in 1951 from the University of Illinois, the M.S. in physics in 1957 from the University of Michigan, and the Ph.D. in higher education in 1964 from Michigan State University. Prior to joining RCA, where he is manager of Engineering Educational Programs, he was assistant dean for administration at the Indianapolis campus of Purdue University. He was formerly a senior technical instructor at the General Motors Institute and a senior project engineer at General Motors' AC Spark Plug Division.



Phase arrays for radars (page 102)

T. C. Cheston (SM) was born in Vienna, Austria, and received the B.Sc. degree from the University of Edinburgh (Scotland) in 1947. He then joined the Marconi Wireless Telegraph Company Ltd. in England and subsequently was employed by the Canadian Westinghouse Company. During the period from 1954 to 1956 he was on loan to the Royal Canadian Air Force. His present responsibilities at the Johns Hopkins University's Applied Physics Laboratory include various efforts directed toward phase-array developments and investigations being undertaken for advanced radar systems.

Ceramic IF filters for consumer products (page 112)

Franz Sauerland (M) received the Dipl.Ing. from the Institute of Technology, Aachen, Germany, in 1954 and the M.S.E.E. degree from Case Institute of Technology in 1966. As manager of filter engineering at the Clevite Corporation, he is responsible for quartz and ceramic filters. **William Blum** received the B.E.S. in physics from Fenn College in 1964 and the M.S.E.E. degree from Case in 1966. In 1967 he joined Clevite, where he is working on ceramic and quartz filter design and development.



Spark breakdown in air at a positive point (page 127)

Essam Nasser (SM) holds the B.S. degree from Cairo University, Egypt, and M.S. and Dr.Ing. degrees from the Technische Universität of West Berlin. Since 1964 he has been an associate professor of electrical engineering at Iowa State University. Previous experience has included positions with the Siemens Company in West Germany and in West Berlin; the Research Laboratories of Electricité de France in Paris; the University of California, Berkeley; and the University of Detroit. At present he is conducting NSF-sponsored research on gas plasmas, corona, and spark breakdown.

Some practical aspects of digital transmission

Because of its many advantages, the digital system is now coming into widespread use, not only for the transmission of inherently digital signals, but for the general transmission of all the signals encountered in telecommunication

John R. Pierce Bell Telephone Laboratories, Inc.

There are three worlds of digital transmission—digital transmission as an area of academic research, in which theoretical limitations and complicated schemes of modulation and encoding play the starring role; digital transmission in exotic space and military applications, in which it is, or may seem to be, worthwhile to search through the noise in a huge bandwidth in order to receive a few bits per second; and the digital transmission of everyday bulk communication. This article is concerned with that third area, in which apparatus must be economical and reliable; in which the FCC limits the bandwidth we may use in radio transmission; in which the increase in attenuation with frequency limits the bandwidth it is practical to use in cable transmission; in which error rates should be low enough that error correction is an expedient that need be resorted to only in special situations.

In the past, most transmission by wire and radio was analog transmission. In analog transmission, the signal sent over a cable or from antenna to antenna varies smoothly with variations of the original, or baseband, signal to be transmitted. Today, an increasing amount of transmission is digital transmission, in which discrete rather than continuous signals are transmitted.

Digital transmission is ideally suited to sending signals that are inherently discrete, such as teletypewriter signals or any encoding of numbers of letters or words.* However, by means of pulse code modulation, analog signals such as voice signals and television signals can be transmitted over digital transmission systems with any desired degree of accuracy. Further, digital transmission has many advantages. Through the advance of the solid-state art it is increasingly easy to implement digital circuitry cheaply and so that it functions with negligible error. In digital transmission involving many repeaters, it is practical to regenerate the signal at repeater points, so that the effect of noise is not cumulative, though the infrequent errors the noise produces do accumulate. Digital trans-

mission is coming into wide use, not only for transmitting inherently digital signals, but for the general transmission of all the signals encountered in telecommunication.

Today, a digital transmission system called TI is in wide use in the Bell System.¹ This system transmits three-level pulses at a rate of 1.5 megabits per second over cable pairs. The possibilities of higher-rate digital transmission via coaxial cable have been explored experimentally.² Waveguide³ seems particularly well adapted to digital transmission. At frequencies above 10 GHz,⁴ digital transmission may allow more microwave communication in a given geographical area than forms of modulation that are more susceptible to interference. Further, the widespread transmission of data signals over analog systems shows that the need for digital transmission is growing rapidly.

In analog transmission we send a smoothly varying signal, one that can have any of a range of values, and receive a signal that also varies smoothly, but differs from the transmitted signal through the influence of noise, interference, or distortion.

In digital transmission, during a certain period of time we select from among a fixed, finite number of possible signals one or more signals and transmit these. At the receiver we try to recognize exactly which signals were sent. Noise or interference and distortion may make the received signal differ from the transmitted signal enough so that we interpret it as resulting from a transmitted signal other than the one that was sent.

In analog transmission, noise, interference, or distortion results in our receiving a signal a little different from the one that was sent. In digital transmission, noise or interference and distortion can occasionally result in our receiving the wrong signal.

It is in principle possible to approach the limiting rate of transmission or channel capacity defined by Shannon in his book, *A Mathematical Theory of Communication*. To approach this transmission rate one would have to use very long, complicated, random-appearing signals. At the transmitting end we would select among such signals. At the receiving end we would have to compare the received signals with a code book containing very, very many long complicated signals. In practice, this plan is entirely impractical, however efficient it may be in theory.

Digital transmission is usually carried out by sending a

* Sometimes the string of digits that is transmitted over a digital transmission system is an encoding of the actual discrete message by means of an error-correcting code. Within limits, this makes it possible to correct at the receiver errors in transmission of individual digits or groups of digits. In order to attain such error correction, additional check digits must be transmitted along with the message digits.

sequence of short, equally spaced signals, which may be pulses of various discrete amplitudes, or waves of constant amplitude that have one of several allowed phases. Each of the successive signals represents a particular number, or a particular digit of a number, that it is desired to transmit. At the receiving end we look at the amplitude or phase of the received signal at appropriate sampling times, and from the amplitude or phase at such sampling times we decide what number or digit was transmitted. In this article we shall consider particularly transmission by means of pulses having one of a given number of discrete amplitudes.

Inherently, such a process is inefficient compared with Shannon's optimal encoding. Figure 1,⁵ for instance, shows calculated values of channel capacity and bit rate for multilevel pulse transmission for a coaxial cable assuming a given size cable, a given length of cable, a given transmitter power, and a given constant noise power density at the receiving point.

For the example of Fig. 1, Shannon's information theory tells us that we should use signals with a bandwidth of about 150 MHz. The transmitted power density of such signals should be almost uniform. Thus, the received power density will be highest at low frequencies, where the attenuation of the cable is low. With such signals we should be able to transmit about 1600 million error-free bits per second. The large number of bits per hertz means, of course, that quite accurate distinctions of amplitude must be made at the receiver.

If we transmit and detect independently a sequence of band-limited pulses of such shape as to give no intersymbol interference, the attainable information rate is considerably lower, even if we assume that we are able to carry out all processes of modulation and detection and sampling of the received signal perfectly. In order to get a usable signal at the receiver we must transmit most of the power at high frequencies, for which the attenuation of the cable is large, and this is partly responsible for the lower information rate.

Transmission over cable by pulses is particularly inefficient if we use only two sorts of pulses, say, pulses with amplitudes $+1$ and -1 . We are able to send many such pulses per second and still detect them in the presence of noise, interference, and distortion. But the signaling rate is only one bit per pulse. Most of the power must be transmitted at the upper end of a very broad band, where the transmission loss is great. If we adopt an error rate of one in 10^{12} at each repeater (one error in 10^9 for 10 000 repeaters), then, ideally, we can send around 260 megabits per second by means of binary pulses, using a bandwidth of about 130 MHz.

If we could satisfactorily implement a multilevel system, ideally we would be able to send some 700 megabits per second with the same error rate. Figure 1 shows that the required optimum number of levels would be 52. This rate is almost half the maximum possible rate for error-free transmission whereas the rate for the two-level system is only about one sixth the maximum.

There is an inevitable loss of efficiency in encoding digital signals by means of pulses that are sought and detected individually at the receiver. Nevertheless, practical digital transmission seems to be committed to either this mode of operation or to a related form of phase modulation. We can somewhat improve the efficiency by using multilevel pulses rather than binary pulses, and in

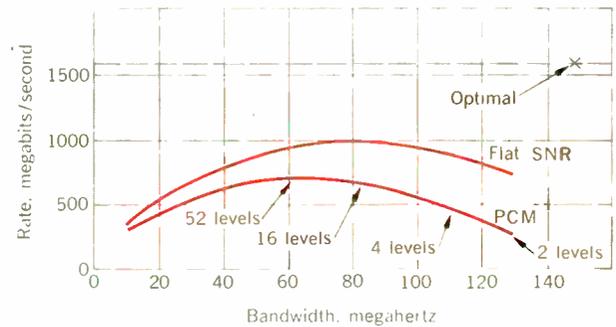


FIGURE 1. Transmission rates in bits/second for a cable whose attenuation in decibels is proportional to $f^{1/2}$, for a prescribed total power and a prescribed flat noise spectrum at the receiving end. "Optimal" is channel capacity for a signal spectrum shaped to give the maximum rate; nothing is transmitted above the frequency indicated. "Flat SNR" is the computed ideal rate for a signal spectrum such that the signal-to-noise ratio is flat at the receiver. The lower curve is for transmission by pulses without intersymbol interference and an error rate of 10^{-12} . (Source: Ref. 5)

some cases this is sure to be done.

We all know that the shorter the pulse the greater the bandwidth. If we are to distinguish the pulses individually at the receiving point, it would appear that we should somehow make the pulses short enough that the tails of one pulse do not overlap the positions at which other pulses are to be received. One solution is to use very short pulses. But this requires an excessive bandwidth.

Nyquist made this clear in a very important paper published in 1928.⁶ He showed that there is a particular sort of pulse which can have unit value at time t_0 and which is zero at times t_0 plus or minus $n\tau$, where n is an integer. Such a pulse has a flat amplitude spectrum from frequency 0 to frequency B , and a linear phase. In terms of the bandwidth B , the time τ between successive pulses should be $\tau = 1/2B$. Such a pulse is shown in Fig. 2.

If we send such pulses at intervals τ apart, and if we sample the received signal at appropriate times separated by τ , each sampling will measure only one transmitted pulse, for the tails of all other transmitted pulses are zero at the time corresponding to the center of any chosen pulse. Using such band-limited pulses we can, in principle, completely eliminate intersymbol interference. The lower curve in Fig. 1 was based upon the use of such pulses, which are commonly called sinc pulses.

Are sinc pulses the only pulses whose tails are zero at all other equally spaced sampling times? The answer is no. There is a whole class of pulses that meet this criterion. And these pulses were also described by Nyquist (see Gibby and Smith⁷ for a generalization).

Figure 3 shows a number of sampling times τ apart. It shows two sinusoidal waves. These waves have different frequencies, but each wave has exactly the same value at all sampling points. By merely sampling a wave at intervals τ apart we cannot tell exactly what frequency components are present in the wave.

This ambiguity is illustrated in Fig. 4, which applies to sampling times $1/2B$ apart.

The upper part of Fig. 4 applies if we work with positive frequencies only. If a frequency component f has a

particular sequence of samples, exactly the same sequence of samples can result from a frequency component of frequency $2B - f$, or $2B + f$, or $4B - f$, or $4B + f$, etc. The frequency components that can be the same at all sampling times can be expressed as $2Bn \pm f$.

Sometimes, and especially in connection with complex Fourier transforms of real-time functions, it is desirable to include negative frequencies. In this case a negative frequency component $-f$ always goes along with a positive frequency component $+f$. The set of frequency components that can give the same samples as the pair $+f, -f$ are shown in the lower part of Fig. 4. They are $f \pm 2Bn$.

The fact that components of several different frequencies can have the same value at all sampling points makes it possible to construct various pulses that have the desirable property of the $\sin x/x$ pulse. That is, one can construct other pulses that are zero at sampling points other than the center of the pulse. Figure 5 shows the spectra of two such pulses. It is assumed that the phase of all frequency components varies linearly with frequency. These spectra meet a criterion given by Nyquist: The sum of the amplitudes of components of frequencies f and $2B - f$ are constant (we assume that no frequencies higher than $2B$ are present).

At times we may wish to shift a pulse upwards in frequency by a carrier frequency f_c and later recover the baseband pulse by multiplying the shifted band of frequencies by the carrier frequency. In such transmission it may be convenient to transmit a little of the lower as well as most of the upper sideband. In demodulating by means of a carrier of frequency f_c , a frequency $f_c - f$ and the frequency $f_c + f$ produce the same output. Figure 6 illustrates another criterion given by Nyquist. A recovered signal will be correct if the sum of transmission at $f_c - f$ and $f_c + f$ is a constant.

If we apply these two criteria to the transmission of pulses shifted in frequency, we find that the condition shown in Fig. 7 will assure perfect transmission. In the vicinity of the carrier frequency, the sum of the transmission at $f_c + f$ and $f_c - f$ is constant. Further, in the vicinity of the band edge $f_c + B$, the sum of the transmission at $f_c + f$ and $f_c + 2B - f$ is constant. It is assumed that the phase is linear. Such a shifted band of frequencies can be demodulated into a baseband pulse that is zero at any instant displaced in time plus or minus $n\tau$ from the center, where τ is equal to $1/2B$.

The $\sin x/x$ and other pulses meeting the Nyquist criterion are important in thinking about digital transmission by means of pulses, and they certainly have importance in practical transmission as ideals toward which we can aim. We must, of course, keep in mind that it is in principle physically impossible to generate exact $\sin x/x$ pulses or other pulses meeting the Nyquist criterion, for such pulses have infinite tails in both directions and hence would require an infinite delay for their exact synthesis.

Even if we could generate and use independent $\sin x/x$ pulses, their employment would lead to certain practical problems. Let us consider, for instance, multichannel digital transmission via radio. The upper curves in Fig. 8 represent the filter characteristics of the combined input and output filters that are necessary to prevent crosstalk in a multichannel radio system. In some ideal sense, these filters could have "square" frequency characteristics, so as to separate the spectrum into contiguous channels of

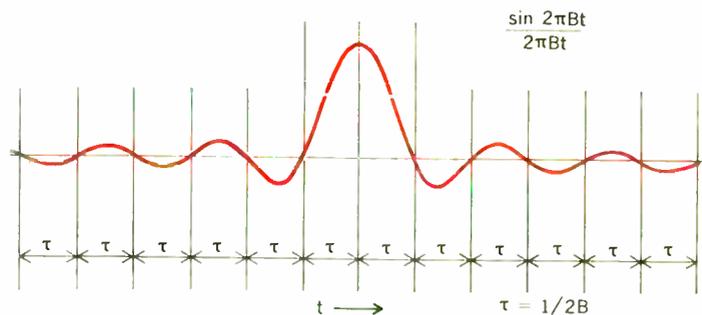


FIGURE 2. A $\sin x/x$ pulse.

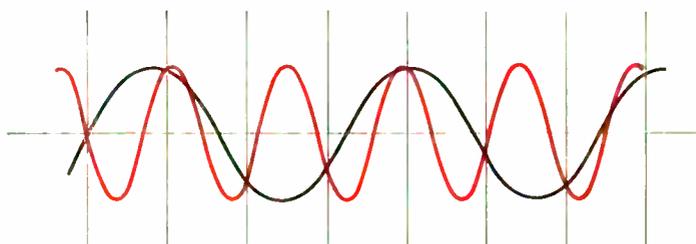


FIGURE 3. Two sine waves of different frequency that have the same values at all sampling times.

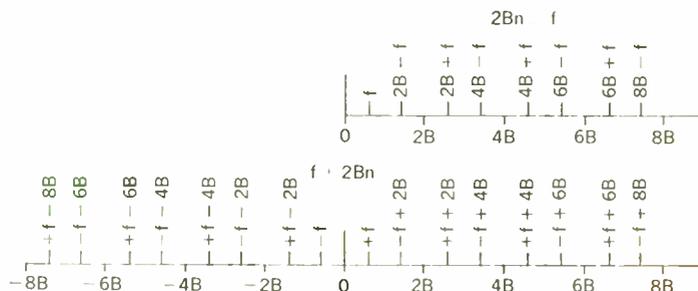
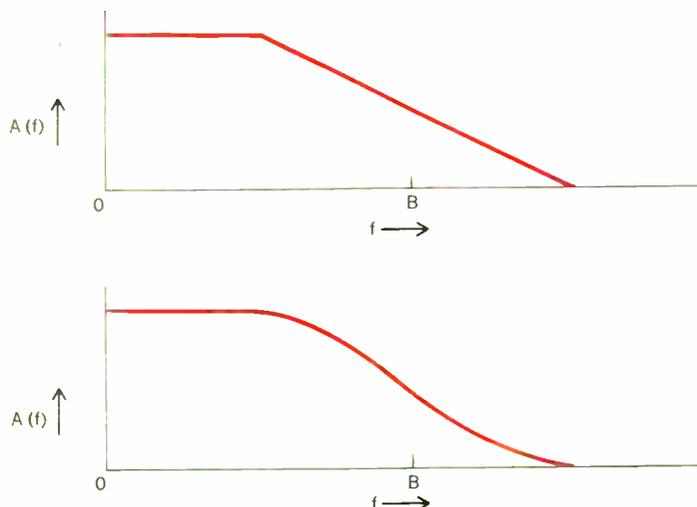


FIGURE 4. Frequencies that can give the same succession of values at sampling points $1/2B$ apart.

FIGURE 5. Amplitude spectra, which, with linear phase, correspond to pulses meeting the Nyquist criterion for sampling times spaced by $\tau = 1/B$. For these spectra, $A(f) + A(2B - f) = \text{constant}$.



width B_c . However, it is impossible to build such filters and uneconomical to approach them. Thus, whereas transmission can be good near the center of the various radio channels, the amplitude and the phase of transmission must become increasingly poor toward the edges.

If we try to transmit $\sin x/x$ pulses with a flat spectrum of width B over such channels, we can achieve excellent transmission if we make the bandwidth B small compared with the channel spacing B_c . But this means an inefficient use of the spectrum. If we make B larger, the frequency components toward the edge of the band will appear with the wrong amplitude and the wrong phase. Thus, the received signal will not meet the Nyquist criterion. Tails of one pulse will appear at the sampling times of other pulses, and we have intersymbol interference.

Figure 9 shows another difficulty with using a flat spectrum of bandwidth B in transmitting pulses by radio. In order to reconstruct the pulses at the receiver, we need the carrier of frequency f_c . We would like to transmit this carrier. Unfortunately, the flat spectrum extends from f_c to $f_c + B$. If we wish to transmit carrier in any practical way we will have to remove the lower part of the spectrum by filtering, and this will alter the pulse shapes of the demodulated pulses from their ideal $\sin x/x$ shape, and so result in intersymbol interference.

Further, we might like to transmit a pilot frequency of frequency $f_c + B$. If we had such a frequency, by taking the differences between this and the carrier frequency we would recover B ; by doubling this we have the timing frequency we need in sampling the received pulse train.

We see that the use of a flat spectrum corresponding to $\sin x/x$ pulses has disadvantages in radio transmission. It also had disadvantages in coaxial transmission. In principle, wire pairs and coaxial cables provide excellent transmission of direct current. In practice, frequencies

near zero cannot be used for several reasons. Among these are earth currents, noise, the transmission of direct current or low-frequency power over the wire pairs or cables, and the practical necessity of using transformers in amplifiers. Thus, we must send over coaxial cables pulses that do not contain any very low frequencies.

In principle, there are several forms of "dc restoration" that might be employed to get the right signal despite the nontransmission of low frequencies.^{8,9} For example, in quantized feedback each received or decoded digit is used to produce a transient equal and opposite to the tail or transient following the received pulse corresponding to the digit. This transient is added to the input of the decoder so as to cancel out the tail of the pulse at all subsequent sampling times.

Other remedies have actually been used in coping with the lack of dc transmission. In the T1 system an excellent remedy called bipolar transmission¹⁰ was used for digital transmission over twisted pairs.

In bipolar transmission, the digit 0 is transmitted as a pulse of zero amplitude and the digit 1 is alternately transmitted as a pulse of amplitude $+v$ or a pulse of amplitude $-v$. Three levels are used redundantly to transmit a binary signal. We have no choice between the amplitudes $+v$ and $-v$; a pulse of amplitude $+v$ conveys the same information as a pulse of amplitude $-v$. In essence, we use a nonrandom or redundant pulse train to transmit the message digits. Because we use a redundant pulse train, the efficiency of transmission is lowered, but in return the pulse train is given a spectrum which goes to zero at zero frequency.

Thus, the developments in digital transmission have been of the following types.

Because it was impractical to implement the highly efficient sorts of codes that Shannon showed could, in

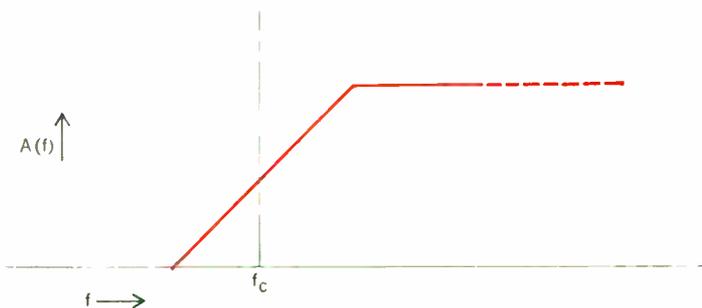


FIGURE 6. Illustration of vestigial sideband criterion $A(f_c - F) + A(f_c + f) = \text{constant}$.

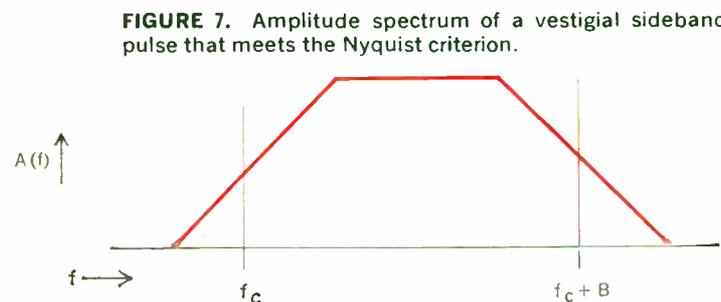
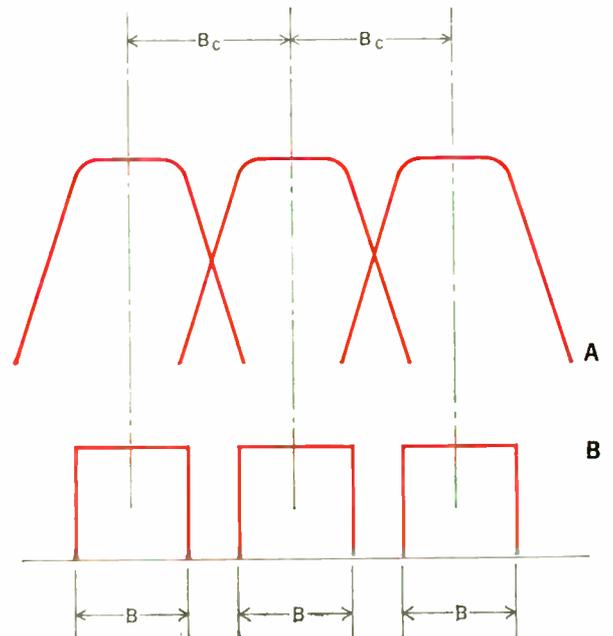


FIGURE 7. Amplitude spectrum of a vestigial sideband pulse that meets the Nyquist criterion.

FIGURE 8. Realizable filter characteristics make necessary a greater than "ideal" spacing of channels. A—Attenuation of transmission channels. B—Spectra of SSB $\sin x/x$ pulse trains.



principle, realize the full channel capacity of the circuit, people went to the transmission of digital information or other information by means of pulses generated and detected individually. Thus, transmission necessarily falls short of channel capacity of the medium.

Now we go further and send out, not independent pulses whose amplitudes we are completely free to choose, but a redundant pulse train, in order to gain some control over the spectrum of all the pulses transmitted. Thus, we can make the spectrum go to zero where the transmission is poor: near direct current, for instance, in the case of transmission over wire or cable. More generally, by using redundancy in the pulse train we can make the transmitted power go to zero where the transmission is bad in amplitude or phase.

This can be a powerful technique. We used Fig. 9 to illustrate some of the problems of radio transmission. We see that if we can make the spectrum of the frequency-shifted pulses go to zero at the carrier frequency f_c , and at $f_c + B$, we can transmit both the carrier and a pilot frequency $f_c + B$ without damaging the spectrum of the transmitted signal.

Consider also Fig. 8. We see that if the spectrum were not rectangular, but fell to zero at the edges of a band of width B , bad transmission gain and phase near the band edge would not seriously affect the received signal.

How can we control the spectrum of a train of pulses, assuming that each individual pulse has a $\sin x/x$ shape, and hence a flat spectrum? The simplest way is called partial response transmission,¹¹ as shown in the following:

$$0 \rightarrow -1, 0, +1$$

$$1 \rightarrow +1, 0, -1$$

Here we encode a zero of an original binary train as a sequence of three $\sin x/x$ pulses, of amplitudes $-1, 0,$ and $+1$, and we encode binary digit 1 as a sequence of three pulses $+1, 0,$ and -1 . These sequences overlap so that the pulse rate is unchanged. This results in a sequence of pulses that can have three amplitudes $-2, 0,$ or $+2$. In partial response encoding of a binary string of digits we generate and detect pulses one at a time, and we think of the signal that is generated as made up of $\sin x/x$ pulses. Actually, a $\sin x/x$ pulse cannot appear by it-

self as a component of the signal. The signal is composed of overlapping pairs of $\sin x/x$ pulses.

A $\sin x/x$ pulse is shown at the top of Fig. 10. The pair characteristic of partial response encoding is shown below. The spectrum of the pair goes to zero at frequency zero and at frequency B . Its tails are smaller at a given distance from the central region than are the tails of a $\sin x/x$ pulse. A pulse stream made up of these paired pulses has three possible amplitudes. A stream of non-redundant three-level pulses could be used to transmit $\log_2 3 = 1.58$ bits/pulse. We use these three-level pulses to transmit $\log_2 2 = 1$ bit per pulse. Through adding redundancy to control the spectrum, we have reduced the efficiency of transmission by a factor 0.63, but we have gained quite a practical advantage in that the spectrum goes to zero at zero frequency and at frequency B .

Partial response encoding has a practical disadvantage. In order to recover the binary stream we must have the correct received pulse amplitudes for the whole past signal. This can be overcome by an expedient of pre-coding.

The first use of redundancy in order to control the spectrum of the pulse train, and hence to reduce intersymbol interference, was not partial response but the bipolar used in the TI transmission system, and described earlier. Bipolar encoding differs from partial response encoding in one particular way. Partial response encoding can never result in any frequency component of zero frequency. A positive transmitted pulse is always paired with a negative pulse two pulse positions away. In bipolar encoding, on the average, there are as many positive as negative pulses. However, it is quite conceivable that in bipolar transmission we will transmit all zeros except for a single one. In that case, the only pulse transmitted is a single $\sin x/x$ pulse, which has a flat spectrum that does not fall to zero at zero frequency. However, for a random string of input digits, on the average the spectrum of a bipolar string of pulses does fall gradually to zero at zero frequency.

Figure 11 shows, for a process producing a random string of the input digits, the spectrum for partial response encoding and the spectrum for bipolar encoding. The

FIGURE 9. Carrier and pilot frequencies for SSB transmission of trains of $\sin x/x$ pulses of bandwidth B .

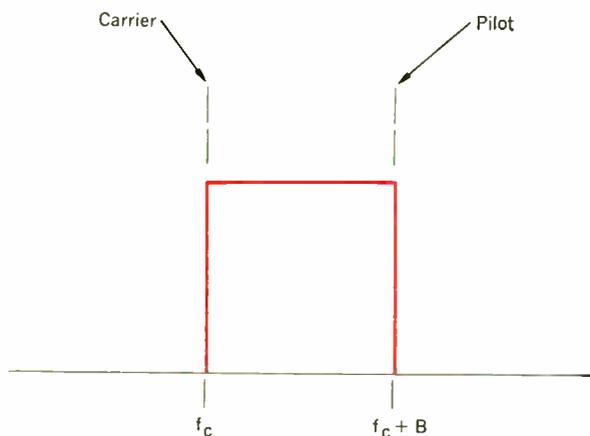
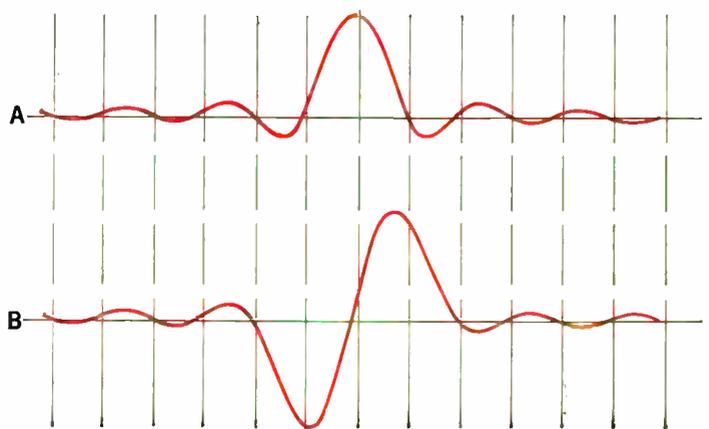


FIGURE 10. Two $\sin x/x$ pulses are combined to give the basic waveform used in partial response transmission. A—Independent $\sin x/x$ pulses. B—Partial response (pairs of $\sin x/x$ pulses). Redundancy, three levels instead of two, $\log_2 2/\log_2 3 = 0.63$.



partial response spectrum goes to zero at zero frequency and at frequency B . It is zero above frequency B because it is assumed that $\sin x/x$ pulses of bandwidth B are used. The bipolar spectrum is zero at zero frequency and a maximum at frequency B . It is zero above frequency B because, again, we have assumed $\sin x/x$ pulses of bandwidth B .

Actually, the effect of the bipolar encoding is to multiply the spectrum of an individual pulse by a raised cosine function with a period $2B$, which, of course, falls to zero at a frequency $2B$.

Maurice Karnaugh¹² devised an ingenious scheme by means of which any encoding that makes the spectrum go to zero at frequency intervals of $2B$, as bipolar encoding does, can be used to attain a spectrum that goes to zero at intervals B , as partial response encoding does. Figure 12 explains this. At the top is a sequence of digits to be encoded. This stream of $2B$ digits per second is divided into two streams of B digits per second. Each stream is encoded as a bipolar stream (or subjected to some other process) and is turned into pulses of bandwidth B . The two bipolar streams are then interleaved to give the transmitted signal. For a random sequence of input digits, the power spectrum falls to zero at zero frequency and at frequency B .

There are other ways of producing a redundant train of pulses such that the spectrum goes to zero at zero frequency and at a frequency $2B$. Among these are pair-selected ternary.¹³ The ingenious trick of Karnaugh can be applied to all such methods to give a spectrum that goes to zero at both zero frequency and at the upper end of the Nyquist band, the frequency B . Hence, we are equipped to make spectra that suit our needs.

There is still, however, the matter of computing the effect of the transmission characteristics on intersymbol interference. If the received and equalized pulse does not meet the Nyquist criterion, the tail of the pulse will be different from zero in many time slots. Intersymbol interference is the summation in a given time slot of tails of pulses assigned to other time slots. How can we calculate intersymbol interference?

If the pulses used have a bandwidth considerably greater than the Nyquist bandwidth, and if the spectrum falls gradually toward zero at the upper end of the band,

the pulse tails will be reasonably short and will cause intersymbol interference only in a few nearby time slots. Such pulses are wasteful of bandwidth, however, and are scarcely suitable for efficient radio transmission. These broadband pulses could be used in transmission over cables or wires but there would be a power penalty in transmitting power at unnecessarily high frequencies at which attenuation is high.

If an effort is made to make the pulses broader, and hence more efficient in the sense of conserving frequency, the tails differ from zero at many time slots. If we try to compute intersymbol interference by making a Fourier transform of the spectrum, or by simulation, we will have to examine the sums of many tails in distant time slots. Is there any way of getting the required information simply from the transmission characteristic and the spectrum of the transmitted pulse train?

There is, indeed, a way of computing the *average* intersymbol interference entirely in the frequency domain.¹⁴ This method takes into account the pulse, the characteristics of the transmission circuit, and the modification of the spectrum of the pulse train through the introduction of redundancy by bipolar or some other expedient means. The following exposition was provided by D. Slepian; the form of the result is different from, and perhaps simpler to understand than, that given by Smith. We start by expressing the transmitted signal in terms of a sequence of numbers. Let $\dots b_{-2}, b_{-1}, b_0, b_1, b_2, \dots$ be a statistically stationary time series of real numbers that we regard as the stream of digital data to be transmitted.* We use these data to generate a pulse train whose amplitude as a function of time is $s(t)$:

$$s(t) = \sum_{n=-\infty}^{\infty} b_n p(t - n\tau) \quad (1)$$

Here τ is the pulse spacing interval. The waveform of the individual transmitted pulse $p(t)$ has a Fourier transform $P(f)$, so that

$$p(t) = \int_{-\infty}^{\infty} e^{2\pi jft} P(f) df \quad (2)$$

$$P(f) = \int_{-\infty}^{\infty} e^{-2\pi jft} p(t) dt \quad (3)$$

Thus, the pulse train can also be expressed:

$$s(t) = \sum_{n=-\infty}^{\infty} b_n \int_{-\infty}^{\infty} e^{2\pi jf(t-n\tau)} P(f) df \quad (4)$$

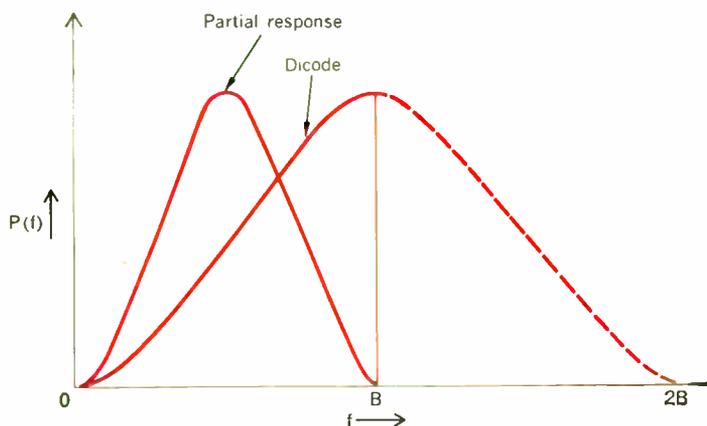
The transmitted signal or pulse train of (1) or (4) is put through a linear transmission system with a transfer function $Y(f)$ that yields a new received signal or pulse train $r(t)$:

$$r(t) = \sum_{n=-\infty}^{\infty} b_n \int_{-\infty}^{\infty} e^{2\pi jf(t-n\tau)} P(f) Y(f) df \quad (5)$$

To recover the data from $r(t)$ we sample this received wave at times $t_0 + n\tau$. Here t_0 is a time displacement that may be adjusted for least intersymbol interference. We multiply the sample amplitudes by a constant factor A so as to make their amplitudes correspond as nearly as

* Although we apply the expressions given to digital or quantized data, the b 's need not be quantized and the results apply to samples of a continuous analog signal.

FIGURE 11. Power spectrum densities for partial response and bipolar transmission.



The binary stream 0 1 1 0 1 0 0 1 0 1 1 1

Is divided into two streams $\left(\begin{array}{cccccccc} 0 & & 1 & & 1 & & 0 & & 0 & & 1 \\ & 1 & & 0 & & 0 & & 1 & & 1 & & 1 \end{array} \right)$

Each is made bipolar $\left(\begin{array}{cccccccc} 0 & +1 & -1 & & 0 & & 0 & +1 \\ +1 & & 0 & & 0 & -1 & +1 & -1 \end{array} \right)$

The two bipolar streams are interleaved 0 +1 +1 0 -1 0 0 -1 0 +1 +1 -1

FIGURE 12. How to use bipolar transmission and make the spectrum fall gradually to zero at a frequency B.

possible to the coefficients b_n that determine the amplitudes of the transmitted pulses.

It can be shown (from Ref. 14) that the mean square difference ϵ^2 between the samples of the received wave and the coefficients b_n of the pulses of the transmitted wave is given by

$$\epsilon^2 = \tau \int_{-B}^B Q(f) \left| 1 - \frac{A}{\tau} T(f) \right|^2 df \quad (6)$$

$$T(f) = \sum_{n=-\infty}^{\infty} P(f + 2Bn) Y(f + 2Bn) e^{2\pi j t_0 (f + 2Bn)} \quad (7)$$

In order to attain the least possible intersymbol interference for the transmission system, the constants A and t_0 should be chosen so as to minimize ϵ^2 of (6). This can be done conveniently with the aid of a digital computer.

The quantity $Q(f)$ in (6) is the power spectrum of the data sequence, the sequence of b 's. We write for the covariance ρ_l of the data sequence:

$$\rho_l = E b_m b_{m+l} = \rho_{-l} \quad m, l = 0, \pm 1, \pm 2, \dots \quad (8)$$

Here $E b_m b_{m+l}$ means the expected value of $b_m b_{m+l}$. Then

$$Q(f) = \sum_{n=-\infty}^{\infty} \rho_n e^{-2\pi j n \tau f} \quad (9)$$

The power spectrum $Q(f)$ is periodic with a period $2B = 1/\tau$. For a random sequence of pulses with amplitudes $+1$ or -1 , $Q(f)$ is constant and equal to 1. For a bipolar or other sequence $Q(f)$ is not constant, but it is always periodic with a period $2B$. By means of the interleaving trick of Karnaugh we can make $Q(f)$ have a period B rather than $2B$.

The summation of (7) appears because, as we have already seen in connection with Figs. 3 and 4, frequency components $f \pm 2nB$ are indistinguishable at sampling

times spaced $\tau = 1/2B$ apart. Hence, we sum over these indistinguishable (at sampling times) components before comparing a received sample with a coefficient b_n .

The quantity $P(f)Y(f)$ represents the complete transmission function of the system excited by the input coefficients b_n , as we see from (5). Quantity $(A/\tau)T(f)$ is a spectral component, adjusted in amplitude by A and in time by t_0 . If the Nyquist criterion is met, $T(f)$ can be made equal to unity by properly choosing A and t_0 , and hence ϵ^2 can be made to go to zero. We can regard the deviation of $(A/\tau)T(f)$ from unity as a deviation from the Nyquist criterion that for zero intersymbol interference,

$$\sum_{n=-\infty}^{\infty} P(f + 2Bn) Y(f + 2Bn)$$

must have constant amplitude and linear phase.

Finally, $Q(f)$ appears in order to take into account the fact that redundancy in the input sequence of b 's can reduce the frequency content of the transmitted and received signals in some frequency ranges and augment it in others.

ϵ^2 is, of course, the mean square error due to intersymbol interference. The distribution of errors is also vitally important. This depends on the nature of the process used in generating the sequence of coefficients b_n that determines the amplitudes of pulses fed into the transmission system. Some work has been done on error bounds due to intersymbol interference,¹⁵ and this important topic deserves further consideration.

Let us now consider the implications and uses of these expressions concerning intersymbol interference.

In designing a transmission system, we will in general expect least intersymbol interference if we use pulses that approximate $\sin x/x$ pulses or other pulses whose spectra meet the Nyquist criterion. Such pulses may have spectra considerably broader than the Nyquist bandwidth, B . Exceeding the Nyquist bandwidth may

be acceptable in cable transmission, but it is undesirable in multichannel radio transmission, where spectrum should be conserved.

It is impossible to produce exactly pulses meeting the Nyquist criterion and difficult to approach such pulses. However, with sufficiently complicated equalization it is possible to make intersymbol interference very small. Indeed, virtually all intersymbol interference can be removed by using a self-adjusting equalizer at the receiver^{16,17} in which a control circuit adjusts a transversal filter so as to make the response zero at times $\pm nT$ from the center of the pulse.

In some cases, a self-adjusting equalizer may be unduly complicated or expensive. A moderate degree of equalization may be practical, but one must use networks of reasonable complexity. In regions where the spectrum is cut off or shaped, the phase will be somewhat nonlinear.

Equation (6) should be a useful guide in arriving at practical networks that will give satisfactory intersymbol interference. The quantity $|1 - (A/\tau)T(f)|^2$ in (6) is a measure of departure of the spectrum from the Nyquist criterion. In some regions this quantity will have an appreciable value. Where this is so, $Q(f)$, the spectrum determined by the redundancy of the pulse train, can be of use. Over some parts of the frequency range we can make $|1 - (A/\tau)T(f)|^2$ small; over other parts we can make $Q(f)$ small. This approach should be a useful resource in reducing average intersymbol interference.

In the exploration that led to this article, the writer's aim was to understand some of the problems of practical efficient digital transmission by radio or cable. He believes that these problems have been reasonably set forth, and that some of the tools useful in exploring and designing such transmission have been discussed. Much more work needs to be done in order to choose and implement a good transmission scheme.

Phase modulation (a form of double-sideband transmission) is particularly attractive for radio and waveguide. Surely, it is important to compare the advantages of phase modulation with those of single-sideband transmission using encodings that make the signal spectrum go to zero in the vicinity of the carrier frequency.

Only a few examples have been given of encodings that make the spectrum of a pulse train go to zero at zero frequency, and at the Nyquist frequency B . It is clear that in all such encodings we must pay a price for shaping the spectrum of the pulse train. Many encoding schemes have been published, and there is surely much unpublished work. Among various coding schemes, which are most efficient? And among the reasonably efficient schemes, which are most practical? A thoroughgoing comparison seems called for.

Further, let us consider the matter of errors. Noise, interference, jitter in the time at which the received wave is sampled, and intersymbol interference can all lead to errors. In radio systems, interfering transmitters may be more important than Gaussian noise. It is not easy to take all sources of error into account at once, even in unreasonably simple cases. Yet, if analysis is to help us, it must be applied to reasonable but somewhat complicated transmission characteristics, whether these be the attenuation of real cables or the transmission characteristics of plausible, practical microwave filters.

In examining the literature, the writer has sometimes

encountered papers in which everything (really not everything—usually interference has been left out) is included in complicated expressions, but numerical results are given only for trivial and unrealistic examples, such as RC networks, which afford little insight into actual problems. Might it not be better to explore effects one by one in more realistic cases?

Because we know the probability distribution of Gaussian noise, if we know the signal strength and mean square noise, we can compute the error rate. Presumably, in the case of interfering radio transmitters it should be possible to calculate the error rate in some reasonably straightforward manner. The case of intersymbol interference is more difficult. Whereas a bound has been found and applied to the case of random $\sin x/x$ pulses,¹⁵ there is still something to be done.

Having called attention to some of the problems that lie between us and an understanding of what is really possible and practical in digital transmission by radio or cable, the writer will happily leave in more energetic and competent hands the monumental work necessary to resolve these problems.

Essentially full text of a paper presented at a meeting of the Sendai Chapter of the Institute of Electronics and Communication Engineers, Sendai, Japan, April 12, 1968.

REFERENCES

1. Fulz, K. E., and Penick, D. B., "The T1 carrier system," *Bell System Tech. J.*, vol. 44, pp. 1405-1451, Sept. 1965.
2. Dorros, I., Sipress, J. M., and Waldhauer, F. D., "An experimental 224 Mb/s digital repeatered line," *Bell System Tech. J.*, vol. 45, pp. 993-1043, Sept. 1966.
3. Hubbard, W. M., Goell, J. E., Warters, W. D., Standley, R. D., Mandeville, G. D., Lee, T. P., Shaw, R. C., and Clouser, P. L., "A solid-state regenerative repeater for guided millimeter-wave communication systems," *Bell System Tech. J.*, vol. 46, pp. 1977-2018, Nov. 1967.
4. Hogg, D. C., "Millimeter wave communication through the atmosphere," *Science*, vol. 159, pp. 39-46, Jan. 5, 1968.
5. Pierce, J. R., "Information rate of a coaxial cable with various modulation systems," *Bell System Tech. J.*, vol. 45, pp. 1197-1207, Oct. 1966.
6. Nyquist, H., "Certain topics in telegraph transmission theory," *AIEE Trans.*, vol. 47, pp. 617-644, Apr. 1928.
7. Gibby, R. A., and Smith, J. W., "Some extensions of Nyquist's transmission theory," *Bell System Tech. J.*, vol. 44, pp. 1487-1510, Sept. 1965.
8. Bennett, W. R., and Davey, J. R., *Data Transmission*. New York: McGraw-Hill, 1965.
9. Simon, M. K., "Extensions to the analysis of regenerative repeaters with quantized feedback," *Bell System Tech. J.*, vol. 46, pp. 1831-1851, Oct. 1967.
10. Barker, R. H., "Electrical signaling and/or amplifying systems," U.S. Patent 2 700 696, Jan. 25, 1955.
11. Kretzmer, E., "Binary data communication by partial response transmission," Paper G19-65-419, *Conf. Record Cat. 19-C11*, First Ann. IEEE Commun. Conv., Boulder, Colo., June 7-9, 1965, pp. 451-455.
12. Karbaugh, M., "Three-level binary code transmission," U.S. Patent 3 214 749, Oct. 26, 1965.
13. Sipress, J. M., "A new class of ternary pulse transmission plans for digital transmission lines," *IEEE Trans. Communications Technology*, vol. COM-13, pp. 366-372, Sept. 1965.
14. Smith, J. W., "The joint optimization of transmitted signal and receiving filter for data transmission systems," *Bell System Tech. J.*, vol. 44, pp. 2362-2392, Dec. 1965.
15. Saltzberg, B. R., "Intersymbol interference error bounds with application to ideal bandlimited signaling," *IEEE Trans. Information Theory*, to be published.
16. Lucky, R. W., "Automatic equalization for digital communication," *Bell System Tech. J.*, vol. 44, pp. 547-588, Apr. 1965.
17. Lucky, R. W., "Techniques for adaptive equalization of digital communication systems," *Bell System Tech. J.*, vol. 45, pp. 255-286, Feb. 1966.

Linear integrated circuits in communication systems

This article should prove instructive to the applications engineer in that it presents current examples of the complexity that is economically achievable within the framework of integrated circuitry

*Robert A. Hirschfeld
National Semiconductor Corporation*

Advances in monolithic technology are broadening the applications areas in which integrated circuits are becoming economically feasible. This article demonstrates the design approaches that permit the realization of communication circuitry functions while simultaneously meeting the constraints imposed by MSI and LSI techniques.

Monolithic technology has already emerged, in the digital area, from an initial period of novelty and a subsequent period of superior performance, into the current realization that discrete transistor computing systems can no longer be competitive with systems employing medium- or large-scale integration. While lagging digital circuits, two-way speech communication systems have reached sufficient usage and complexity to become the logical targets for a similar large-scale transition to monolithic subsystems. Although there is considerable diversity in transmission media, modulation methods, etc., sufficient similarity exists between systems to allow nearly all circuit functions to be performed with present and currently emerging monolithic techniques.

The transition to monolithic communication systems will not be as smooth as was the corresponding change in computers, because considerably more functional diversity exists in communications, and also because designers of such systems are more intimately involved in the relationship between circuit elements and the system than are digital designers, who were dealing with functional blocks before the advent of microcircuits.

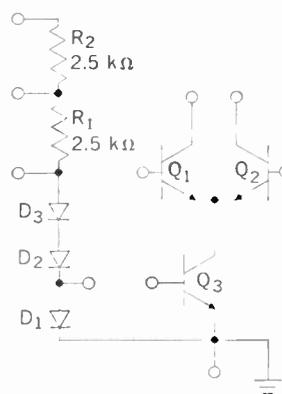


FIGURE 1. RF/IF amplifier circuit, type LM171.

The economics of communication systems require not only competitive cost, but superior performance, compared with present discrete circuits, to make a transition worthwhile. To this end, the potential producer of monolithic circuits must not only combine as many circuit functions on a chip as possible, but also circumvent the use of external components wherever feasible. Thus, economic requirements will force increasing circuit specialization and a firmly committed partitioning of circuit functions.

Despite large-scale needs for multifunction monolithic communication chips, the transitional period will also require a number of high-performance single-function circuits, designed to be as versatile and inexpensive as possible, to fill system gaps while the availability of complete subsystems increases. Such a versatile circuit, exemplified by the National Semiconductor Corporation's LM171, which is shown in Fig. 1, can serve as tuned or untuned amplifier, limiter, mixer, oscillator, detector, etc.^{1,2} while justifying its inclusion in the system through self-contained dc biasing, improved performance over single transistor stages, and simplified system component procurement.

System requirements

Whether a communications system uses radio frequency, a laser beam, or more prosaic direct wire as a transmission medium, basic requirements are sufficiently universal to permit a few generalizations. For either speech or video information, some form of transduction into electrical signals is necessary. Because the original information may be unpredictable and noisy, one class of microcircuit is needed to process and optimize the signal, so that it may be ready for superimposition on the transmission medium. This class includes preamplification, squelch, automatic gain control, and even servo control of the input transducer's reception pattern.

Signal gain is generally needed ahead of the transmission—followed, in most systems, by modulation onto a carrier. These functions fall naturally into another class of microcircuit, which may contain sufficient power output to drive some transmission media directly, and, because of the inexpensive complexity possible in monolithic circuits, may contain internal control and self-linearization of what may normally be a nonlinear modulation process. This class includes analog-to-digital conversion, and pulse coding in digital communication

systems, which may be readily included on one chip with associated linear functions.

On the receiving end of the transmission channel, the first order of business is usually amplification, to restore channel losses, and selectivity, if more signals than one use the same transmission medium. The first stage of amplification generally is more critical, with respect to internally generated noise and cross-channel intermodulation problems, than other amplifiers in the system. To some extent, the local oscillator and mixer, if used to translate to a standard intermediate frequency, share similar problems, so that critical processing requirements may make it attractive to design a receiver "front end" into a single block.

Present radio-receiver practice places most of the selectivity function between stages of a discrete transistor IF strip, where it can not only shape receiver bandpass, but provide interstage dc decoupling. Functionally, however, it is more efficient to treat selectivity as a single, lumped system function, ahead of the IF strip. Since economical bandpass filtering is at present impossible with mass-produced monolithic circuits, the mixer/bandpass/IF-strip interface is a logical point for chip partitioning.

Where information has been modulated onto a carrier, the next function is demodulation. Most simple detectors require external tuned components, whether they be diode AM, ratio-detector or discriminator FM, or single-sideband (SSB) product detectors. Monolithic techniques allow a rethinking of the detection problem in terms of more complex, direct-coupled, untuned circuits, so that detection could be included in the IF amplification function directly.

After demodulation, additional signal processing may be needed to drive appropriate audio or video transducers. This last class of microcircuit must deal with higher power levels, and may also include automatic control functions—such as automatic gain control

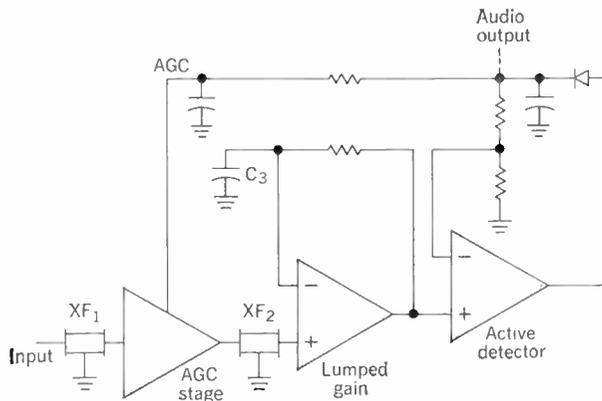
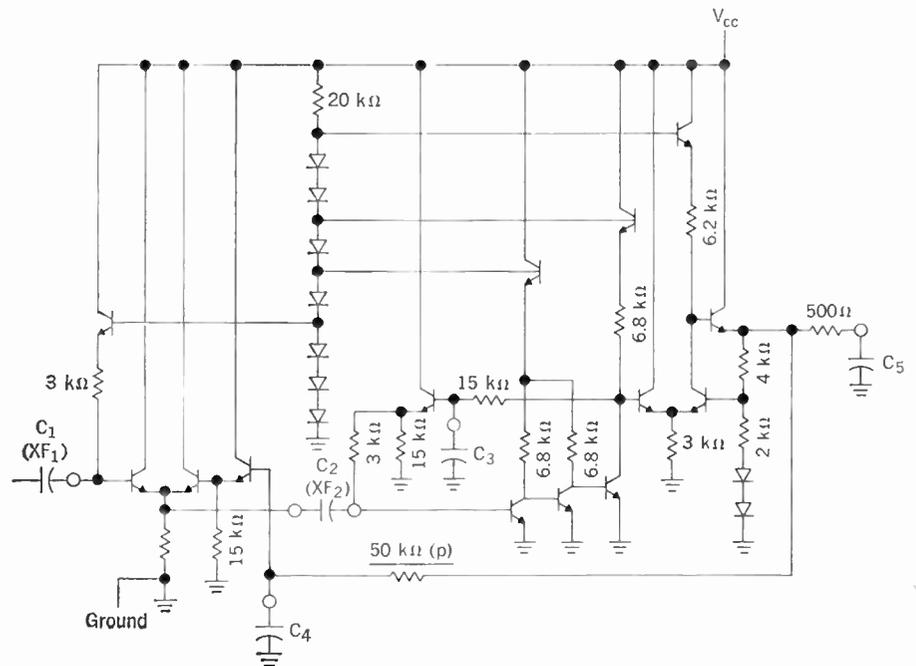


FIGURE 2. Block diagram of amplitude-modulated IF strip.

FIGURE 3. Schematic diagram of LM172 IF strip, showing external capacitors.



(AGC), squelch, or automatic frequency control (AFC)—to compensate for undesired transmission channel variations.

Monolithic-circuit requirements

Limitations and advantages of current monolithic technology are well known^{3,4} and will not be treated here in detail. Communication applications make large technological demands in only a few circuit functions, generally those where the microcircuit interfaces with a transducer or the transmission channel, and these challenges are already being met in the areas of medium-power audio amplifiers, low-noise front ends, etc. Small-signal gain and signal processing have been routinely performed by single-function microcircuits for several years. If these circuits are to compete economically, judicious choice of system function, as well as circuit design based upon known monolithic superiorities rather than traditional techniques, is necessary.

For the microcircuit manufacturer, the use of a minimum number of pins allows inexpensive packaging, as well as easy printed-circuit insertion by the user. But more important, minimization of pins implies less interstage coupling external to the chip and a drastic reduction in the number of external components required to sustain operation. Just as large-scale-integrated digital circuits become most attractive in very complex shift registers and memories, where interconnection is mostly on the chip and only inputs, outputs, and supplies need be brought out as pins, communication microcircuits become most economical when large numbers of system functions can be performed internally without any external intermediate connections. Thus, such subsystems are likely to become radical, direct-coupled, broadband departures from current practice.

Since monolithic subsystems inevitably deprive the system designer of some of his former control over components, predictable reliability and meaningful testing by the microcircuit manufacturer are needed. There is little doubt that the reduction of printed-circuit interconnections and the elimination of a number of reactive and tuned components will lead to more reliable systems. Although replacement and stocking of components will also be simplified, the repairman will be forced into modular troubleshooting and will know less about the actual internal workings of each module. Modular repair techniques, already effective in military equipment, will be economical for commercial and consumer applications as well. Well-designed monolithic circuits can be tested with fast, automatic dc machines and can usually be reliably extrapolated to known high-frequency guaranteed parameters by the manufacturer.

Practical communication subsystems

Two monolithic subsystems will be used as illustrations. The first, a complete amplitude-modulated IF strip, is a newly introduced replacement for all IF components from the mixer output to the power amplifier input of a superheterodyne radio receiver, and embodies the subsystems approach previously outlined. The second is an audio AGC/squelch system, which allows automatic speech control and audio level regulation in inexpensive, two-way, broadcast and public-address applications, and includes a number of functions on one chip.

The IF strip illustrates a minimum-pin functional

block, clearly partitioned; the AGC/squelch amplifier trades pins for versatility.

A complete monolithic IF strip

In current discrete-component practice, the most efficient utilization of available power gain, from a limited number of active devices, is obtained by matched, tuned interstage networks. Because of the limited AGC range that is obtainable by varying, for example, dc emitter current in a conventional common-emitter stage, several stages must receive AGC voltage from the detector, and require dc decoupling to eliminate effects of changing dc operating points with AGC.

Economics usually dictate the simplest diode detectors, biased from tuned transformer secondaries, in present strips, and the large tuned gain and often marginal stability necessitate power-supply decoupling for each stage. Conventionally tuned strips can be simulated by using a number of monolithic amplifiers on one chip, with each input and output brought out to the interstage network.² However, neither stability nor economy has been achieved.

Using a systems approach (Fig. 2) gives a much more efficient arrangement. All gain is attained in a single, direct-coupled, lumped-gain stage. To avoid AGC disturbance, the high-gain amplifier is left running at maximum gain at all times. The AGC function is performed by a single, efficient, gain-control stage, consisting of a voltage-variable attenuator, with a single decoupling capacitor (or ceramic filter) taking the place of individual stage-decoupling elements in conventional strips. Bandpass shaping is done by an external filter, ahead of the strip, which also serves to couple the strip to the receiver "front end." Rather than demand dc stability of the lumped-gain stage, it is designed for maximum RF effectiveness. The stage is stabilized by a dc feedback loop, which sets all bias points at optimum level, regardless of variations in the monolithic process. Finally, the circuit employs an AM detector, which can be directly coupled to the gain stage output. The detector is insensitive to dc bias, reliably provides the correct AGC voltage, and, by its "active detector" construction, is capable of audio voltage gain, unlike simple diode detectors.

The type LM172 IF strip is shown schematically in Fig. 3. Examination of the AGC stage shows it to be an emitter couple, in which minimum attenuation occurs when the right half is cut off, and the left half becomes a simple emitter follower. As the AGC voltage rises, the input emitter follower is progressively "robbed" of current from its emitter resistor by the right-half emitter follower, and in the limit is cut off, while the right half conducts fully. Thus, the gain control is a series-shunt arrangement with a larger range than that usable in a single common-emitter-tuned IF stage.

The lumped-gain stage consists of a common-emitter triplet, with no emitter degeneration. Such a circuit would be difficult to bias in discrete form, but is easily controlled in the LM172 by the dc feedback loop, rolled off by external capacitor C_3 . For correct biasing, collectors of the first and second common-emitter stages must operate at one forward base-emitter drop, V_{be} . Consequently, these devices must have excellent high-frequency geometries and low saturation voltages. Because the three 6.8-kilohm collector load resistors

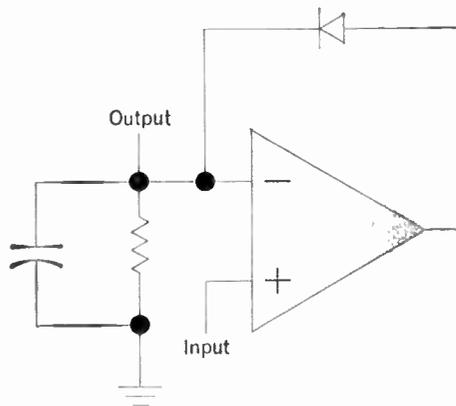


FIGURE 4. Active-detector circuit.

FIGURE 5. Active detector with audio gain.

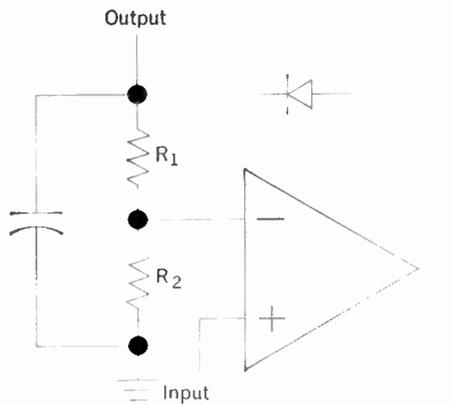
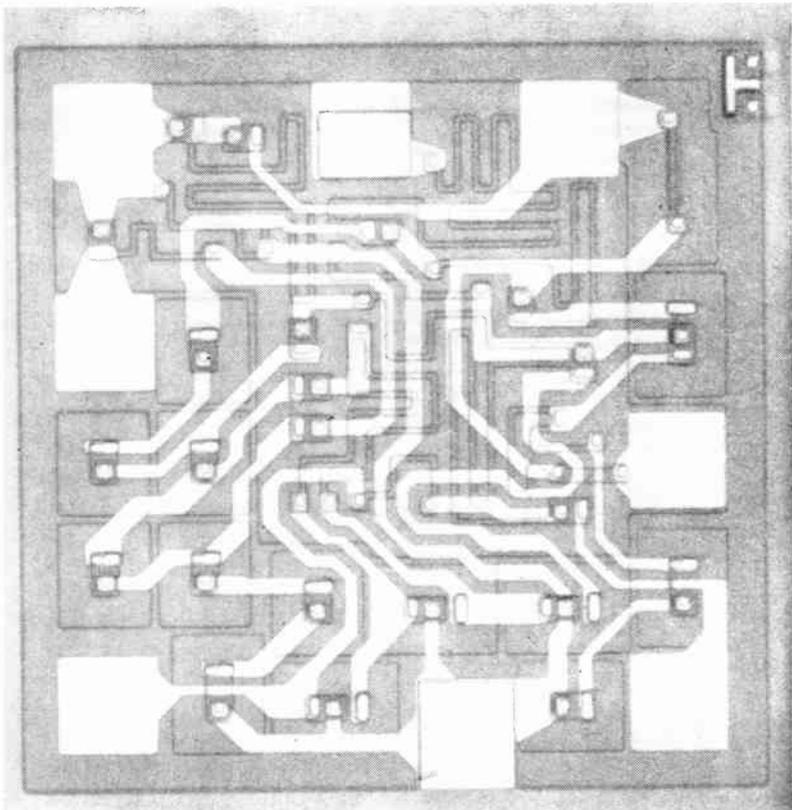


FIGURE 6. IF monolithic die, type LM172.



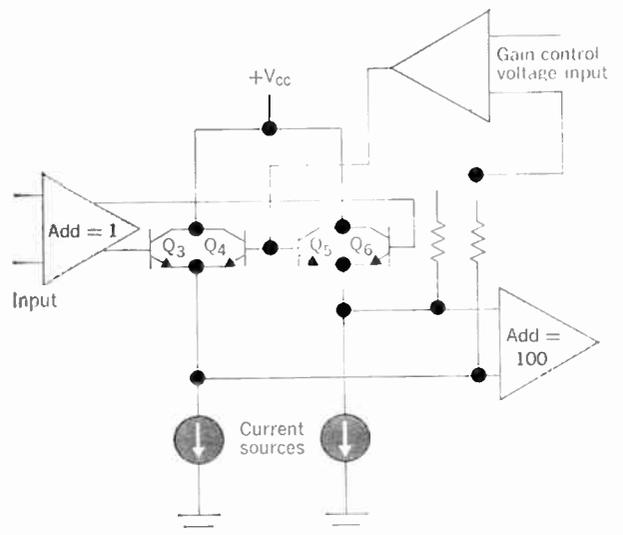
are biased from their own emitter-follower supply regulators, individual supply decoupling capacitors are unnecessary; moreover, the active part of the circuit is allowed to consume constant current from any supply voltage. Since each part of the LM172 is similarly supply-regulated, the circuit is efficient at both low and high voltages.

Operation of the active detector may be more easily seen from the circuit diagram of Fig. 4, which is commonly found in operational amplifier handbooks. It offers a number of improvements over simple diode detectors, which must be slightly forward-biased by external circuitry to respond to small signals, and which are generally inefficient, giving less audio output than is available from the IF strip's output modulation envelope. The circuit of Fig. 4, with a high-gain amplifier, becomes a nearly ideal peak detector, with automatic optimum biasing of the diode in its feedback loop. If a resistive divider is added to the active detector, as in Fig. 5, audio output will equal the modulation envelope multiplied by the ratio $(R_1 + R_2)/R_2$. In the LM172, the ratio is set to give an audio voltage gain of three; thus about one volt peak-to-peak is obtained from a 100 percent modulated signal within the automatic-gain-control range of the circuit.

In Fig. 3, external capacitor C_3 serves to smooth carrier ripple from the demodulated audio output, and to shape audio-frequency response. C_4 , in conjunction with a 50-kilohm feedback resistor, determines the AGC attack and release time constant. C_1 may be replaced by a Murata SF455D ceramic filter, which has input and output impedances suited to the LM172, as an example of a simple bandpass network. In a typical 455-kHz application, the strip exhibits an AGC range that extends from 50 μ V into the AGC stage input, to 50 mV, or a 60-dB control range.

The LM172, then, is a complete IF strip, requiring only external tuning and bypass capacitors, in an eight-pin T05 can. Inspection of the schematic diagram, and of the actual chip (Fig. 6), reveals that despite the use of 23 junction devices, the circuit fits on a very

FIGURE 7. Block diagram of LM170.



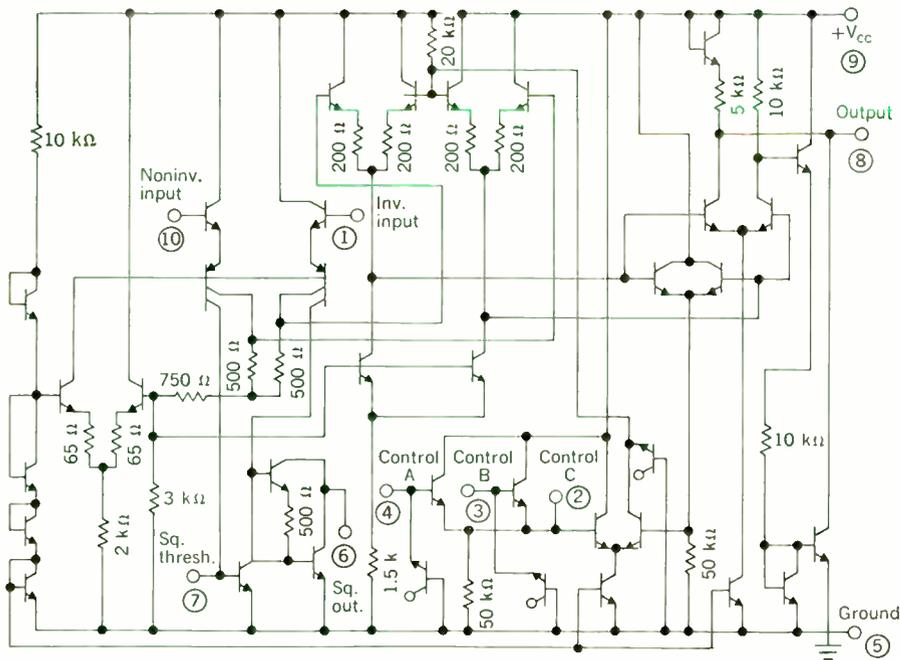
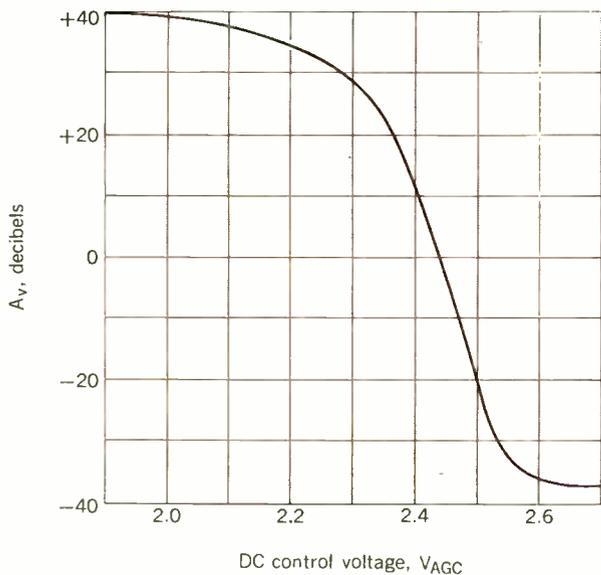


FIGURE 8. Schematic diagram of LM170 AGC/squelch amplifier. Circled numbers are pin connections.

FIGURE 9. Voltage gain vs. control voltage for LM170.



small die, 33 by 33.5 mils (0.84 by 0.85 mm), and can consequently be very competitive with the inexpensive transistors and IF transformers it is intended to replace. Although very inexpensive ceramic filters are the most attractive tuning scheme, LC or crystal lattice filters are suitable in the range from 50 kHz to 2 MHz. Because it is a broadband subsystem, it may even be used alone, or with slight additional gain, in TRF applications.

It is anticipated that communication subsystems having as high a potential usage as the AM strip will continue to receive attention from microcircuit manufacturers, as complete functional blocks with high reproducibility but having the lower degree of versatility characteristic

of “designed for one application” high-volume circuitry. Pin standardization may turn out to be a matter of historical precedent, as it was with the type 709 operational amplifier and its successors.

A monolithic AGC/squelch subsystem

In contrast to the IF strip, another subsystem to be exemplified is one that is not as likely to be used in very high volume in a single application, but can be useful in a range of related circuitry, as in speech-controlled transmitting and receiving terminals, and in improving intelligibility through AGC of rapidly fluctuating transmitted signals. For this reason, more pins are used, and more of the internal portions of the circuit are made available to the system designer. It is still important to minimize external interstage components; thus, the circuit is essentially a variable-gain dc amplifier, with direct coupling throughout (Fig. 7).

This is a departure from conventional variable-gain circuits, in that the inevitable change in dc output voltage, as the operating point of any commonly used gain-changing element is varied, is canceled through differential structures. The need for dc decoupling is therefore eliminated. Such a system also cancels the transients in the output that result from rapid gain changes. These transients are not eliminated in conventional capacitor or transformer AGC circuits.

A number of points of interest appear in the complex schematic, type LM170, of Fig. 8. The circuit, in a ten-pin TO-5 package, has essentially the configuration of an operational amplifier, with differential inputs and a single-ended output that lies halfway between ground and the positive supply—or at ground, when two supplies are used. The circuit’s maximum gain is only 40 dB, and thus it cannot be strictly considered an operational amplifier. Pins 3 and 4 are emitter-follower gain-control inputs, with control action occurring

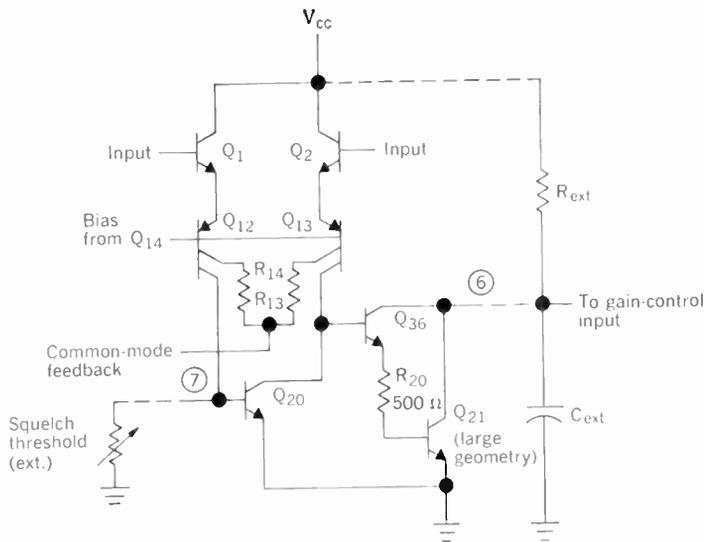


FIGURE 10. LM170 internal squelch detector.

FIGURE 11. Fast-attack, slow-release squelch, with $C_{ext} = 25 \mu F$, $R_{ext} = 100 k\Omega$. Upper trace is input, switched above and below threshold (25 mV per division). Lower trace is output, first zero gain, then abrupt turn-on, followed by constant-gain period, then smooth transition back to zero again. Lower-trace scale is 2 volts per division. Horizontal scale is 200 ms per division.

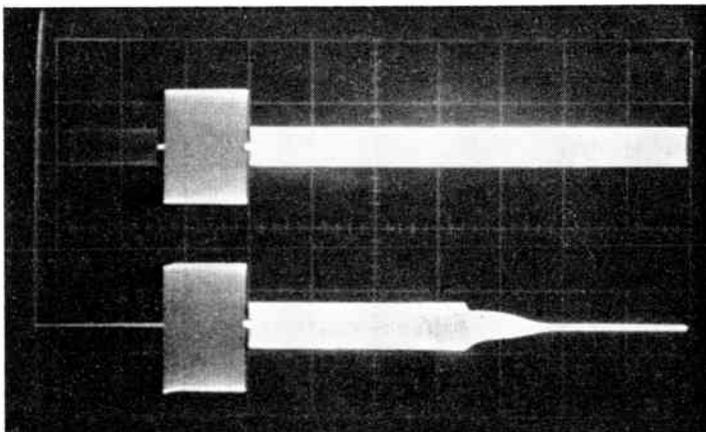
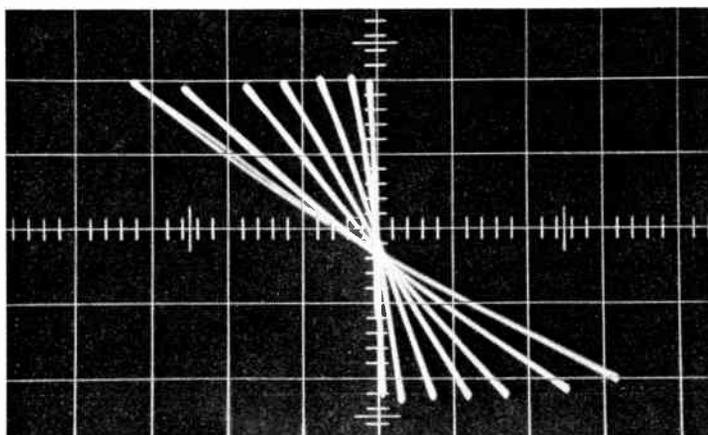


FIGURE 12. Transfer characteristics of LM170, connected in a closed-loop AGC system. Vertical axis is output; horizontal is input. Both scales are 10 mV per division.



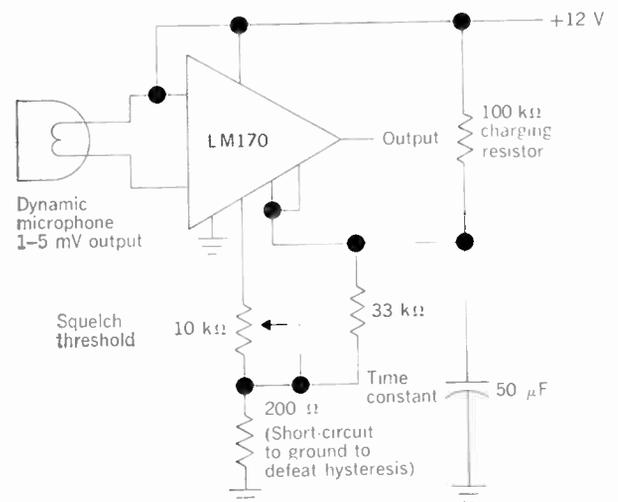
between +2.1 and +2.5 volts, as shown in Fig. 9. If pin 2 is bypassed, pins 3 and 4 become detectors, capable of providing control signals from an ac input.

In addition to the variable gain and preamplification performed by the LM170, a second, high-gain amplifier-detector precedes the balanced variable attenuator, with output at pin 6 and threshold control at pin 7. This section performs as a squelch detector, which can drive pin 3 or 4 directly with a direct voltage that either shuts off all gain in one state, or turns the amplifier fully on, when the incoming signal exceeds a level chosen by setting a potentiometer at pin 7.

For clarity, the squelch-detector portion of the LM170 is shown in Fig. 10, along with the differential input amplifier, which exhibits the same versatility and tolerance to abuse found in the LM101 operational amplifier input circuit. Q_{20} is brought out of saturation when a negative input peak greater than the set threshold appears. Because of its high-impedance collector load, Q_{20} forms a very-high-gain stage, which turns on Q_{36} and Q_{21} . A large external transistor has been charged above the control input's turnoff point by a large external resistor (or potentiometer, if variable timing is needed). When Q_{21} , a large-geometry transistor, suddenly discharges the external capacitor, control voltage is abruptly brought below the value needed to turn the amplifier fully on. In the absence of input signal, however, turnoff is delayed until the capacitor slowly charges to and through the gain-control region. The resulting fast-attack, slow-release squelch action is illustrated in Fig. 11.

In audio AGC applications, the LM170 is driven by the system's audio output. One of the two internal detectors is used to provide an "error" voltage if positive peak output exceeds a preset maximum, and amplifier gain is reduced as needed. The family of transfer characteristics (Fig. 12) shows how peak-to-peak output is kept nearly constant for widely varying inputs. The AGC system, because of its emitter-follower detectors, also has a fast-attack, slow-release action and can be made to respond only to average, rather than instantaneous, speech inputs.

FIGURE 13. Directly driven squelch circuit with hysteresis.



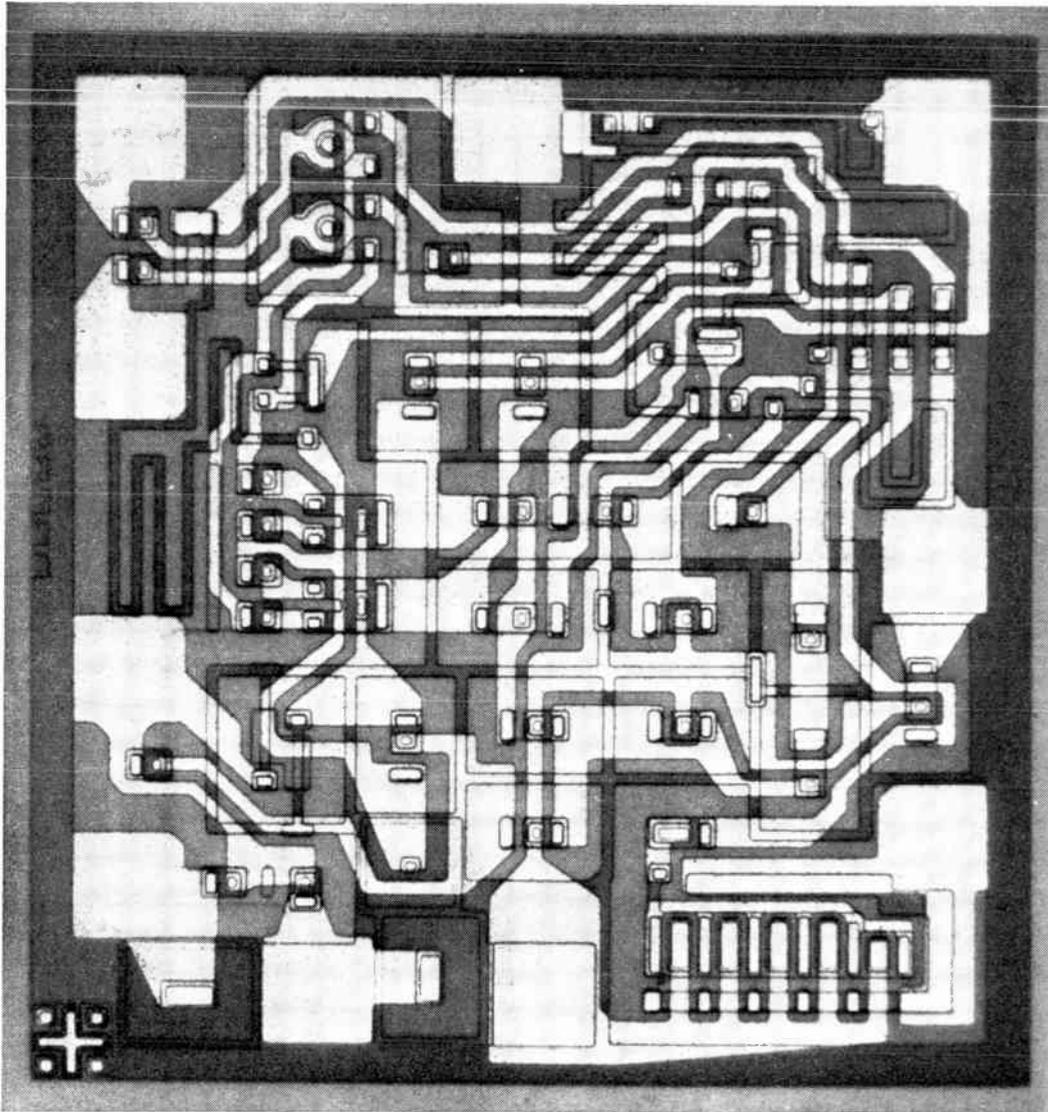


FIGURE 14. Type LM170 AGC/squelch-amplifier chip.

Because its input configuration allows common-mode voltages equal to the positive supply, such transducers as dynamic microphones may directly drive the LM170, with no external biasing or decoupling components required; see Fig. 13. The slightly positive, two-stage feedback shown creates a controlled amount of hysteresis, which, together with the fast-attack, slow-release squelch action, immunizes the circuit against erratic speech patterns and noise. Such a circuit might be built into a mobile microphone case, for example, giving self-contained speech control. At the same time, the squelch-detector output could become a "vox" control, automatically causing a two-way system to go into the "transmit" mode at the first syllable of speech and might even be connected for simultaneous AGC action by driving the second gain-control input from the system output. The LM170 chip is larger and more complex than the LM172; however, its 39- by 42-mil (1.0- by 1.07-mm) size still allows reasonably low cost, compared with the LM101, for example, which is 45 mils (1.15 mm) square; see Fig. 14.

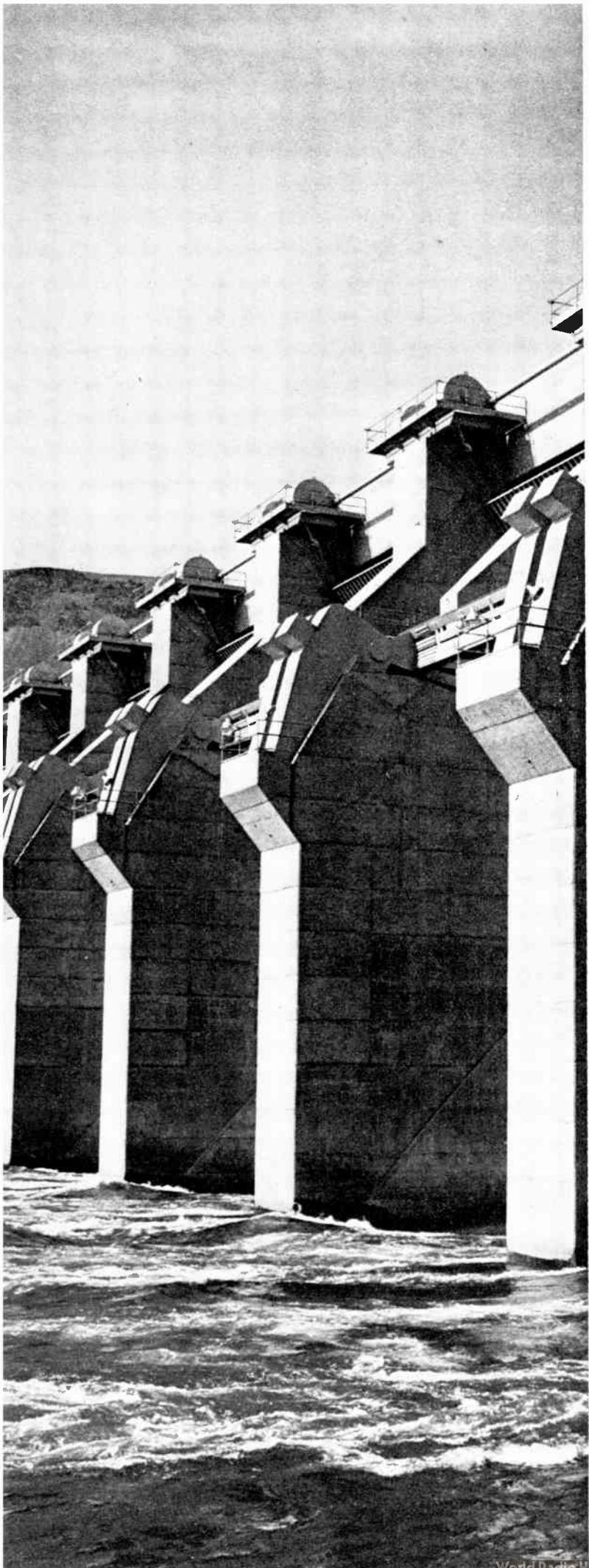
Conclusion

Monolithic circuits are likely to become major forces in the communication industry within the next few years, both in versatile, general-purpose elements and in specialized monolithic subsystems. Ultimately, it appears that most mass-produced communication equipment will use standardized one-chip subsystems, which will be produced competitively within the semiconductor industry. The resultant economies will allow commercial and consumer equipment to make full use of the sophistication and automatic control presently restricted to relatively expensive military systems.

Essentially full text of a paper presented at WESCON, Los Angeles, Calif., August 20-23, 1968.

REFERENCES

1. Hirschfeld, R. A., "Tuned circuit design using monolithic RF/IF amplifiers," *Electron. Design*, May 23, 1968.
2. Eimbinder, J., ed., *Linear Integrated Circuits*. New York: Wiley, 1968, chap. 14.
3. Robertson, J. J., and Hirschfeld, R. A., "Application of microelectronics to IF amplifiers," *IEEE Trans. Vehicular Communications*, vol. VC-14, pp. 162-172, Mar. 1965.
4. Widlar, R., "Future trends in IC operational amplifiers," *Elec. Design News*, June 10, 1968.



The story of Bonneville Power: 1937-1968-1987...

Dams of the Columbia River Basin

Bonneville Dam, the original "make work" project built in the depression days of the 1930s, was the forerunner of one of the world's most extensive hydro power projects. Today—31 years and 23 dams later—the Columbia River Basin is reaching toward its ultimate development as a vast source of electric energy

Gordon D. Friedlander Staff Writer

Realization of the vast hydro power and flood-control potential of the Columbia River, its tributaries, and watershed basin dates back more than half a century. As early as 1923, the U.S. Department of the Interior and the Army Corps of Engineers presented comprehensive testimony to the Congress on the feasibility of such eventual development. And, in 1937, B. E. Torpen, a hydroelectric engineer, presented to the ASCE an outline of the power possibilities of the Columbia River and its tributaries. This paper, entitled "Where Rolls the Oregon," was an accurately prophetic blueprint for the optimum development of the Columbia River Basin over a 50-year period. In it, he anticipated the need for the Canadian Treaty to provide upstream storage and to interconnect the great hydro energy in the future of British Columbia. His dream and his predictions are well on their way toward ultimate fulfillment.

From 1937 to the present time, the population of the Pacific Northwest has increased by 100 percent. Six million people now live in the states of Oregon, Washington, Idaho, and the portion of Montana that is west of the Continental Divide. Thirty-one years ago the aver-

age residential consumer in the four-state area required approximately 1200 kWh per year. As of 1967, the average per consumer increased tenfold—to more than 12 000 kWh annually.

The Columbia River—background to power

The mighty Columbia rises in the Selkirk Range of the Canadian Rockies in the Province of British Columbia. Its source is Columbia Lake, a 21-km-long, deep, glacial-fed pool that is situated 800 meters above sea level. From here, flowing southwestward 2000 km to the Pacific Ocean, the initially small stream is joined by more than 150 tributaries large enough to be designated as rivers. And on this long journey to the sea, the Columbia develops into one of the world's most powerful rivers. It is the fourth largest river in North America. Although the Mississippi, St. Lawrence, and the Mackenzie are larger streams, the Columbia easily outpaces them as a source of energy. The river and its tributaries contain 30 percent of the hydroelectric potential of the North American continent.

The Columbia River Basin drains an area larger than France. The watershed includes part of British Columbia and portions of Oregon, Washington, Idaho, Montana, Wyoming, Nevada, and Utah.

The flow of the Columbia and its tributaries, however, seasonally fluctuates over a wide range. For example, at the Canada-United States boundary, the Columbia's largest recorded flow was 15 400 m³/s; its smallest was 390 m³/s. The great river enters the United States about 140 km north of Spokane, Wash. At the international boundary, it has descended 410 meters and has traveled 800 km from its source.

The volume of water and its varying flow can create problems in flood control and power production. When heavy springtime precipitation is accompanied by a normal-rate melting of the snowpack on the mountain slopes of the basin, there will be adequate runoff for power production, irrigation, potable water, and fish breeding. But calamitous floods can ensue if warm, torrential rains cause a rapid melting of snows in the watershed area. At the other extreme, a lack of rainfall can reduce the runoff in the basin to that point where power production becomes critically affected; it may also sharply reduce the amount of water that can be diverted for agricultural irrigation.

Historical background

The early dams. The first sizable dams in the Columbia River Basin were built on the Upper Snake River in Idaho by the Bureau of Reclamation of the U.S. Department of the Interior. The Minidoka Dam, a 30-meter-high

structure, was the initial project. It began delivering power to nearby farms in 1909. This dam still serves its area with a generating capability rated at 13 400 kW. In 1912, the Boise Diversion Dam added 1500 kW of hydro power; and, in 1925 and 1927, respectively, Black Canyon and American Falls raised the generating capacity along the Bigwood and Snake Rivers by 38 500 kW. The generators at the earlier dams furnished electric power primarily to operate the irrigation pumps of federal projects.

Not all of the early hydro projects were federally sponsored, however. Private and public utilities in the Pacific Northwest completed the Twin Falls Dam (13 500 kW) on the Snake, in 1912; the Long Lake Dam (70 000 kW) on the Spokane River (Wash.), in 1918; the Bull Run Dam (21 000 kW) on the Bull Run River (Oreg.), in 1928; the Leaburg Dam (13 500 kW) on the McKenzie (Oreg.), in 1928; and the Lewiston Dam (10 000 kW) on the Clearwater, in the same year. The year 1927 saw

FAR LEFT. Spillway section of the John Day Dam, presently under construction. This facility will generate a minimum of 2160 MW in its first-phase service, scheduled to go on the line in 1970.

FIGURE 1. The nonfederal Merwin arch-gravity dam on the Lewis River in Washington was the first structure of this type to be built in the Pacific Northwest. Capable of generating 135 000 kW, electric energy from this facility was put on the line back in 1930.



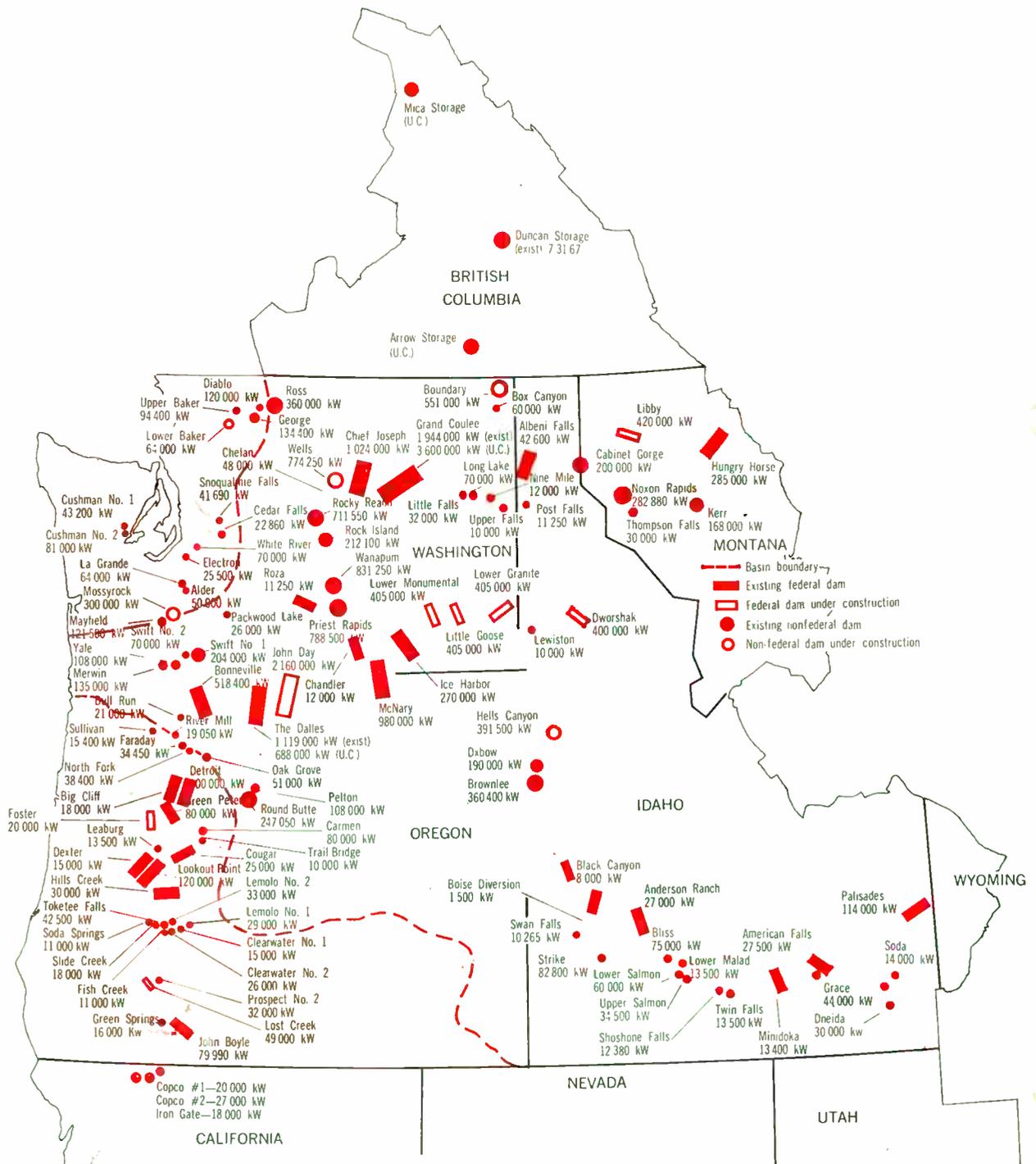
the completion of the American Falls (27 500 kW) federal dam on the Snake. In 1930, the 100-meter-high non-federal Merwin arch dam (135 000 kW) on the Lewis River in Washington (Fig. 1) was completed. There were also a number of other notable dams built in the basin area for flood control and irrigation purposes only in the 1910–1930 period. Figure 2 shows the existing and projected federal and nonfederal power dams in the

Pacific Northwest, both within and outside of the Columbia River Basin. It may be used as a ready reference throughout this article.

Bonneville: the first of the great federal dams

In that grimmest year of the Great Depression, 1933, the initial large-scale power development of the Columbia River at Bonneville, Oreg., 65 km east of Portland, began

FIGURE 2. Federal and nonfederal hydroelectric dams in the Pacific Northwest, lying both within the outside of the boundaries of the Columbia River Basin. Quantities shown represent both existing generation and that which is being installed (where applicable). Symbols indicate federal and nonfederal projects authorized, licensed, or under active consideration, as well as potential additions.



on September 30, under the provisions of the National Recovery Act. For the short term, it was admittedly an emergency public works project to ease the huge unemployment problem then existing in the Pacific Northwest. The formal authorization for construction was made by Congress under the River and Harbor Act of 1935. For the long term, however, it represented the first step in a 50-year program.

The entire project was constructed—and is being maintained and operated at present—by the U.S. Army Corps of Engineers.

Details and statistics. The Fig. 3 map shows the immediate dam area and appurtenances. The Columbia

River reservoir runs east of Bonneville to The Dalles Dam—the next great low dam in the system, situated 75 km upstream.

As may be seen in Fig. 3, the powerhouse portion of the dam is situated between Bradford Island and the Oregon shore. The original construction provided for two hydro generating units (plus a 4000-kW station service unit), each rated at 43 200 kW, for a total generation capacity of 86 400 kW with substructures for four additional units. The subsequent installation, completed in 1943, provided eight additional generators of 54 000 kW each—to bring the present total generating capacity to 518 400 kW.

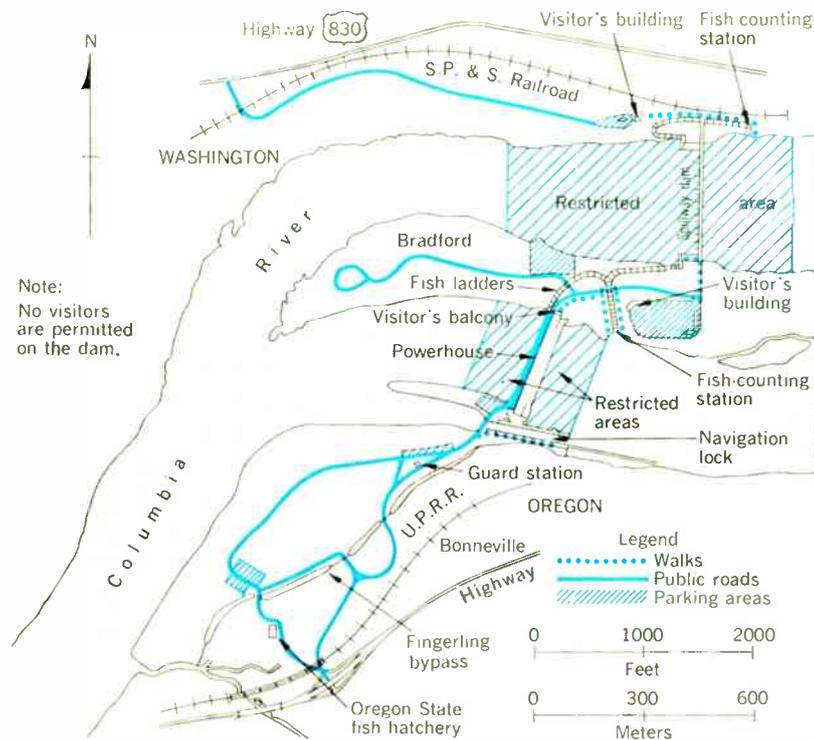
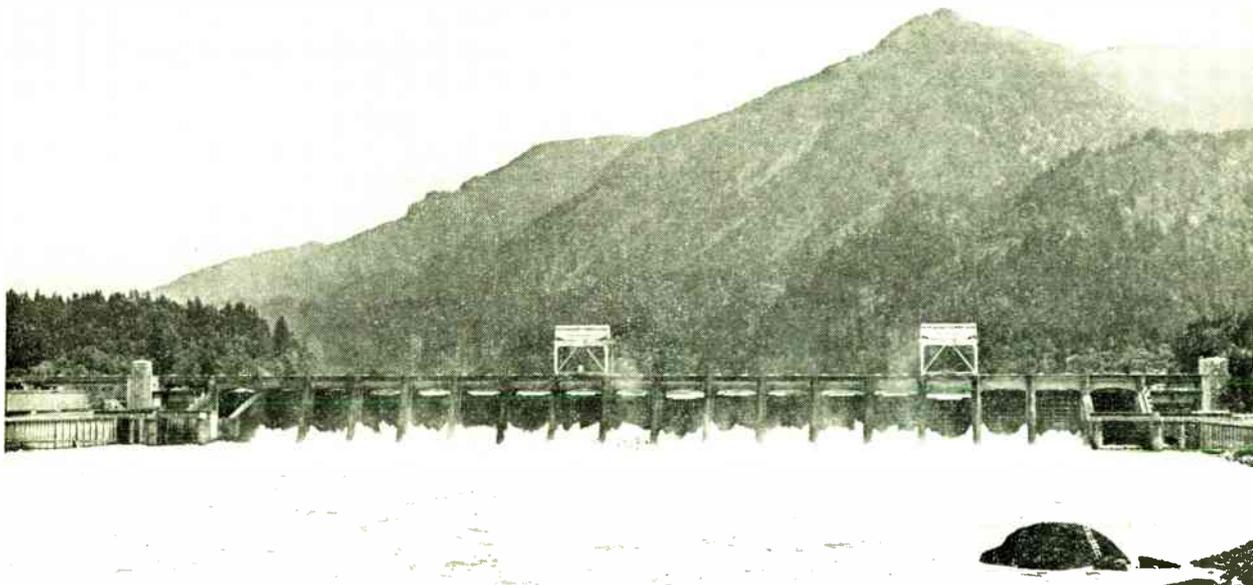


FIGURE 3. Map of the immediate area of Bonneville Dam and its appurtenances.

FIGURE 4. View of the spillway portion of the Bonneville Dam. This section forms a barrage between the Washington shore and Bradford Island.



I. Bonneville Dam statistics

Powerhouse

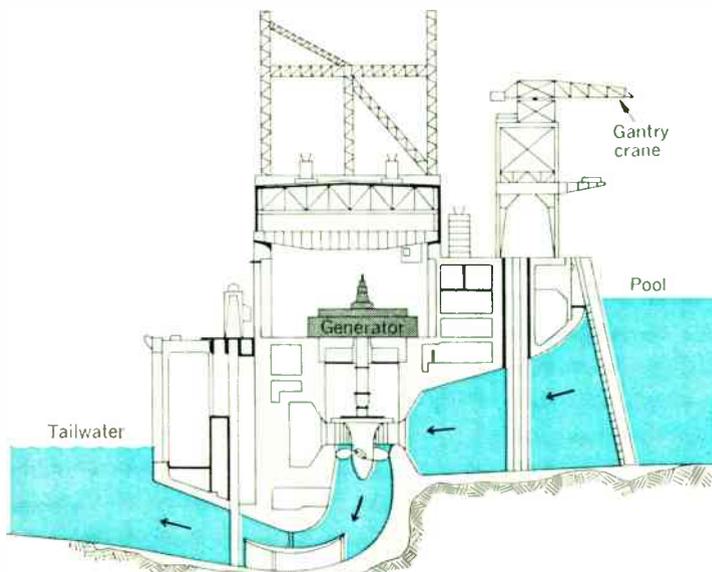
Length.....	313 meters
Width.....	58 meters
Height above lowest bedrock....	58 meters
Station service unit.....	4000 kW
Number of hydro-generating units.....	2 @ 43 200 kW 8 @ 54 000 kW
Total rated capacity.....	518 400 kW
Generators.....	13 800 volts
Generator housing diameter.....	14.6 meters
Transmission voltage.....	115 000 and 230 000 volts
Kaplan turbines (adjustable blade).....	2 @ 49 000 kW, 15-meter head 8 @ 55 000 kW, 18-meter head
Revolutions per minute.....	75
Discharge per turbine, in cubic meters per second....	386
Propeller.....	5 blades, 7.1 meters diameter

Navigation Lock

Length.....	153 meters
Width.....	23 meters
Lift.....	9.2 to 21.4 meters
Capacity.....	approx. 8000-tonne ship

The spillway dam, used primarily for river flow and navigation control, is situated between Bradford Island and the Washington shore (Fig. 4). The Bonneville Reservoir pool is normally maintained between elevations 72 and 74 feet (22.0–22.6 meters) above mean sea level, and it is regulated to provide optimum power and navigation benefits with minimum obstruction to the upstream movement of migrating salmon and other

FIGURE 5. A typical powerhouse cross section at a generator in Bonneville Dam. With minor variations in the forebay and tailrace construction, this section is applicable to most of the concrete low dam installations in the Northwest.



fish. Daily fluctuations in water level are generally small; however, in emergency situations, the drawdown may be as great as 61 cm (2 feet) per day. During the annual spring high water, elevation 74 is usually exceeded at the crest of the flood. In the upper reaches of the reservoir, significant seasonal fluctuations are experienced as the result of varying tributary river flows. The backwater effect of these streams increases with the magnitude of their flow and their distance upstream of the dam. Flood flows and resultant high backwater elevations normally occur during May and June.

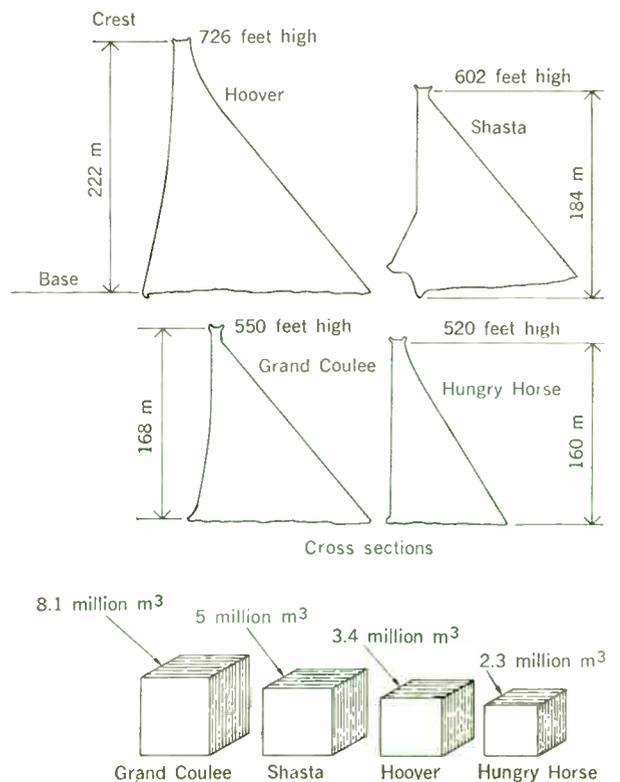
Table I gives the quantitative statistics for the powerhouse and navigation locks; and Fig. 5 shows a typical powerhouse cross section at a generator.

Bonneville Project Act: birth of BPA. On August 20, 1937, President Franklin Roosevelt signed the Bonneville Project Act, and Congress subsequently created the Bonneville Power Administration to act as the *marketing agent** for the power to be generated by the Bonneville Dam, starting in 1938. On November 1, 1937, James D. Ross was appointed as the first Bonneville Power Administrator.

The Bonneville Dam was dedicated in June 1938, and the BPA completed its first transmission line (13 800 volts) from the dam to the nearby town of Cascade Locks, Oreg., on July 9, 1938.

* All of the federal dams of the U.S. Columbia River System were built, and are operated, either by the Bureau of Reclamation or the Corps of Engineers. Table III gives statistics on all existing and proposed dams and their respective operating agencies.

FIGURE 6. Cross sections of the four largest concrete dams in the United States, two of which are in the Pacific Northwest. Cubes indicate a graphic comparison of the volumes of concrete in each structure.



Public vs. private power. Bonneville effectively rekindled the 40-year-old simmering feud between the public and private power interests. In 1937, the industrial, commercial, and domestic consumer loads were adequately served by the existing public and private utilities. Many of the more vocal critics snorted: "What the hell are we going to do with an additional 300 000 kW of electric energy? We won't need that much by the end of the 20th century!" Thirty years later, the power dams of the Columbia River Basin were generating more than 16 000 MW (about 50 percent of this at federal dams)—with no end to the ever-increasing demand in sight.

War, progress, and Grand Coulee

When Administrator Ross died in March 1939, he left to his successor, Dr. Paul J. Raver, a master plan for the future growth of the BPA system. In 1939 and 1940, the first 115 000-volt transmission lines were erected from the Bonneville Dam westward to the J. D. Ross Substation in Vancouver, Wash. (directly across the Columbia from Portland), and from the dam eastward to The Dalles. The Ross Substation subsequently became the "power junction" for southwest Washington, western Oregon, and the Willamette Valley.

II. Grand Coulee Dam statistics

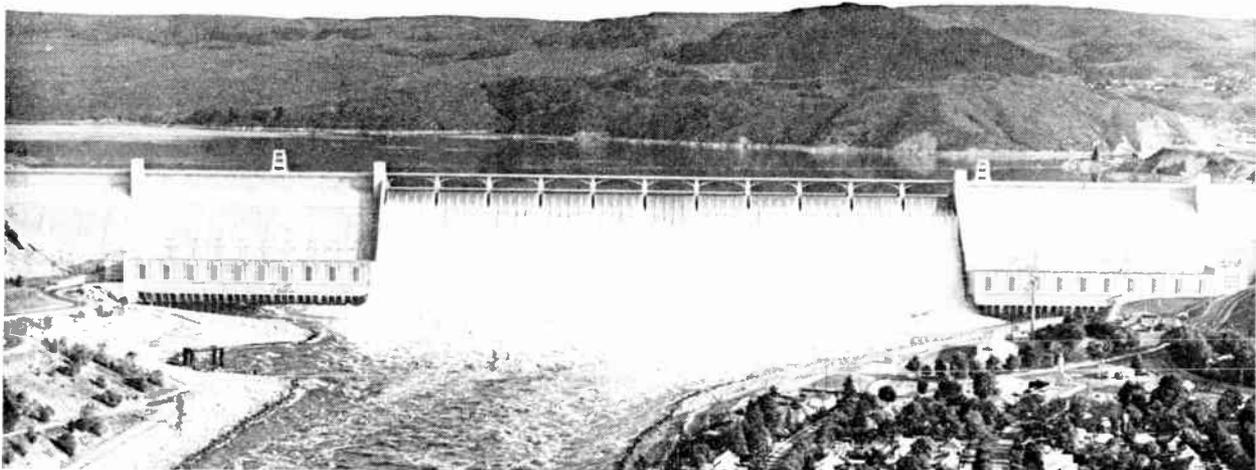
Length at crest.....	1272 meters
Volume of concrete (including appurtenances).....	8.1 million m ³
Length of reservoir.....	236 km
Max. depth of reservoir.....	153 meters
Total storage capacity.....	9.5 million acre-feet
Ultimate installed generating capacity.....	9435 MW

Meanwhile, World War II broke out in Europe and, in late 1940, a state of national emergency was declared in the United States. As part of our accelerated defense effort, work was speeded on meeting the power requirements for the expanding industrial complexes in the Pacific Northwest, and on the completion of the world's most ambitious hydro project up to that time . . .

The Grand Coulee Dam. This immense undertaking, regarded by skeptics as a white elephant and the greatest "boondoggle" of all federal projects, began in 1937 and was completed in 1942. To this day, in sheer volume of mass concrete—10.6 million cubic yards, or 8.1 million cubic meters (including the appurtenance structures)—Grand Coulee stands as the largest straight gravity dam in the world, and it will hold this distinction until the Aswan Dam, on the River Nile, is completed in the 1970s. Figure 6 shows the comparative volumetric quantities of the world's four largest dams and their relative crest heights. Note that Hungry Horse, another dam (to be discussed later in this article) of the Columbia River Basin system, is in this category. Like the other three in the figure, Grand Coulee is classified as a "high dam." Although its height is about 54 meters less than that of Hoover Dam, its crest is almost four times longer. Thus Coulee contains three times more volume of concrete than Hoover. Other construction statistics and reservoir information are contained in Table II.

Grand Coulee stands on a granite base, exposed in early geological times when the cordilleran icecap cut into the lava plateau to form the 500-meter-wide canyon in which the Columbia River flows (Fig. 7). The dam is situated 150 km northwest of Spokane and 390 km east of Seattle. The south end of this reservoir is 960 km from the mouth of the Columbia, and the normal surface elevation of the lake is 1288 feet (393 meters) above mean sea level. The water stored in this huge upper pool is used to irrigate more than 500 km

FIGURE 7. View of the downstream face of the Grand Coulee high dam and powerhouses. This facility, originally completed in 1942, will have an additional powerhouse by 1987, to bring its total generating capacity to more than 9400 MW.



of farmland, to regulate the flow of the Columbia River, and to develop a large block of electric energy that is used for local pumping and irrigation. This energy is also transmitted to distant load centers.

An extensive hydraulic distribution system lifts water needed for irrigation from the Columbia River and transports it through an 8-meter-wide concrete-lined flume for a distance of 96 km to the northerly ridge of the irrigable areas. South of the dam, a 44-km-long equalizing reservoir (which acts as a retention pool to store surplus water from the Columbia during flood season) was formed by earth- and rock-fill dams across an abandoned watercourse called the Grand Coulee.

The major portion of the electric energy produced at the dam is carried over BPA's 230-kV lines to substations near Spokane, Portland, and Seattle for distribution to industrial plants and to private and municipal power utilities of the Northwest Power Pool.

Power from Grand Coulee was first put on the line in 1942, when it was transmitted over 115-kV lines eastward to Spokane and westward, over the Cascade Range, to the Puget Sound area. This historic event actually represented the birth of the Northwest Power Pool, in which all of the power resources of the Pacific Northwest existing at that time were "pooled" for the use of industry in the massive war production effort of the United States. Bonneville and Grand Coulee Dams con-

tributed 26 billion kilowatt-hours as their share of this task during the war years.

The present power plant was completed in 1951. It contains 18 main hydro generating units, each with a nameplate rating of 108 MW (plus three station service generators, rated at 10 000 kW each), for a total nameplate generation output of 1944 MW. Each of the generators, however, is capable of a short-time maximum output of 120 MW.

A third powerhouse, presently under construction, will boost the existing generating capacity by 3600 MW. This installation will consist of six generating units, each with a nameplate rating of 600 MW. Authorized for future construction is the installation of six more 600-MW generators, plus six pump-turbine units (to be used in a pumped-storage operation) for peaking power that will generate 291 MW, for an eventual grand total generating capacity of 9435 MW at Grand Coulee.

The postwar era begins:

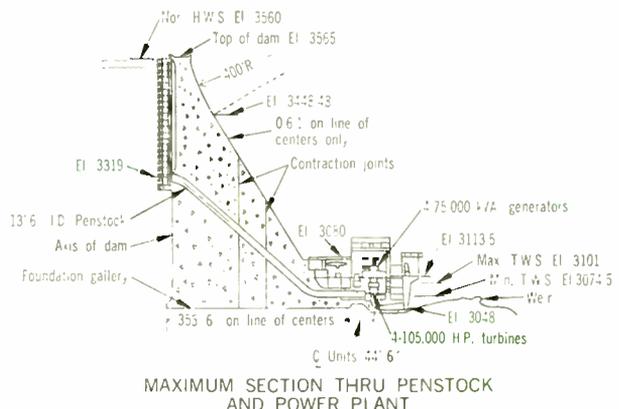
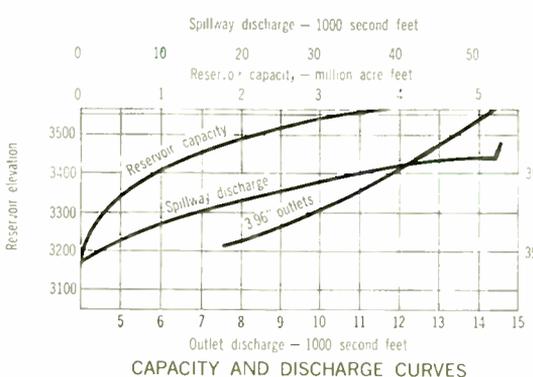
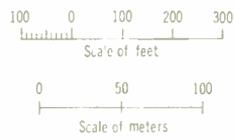
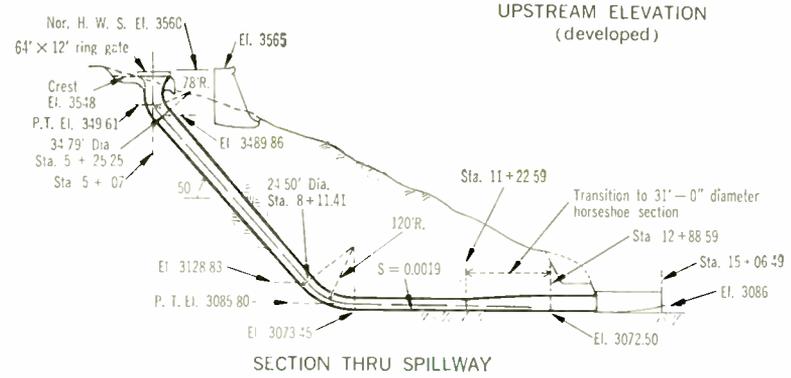
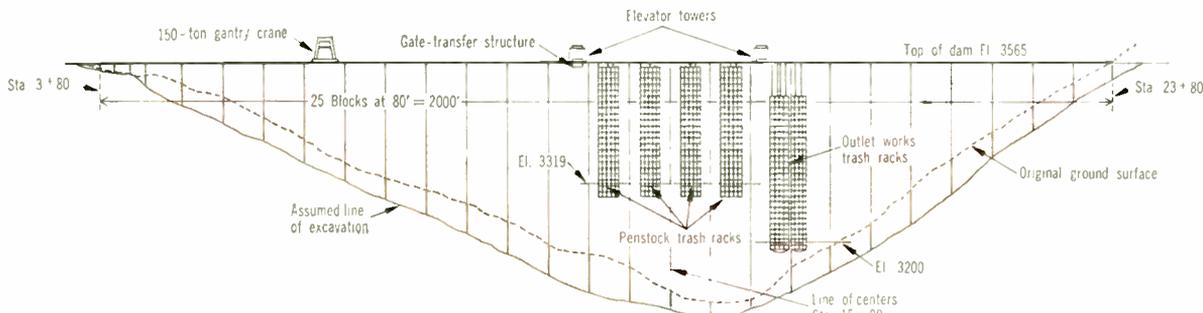
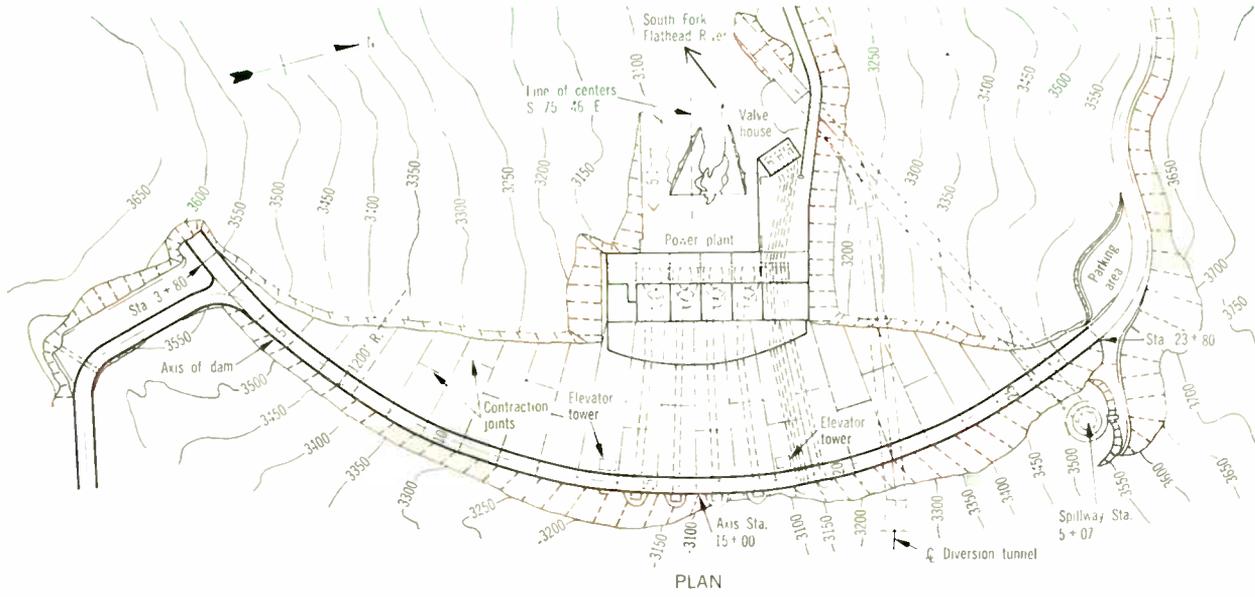
McNary and Hungry Horse Dams

McNary. Two years after the end of World War II, ground was broken for the construction of the Corps of Engineers' second Columbia River project, the McNary

FIGURE 8. Hungry Horse Dam, a graceful high dam structure of the arch-gravity type, contributes 285 MW of generated power to the BPA transmission system.



FIGURE 9. Construction plan, upstream elevation, and sections of the Hungry Horse Dam.



Dam. Situated 470 km upstream from the mouth of the river, this low dam impounds a 100-km-long pool whose normal surface elevation is 340 feet (104 meters) above mean sea level.

McNary was completed in November 1953, and power was put on the line (230-kV transmission) during the following month. The 435-meter-long powerhouse contains fourteen 70 000-kW generator units for a total (nameplate) generation capacity of 980 MW. The appurtenances of the 28-meter-high dam include a 400-meter-long spillway, and a 206- by 26-meter navigation lock, capable of lifting vessels 28 meters. The customary fish ladders were also constructed to permit the upstream migration of salmon and steelhead trout.

The cost of the McNary project was almost \$287 million.

Hungry Horse. This fourth great federal dam (Fig. 8) built in the BPA era is situated on the South Fork of the Flathead River near Kalispell, Mont. (see Fig. 2 map). A high dam (160 meters from base to crest), Hungry Horse differs from the Grand Coulee (a straight, gravity dam) in that it is of the arch-gravity type. This construction, in which the dam is curved upstream in plan (Fig. 9), not only offers maximum resistance to the sliding and overturning lateral forces of the impounded water (by means of the structure's great mass and weight), but additionally affords a large measure of inherent stability by transmitting a portion of the hydraulic pressure or load by arch action into the canyon walls. The arch-gravity design was selected because the site at Hungry Horse was intermediate between the narrow rock gorge condition that provides the ideal location for an arch dam, and the wide, flat valley terrain for which a gravity dam is more suitable.

Construction on Hungry Horse began in 1948 and was completed in October 1952. Built by the Bureau of Reclamation, it is a key project in the Interior Department's long-range program for multipurpose develop-

ment of the water resources of the Columbia River Basin. As a major upstream storage dam on the Columbia River system, it provides power benefits that extend from the Continental Divide westward almost to the Pacific Ocean. Operated in coordination with the downstream dams and power plants, the facility contributes more than 600 MW of prime power to the BPA transmission system.

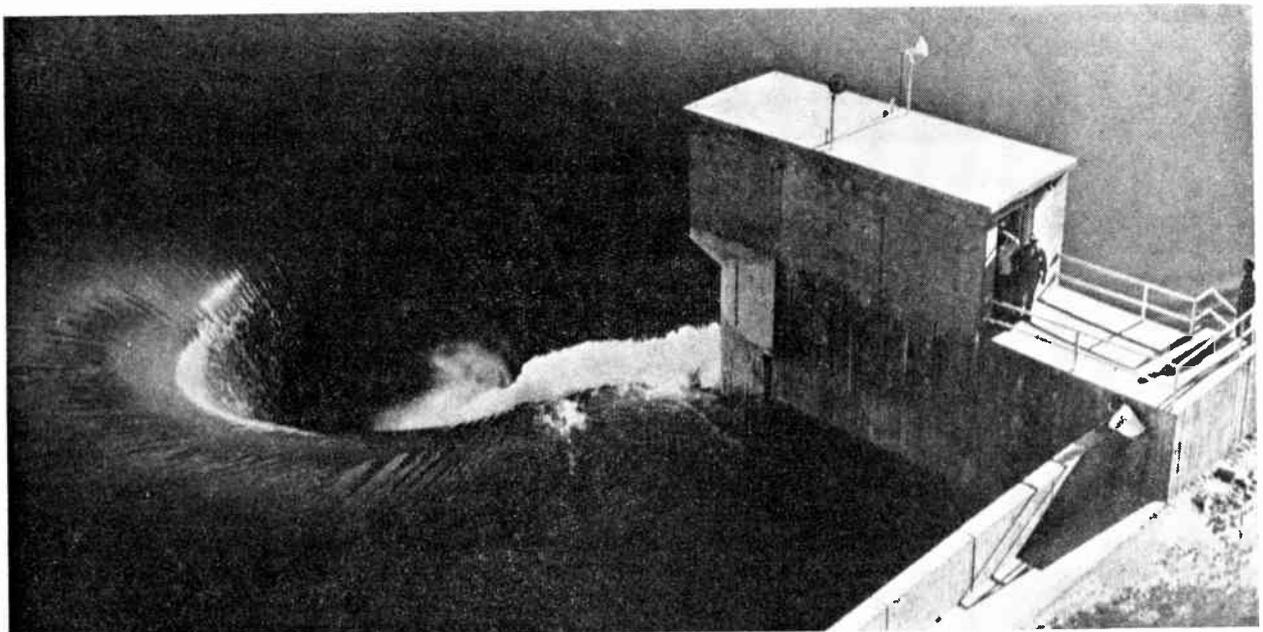
As one of its multiple-purpose features, the Hungry Horse Dam contributes materially in flood control on the Columbia River and its tributaries. It helps eliminate floods in the Flathead Valley and reduces peak discharges between there and Grand Coulee Dam by 10 to 25 percent, and at Portland, Oreg., by about 5 percent. Approximately two million acre-feet of the reservoir storage (acting as a basin runoff retention pool) can be used for flood-control purposes, when necessary.

The power-generating facilities (Fig. 8) are housed in a reinforced concrete structure, 120 meters long by 23 meters wide by 48 meters high, built athwart the river channel at the downstream toe of the dam. Four 4.1-meter-diameter by 137-meter-long penstocks carry water under pressure from the reservoir to the turbines. The maximum operating head (drop from the surface of the upstream reservoir to the river level below the powerhouse) is 147.5 meters. When the turbogenerators operate at maximum capacity, 73 tonnes of water pass through each turbine per second.

The power plant consists of four 71 250-kW (nameplate capacity) generators, with a total output of 285 000 kW. The power, generated at 13 800 volts, is stepped up to 115 and 230 kV for long-distance transmission by eight transformers located on the downstream deck of the powerhouse. High-tension lines carry the power from the transformers to the switchyard, situated 370 meters downstream.

A unique feature at Hungry Horse is the world's highest "glory-hole" (spiral) spillway (Fig. 10). Flood

FIGURE 10. "Glory-hole" spiral spillway is a unique feature in the Hungry Horse Dam's construction.



water from the reservoir, cascading over the circular rim of the spillway, drops a maximum of 150 meters through a 10.7-meter-diameter concrete-lined tunnel over a distance of 344 meters to the outlet portal. The maximum discharge capacity of the spillway is 1480 m³/s.

Other interesting dimensional statistics of the world's fourth largest dam are: length at crest, 645 meters; width at base, 100 meters; width at crest, 12 meters. The elevation of the crest of the dam is 3565 feet (1088 meters) above mean sea level. The total cost of the project was \$102 million.

The new decade:

Korea, and the demand for more power

The end of the 1940s witnessed the authorization of three new earth- and rock-fill multipurpose dams: Detroit and Lookout Point, in the Willamette Basin of Oregon (100 000 and 120 000 kW capacity, respectively); and Anderson Ranch (27 000 kW), on the South Fork of the Boise River, Idaho. Anderson Ranch was built by the Bureau of Reclamation; the others were constructed under the aegis of the Corps of Engineers. Earth- and rock-fill construction is often specified for smaller dams that are erected across narrow river gorges with a hydraulic head range of 60–90 meters.

In 1950 the Korean War began, and hydroelectric energy requirements for war material production soared to unprecedented levels. In that year, seven new dams were authorized by Congress: the low concrete dams at Albeni Falls, Idaho, and at The Dalles (Oreg.–Wash.). Construction of the Chief Joseph Dam in Washington was also approved. The Congressional action also provided for building the Big Cliff, Dexter, Chandler, and Roza earth- and rock-fill dams.

The 'brownouts' of 1952–1953. With the completion of Hungry Horse in 1952, 220 MW of additional hydro generation was made available for long-distance transmission in 1953. The at-site storage capability of the high dam also increased production at the federal and nonfederal

downstream dams by 832 MW. But even this large block of power, plus the generation of all available thermal plants in Utah and hydro power imported from British Columbia, did not fill the power deficiency caused by low winter streamflows and frequent periods of sub-normal temperatures.

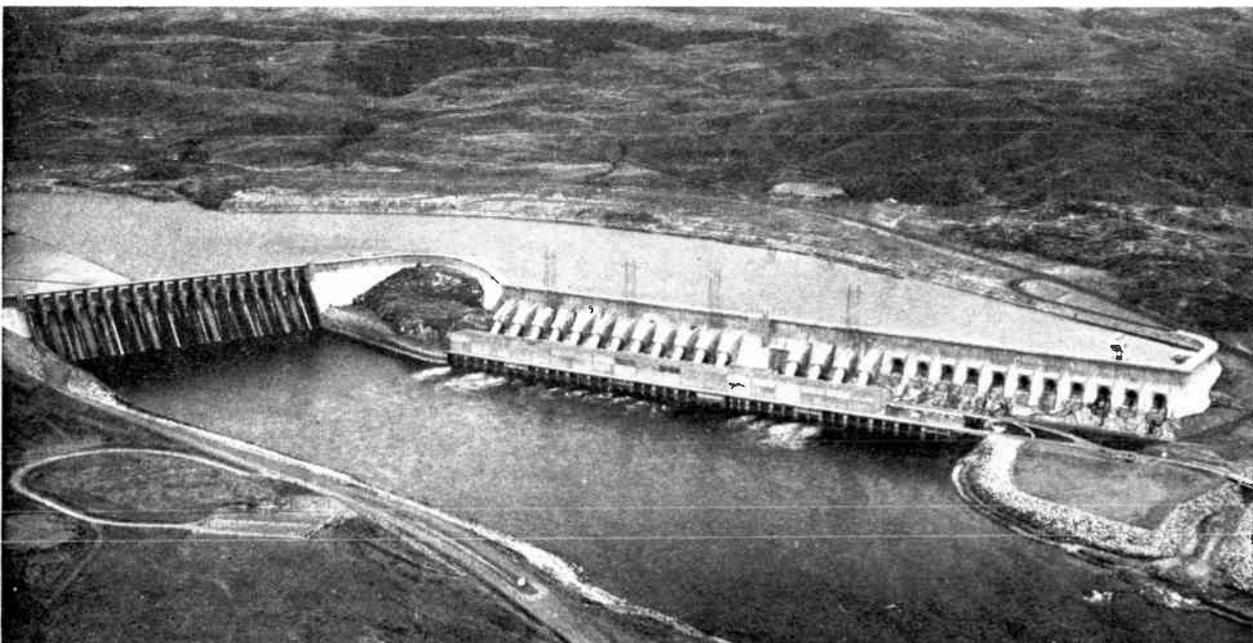
During the winter of 1952–1953, lights often dimmed during peak-load periods. Thus power to aluminum plants in the Portland area and heavy industry in the Seattle–Tacoma–Everett complex was curtailed to prevent the brownouts from affecting commercial business, farm, and residential customers. And the utilities requested their customers to limit their use of electricity, particularly during the morning and evening peak-load hours.

'Doldrums' of 1954; three nonfederal dams—and a new administrator

In 1954, new authorizations for federal dam projects came to a halt. Public agencies (public utility districts and municipalities) quickly filled the vacuum created by the lack of projected federal generation by requesting licenses for three large mid-Columbia hydro facilities—Priest Rapids (789 MW), Wanapum (831 MW), and Rocky Reach (712 MW)—to be built along the 450-km stretch between McNary and Chief Joseph.

On January 15, 1954, Dr. William A. Pearl, a former director of the Washington State Institute of Technology, became BPA's third Administrator. In that year a series of opinions from the U.S. Solicitor-General, based upon the power interchange provisions of the Bonneville Act, made it possible for a nonfederal utility to use BPA's interconnected regional power grid to transmit electric energy from an isolated generating facility to distant load centers. This so-called "wheeling" program brought three significant benefits to the Pacific Northwest:

FIGURE 11. Aerial oblique view of the Chief Joseph Dam and powerhouse. Note the 11 openings for the installation of future penstocks and generators.



1. Power from nonfederal projects could be marketed at distant load centers.
2. Costly duplication of transmission facilities was avoided.
3. Power could be transmitted at a lower cost.

In 1955, BPA completed the first EHV transmission system west of the Rocky Mountains, a 345-kV line to carry power from McNary Dam to Portland and Vancouver (Wash.).

Albeni Falls and Chief Joseph

The year 1955 saw the completion of the Corps of Engineers' concrete, gravity-type low dam at Albeni Falls on the Pend Oreille River of northern Idaho. At this relatively small dam, three hydro-generating units, each of 14 200-kW nameplate rating, contribute a total of 42 600 kW of generation—plus downstream prime power storage benefits—to the BPA power system.

The dam impounds the 110-km-long Pend Oreille river and lake reservoir to provide a normal usable storage of 1.15 million acre-feet. The maximum controlled pool elevation is 2062.5 feet (630 meters) above mean sea level.

Chief Joseph. This important Corps of Engineers' dam, named for the famous warrior chief of the Nez Perce Indians, is situated on the Columbia River about 80 km west of Grand Coulee, at a strategic point that takes full advantage of the hydraulic energy release from the high dam upstream. Chief Joseph is an L-shaped structure (see Fig. 11), with a 281-meter-long spillway section built athwart the mainstream flow of the river, and a 622-meter-long intake and powerhouse structure skewed at nearly a right angle (slightly obtuse) to the spillway and almost parallel to the stream flow.

Connecting the two sections is a horizontal arch abutment whose foundations bear on a solid rock outcropping in midstream. Another abutment section, at right angles to the powerhouse (shown in the lower right foreground of the illustration), serves to form an enclosed backwater pool for the sixteen 7.6-meter-diameter penstocks presently installed in the intake section. Provision has also been made for the future installation of 11 more penstocks and a corresponding number of generating units. The total length of the abutments is 410 meters.

The present power-generating equipment consists of 16 Francis-type turbines and 16 generating units, each with a nameplate rating of 64 000 kW, for a total generating capability of 1024 MW. Future plans call for the installation of 11 more units for an additional capacity of 704 MW. The rated hydraulic power head on the turbines is 50 meters.

Chief Joseph impounds an 82-km-long reservoir (Rufus Woods Lake), with a usable storage of 36 000 acre-feet. The normal surface level of the pool is 946 feet (289 meters) above mean sea level.

Transmission lines at 345 kV carry power 360 km westward from the dam (and also from Grand Coulee) to the principal load center in Seattle.

The cost of the Chief Joseph project from its inception to June 30, 1967, totaled \$144.4 million.

The Dalles Dam

The Dalles Dam, a major facility authorized in 1950, begun in 1952, and completed in 1957, is an important multipurpose project that provides

1. A 40-km-long slackwater pool for navigation upstream to the site of the John Day Dam (now nearing completion).

2. The necessary additional generating capacity to meet the present and future demands of the Northwest Power Pool, and particularly for the nearby Portland area.

3. Reduction in the pumping lift required for irrigation projects on lands along the reservoir.

4. Additional recreational facilities to area residents.

The design and construction aspects of The Dalles are unique in that the contiguous dam structures are of a heterogeneous type. The 37-meter-high, concrete, 23-gate gravity spillway section (Fig. 12) runs from the navigation lock along the Washington shore athwart the stream flow of the Columbia River to a right-angle intersection return with a 61-meter-high nonoverflow concrete gravity dam section that is integral and monolithic to the powerhouse portion, running parallel to the Oregon shore. A rock-fill closure section runs from the concrete barrage at an open angle, across the main channel, to the Oregon shore.

The Dalles Dam is situated about 145 km east of Portland and approximately 72 km upstream from Bonneville. The normal elevation of the reservoir pool is 160 feet (49 meters) above mean sea level. The pool drainage area (watershed) is 61 000 km². A single-lift navigation lock, measuring 26 by 206 meters clearance in plan dimensions, provides a normal vessel lift of 26.7 meters.

The initial power facilities consisted of 16 generating units with a total nameplate generating capacity of 1119 MW. Presently under way is the installation of eight additional generators, rated at 86 000 MW each, to bring the ultimate generating capacity of The Dalles powerhouse to 1807 MW.

Temporary dislocations, and a consolidation of power

Because new projects—federal and nonfederal—were not coming on the line in an orderly scheduled manner, BPA, in 1958, was confronted by a temporary surplus of firm power that could not be offered for sale; yet, it had to be held to meet the normal load growth of preference customers. And there were no short-term customers for this power. Thus the system began to waste large surpluses of secondary, or seasonal, power for want of a Northwest market. In the course of this crisis, BPA began incurring annual operating deficits for the first time in its history. Its surplus of \$80 million, accumulated in previous years, was reduced during this period.

In 1961, Charles F. Luce (presently chairman of the Board of the Consolidated Edison company of New York, Inc.) was appointed as the fourth BPA Administrator. And in that year, BPA, the Corps of Engineers, and nine private and public owners of hydro facilities made power history by the signing of a one-year coordination agreement to ensure maximum power production for all existing plants in the Columbia River Basin. By means of such an agreement, the supply of firm power was substantially increased.

In 1962, Congress authorized the construction of two new (Corps of Engineers) major dams—Little Goose (405 000 kW) and Lower Monumental (405 000 kW)—on the Snake River, and approved the BPA proposal for construction of secondary thermal generating facilities



FIGURE 12. Downstream face of the spillway section of The Dalles Dam. Powerhouse portion is visible at extreme right of photo. This Z-shaped contiguous structure features both concrete and earth- and rock-fill construction.

FIGURE 13. Ice Harbor Dam, showing the powerhouse section at the right, spillway in the center, and fish ladders and navigation lock to the left.



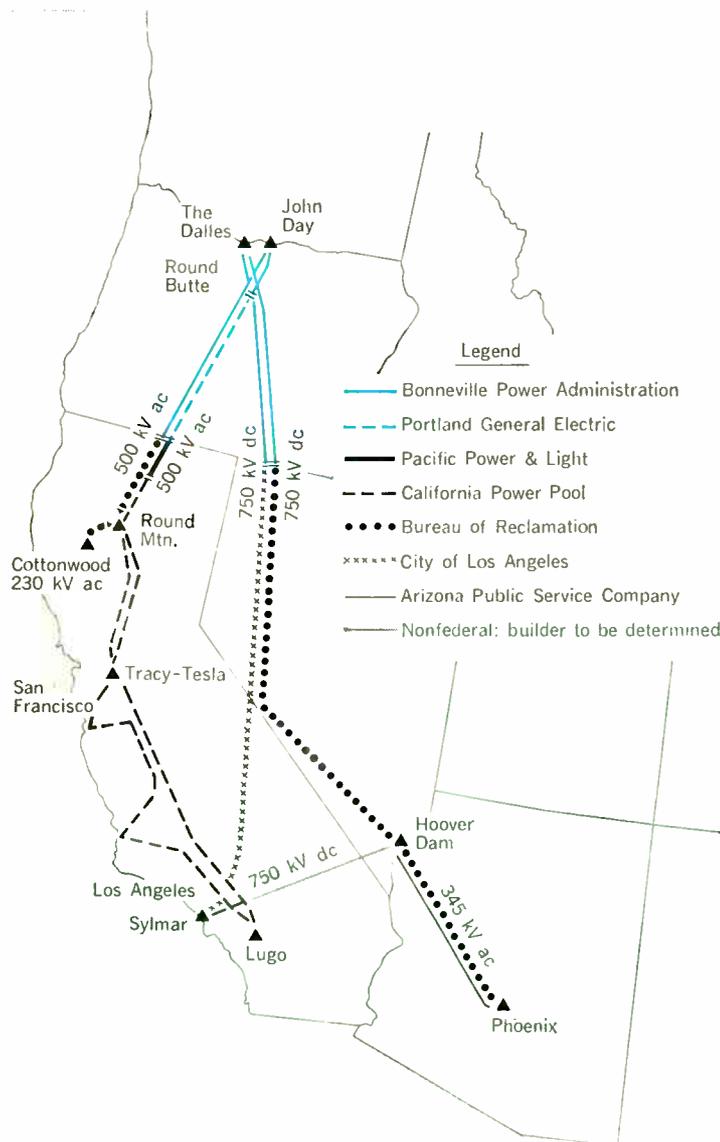
at the Hanford (Wash.) nuclear project of the AEC, to be built and operated by the Washington Public Power Supply System. In the following year, BPA became the power marketing agency for all of the Columbia River Basin federal power projects, including southern Idaho's Upper Snake River drainage area.

Four more dams for the sixties. In the six-year period between 1961 and 1967, one major concrete low dam, Ice Harbor (Corps of Engineers), was completed. Situated on the Snake 16 km from the confluence of the Snake and Columbia Rivers, Ice Harbor (Fig. 13) put an additional 270 000 kW from three generating machines on the BPA transmission lines. A projected future expansion will double this power output. The initial project also included a single-lift navigation lock (31-meter rise) to handle the two million tonnes of waterborne freight being shipped up the Columbia and Snake Rivers annually to Lewiston, Idaho.

This period saw the completion of three earth- and rock-fill power dams along Oregon rivers: Hills Creek (30 000 kW), on the Main Fork of the Willamette; Cougar (25 000 kW), on the South Fork of the McKenzie; and Green Peter (80 000 kW), on the Middle Santiam.

The Canadian Treaty. In 1964, The Treaty with Canada for Joint Development of the Columbia River was implemented. Under its terms, three large storage dams, to be built on the upper reaches of the Columbia River in British Columbia, will add 2800 MW of firm power at downstream dams in the United States, and also will provide much-needed flood control. One of these Canadian dams (Duncan Storage) has been completed and the other two—Arrow Storage and Mica Storage—are under construction. The sites of these projects are shown in the Fig. 2 map of the Columbia River Basin area. In accordance with the agreement, half of the power, 1400 MW, goes to each nation. The Canadian share, however, is tem-

FIGURE 14. Diagram of the eventual Pacific Northwest-Pacific Southwest Intertie, portions of which are presently under construction in a five-year program.



III. Columbia River power system as of June 30, 1967

Project	Operating Agency*	Location	Stream
Bonneville	CE	Oreg.-Wash.	Columbia
Grand Coulee	BR	Washington	Columbia
Hungry Horse	BR	Montana	S. Fk. Flathead
Detroit	CE	Oregon	North Santiam
McNary	CE	Oreg.-Wash.	Columbia
Big Cliff	CE	Oregon	North Santiam
Lookout Point	CE	Oregon	M. Fk. Willamette
Albeni Falls	CE	Idaho	Pend Oreille
Dexter	CE	Oregon	M. Fk. Willamette
Chief Joseph	CE	Washington	Columbia
Chandler	BR	Washington	Yakima
The Dalles	CE	Oreg.-Wash.	Columbia
Roza	BR	Washington	Yakima
Ice Harbor	CE	Washington	Snake
Hills Creek	CE	Oregon	M. Fk. Willamette
Minidoka	BR	Idaho	Snake
Boise Diversion	BR	Idaho	Boise
Black Canyon	BR	Idaho	Payette
Anderson Ranch	BR	Idaho	S. Fk. Boise
Palisades	BR	Idaho	Snake
Cougar	CE	Oregon	S. Fk. McKenzie
Green Peter	CE	Oregon	Middle Santiam
Foster	CE	Oregon	South Santiam
John Day	CE	Oreg.-Wash.	Columbia
Lower Monumental	CE	Washington	Snake
Little Goose	CE	Washington	Snake
Lower Granite	CE	Washington	Snake
Teton	BR	Idaho	Teton
Lost Creek	CE	Oregon	Rogue
Dworshak	CE	Idaho	N. Fk. Clearwater
Strube	CE	Oregon	S. Fk. McKenzie
Libby	CE	Montana	Kootenai
Asotin	CE	Wash.-Idaho	Snake

Total number of projects and installed capacity

*CE—Corps of Engineers; BR—Bureau of Reclamation.

†Nameplate rating.

‡Includes a recommended 3 600 000 kW at the third powerhouse in add

porarily being sold to utilities in the United States.

Finally, the Treaty gives the U.S. the right to build the Libby Dam on the Kootenai River in Montana, the reservoir of which will impound backing water 68 km into Canada. Libby will add 750 MW of firm power at the site and downstream in the United States.

In conjunction with the Canadian Treaty, the original one-year coordination agreement was extended 39 years and expanded to include BPA, Corps of Engineers, and 14 major generating utilities in the Pacific Northwest. This auxiliary agreement was necessary because the Canadian share of the power benefits was calculated on the basis of maximum coordinated international power output. Although voluntary coordination existed in the Northwest for many years, the new joint agreement assured that coordination would be extended and all facilities operated for maximum benefit to the region.

New construction to the present time

In 1965, the construction of power plants and transmission lines reached an all-time high: eight federal and four nonfederal dams were under construction; and, after years of preliminary study and planning, the Pacific Northwest–Pacific Southwest Intertie (500 kV ac and 750 kV dc), approved by Congress in 1964, were under construction. These lines, the largest single transmission program ever undertaken in the U.S., will eventually interconnect the power generation sources at The Dalles, John Day, and Hoover Dams (Fig. 14) with load centers at San Francisco, Los Angeles, and Phoenix, Ariz. This huge program will be discussed in more detail in the second installment of this article.

In 1966, Congress authorized the third power plant for Grand Coulee Dam. Completion of this installation will eventually make the dam one of the world's largest power producers (9435 MW). In the same year, David

Initial Date in Service	Existing		Under Construction		Authorized		Total	
	Number of Units	Total Capacity, kW†	Number of Units	Total Capacity, kW†	Number of Units	Total Capacity, kW†	Number of Units	Total Capacity, kW†
June 1938	10	518 400	—	—	6	324 000	16	842 400
Sept. 1941	18	1 944 000	6	3 600 000	6	291 000	36	9 435 000‡
Oct. 1952	4	285 000	—	—	—	—	4	285 000
July 1953	2	100 000	—	—	—	—	2	100 000
Nov. 1953	14	980 000	—	—	6	420 000	20	1 400 000
June 1954	1	18 000	—	—	—	—	1	18 000
Dec. 1954	3	120 000	—	—	—	—	3	120 000
Mar. 1955	3	42 600	—	—	—	—	3	42 600
May 1955	1	15 000	—	—	—	—	1	15 000
Aug. 1955	16	1 024 000	—	—	11	704 000	27	1 728 000
Feb. 1956	2	12 000	—	—	—	—	2	12 000
May 1957	16	1 119 000	8	688 000	—	—	24	1 807 000
Aug. 1958	1	11 250	—	—	—	—	1	11 250
Dec. 1961	3	270 000	—	—	3	270 000	6	540 000
May 1962	2	30 000	—	—	—	—	2	30 000
May 1909	7	13 400	—	—	—	—	7	13 400
May 1912	3	1 500	—	—	—	—	3	1 500
Dec. 1925	2	8 000	—	—	—	—	2	8 000
Dec. 1950	2	27 000	—	—	1	13 500	3	40 500
Feb. 1957	4	114 000	—	—	—	—	4	114 000
Feb. 1964	2	25 000	—	—	1	35 000	3	60 000
June 1967	2	80 000	—	—	—	—	2	80 000
	—	—	2	20 000	—	—	2	20 000
	—	—	16	2 160 000	4	540 000	20	2 700 000
	—	—	3	405 000	3	405 000	6	810 000
	—	—	3	405 000	3	405 000	6	810 000
	—	—	—	—	2	22 000	2	22 000
	—	—	2	49 000	—	—	2	49 000
	—	—	3	400 000	3	660 000	6	1 060 000
	—	—	—	—	1	4 500	1	4 500
	—	—	4	420 000	4	420 000	8	840 000
	—	—	—	—	4	540 000	4	540 000
	22	6 758 150	8	8 552 000	16	5 459 000	33	24 369 150

tion to that under construction, and 291 000 kW from six authorized pump–turbine units.

S. Black, a former FPC Commissioner, became BPA's fifth Administrator.

Power system automation. A comprehensive plan for the automation of the U.S. Columbia River Power System was initiated in 1967. The plan proposes to integrate thermal and hydro generation, irrigation, flood control, navigation, and other river functions to optimize the use of the region's total resources for all federal agencies and public and private utilities. System automation is part of a long-range plan, looking ahead to 1975 and 1985. A detailed discussion of the program will appear in Part II.

John Day Dam. To bring this portion of the first installment up to date insofar as dam construction is concerned, the writer must conclude this phase with John Day—the final key link in the important chain of lower Columbia River projects (title illustration).

The \$410 million, multipurpose project, now producing more than 300 MW from the first two units in operation, is essentially similar in basic design to McNary, 122 km upstream. Situated 350 km east of the mouth of the river, the dam impounds a 122-km-long reservoir that completes a 525-km slack-water pathway up the Columbia from the Pacific Ocean to a point above Pasco-Kennewick (Wash.), and thence up the Snake River to Lower Monumental Dam. Completion of the four approved dams on the Snake River—Lower Monumental, Little Goose, Lower Granite, and Ice Harbor (second phase)—will add 227 km to this navigable waterway and make slack-water navigation possible above Lewiston, Idaho.

The initial 16 generating units to be installed in the powerhouse each have a nameplate capacity of 135 000 kW, with an overload rating of 155 250 kW. Thus the John Day total capacity in the first phase will be a minimum of 2160 MW. Ultimately, a total nameplate generating capacity of 2700 MW will be available by the installation of four more units.

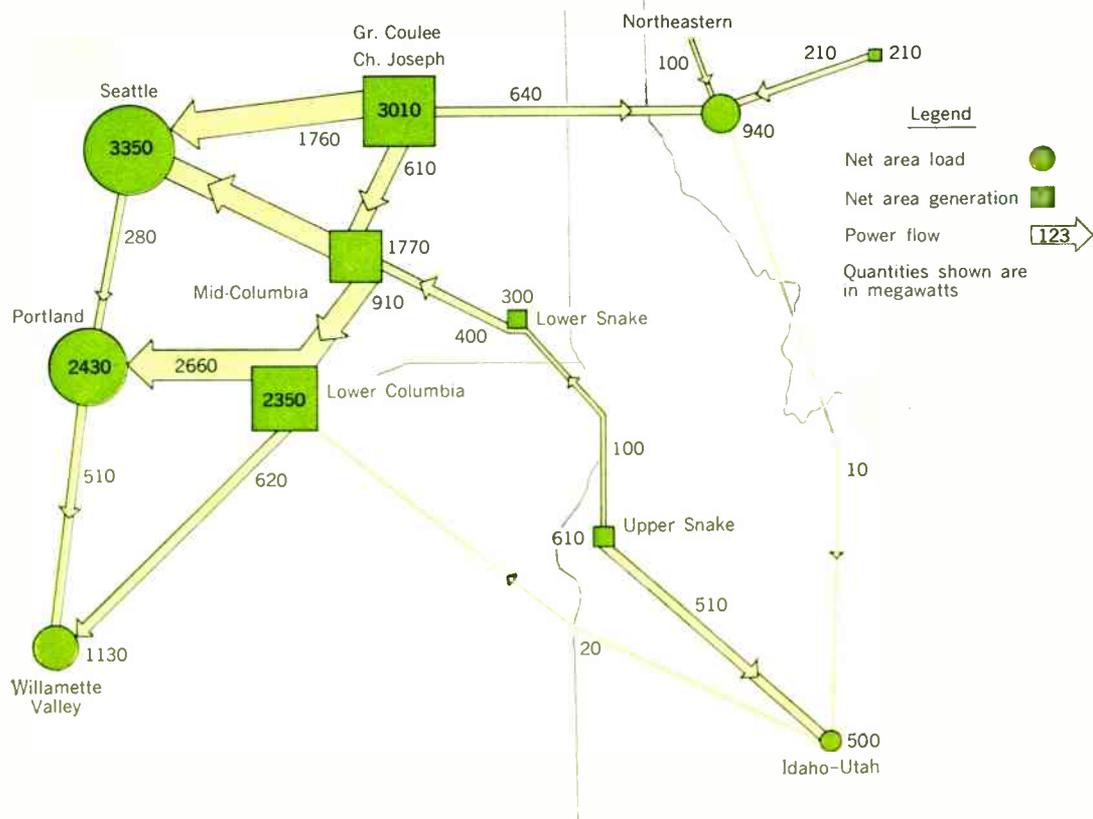
The 1800-meter-long straight gravity dam consists of a spillway and powerhouse section, running normal to the mainstream flow of the river from the navigation lock along the Washington shore to the fish-ladder appurtenances on the Oregon side. The 206- by 26-meter single-lift navigation lock will raise vessels 34.5 meters.

As previously noted, both John Day and The Dalles Dams have become particularly significant in the BPA system as the northern generating terminals for the 500-kV and 750-kV ac and dc transmission lines of the Northwest-Southwest Intertie.

Industrial growth and power marketing

The history of BPA is essentially interlocked and dependent upon the classical chronicle of supply and demand, coupled with the phenomenal growth of industry in the Pacific Northwest during the past 30 years. Today, it is estimated that 70 percent of the electric power purchased by industries in the Northwest comes from the large federal dams of the Basin. This electric energy is sold and delivered by the BPA, the marketing agent. A portion of this electricity goes to public and private utilities who resell it to their industrial customers. But

FIGURE 15. Diagram of power flows and load center demands in the Pacific Northwest as of January 1967.



the largest blocks of power go to the heavy electroprocess industries, such as aluminum-reduction plants, who take delivery of the power directly from BPA. Many of these plants were originally located in the Northwest for the primary reason of obtaining low-cost power from the dams as they were completed.

The Aluminum Company of America, for example, became BPA's first industrial customer shortly after power from Bonneville Dam was put on the line. When Alcoa completed its plant at Vancouver, Wash., in 1938, it initially received 26 000 kW—about 5 percent of the dam's generation.

By mid-1941, BPA had signed contracts with six companies to provide a total of 265 500 kW.

Industrial power sales have soared over the years. By May 1967, sales to major industrial customers reached 2140 MW. Of this quantity, 1698 MW are covered by 20-year contracts for firm power. The balance of the power is sold on an interruptible basis, contingent on stream-flows. By 1967, 30 percent of the primary aluminum produced in the United States came from Northwest smelters. In addition to aluminum, the products manufactured by the area's electroprocess plants include silicon carbide, ferronickel, calcium carbide, and ferromanganese. Today, 39 percent of BPA's power sales are made directly to heavy industry.

Load growth—present and future

To those who wondered, in 1937, what the Pacific Northwest would do with that 300 000 kW of electric

energy transmitted from Bonneville Dam, the example of power flow in the Pacific Northwest (Fig. 15) as of January 1967 may provide some answers! At that date, the load centers in Seattle, Portland, the Willamette Valley, western Montana, and Idaho—Utah were drawing a total of 8350 MW over BPA transmission lines from the hydro and thermal generation capacity of the Columbia River Basin facilities.

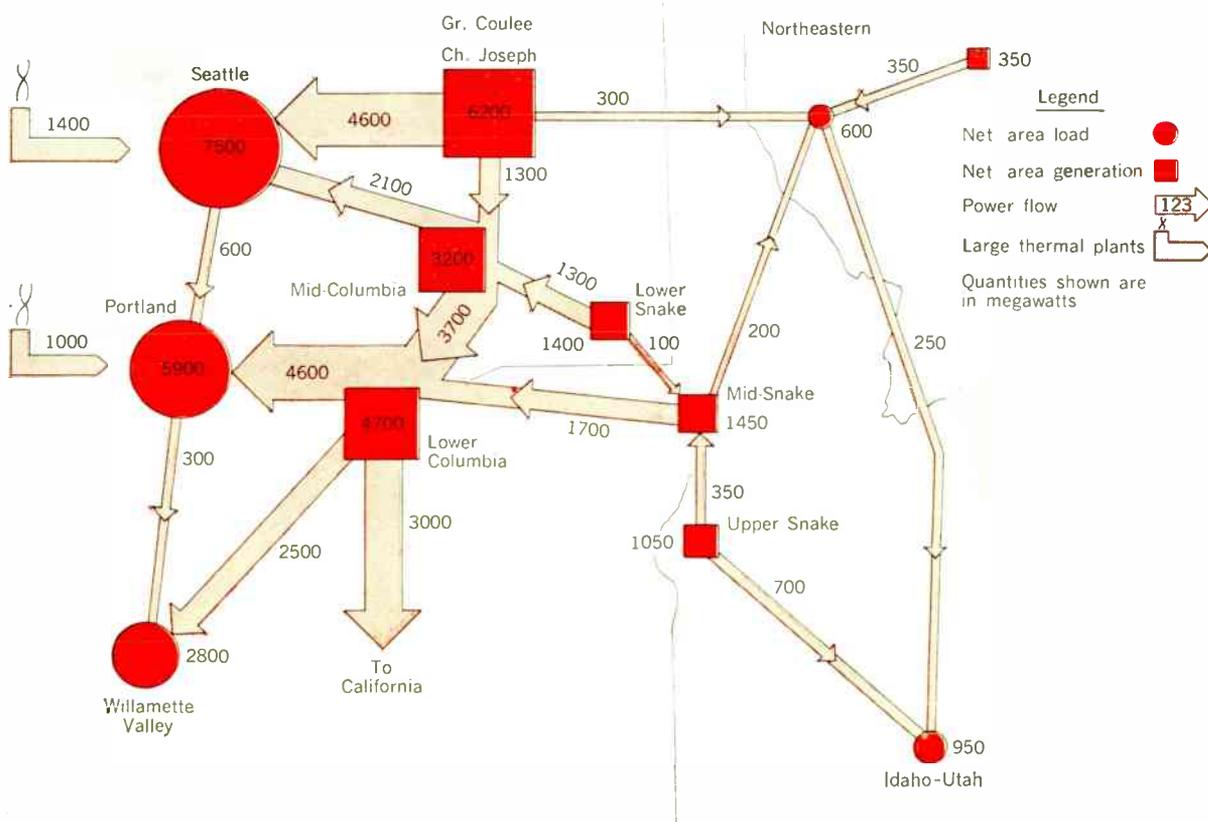
Table III shows the nameplate generating capacity of the federal dams of the Basin as of June 30, 1967, the generating capacity presently under construction, and the authorized future installations that will bring these facilities up to their ultimate capacity by the target date of 1987. Note that the deficiency of power generation between Table III and Fig. 15 is met by generation from nonfederal dams and thermal plants.

Figure 16 is a forecast of the power flows in the Northwest by January 1977. It includes 2000 MW in reserve from six new thermal (nuclear) plants that have been recommended for construction west of the Cascades, the transmission lines from which would be built as an extension of the existing BPA system. At this future date, a total of about 17 750 MW—more than twice the present demand—will be required by the same five load centers indicated in Fig. 15.

Finally, Fig. 17 indicates the power flows in the Northwest by 1987, when the projected ultimate generating capacity of the entire Basin (U.S. and Canada) will be realized.

The diagrams, however, should be considered only as

FIGURE 16. Diagrams of power flows and load center requirements in the Northwest anticipated by January 1977.



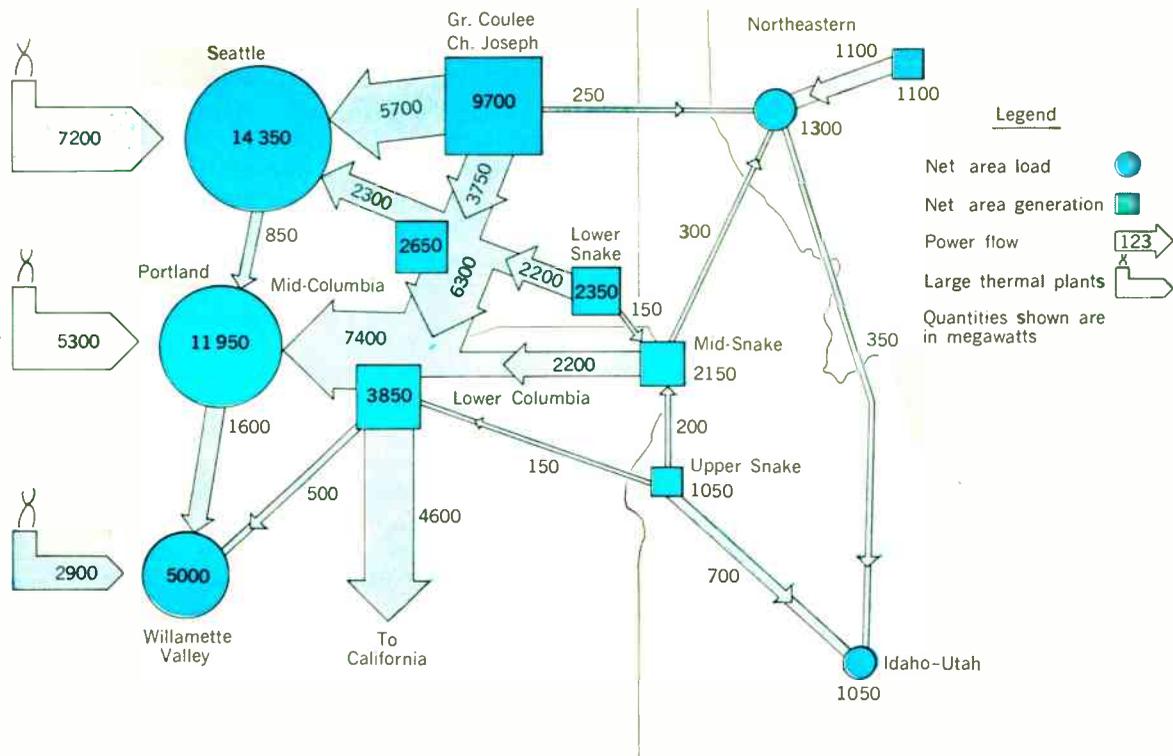


FIGURE 17. Anticipated power flows and load demands in the Northwest as of 1987. Large thermal plant sitings west of the Cascade Mountains are indicated.

representative examples, since any modifications in resource development schedules or load patterns will cause changes in the power flow quantities. Nevertheless, when one considers that projected peak loads will more than triple in the next two decades, one can see the necessity of going to higher-voltage transmissions, using existing rights of way whenever possible. BPA's 500-kV grid overlay, which permits the transmission of large blocks of power at a low unit cost, has been a major factor in maintaining its low rate schedule despite rising labor and material costs. The agency hopes that similar economies can be achieved by means of still higher transmission voltages, when the transmission system and the interconnected generating sources have been developed to the degree that the sudden loss of several thousand megawatts will not affect system stability. Studies are presently under way to determine the feasibility of utilizing transmission voltages up to 1000 kV.

Years of planning—centuries of progress

Like the grand-scale works of antiquity—the Great Pyramid, the Roman aqueducts, and the Gothic cathedrals—the mighty power dams of the Columbia River Basin should benefit mankind for centuries to come. Their creation was the result of years of coordinated scientific planning and study—meticulous engineering calculations concerning volumes of drainage and runoff, streamflows, historical meteorological data and long-range clima-

tic forecasts, future power demands, and the optimum hydro generation development possible in the vast watershed area. In addition, exhaustive geological and seismological analyses were required for the selection of the most strategic damsites, navigation locks, and other appurtenances. In recent years, network analyzers and computers have been of great assistance in accumulating large blocks of data for this gigantic undertaking.

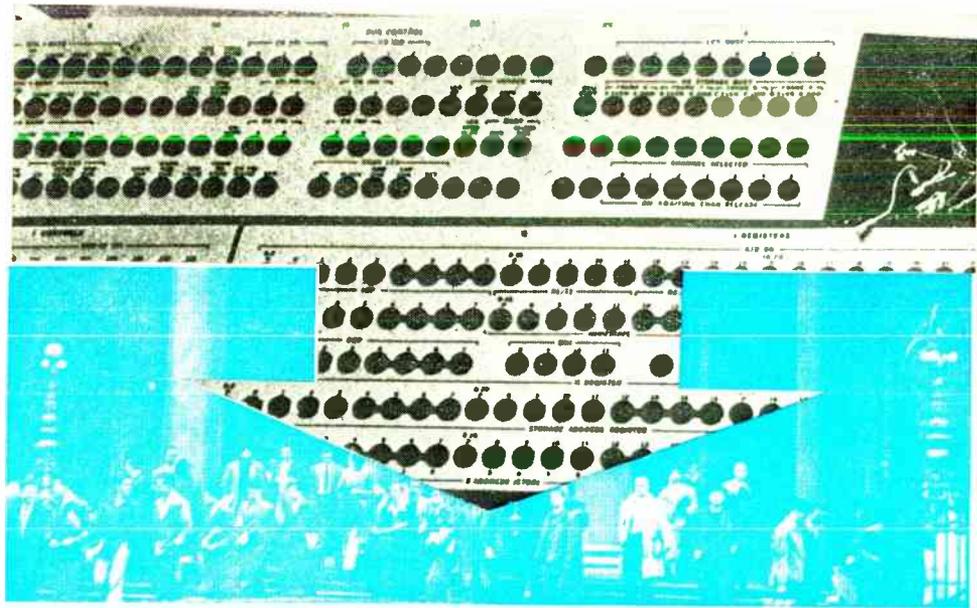
As in every long-term developmental project, the full potential use of available resources is inevitably reached. Thus there are just so many power dams, generator units, and transmission lines that can be installed or built. This saturation point, it is believed, will come by 1987—some 50 years after the creation of the Bonneville Power Administration.

Postscript to the present: a preview of Part II

In July 1967, Mr. Black was named Under Secretary of the Interior; H. R. Richmond was subsequently appointed as the sixth BPA Administrator.

The concluding installment of this article, to be published in the December issue, will discuss the advance programs of the BPA in ac and dc EHV transmission, the present and proposed interties, the construction of control and converter stations, the "hydromet data management system," and the role of future thermal plants as part of a combined and coordinated generation complex.

Photos, diagrams, and illustrations of federal dams used in this article are based on material supplied by the Bonneville Power Administration, U.S. Bureau of Reclamation, and the U.S. Army Corps of Engineers.



Continuing education for the engineering manager

Industry has found it necessary to implement continuing education programs of its own design to prevent educational obsolescence in upper-level engineering and technical personnel

Joseph M. Biedenbach Radio Corporation of America

Many universities feature postgraduate programs intended to keep engineering executives technically updated, but these have proved to be only partially successful. Engineers in managerial positions are engrossed in their careers and personal lives; they frequently find returning to a campus environment uninspiring, inconvenient, and sometimes even distasteful. Universities often find it financially unfeasible to design specific programs geared to the varied, constantly changing needs of different industries. RCA's answer to this problem is Current Concepts in Science and Engineering, a program designed exclusively for top RCA personnel, taught by experts in specialized fields from within the company, and employing new approaches to presentation designed to impart the most information in the least amount of time.

In order to minimize the educational obsolescence in their expanding technical staffs, industrial firms have developed independent programs of their own. Many companies are currently evaluating their top engineering personnel to determine the need to reinforce skills learned long ago but not used because of increased administrative responsibilities. Many of these men are near the top in salary scale and represent valuable experience within specific fields. Their worth represents a large investment

of themselves to the company, and vice versa. Informal methods of self-education, such as participating in lectures, meetings, and technical societies, as well as reading current periodicals, are necessary and helpful, but often inadequate.

A unique, modern continuing education program has been designed by RCA to help its engineering managers remain up to date technically. All the participants are employed by RCA, have recognized administrative ability, and have managerial potential within the company. They also have another common denominator: the need to update and to advance their technical information within generalized areas, instead of overspecializing in specific technologies. This need for generalization rather than specialization usually eliminates the typical evening graduate education programs available at most universities.

A new approach

The Current Concepts in Science and Engineering program was conceived early in 1963 by RCA Corporate Staff Personnel and Product Engineering. A program coordinating committee was formed with representation from all engineering locations throughout the company; and this group planned the program objective, format, and curricula. In September 1964 a pilot program was field-tested in the Camden, N.J., area.

The original class had 28 engineering managers and met biweekly for 12 weeks in two-day sessions. Since the program was enthusiastically received by the pilot group, plans were made to offer it on a broader base to engineering managers throughout the corporation who had been out of school for ten or more years. Approximately 600 managers have taken advantage of this program to date.

Establishing the program environment

Current participants are organized into classes ranging in size from 32 to 55 students. Each class meets for 12 two-day sessions held at three-week intervals throughout the academic year at a location offering overnight accommodations. Experience has shown that 24 days allows adequate treatment of a wide range of material yet does not seriously divert the engineering manager from his routine activities. It also has been determined that managers can break away from normal business duties for two-day periods to take part in an intensive training program without reaching their saturation limit on the learning curve. The three-week interval between class sessions can be used for reading assignments and other homework required of the participants.

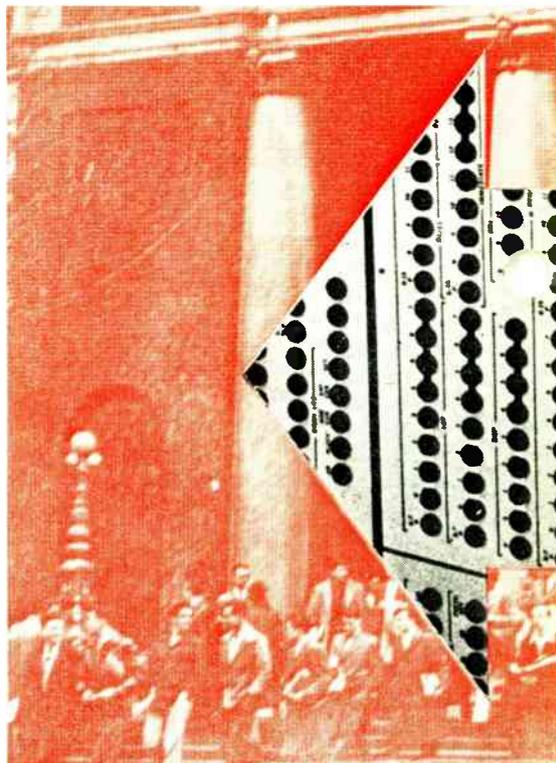
Approximately half of the CCSE students find it necessary to travel moderate distances and to remain overnight at the training center location during each session. They feel more than adequately compensated for this inconvenience by the benefits derived from informal discussion sessions with other course participants. During lunches, students are seated in small groups at individual tables. The informal conversations that result help managers become better acquainted while discussing the information learned in the class sessions.

Procedural format—a new approach

A significant amount of preparatory work precedes each class session in the CCSE program; each session is based on a combination of classwork and outside study assignments. Class activities include lectures, films, class discussions, video-taped lectures, and an occasional workshop period. Study guide material for each class session is developed and distributed to the students two weeks prior to their arrival. Homework assignments are given to the group following each session and students are encouraged to complete these assignments. Approximately half the participants make extensive use of the homework.

The standard classroom used for the program is equipped with an electronic lectern flanked on each side by an overhead projector, a motion picture projector, and a 35-mm slide projector. All lecture sessions are tape recorded; and the tapes are used by the staff to help the speaker to evaluate his presentation prior to the next lecture. The overhead projectors are satisfactory for use before large groups of 30 to 50 students seated at tables. A continuous roll of clear acetate film enables the lecturer to use the projector as a chalkboard while facing his audience. A copier is available in the room so that instructors can make overhead projection transparencies quickly in order to answer questions or cover an unexpected topic.

The staff has found that a great deal of class time can be saved by the effective use of audio-visual aids in the form of 8- by 10-inch (20- by 25-cm) overhead projection transparencies, and by distributing notes pertaining to



the lecture before the session begins. Study guide material is prepared so that students need not be distracted by note-taking while the lecture is being given. Textbooks are distributed to each participant during the course. When the course is completed, each man has about 12 textbooks and a study guide for each session of the CCSE program.

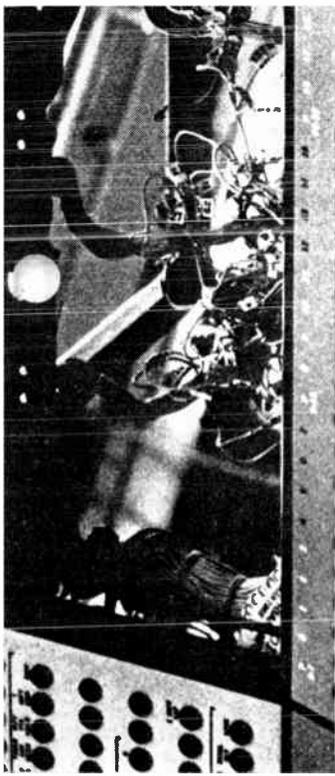
Syllabus and presentation

The subject matter for the CCSE program is organized into five basic courses conducted simultaneously: modern physics; semiconductors; electronics engineering; computers and their utilization; and applications. During each of the two-day sessions one or two topics, from at least four of the five courses, are developed. The students welcome the change in pace provided by moving from one subject area to another.

The modern physics portion of the course is designed to enable the manager to "talk the same language" as the young graduates who join RCA's engineering groups each year. Much emphasis is placed on quantum mechanics at the atomic and subatomic levels.

The semiconductor section of the program is typical of the depth and breadth of the material presented. Topics covered include crystallography, band theory, electrons and holes in solids, transport properties and elements of statistical mechanics, junctions, physics of semiconductor devices, transistors, integrated circuits, linear integrated circuits, and integrated-circuit applications.

Each of these topics is presented by the lecturer as an entity in and of itself. For example, the crystallography lectures concern the elastic properties of crystals, chemical bonding, X-ray diffraction and crystal planes, the crystal-



line structure of matter, states of matter, and properties and microscopic theory of the solid state. Using Holden's book *The Nature of Solids*¹ and Sproull's *Modern Physics*,² the participants review the band theory of solids, the free-electron theory of metals, and the properties of conductors, insulators, and intrinsic semiconductors.

Included in the theory of transistor operation are topics dealing with grown-junction, alloy-junction, drift-field, MADT, mesa, and planar transistors.

Applying mathematics appropriately

Each of the topics is presented at the mathematical level appropriate to graduate engineers. Although the program initially had a section devoted to pure mathematics, this was later dropped. Mathematical topics are retained in the program but are treated in the context of material presented by lecturers in the physics and engineering portions of the course.

Review material in mathematics, as well as in physics and basic electronics, is mailed to each registered student three months in advance to prepare him for the program. During the course, topics in calculus and elementary differential equations are reviewed as needed by the various speakers to assure the best possible "impedance match" with the audience. By diffusing the mathematical concepts throughout the course, presenting them as needed, the participants' interest and motivation are enhanced; and their comprehension of the pertinent application of mathematics to specific problems is reinforced.

The electronics engineering section of the course places much emphasis on digital computers. Fortran is taught to the group by a "workshop" method, with the partici-

pants writing Fortran programs that are processed subsequently by a computer during intervals between class sessions. Solid-state devices and switching circuitry, information theory, microwave solid-state devices, and electromagnetic theory are some of the topics discussed.

The portion of the program concerning applications enables the staff to interrelate the theoretical aspects of physics, semiconductors, computers, and electronics engineering topics presented in the various courses. Managers from different areas of specialization are able to develop an awareness of the interrelations that exist between one discipline and another. Lecturers who present the applications topics derive the illustrations of applied sciences from their own work experience.

The final analysis

Educational programs with objectives similar to those of the CCSE project have been established in several educational institutions, such as U.C.L.A. and the Polytechnic Institute of Brooklyn, and by several other industrial corporations. The CCSE program differs from the others in that

1. Company resources are used for most lecturers, enabling many competent RCA scientists and engineers, who are currently working on the "cutting edge" of technology, to present their technical specialties to RCA management.

2. A definite attempt is made to conserve the participants' time. Students are relatively high-salaried executives who have advanced management positions; and the time they have available for training is limited. Class schedules are arranged to minimize disruptive effects upon participants' normal work routine.

3. Teaching is reinforced by many carefully prepared audio-visual aids produced well in advance of the actual class sessions to help conserve time. Overhead projectors are used extensively in lieu of typical blackboard presentation.

4. The whole program, conceived from an engineering design viewpoint, is dynamic in character. Evaluation techniques are applied to every session by the lecturers, administrators, and the participants themselves in order to constantly improve and update the curriculum. Lecturers frequently incorporate changes in their presentations on the basis of these evaluations.

The Current Concepts in Science and Engineering project, to date, has attained the various goals proposed for it by the CCSE Planning and Coordinating Committee. It has enabled the RCA executive to update his technical knowledge, recognize major technical trends affecting future company business, and appreciate the significance of the major unifying concepts common to many scientific and engineering disciplines. The CCSE participant has been able to acquire a greater knowledge of RCA resources while simultaneously widening his circle of acquaintances within company engineering management. Finally, he has been able to develop a knowledge of and confidence in new engineering methods and technological approaches, especially those based on the increasing use of computers.

REFERENCES

1. Holden, A., *The Nature of Solids*. New York: Columbia University Press, 1965.
2. Sproull, Robert L., *Modern Physics: the Quantum Physics of Atoms, Solids, and Nuclei*, 2nd ed. New York: Wiley, 1963.

Phased arrays for radars

The development of phased arrays, in various configurations, has resulted in some unique radar capabilities, in that it is now possible to achieve fast, inertialess scanning and reshaping of the radiated beam

T. C. Cheston

The Johns Hopkins University Applied Physics Laboratory

An array antenna can enlarge the capability of a radar system by contributing beam-pointing agility as well as providing a way of handling large power. The fundamental theory of such antennas was outlined in IEEE Spectrum exactly four years ago.¹ Advances in the art have been so impressive that this month's Proceedings of the IEEE is a special issue on Electronic Scanning. The present article discusses the principles of operation and the most general applications of one very important class of array antennas.

Early radar systems used antenna arrays formed by the combination of many individual radiators. As radars progressed to shorter wavelengths, arrays were displaced by simpler antennas, such as parabolic reflectors. For modern radar applications the development of simple electronically variable phase shifters has once more directed attention to arrays where now the phase of each element is controlled by electronic means; these phased arrays permit fast, inertialess scanning of the radiated beam in three dimensions.

With the capability of rapidly and accurately positioning beams, it has become possible to develop radars that perform multiple functions interlaced in time. A phased-array radar may track a great number of targets while also carrying out a search for new targets. It may perform further functions such as illuminating targets and directing missiles toward them or acting as a communication system directing high-power beams toward distant receivers and transmitters. The phase shifters may be freely adjusted to broaden the radiated beam pattern to search some areas more rapidly, analogously to a zoom lens that retreats to show a wider view. Within certain bandwidth limitations, frequency is a free parameter and may be varied at will. This is in contrast to frequency-scanned

radar antennas in which the beam is scanned by the use of enhanced frequency-sensitive characteristics. Phased arrays can give radars the flexibility needed to perform all the various functions in a way best suited for a particular situation, within the constraint set by the total use of time. The functions may be programmed adaptively to the limit of one's capability of exercising effective control.

Phased arrays consist of an aperture filled with many individual radiating elements. Each radiator is connected to a phase shifter where the active phase control is exercised; finally, a feed network connects the transmitter or receiver to the phase-shifter section.

Typical phased-array radars have pencil beams some one to three degrees wide, requiring, respectively, about 10 000 and 1000 radiating elements and phase shifters. An array of this type may be scanned to a maximum value of perhaps ± 60 degrees; therefore, three or four arrays are necessary for hemispherical coverage. Figure 1 shows how four such phased arrays may be arranged on a ship to permit a view that is unimpeded by the central superstructure.

The frequencies used by phased-array radars primarily extend from 1000 to 10 000 MHz. The physical size of the array is inversely proportional to the product of beam width and frequency. In the S band (3000 MHz), a one-degree beam requires an aperture diameter of about 8 meters.

Principles of operation²

The phased-array radar antenna usually has a planar radiating aperture that has been assembled from a great number of similar radiating elements, such as slots or dipoles, each element being individually controlled in phase. It is designed to generate a narrow, electronically

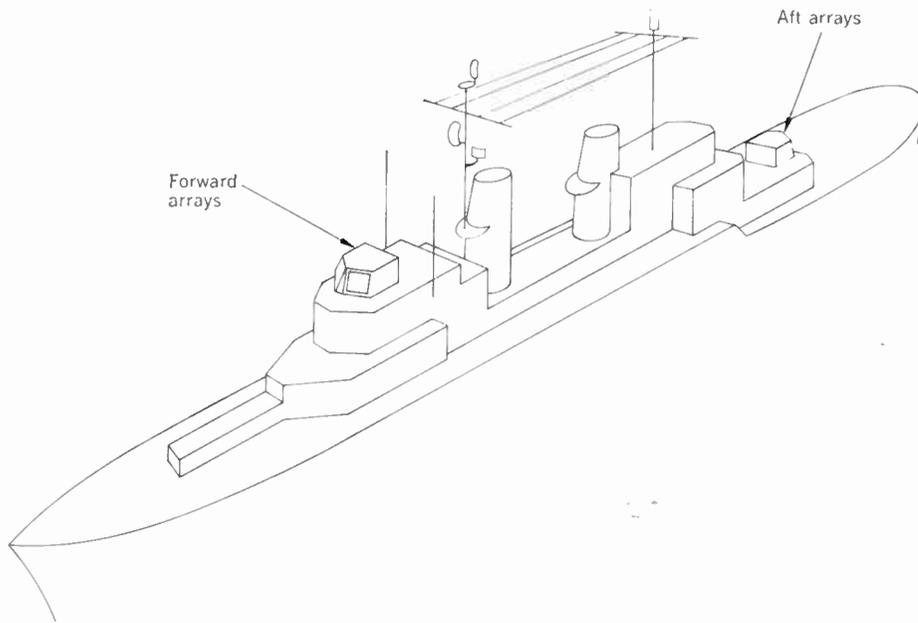
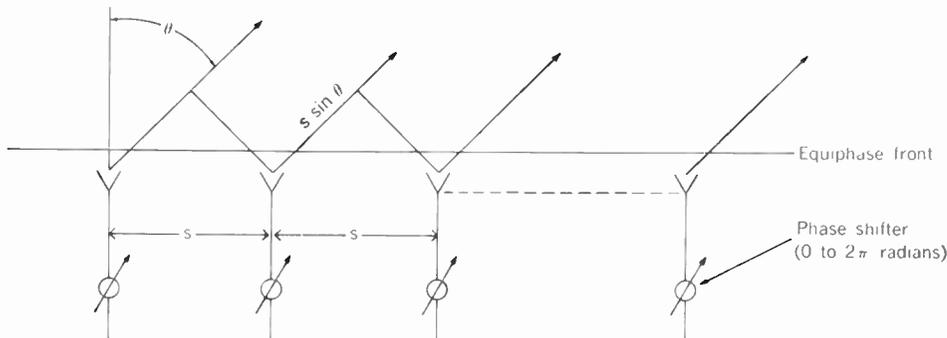


FIGURE 1. Shipborne phased-array antennas.

FIGURE 2. A linear phased array.



steerable beam to resolve targets in angle and to supply antenna "gain"—the gain being the ratio of power in the direction of the beam maximum to the power that would be there if the antenna radiated with equal intensity in all directions. The antennas are capable of operating with high efficiency and giving accurately calculable beam pointing.

For ease of explanation, a line array will be considered, as shown in Fig. 2, with a number of phase shifters adjustable between 0 and 2π radians. The radiators are spaced equally by a distance s . The system is

assumed to be transmitting although, by reciprocity, the results may be applied equally for receiving. If all the radiators are excited in phase, so that there is an equiphase front parallel to the aperture, then a radiation maximum will occur along the direction normal to the aperture (broadside), since at a distance in that direction the contributions from all the elements will add in phase. In some other direction θ , the contributions from the elements will add with a relative phase that changes from element to element and that is determined by the additional distance $s \sin \theta$ that has to be covered. The corre-

responding phase change between elements is $(2\pi s/\lambda)s$ in θ radians, where λ is the wavelength. The vector summation of the contributions of all the N radiators as a function of the angle θ gives the radiation pattern. With constant amplitude distributed to the elements, this radiation pattern is

$$E(\theta) = \frac{\sin [(N\pi s/\lambda) \sin \theta]}{N \sin [(\pi s/\lambda) \sin \theta]}$$

With many closely spaced elements the pattern has a half-power beam width θ_B of

$$\theta_B \approx \frac{50\lambda}{a} \text{ degrees}$$

where $a = Ns =$ length of aperture.

The radiation pattern or beam of the antenna may be steered to an angle θ_0 by applying linearly progressive phase increments from element to element so that the elements will all add in phase in the direction θ_0 . This requires that the phase shift between adjacent elements compensate for a difference in distance of $s \sin \theta_0$; that is, a phase shift (modulo 2π) of $(2\pi s/\lambda) \sin \theta_0$ radians is needed. The radiation pattern then has the form

$$E(\theta) = \frac{\sin [(N\pi s/\lambda) (\sin \theta - \sin \theta_0)]}{N \sin [(\pi s/\lambda) (\sin \theta - \sin \theta_0)]}$$

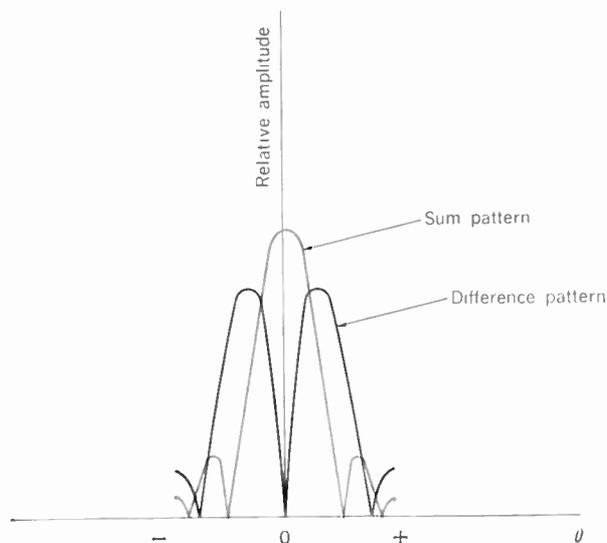
With many closely spaced elements, the half-power beam width is then

$$\theta_B \approx \frac{50\lambda}{a \cos \theta_0}$$

Scanning has increased the beam width by the cosine of the scan angle; that is, the effective aperture size has been reduced to its projection onto the equiphase front.

An examination of the equation for the scanned radiation pattern shows that several maxima can occur. These additional beams, called "grating lobes," occur when the denominator goes to zero. The desired suppression of these beams is achieved by spacing the radiating elements by about $s = \lambda/2$. When $s = \lambda/2$, with N radiating elements in a planar array forming a pencil beam, the

FIGURE 3. Sum and difference radiation patterns.



half-power broadside beam width θ_B is related to N by

$$N = \frac{10\,000}{(\theta_B)^2}$$

or

$$\theta_B = \frac{100}{\sqrt{N}} \text{ degrees}$$

and the antenna gain $G(\theta_0)$ of the array with the beam scanned to θ_0 is proportional to N and varies with the projected aperture as follows:

$$G(\theta_0) = \pi N \cos \theta_0$$

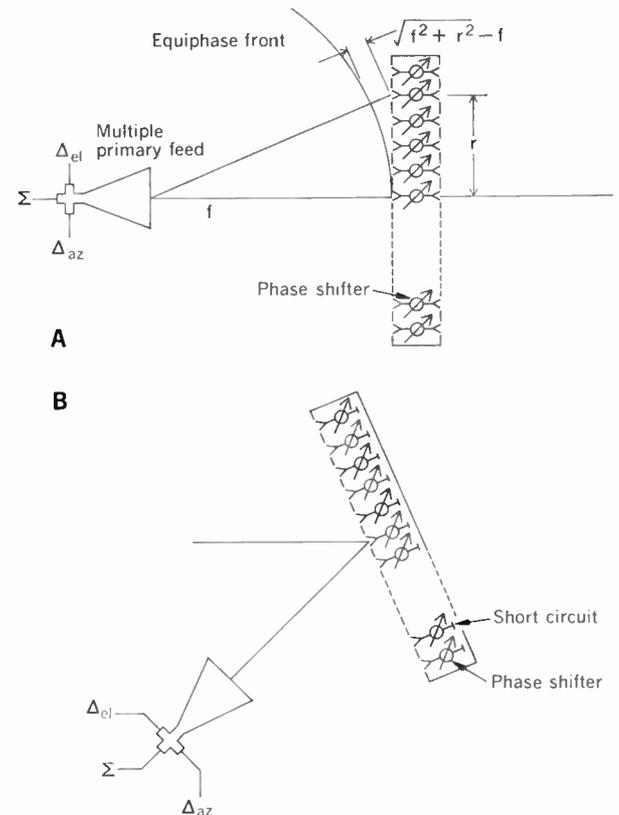
A one-degree beam has about 10 000 elements and a broadside gain of about 30 000. When scanning is to ± 60 degrees, the beam width in the plane of scan has been widened to about 2 degrees and the gain becomes 15 000.

When the beam is scanned, the required phase between elements is $(2\pi s/\lambda) \sin \theta_0$. This expression is seen to be wavelength-dependent, and is therefore frequency-dependent. For a given setting of the phase shifters, the beam scans toward broadside as the frequency is increased, and the bandwidth is limited. With scanning to ± 60 degrees, a reasonable value of this limitation is given by

$$\text{Bandwidth (percent)} = \text{Beam width (degrees)}$$

where the beam width is the broadside ($\theta_0 = 0$) beam

FIGURE 4. Optical feed systems. A—Lens. B—Reflector.



width of the antenna. If the frequency is changed sequentially, the phase shifters may then be reset for each frequency to give the correct beam-pointing direction.

Feed networks

Requirements. Various methods of feeding the phase shifters and radiating elements are possible. When the radiation pattern was derived, it was assumed that all elements were excited with equal amplitude. The resulting field-strength pattern was of a form that approximated a $(\sin x)/x$ function, with sidelobes whose maxima reach 21 percent (-13 dB) of the peak of the beam. Smaller sidelobes are often desired, and may be obtained from a feed system in which the amplitude distribution is tapered to some lower value at the edges. For the radar-tracking functions on receiving, a different amplitude distribution is desired, together with a phase reversal in half of the aperture. This gives rise to a radiation pattern that has a null directed toward the target; accurate tracking is obtained by moving the beam to keep the target in the null. The pencil-beam pattern is also (and simultaneously) required for tracking, to provide a signal from the target as reference. These two beams, the normal pencil beam and the split-track beam, are shown in Fig. 3. They are referred to as sum and difference radiation patterns, respectively, since they are derived by taking the sum or the difference of the patterns from the two halves of the aperture. Three-dimensional tracking requires difference patterns in both azimuth and elevation.

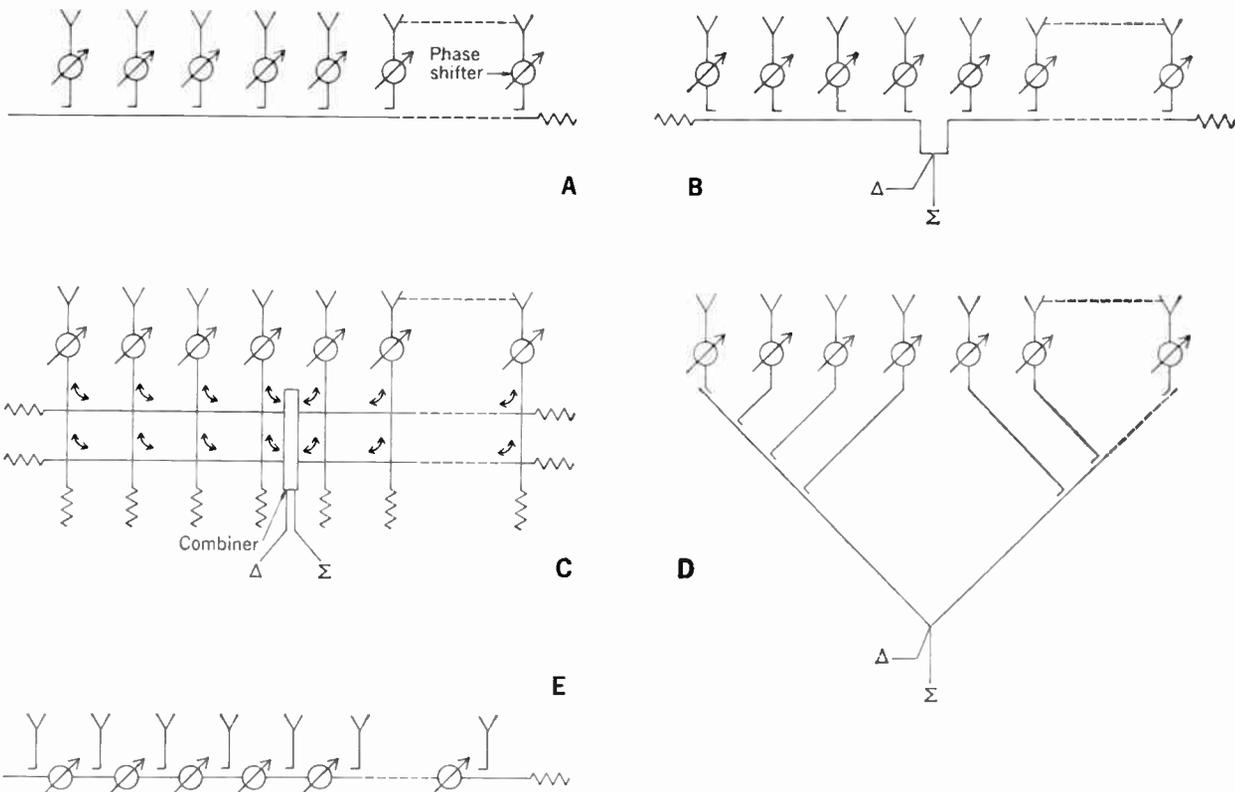
Optical feed system. The simplest method of energizing a phased array is with an optical feed system, which may take the form of a lens or reflector, as shown in Fig. 4. The primary feed has the same design as used with conventional lens antennas or reflectors like the common parabolic reflector. It radiates with a radiation pattern that suitably illuminates the inputs to the phase shifters for the sum (Σ) mode (pencil beam) as well as with patterns suitable for the difference modes for elevation and azimuth (Δ_{e1} and Δ_{a2}). The distance from the primary feed is different for each element of the array and has to be known and taken into account when setting the phase shifters.

In the case of the lens, there are two array surfaces, one on each side of the phase shifter: one receiving and the other transmitting. There is relatively little space for the phase-shifter control circuitry.

In the case of the reflector, the same element collects and reradiates after reflection. Ample space for control equipment is available behind the reflector. To avoid aperture blocking, the primary feed may be offset as shown. The phase shifter is traversed twice and has to be reciprocal so that there is a net controllable phase shift after passing through the device in both directions. This requirement rules out a class of nonreciprocal phase shifters.

Constrained feeds. Figure 5 shows several types of constrained series feeds, where the radiators are coupled with taps along a transmission line such as waveguide, coaxial cable, or strip line. In most cases the path length to each radiating element is different and has to be taken into account when setting the phase shifters. Figure 5(A) is an end-fed array. The feed system itself is frequency sensitive and, therefore, reduces the available instan-

FIGURE 5. Series-feed networks. A—End feed. B—Center feed. C—Center feed with separately optimized sum and difference channels. D—Equal-path-length feed. E—Series phase shifters.



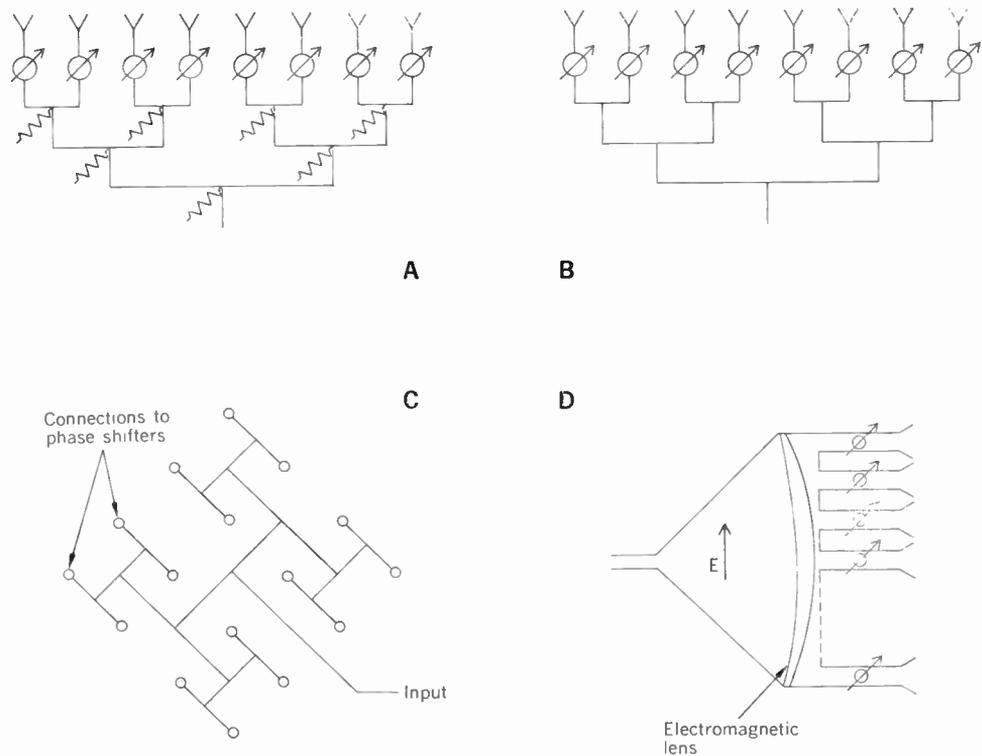


FIGURE 6. Parallel-feed networks. A—Matched corporate feed. B—Reactive corporate feed. C—Strip-line reactive feed. D—Multiple reactive power divider.

taneous bandwidth of the antenna system. Moreover, only a single radiation pattern is available; normally the sum pattern is chosen. The amplitude distribution across the aperture can be adjusted for low sidelobes by a suitable choice of coupling coefficients. The network of Fig. 5(B) is center fed and has sum and difference pattern outputs. The coupling to the elements again controls the aperture distribution and may be adjusted for either best sum or best difference pattern, but no reasonable compromise seems possible. In Fig. 5(C) this disadvantage is overcome at the cost of some additional complexity. Directional couplers are used with two sets of feed lines, which couple in the direction indicated by the arrows. Sum and difference amplitude distributions may be adjusted independently, making this configuration the most useful of the series feeds. Figure 5(D) shows a series feed that gives equal path length to all elements. The calculation of phase is a little simplified, but on the whole very little advantage is gained at the cost of a considerable increase in size and weight. The feed network of Fig. 5(E) permits simple programming, since each phase shifter requires the same setting. The insertion loss increases with successive radiators, and close tolerances are required for setting the phases.

Figure 6 shows some constrained parallel-feed networks in which the radiating elements are combined into subarrays with transmission paths of equal electrical length. Amplification for both transmitting and receiving is normally carried out at the subarray level. For transmitting, therefore, a number of relatively low-power amplifiers are used rather than a single high-power one and very high levels of RF power may be achieved.

The subarrays are combined to form sum and differ-

ence radiation patterns, and the amplitude weighting may be chosen separately for the two functions so that each may be adjusted independently for best performance. Since this combination is done before amplification when transmitting or after amplification when receiving, the networks may contain lossy elements. The amplitude weighting has a granularity that depends on the size and shape of the subarrays. Figure 7 shows how the subarrays may be shaped and interlaced to give smooth distributions.

Hybrid systems are possible, formed by the combination of series- and parallel-fed networks; series feeds may also be combined into subarrays.

Phase shifters

For array radars there are two general types of electronically operated phase shifters: ferrite and diode. Because the insertion loss of diodes tends to increase and reliability tends to decrease as the frequency is increased, ferrite phase shifters are more common at higher frequencies and diodes at lower frequencies, with the meeting ground somewhere in the region of 2000 to 3000 MHz. At the same time, at higher frequencies ferrite materials with higher values of magnetization can be used without causing lossy magnetic resonances. The physical size, therefore, decreases rapidly with an increase in frequency. The expression "ferrite" includes garnets, which are also ceramic materials with magnetic properties. Only the two most common types of ferrite phase shifters are described; both are nonreciprocal. Other types, including reciprocal types, have been developed, but they are not commonly used in phased arrays.

Ferrite phase shifters.³ Figure 8 shows a digital latching

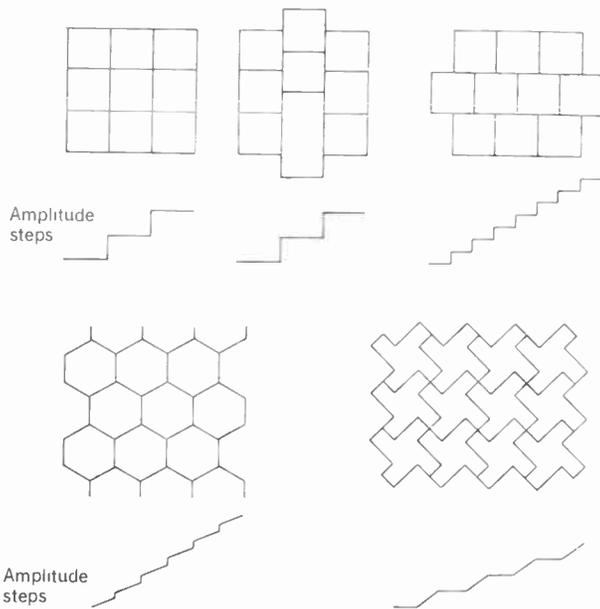


FIGURE 7. Interlaced and shaped subarrays and resulting steps in the amplitude distribution.

FIGURE 8. C-band digital latching ferrite phase shifter.

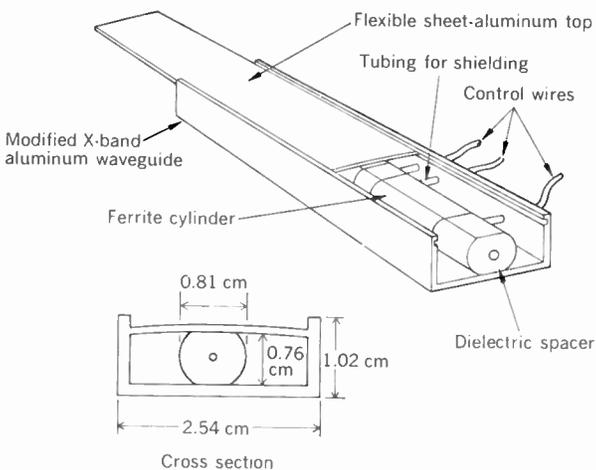
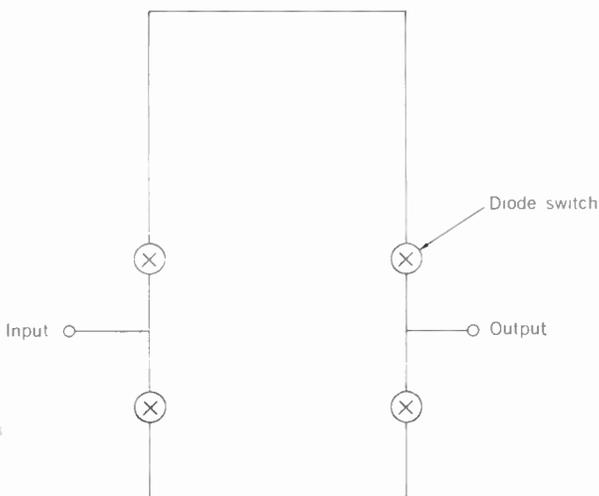


FIGURE 9. Switched line-length phase shifter (one bit).



ferrite phase shifter, which contains a ferrite or garnet cylinder enclosed in a waveguide. A wire runs through the center of the ferrite and carries a positive or negative current pulse that magnetizes the ferrite to saturation. After the pulse has been passed, the ferrite retains most magnetization (that is, it “latches”).

The two remanent states of magnetization due to the positive or negative pulse correspond to a different electrical length, the difference being the differential phase shift. The electrical length for a forward traveling wave “latched” to one state is the same as that for a backward traveling wave latched to the opposite state. This nonreciprocal property, which is responsible for the differential phase shift, also requires switching between the transmitting and receiving modes of the radar. The complete digital phase shifter contains several lengths of ferrites, adjusted to give differential phase shifts of 180, 90, 45 degrees, etc., depending on the number of bits required. For phased-array application there are usually three or four bits, giving a phase setting accurate to within ± 22.5 or ± 11.25 degrees.

The phase shifter in Fig. 8 shows spacers between adjacent ferrite bits to reduce interactions. Since these spacers are not always necessary, they are frequently omitted. It is extremely important for good contact to exist between the waveguide walls and the ferrite. The geometry shown achieves this by making the waveguide top a weak spring that conforms to the exact shapes of the ferrite. Use of this technique relieves the requirement of close tolerances.

Switching takes only a few microseconds. To switch between transmitting and receiving merely requires the complementary digital state, wherein the polarity of all pulses is reversed.

Flux-drive ferrite phase shifters are the analog equivalent of the digital latching phase shifter. Only one length of ferrite is used, long enough to give a phase shift of 360 degrees, and the amount of residual flux in the ferrite (to which the phase shift is proportional) is varied. This variation may be achieved, for example, by driving a current through the control wire from a constant-voltage source for a length of time proportional to the required phase shift. Before each setting, the phase shifter has to be reset to a known state—namely, saturation. Therefore, it requires twice as many switching operations as the digital phase shifter.

The driver and logic for the flux-drive phase shifter are much more complex than for the digital phase shifter; however, only one circuit is required, whereas the digital phase shifter requires a separate driver for each bit.

The flux-drive technique has been developed by the M.I.T. Lincoln Laboratory, which has also shown that the technique has the potential of providing compensation for the variations of phase shift that result from temperature changes. Temperature sensitivity of many ferrite materials can otherwise pose problems, especially in those cases in which the transmitter power is applied unevenly to the aperture, thereby causing temperature gradients.

Diode phase shifters.⁴ Semiconductor p-i-n diodes can be used to actively change the phase through a microwave circuit containing relatively high RF power. They are switched between two states, with forward or reverse bias, and give digital type of operation. These diodes can switch up to about 10 kW of peak power and

200 watts of average power. Other types of diodes are also suitable, but they have lower power capabilities.

In one configuration, shown in Fig. 9, the diodes are used to switch a transmission-line circuit so that input and output are connected either by a short transmission line or one that is longer. The difference in length gives the differential phase shift; one circuit is required for each bit. An alternative principle, which can be used in various ways, changes the phase of a wave after reflection. The reflection point is either a short circuit or a diode that is in front of the short circuit and is either transparent or reflecting, depending on its bias.

A third method of using diodes as phase shifters is to use many diodes spaced along a transmission line. Switching a diode changes the coupled reactance, which, in turn, changes the phase of transmission.

Beam-steering control⁵

The required amount of phase shift for each of the very many elements of a phased array has to be calculated, and the phase shifters have to be set to the proper value.

Conceptually, the simplest form of beam steering involves the use of orthogonal row and column commands, which are in the form of $m T_{xs}$ and $n T_{ys}$ for the m th row and n th column. T_{xs} and T_{ys} are the computed values for the element-to-element phase increments for the particular beam-pointing direction. With a square aperture of N elements a row or column contains \sqrt{N} elements, so the total number of calculations is $2\sqrt{N}$. At each element the row and column phases have to be added to form the proper command for the phase-shifter driver. In some cases additional phase corrections have to be applied and may be added at the phase-shifter driver. For example, compensations are necessary if the feed system (such as optical and series feeds) does not provide equal phase excitation at the input to each phase shifter. Compensation may be added for known errors such as those caused by components, by environment (for example, temperature), or by frequency dependence.

If the antenna is divided into a number of similar units or subarrays (such as those shown in Fig. 7), a different form of beam steering becomes possible. This is indicated in Fig. 10. Steering commands are calculated for all the elements of one subarray (by rows and columns of the subarray, for example), and these commands are then applied to all subarrays. All subarrays then give rise to the same sloping phase front, as shown. What is still necessary is to add a further phase correction so that the phase fronts line up as shown by the dashed lines. This can easily be accomplished by the addition of a phase shifter behind each subarray. The additional insertion loss of the phase shifter does not matter, since amplification takes place between it and the aperture. The phase-steering command for this phase shifter may be calculated as before by considering the subarray as one element in an array of subarrays.

The width and shape of the radiation pattern may be controlled by the phase at the aperture. Beam broadening in particular is useful for many radar applications, and may be obtained with gablelike or spherical phase distributions. These desired distributions may be approximated by phasing that can be achieved with row-and-column or subarray beam-steering methods.

The choice of beam-steering system may have a de-

cisive effect on the overall cost of the phased array. This point is discussed in more detail in the section on cost considerations.

Aperture design^{2,6}

Matching. The radiating elements are, in effect, transducers that change the mode of propagation from transmission line or waveguide to radiation into space. Isolated and widely spaced elements produce radiation patterns with multiple beams; a phased-array antenna, therefore, requires elements that are closely spaced. The elements are then necessarily mutually coupled, and when matched to free space they share the aperture and accept the incident power on receiving and cause no reflections on transmitting. The main problem of aperture design is how to achieve a reasonable match, not only for efficient operation but also because the reflected energy may reappear in the form of undesired sidelobes or, on transmitting, may cause undesirable effects on the power amplifier and its associated circuits.

Mutual-coupling effects significantly change the element impedance from that of an isolated element and have to be taken into account in matching the aperture. A matched element accepts all the power that falls upon its share of the aperture. When the beam is scanned, the power density on the projected area is maintained, but on the actual area the power density is reduced by the factor $\cos \theta_0$. The aperture impedance varies for this reason with scanning, approximately as $\cos \theta_0$ in the plane of polarization but as $1/\cos \theta_0$ in the plane at right angles to it—that is, the magnetic plane. Thus, the problem of matching is further complicated, since the aperture should be matched for all conditions of scanning, and there are, apparently, contradictory requirements for matching in the two principal planes. In spite of all these difficulties, acceptable aperture matches can be achieved with reasonable bandwidth. For scanning to ± 45 degrees detaching in the range of 1.5–2 to 1 seems reasonable, deteriorating to perhaps 2–3 to 1 with ± 60 degrees of scan.

The aperture-matching problems have been analyzed and are, on the whole, well understood. Some measure of compensation for the impedance changes on scanning have been achieved with a thin layer of high-dielectric-constant material in front of the aperture. The reflection coefficient from a dielectric interface depends on both angle of incidence and polarization, as does the impedance of the aperture.

A technique that permits a simple experimental approach to the matching problem without requiring all the interacting elements to be present has been developed by Wheeler Laboratories. A single radiating element is placed in a waveguide that is so proportioned that in its walls are seen the mirror images of the element, forming an infinite array with the proper element spacing. The impedance the element looks into is then the same that an element sees in an infinite array. Therefore, it is adequate to devise a way of matching the single element to the waveguide. The conditions of this simulation are for no scanning in the plane of polarization, but in the magnetic plane a phase reversal results upon reflection so that the simulated array, in that plane, has alternate elements in antiphase. This amounts to scanning to an angle that is the same as the waveguide “zig-zag” angle. A number of different but distinct scan angles can be

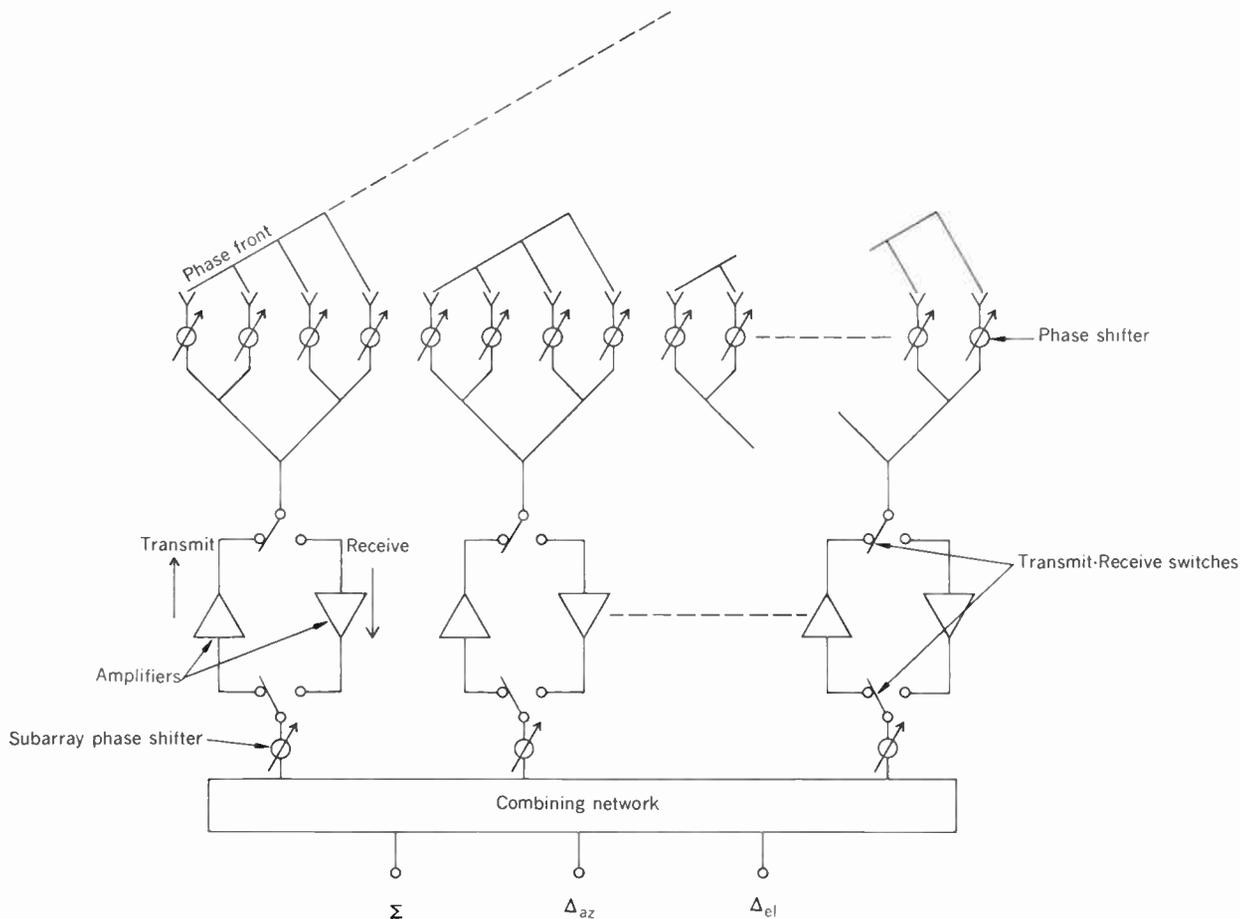


FIGURE 10. Phased array with subarrays.

simulated by placing more than one element into the waveguide and by adjusting the waveguide size, shape, and modes. Normally, simulators are most useful for matching with the beam scanned to near broadside. The array may then be checked for proper functioning under conditions of scan by measuring the element pattern as described in the following discussion.

With well-matched radiating elements the scanned beam, at its peak, contains the (in-phase) addition of all the contributions of all elements. With equal weighting of N such elements the gain pattern of one element, surrounded by other similar but resistively terminated elements, will be $1/N$ times the gain of the whole antenna. The element pattern will vary with angle in the same way that the radiated beam from the total antenna varies when successively scanned to all the scan angles. The well-matched element, therefore, has a power pattern that is approximately a cosine function, with zeros at ± 90 degrees. For experimental purposes it is usually sufficient to surround one element by terminated other elements to a radius of one to two wavelengths to adequately simulate the conditions of the array.

For example, at 60 degrees the power pattern of the element should be reduced to half its maximum value ($\cos 60^\circ = 0.5$). If the value is smaller, then it can be predicted that this additional loss will occur when the antenna is scanned to 60 degrees and will be caused by mismatch at the aperture or equivalent misdirection of energy into undesired regions. If dips or nulls appear in

the element pattern, then these will be reproduced when the whole array is scanned to the direction of the dip or null. Some matching structures have been known to generate surface waves under certain resonance conditions corresponding to some angles of scan. Under those conditions beams cannot be pointed in the specific direction since the power is diverted to the surface wave. The element pattern will reveal all those conditions by indicating a null at that angle. A check of the element radiation pattern, surrounded by terminated elements, is therefore an important precautionary measurement that gives an indication of how the array will behave under the various conditions of scan, before the array is actually built.

Element spacing. The elements are spaced by approximately half a wavelength to prevent the formation of grating lobes. They are normally arranged in either a square or triangular matrix. The triangular configuration has the advantage of requiring about 15 percent fewer elements for equal suppression of grating lobes.

The array may be "thinned" by removing active elements from it and replacing them with resistively terminated dummies to avoid upsetting the match condition. Since the removal is substantially random, the grating lobes will not form in coherent fashion. The gain still corresponds to that of the active elements; the beam width, however, is approximately unchanged, as if no element had been removed, and a substantial amount of power is radiated into the sidelobe region. For example,

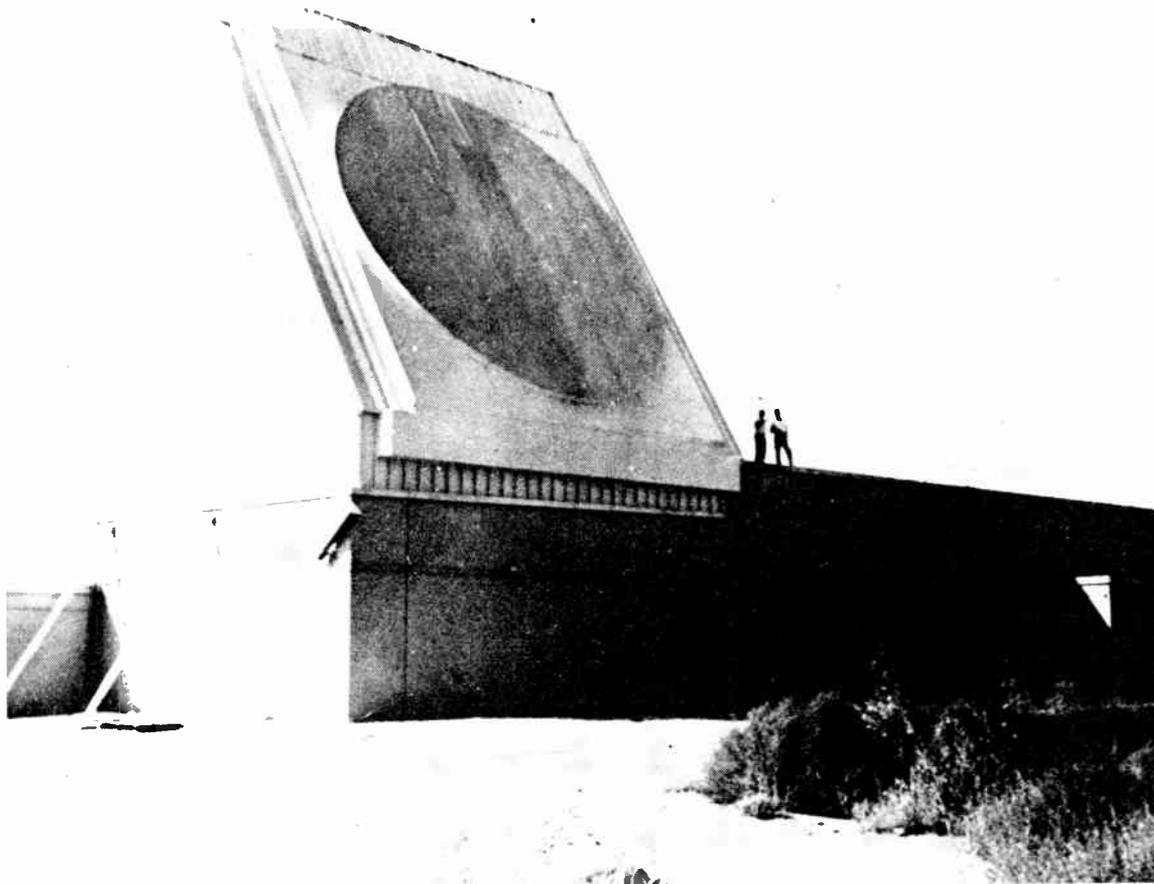


FIGURE 11. HAPDAR building, showing phased array. (Photo courtesy Sperry Gyroscope Company)

if the array has been thinned so that only 10 percent of the elements are actively used, then the gain will have dropped to 10 percent and the beam width will have stayed the same, but about 90 percent of the transmitted power will be delivered to the sidelobe region.

HAPDAR

The phased-array system described here is just one of many phased-array configurations that have been built. It is at a low frequency and therefore its antenna system is large.

HAPDAR (Hard Point Demonstration Array Radar)⁷ was sponsored by the Advanced Research Project Agency and developed by the Sperry Gyroscope Company Division of the Sperry Rand Corporation. It performs search-acquisition-track radar functions on a time-sharing basis. The antenna system uses an electronically steered phased array in the form of a lens with a thinned aperture. The phase shifter has three bits and uses switching diodes. Figure 11 shows the building with the antenna, which is tilted 30 degrees from the vertical.

The principle of the antenna is shown in Fig. 12. The lens has a full complement of elements on the side of the primary feed. Either a single element connects directly to a phase shifter, or two elements are combined in phase

(with a fixed bias if necessary) and then connected to a phase shifter. Each phase shifter in turn connects to an active element at the radiating aperture. In this way substantially all the transmitter energy incident on the feed side of the lens is utilized and distributed to a more economic aperture, which has a reduced number of phase shifters. Terminated dummy elements at the aperture maintain the proper impedance environment. The lens aperture is not severely thinned; only about 40 percent of the radiating elements have been replaced by dummies. There are 2165 active radiators, giving a corresponding gain. The full aperture contains 3750 element positions and gives a corresponding beam width. The elements are spaced in a triangular grid.

A printed circuit card contains the phase shifter and the dipole(s) facing the feed. The card plugs into position and connects with waveguide horns that form the elements at the radiating aperture. The aperture match gives a voltage standing-wave ratio of about 2 to 1, approximately unchanged with scanning. At the feed side no scanning occurs and the match is better.

Beam steering is achieved in steps of 1/32 of a beam width and interpolated to 1/500 of a beam width. Automatic testing of the phase shifters is available with computer printout of the location of faulty diodes.

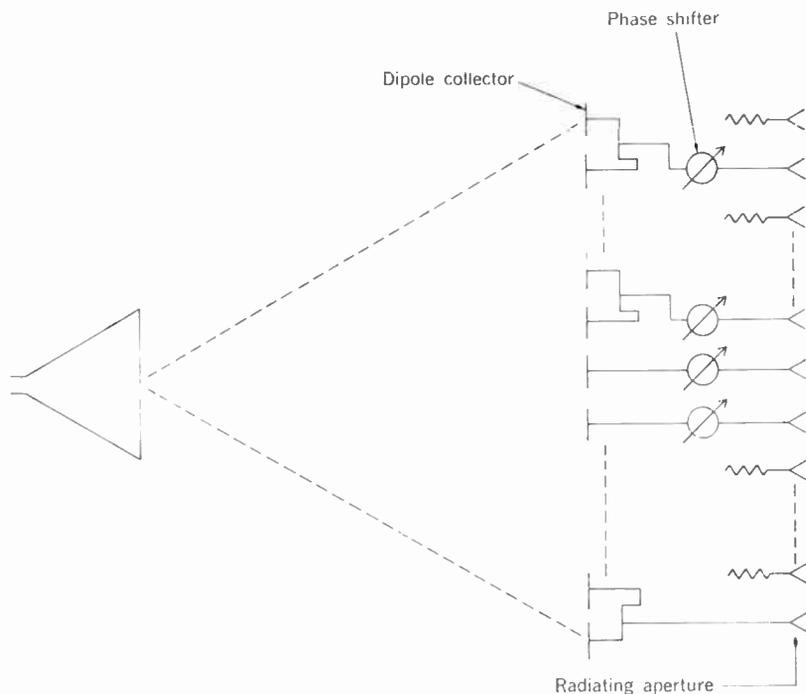


FIGURE 12. HAPDAR thinned lens.

Cost considerations

Many different types of phased arrays have been constructed and tested. Probably all the different configurations described here have been built or at least attempted on an experimental scale. But all phased arrays and phased-array radars were built as one of a kind to prove the technique that was used. What has not been established is the cost reduction that can be achieved in production, where one radar alone uses perhaps four antenna faces and a total of 20 000 elements or more.

Considerable differences in cost may result from following approaches that are similar and that differ only in detail. The most costly items in phased arrays are the phase shifters and their drivers. In the case of ferrite phase shifters, the cost of the active material, ferrite or garnet, is presently high because it is produced in relatively small quantities for the individual job at hand. Cost reduction in this area can be anticipated when production techniques become available. The tight tolerances normally required can be relaxed if the phase shifters are graded; then a correction can be added at the subarray level or the phase shifters can be placed in the array so that advantage is taken of their known performance characteristics. The relaxation of tolerances in this and other areas should lead to further reductions in cost.

The cost of the phase-shifter driver can vary by large factors depending on how the beam is steered. If row and column steering is used, then one complete driver circuit is required per phase shifter. The cost of the driver may then well be the most expensive item of the phased array. The alternative method that was described becomes possible if the antenna is divided into subarrays, each with its own amplifiers. This technique reduces the number of drivers and driver logic by a large factor. To show what savings can be achieved under certain favorable conditions consider, for example, an array face of 32 subar-

rays, each with 128 elements, and assume a flux drive phase shifter. With row and column steering, $128 \times 32 = 4096$ separate drivers, complete with adding networks and logic, are required. On the other hand, 128 drivers with adding networks and logic are needed to drive one subarray. The other subarrays still have to be driven, but in identical manner to the first subarray and only driver amplifiers are needed for the remaining elements. A further 32 drivers with adding networks and logic are then needed to drive the phase shifter that precedes the RF power amplifiers (Fig. 10).

These savings are possible with a system that uses a number of RF power amplifiers rather than a single high-power tube. This may be desirable for other reasons, but the cost advantage of a single high-power amplifier may outweigh the described savings. The method may still be used with a single amplifier, but now the subarray phase shifters will have to operate with high RF power and their insertion loss of typically 20 percent will reduce the overall efficiency.

REFERENCES

1. Allen, J. L., "Array antennas: new applications for an old technique," *IEEE Spectrum*, vol. 1, pp. 115-130, Nov. 1964.
2. Hansen, R. C., *Microwave Scanning Antennas*, vols. I, II, and III. New York: Academic Press, 1964.
3. Pippin, J. E., "Microwave ferrite devices—1968," *Microwave J.*, vol. 2, pp. 29-45, Apr. 1968.
4. Ryder, R. M., Brown, N. J., and Forest, R. G., "Microwave diode control devices, Part 2—Duplexers, switches and phase shifters," *Microwave J.*, vol. 2, pp. 115-122, Mar. 1968.
5. Von Aulock, W. H., "Properties of phased arrays," *IRE Trans. Antennas and Propagation*, vol. AP-9, pp. 1715-1727, Oct. 1960.
6. Hannan, P. W., "Element-gain paradox for a phased-array antenna," *IEEE Trans. Antennas and Propagation*, vol. AP-12, pp. 423-433, July 1964.
7. Kahrilas, P. J., and Jahn, D. M., "Hardpoint demonstration array radar," Supplement to *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-2, no. 6, pp. 286-289, 1966.

Ceramic IF filters for

Ceramic filters have been in use for a number of years in military and commercial equipment where exact interstage alignment and frequency stability are more critical than cost. Recent developments show promise of consumer applications in those cases in which cost is an extremely important factor

Franz Sauerland, William Blum Clevite Corporation

Ceramic filters offer a number of performance and compatibility advantages in solid-state consumer IF circuits; and they have recently become economically competitive with the conventional LC approach. This article surveys the different types of ceramic filters, as well as their properties and applications, including reports of some new network approaches regarding spurious suppression and application in integrated circuits.

Over the past decade, solid-state technology has changed the design of receiver circuits radically. With the impending introduction of varactor tuning, one of the few conventional components remaining—and certainly one of the most cumbersome because of size and frequency alignment—is the wound inductance in intermediate-frequency (IF) transformers.

Considerable effort has gone into finding new and more compatible methods for IF filtering. Many solutions have been attempted, including RC-active (by feedback, NIC, gyrator, phase-locked loop), electromechanical (resonant gate transistor, tunistor, magnetostriction), digital, and sampling filtering techniques. At present none of these methods can match the performance and economy of the conventional wire-wound IF transformer at the intermediate frequencies of principal interest, namely, 455 kHz for AM radios, 4.5 MHz for television audio reception, and 10.7 MHz for FM receivers.

The characteristics of ceramic filters are such that, if applied properly, this type of filter performs better than the conventional IF transformer and is more compatible with modern receiver circuitry. Ceramic filters for high-volume consumer application in recent months have become competitive economically with the conventional approach. They produce equal or better performance than conventional devices at equal or lower overall price, and are compatible with present trends in solid-state circuitry with regard to size, reliability, and fixed tuning.

A number of manufacturers, both in the U.S. and elsewhere, are beginning to offer various types of economical

ceramic filters. Some have been designed into receivers that are now, or will shortly be, on the market. For reference and comparison with familiar ground, in this article the ceramic filter response frequently will be compared with the well-known performance of tuned LC circuits.

In the following, most of the notes on performance and application relate to filters consisting of one or more individual two-electrode resonators operating at center frequencies between 0.3 and 1 MHz. This type of resonator has been the most promising building block for a true, versatile, and well-proved “economy” filter. Recently, however, various thickness-mode resonators, and especially multiresonator configurations arranged on a single ceramic wafer, are being offered for operation at higher frequencies, especially at the 4.5- and 10.7-MHz IFs. These filters are certain to have a major impact on modern IF circuitry, and although no long-term or large-scale performance data are available to date, their principles of operation and equivalent circuits will be described.

History

The basic building block of ceramic filters is the ceramic resonator. Good ceramic bandpass filters became feasible about a decade ago with the introduction of a stable, high- Q piezoelectric ceramic material¹ that is still the basis for all ceramic filters on the market today.

Until recently, the main application of ceramic filters has been in high-performance military and commercial communication and navigation equipment, where high selectivity, stability, ruggedness, and small size are of prime importance. This type of filter is well established,²⁻⁵ with standard units ranging in fractional bandwidths from approximately 0.05 to 20 percent and 60/6-dB skirt ratios as low as 1.1. The majority lie in the frequency range between 200 and 700 kHz, but continuing efforts have stretched the range of specific filter types down to a few kilohertz⁶ and to as high as 10 MHz.⁷

Since the inception of ceramic filters, attempts have been made to use them in place of conventional IF transformers in broadcast radios. This type of ceramic filter

Consumer products

was described in a number of publications.^{8,9} It offered the genuine advantages of fixed tuning, frequency stability, and small size. However, it never got into high-volume application, for a number of reasons. First, the filters were introduced by a single-source supplier at a time when conventional IF transformer prices were decreasing as a result of the manufacture of these transformers in the Far East. In addition, the ceramic filters lacked a dc path between the filter terminals with a resulting reduction in amplifier gain, which, until a few years ago, represented a significant economic disadvantage; and, finally, they had not yet achieved a proved long-term record of reliability.

None of these drawbacks exist now. Transistor amplifier gain is cheap. Millions of ceramic resonators have been operating in both military and commercial equipment for years without failure or deterioration. New material resonator and network approaches, as well as automated mass-production techniques, are being applied to make the use of ceramic filters in broadcast receivers and other high-volume applications economically feasible. In fact, the mass production of ceramic IF filters is inherently simpler than that of wire-wound LC circuits, and should ultimately result in reduced prices for IF circuitry.

The performance of a ceramic filter depends on the quality of the ceramic material and processing, on the geometric configuration of the ceramic resonator, and on the manner of applying the resonator(s) to the IF network. The principal considerations of this article are (1) ceramic resonator structures and their electrical equivalence; (2) single-resonator ceramic filters; (3) multiresonator ceramic filters; (4) inductorless IF networks; (5) circuit applications; and (6) ceramic filters in integrated circuits.

Ceramic resonators

Ceramic resonators rely on the piezoelectric effect¹⁰ for a direct interaction between electric and mechanical energy, and particularly for a direct relation between electrical and mechanical resonances. Since mechanical resonances are related to geometric configuration, one may envision many types of ceramic resonators with a large variety of vibrational modes and electrical resonances. Only a few of the more popular configurations in use will be discussed. Demonstrated will be a distinction between "basic resonators," consisting of one homogeneous resonating body, and "composite resonators" that combine two or more basic resonators in one unit.

A popular basic resonator with two electrodes is shown in Fig. 1. It utilizes the radial mode of vibration, i.e., it expands and contracts radially at the frequency of the electric signal applied to its electrodes, whereas the center has no radial motion. In order to minimize damping of

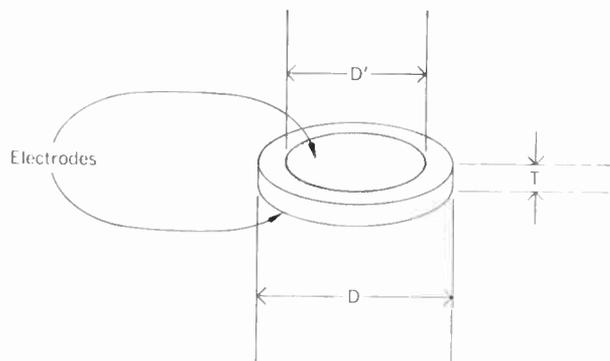
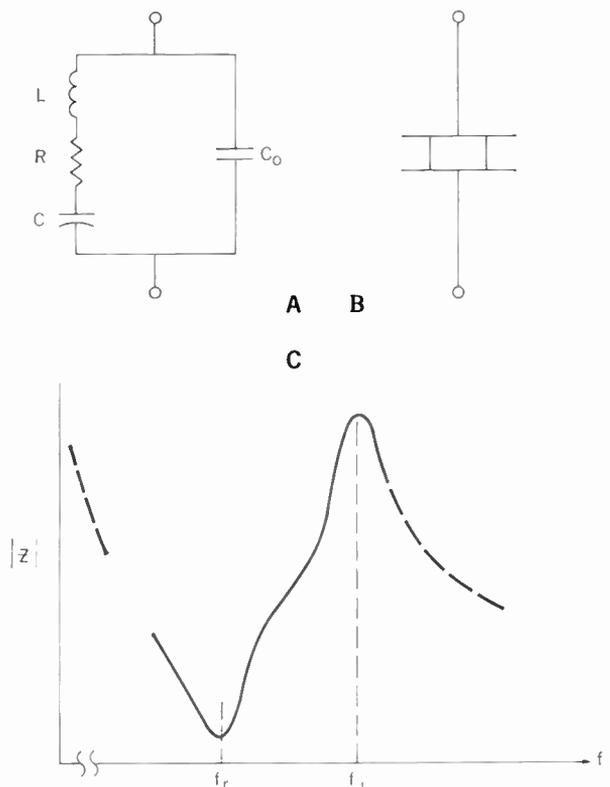


FIGURE 1. Basic resonator with two electrodes.

FIGURE 2. (A) Equivalent circuit, (B) circuit symbol, and (C) impedance-frequency characteristics of basic two-electrode resonator.



its vibrations, the resonator is usually mounted at the center. Electrodes on opposite faces of the disk serve to intercouple electric and mechanical energy.

There are many two-electrode resonator configurations, operating in various extensional, flexural, shear, or thickness modes. They all have the electrical equivalent circuit, circuit symbol, and impedance-frequency characteristics given in Fig. 2. Electrical characteristics are generally specified by either R, L, C, C_0 , or $Q, C_0, f_r, \Delta f$, where

$$f_r = \text{series resonant frequency} = \frac{1}{2\pi\sqrt{LC}}$$

$$f_a = \text{antiresonant frequency} = f_r\sqrt{1 + C/C_0}$$

$$\Delta f = f_a - f_r$$

$$Q = \frac{2\pi L f_r}{R} = \frac{1}{2\pi R C f_r}$$

The electrical resonator parameters may be varied over a range dependent upon the ceramic material and processing and on the resonator geometry. For example, Table I lists some typical properties of radial-mode resonators for two widely used ceramic materials.

The value of the resonator shunt capacitance C_0 depends on the dielectric constant of the material and the dimensions T and D' of Fig. 1. The frequency f_r is equal to the radial mechanical resonance frequency of the disk. For the materials listed, $f_r \approx 254/D$, where the diameter D is measured in centimeters and f_r in kHz. Hence, for both very high and very low operating frequencies, D becomes impractically small or large. At present the preferred frequencies for radial resonators lie between 200 and 700 kHz.

The quantities $\Delta f/f_r$ and C_0 may be adjusted by the manufacturing process to any value within the tabulated ranges. This makes the ceramic resonator a much more versatile electric circuit element than, for example, the quartz resonator, whose equivalent circuit is the same as that of Fig. 2(A) but whose $\Delta f/f_r$ is a material constant of small and fixed value.

The stability of the filter relies on the temperature dependence of the resonator parameters. The quality factor Q remains essentially constant from room temperature down to -55°C , but declines toward higher temperatures. The strong positive temperature coefficient of C_0 can generally be accounted for in the filter design in such a manner that it has no significant effect on the filter performance. The frequency-temperature coefficient can be controlled by processing to be either positive or negative. Frequencies f_r and f_a track closely over the temperature range to maintain an essentially constant Δf .

Compared with the average consumer-type IF transformer, ceramic materials and filters have a better frequency-temperature stability and quality factor Q , the

I. Typical properties of radial-mode resonators for two ceramic materials

	Ceramic A	Ceramic B
$\Delta f/f_r$ range, percent	1-10	0.2-3
C_0 range, picofarads*	20-85	10-400
Relative dielectric constant		
At 25°C	1150	500
-55°C to 85°C	900-1400	460-580
-20°C to 60°C	1000-1300	480-550
Q value		
-55°C to 25°C	450	1400
25°C to 85°C	450-250	1400-900
Maximum total f_r shift, percent		
-40°C to 85°C	0.2	0.2
-20°C to 60°C	0.1	0.1
Maximum f_r time shift, percent†		
within five years	0.2	0.2

* For $f_r \approx 445$ kHz.

† Maximum advertised value. Typical value, 0.1 percent.

latter by an order of magnitude. A typical radial resonator with $f_r = 455$ kHz has a diameter of 0.56 cm and a thickness of 0.038 cm. All equivalent piezoelectric resonator circuits are only valid in the vicinity of the operating frequency. For radial resonators, the equivalent circuit of Fig. 2(A) is accurate for frequencies up to approximately $1.5 f_r$. At higher frequencies, other resonances occur due to overtones of the radial mode and to other vibrational modes. By special resonator geometry it is possible to reduce, eliminate, or enhance a particular overtone. This method may be used to obtain higher operating frequencies without reducing D to impractically small sizes.⁷

For operating frequencies far below and above the radial-resonator range, different resonator configurations and modes are used.^{6,7} For example, ceramic filters at 10.7 MHz use modes whose resonant frequency is determined by the thickness of the ceramic material between the electrodes.

By partitioning one of the electrodes of Fig. 1 into two isolated parts, one obtains a three-electrode resonator whose simplified equivalent circuit and network symbol are shown in Figs. 3(A) and 3(B), respectively. Here C_{01} and C_{02} are the shunt capacitances between terminal pairs 1-1' and 2-2', C' is the coupling capacitance between input and output electrodes, and the transformer ratio is⁹

$$n = \sqrt{\frac{C_{01}}{C_{02}}}$$

Variety in composite resonators is as large as that in basic resonators, but only three versions will be dis-

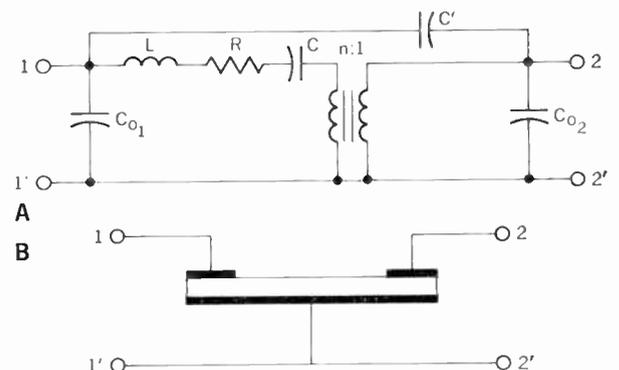
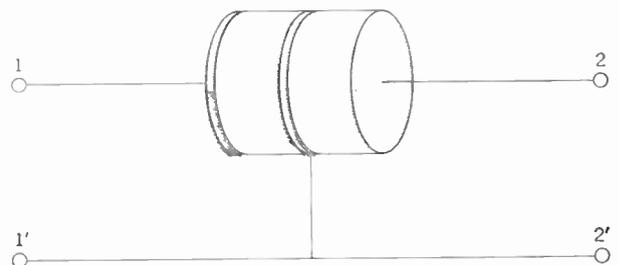


FIGURE 3. Three-electrode resonator. A—Equivalent circuit. B—Network symbol.

FIGURE 4. An example of a composite resonator obtained by bonding together two single resonators. Common center electrode is grounded.



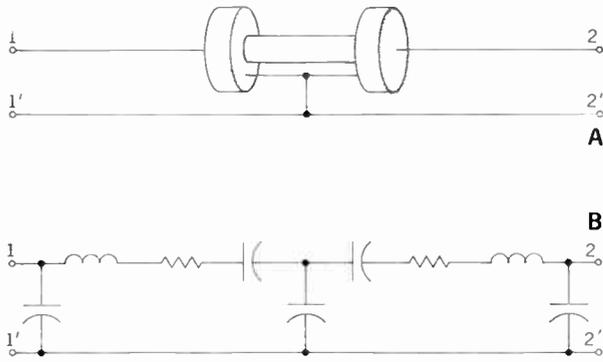


FIGURE 5. Two basic resonators are bonded to an intermediate coupling link, forming a more complex and selective resonator (A), as shown schematically in (B).

FIGURE 6. A—Fundamental filter network employing basic two-electrode, series-connected resonator. B—Typical filter response of the circuit.

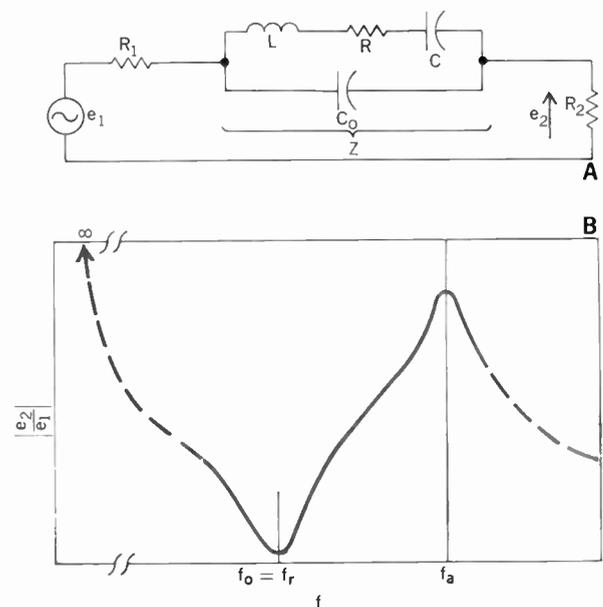
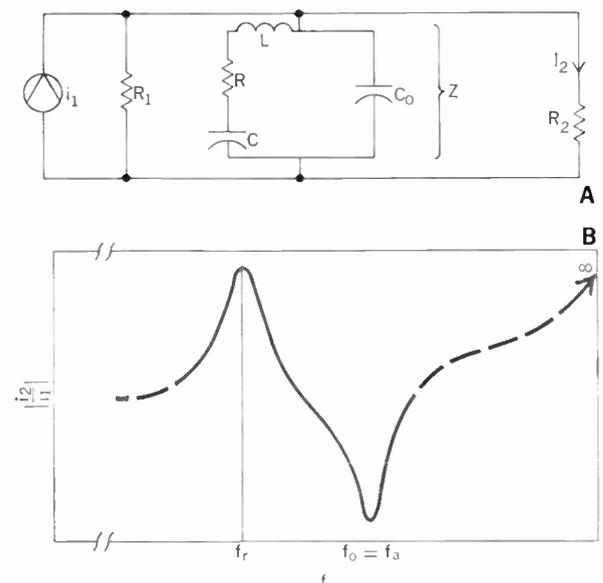


FIGURE 7. A—Another fundamental filter network, but with parallel connection. B—Response characteristics.



cusced. One type consists of a ceramic transducer bonded to a metal resonator. This has the equivalent circuit of Fig. 2(A) with potentially improved Q and stability at the expense of narrow limits on C_0 and Δf .

Another version is obtained by bonding together two single resonators and connecting them in the manner of Fig. 4. This arrangement also has the equivalent circuit of Fig. 3(A). The grounded common center electrode acts as an electric shield between the input and output and greatly reduces the value of C' .

A more complex and selective composite resonator is obtained by bonding two basic resonators to an intermediate coupling link that controls the amount of mechanical energy transferred from one resonator to the other. One form of this and the corresponding equivalent circuit are shown in Fig. 5.

The same principle and equivalent circuit hold for the so-called “coupled mode” or “monolithic” filter, which is becoming especially important for thickness-mode operation in the MHz range. It consists of the combination of two or more separate but acoustically intercoupled resonators on a single, thin ceramic wafer. For n resonators one obtains n resonant circuits—separated by shunt capacitors—in the equivalent circuit. This concept has been applied to quartz filters as well, and has been described in a number of recent publications.¹¹

Single-resonator filters

In general, several resonators are required to shape the IF response. This may be done by arranging individual resonators at separate points in the IF amplifier or by lumping several resonators into composite filter structures.

For ease of understanding and for comparison with the LC approach, the results are related to the performance of a single-tuned LC circuit. We recall that the 3-dB bandwidth of such a circuit is $B = f_0/Q_L$ —where f_0 and Q_L are the center frequency and loaded Q , respectively—and that its frequency response is given by the universal resonance curve.¹² A typical value for the loaded Q of 455-kHz IF transformers is $Q_L = 30$, resulting in a bandwidth of $B = 15$ kHz and the filter response shown in Fig. 10 (in color). For performance comparison, the ceramic filters will be referred to the same 3-dB bandwidth response.

The filter applications of the basic two-electrode resonator all relate to the fundamental networks, Fig. 6(A) or 7(A), where Z represents the resonator impedance and R_1 and R_2 the source and load impedances, which, for our purposes, are assumed to be resistive. Figures 6(B) and 7(B) illustrate pertinent filter responses, where f_r and f_a are the resonator frequencies identified in Fig. 2(C). The passband response approximates that of a single-tuned LC circuit if the value of $\Delta f/f_r$ is relatively large.

For example, in the network of Fig. 6(A), the passband center coincides with the series resonant frequency f_r of the LRC branch of Z . The loaded Q for series resonance is

$$Q_L = \frac{1}{4\pi\Delta f C_0 R_{ser}} \quad (1)$$

where, for all practical purposes, $R_{ser} = R_1 + R_2$. If

$$\frac{\Delta f}{f_r} \geq \frac{1}{Q_L} \quad (2)$$

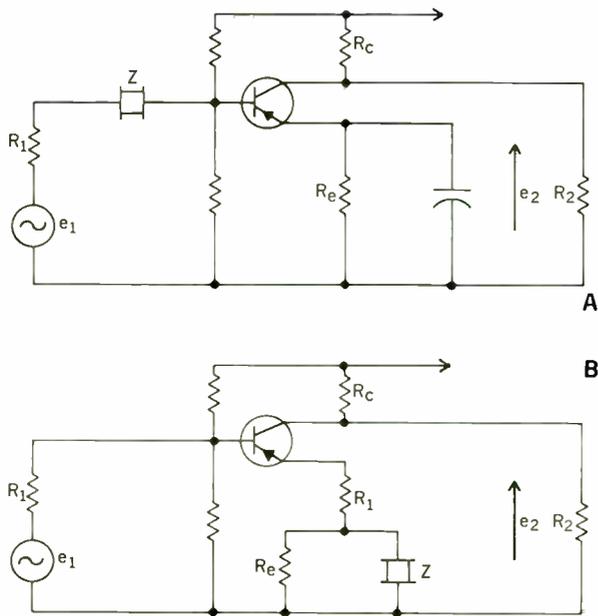
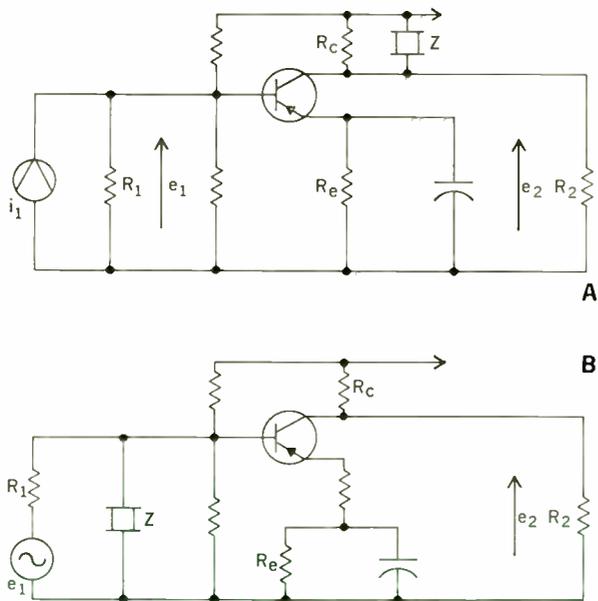


FIGURE 8. Two-electrode resonators incorporated into a common-emitter amplifier stage. A—Series-coupler circuit. B—Emitter-bypass network.

FIGURE 9. Collector-shunt coupler network (A) and base-shunt coupler circuit (B) are additional examples of two-electrode resonators used in a common-emitter amplifier stage.



then the 3-dB bandwidth is

$$B \approx \frac{f_o}{Q_L} \quad (3)$$

The capacitance C_o in Fig. 6(A) contributes the insertion loss peak at f_a (due to parallel resonance of Z) and the declining stopband insertion loss toward higher frequencies.

The network shown in Fig. 7(A) is the dual of that il-

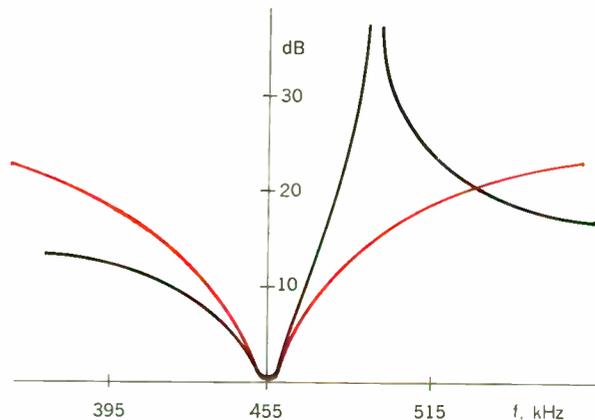
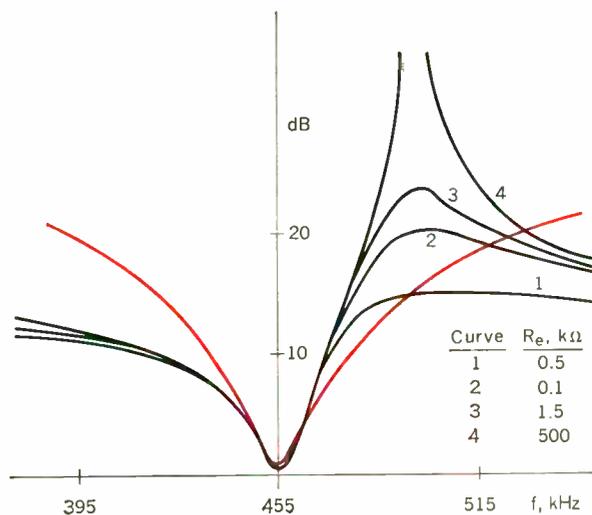


FIGURE 10. Selectivity of a series coupler compared with that of a single-tuned circuit of equal 3-dB bandwidth.

FIGURE 11. Typical emitter-bypass responses for various values of R_e compared with the selectivity of a single-tuned circuit having an equal 3-dB bandwidth.



lustrated by Fig. 6(A), and the response, Fig. 7(B), may be explained in an analogous manner. In particular, it can be shown that Eq. (3) holds if (2) is satisfied, and that

$$Q_L = \frac{\pi C_o R_p}{\Delta f} \quad (4)$$

where, for all practical purposes,

$$R_p = \frac{R_1 R_2}{R_1 + R_2} \quad (5)$$

Figures 8 and 9 show four different examples of incorporating two-electrode resonators in a common-emitter amplifier stage.¹⁴ With proper design, the circuits illustrated in Figs. 8 and 9 produce the responses shown by Figs. 6(B) and 7(B) respectively.

The series coupler [Fig. 8(A)] conforms directly to the previous analysis of the network shown in Fig. 6(A), except that R_2 now corresponds to the combined resistance of the transistor base input and the associated biasing

circuit. Figure 10 compares the actual selectivity of a series coupler with that of a single-tuned circuit designed for the same 3-dB bandwidth of $B = 15$ kHz.

In the emitter-bypass circuit [Fig. 8(B)], the usual emitter-bypass capacitor has been replaced by a resonator. The additional resistor R' in the emitter branch may or may not be necessary to adjust for the desired bandwidth. If R_s is the equivalent resistance of the signal source as seen from the transistor base, one obtains for R_{ser} of Eq. (1)

$$R_{ser} \cong R' + \frac{R_s + h_{ie}}{h_{fe}} \quad (6)$$

where h_{ie} and h_{fe} are the hybrid transistor parameters. For the voltage transfer function one obtains

$$\frac{e_1}{e_2} \cong \frac{R_{ser} + \frac{R_c Z_2}{R_c + Z_2}}{R_i} \quad (7)$$

Figure 11 shows typical emitter-bypass responses (for various values of R_i) compared with the single-tuned-circuit selectivity.

The performance of the collector shunt coupler [Fig. 9(A)] can be derived from Eq. (7). Since the emitter resistor is now bypassed by a capacitor rather than a resonator,

$$\frac{e_1}{e_2} \cong \frac{R_s + h_{ie}}{Z_2 h_{fe}} \quad (8)$$

is obtained, where Z_2 is the total load impedance (including R_c and Z). The frequency response is proportional to $1/Z_a$. The loaded Q is given by (4) if R_c is substituted for R_1 in (5).

If the base shunt coupler, shown in Fig. 9(B), is driven from a constant-current source (this may not be a practical assumption for conventional circuits, but can be achieved by special means), its performance is also described by Eqs. (8) and (4), provided that r_2 in Eq. (5) is replaced by the combined resistance of the transistor base input and the associated biasing network.

Figure 12 shows a shunt-coupler response in comparison with that of a single-tuned circuit.

In summary, the applications of Figs. 8 and 9 result in asymmetric passband responses that individually are not well suited for IF filtering. However, networks of the type shown in Fig. 8 can be combined with those of the type shown in Fig. 9 (either in one amplifier stage or in consecutive stages) such that the individual curves superimpose and yield a symmetric response. For example, Fig. 13 shows the superposition of Figs. 10 and 12; it compares the combination of a series and a shunt coupler [curve (a)] with that of two single-tuned circuits [curve (b)]. Although curve (a) has steeper passband skirts, curve (b) offers better stopband rejection. Neither curve (a) nor (b) is optimal, but a combination of the two would produce a desirable IF response.

Next, consider the basic three-electrode resonator having the equivalent circuit of Fig. 3(A). Generally, C' is negligibly small. If, furthermore, $n = 1$, the network may be redrawn as in Fig. 14, which also includes the generator and load resistances R_1 and R_2 .

Without going into a detailed analysis, one observes that the network corresponds essentially to a single-tuned series circuit whose selectivity can be evaluated

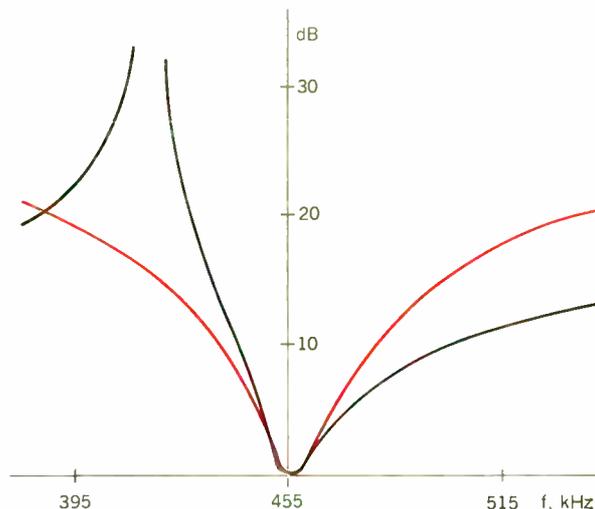


FIGURE 12. Shunt-coupler response (black curve) compared with a single-tuned circuit response (colored curve) with equal 3-dB bandwidths.

FIGURE 13. Response of a combined series and shunt coupler (curve a) compared with the response of two single-tuned circuits (curve b) with equal 3-dB bandwidths.

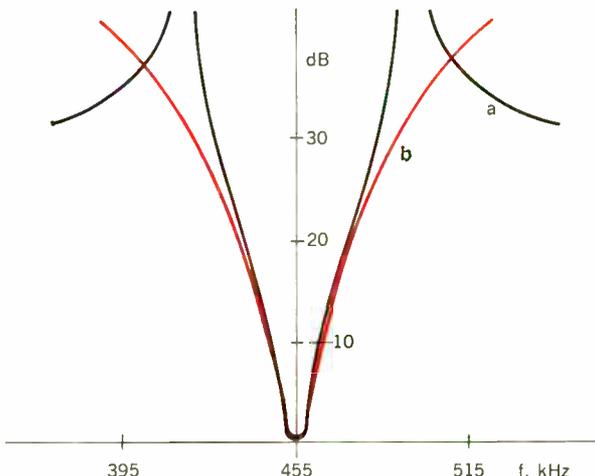
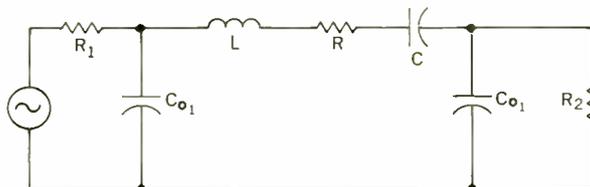


FIGURE 14. An approximate circuit of a basic three-electrode resonator, including generator and load resistances R_1 and R_2 , and with $n = 1$.



in terms of its loaded Q .^{9,13} Note, however, that the center frequency f_o —in contrast to the conventional IF transformer—is critically affected by the loading at both the input and output of the filter. If, for example,

$$R_1 \ll \frac{1}{2\pi f_o C_{01}} \quad (9)$$

and if

$$R_2 \ll \frac{1}{2\pi f_o C_{o1}} \quad (10)$$

then

$$f_o \approx \frac{1}{\sqrt{LC}} \quad (11)$$

whereas if

$$R_2 \gg \frac{1}{2\pi f_o C_{o1}} \quad (12)$$

then

$$f_o \approx \frac{1}{\sqrt{LCC_{o1}(C + C_{o1})}} \quad (13)$$

If $n \neq 1$, the basic filter response is not changed except for an input/output impedance transformation.

The preceding analyses of the two- and three-electrode circuits are based on several simplifying assumptions; in particular, the termination impedances are assumed to be purely resistive. If significant and known reactive components are present, they can be accounted for by adjusting the resonator parameters.

Another general observation regards the lack of a dc path in ceramic resonators and the resulting gain reduction. In the circuits shown in Figs. 8 and 9, for example, the collector direct current is supplied through R_c . For minimum dc voltage drop, R_c should be small; for maximum signal output, however, R_c should be of the order of the collector impedance, which may be several hundred kilohms. This dilemma occurs with both single- and multiresonator ceramic filters, and may be solved by a compromise for the value of R_c and the achievable gain. An alternative is the substitution of R_c by a choke or IF transformer, but this method is becoming less attractive because of the low cost of transistor gain.

Multiresonator ceramic filters

Most ceramic filters in consumer applications consist of multiresonator structures. The trend toward integrated circuits and toward the reduction of the number of leads to the IC package will probably further enhance the lumped-filter concept.

A very important aspect of filters is their sensitivity to tolerances and stability of their network elements. This is especially true for ceramic filters, whose center frequencies cannot be adjusted like those of IF transformers. Obviously, less critical tolerances mean better manufacturing yield and efficiency. For this reason, an economical filter should not only have a minimum number of elements, but its resonators should be as simple as possible. Nevertheless, some multiresonator filters are using composite or three-electrode resonators, and they offer advantages in some applications. However, it seems that with the exception of coupled-mode filters in the MHz range, the simple two-electrode resonator is the most economical and versatile building block for ceramic filters.

The multiresonator filters subsequently described are all directly or indirectly related to ladder or lattice configurations. These are three- or four-terminal structures in which the 3-dB bandwidth is no longer a simple function of center frequency and loaded Q , and cannot be varied in a predictable manner by merely adjusting the termination

impedances. Instead, the filter is designed for the desired performance and must be matched to the predetermined termination impedances.

It is worth noting that for application in integrated circuits, LC filters are also being converted from the separate IF transformer approach to lumped structures. Because of lower Q values, their performance with regard to insertion loss and steepness of filter skirts cannot match that of ceramic filters.

Composite filters using three-electrode resonators have been described.⁹ They are essentially ladder structures consisting of cascades of resonators coupled either directly or by series or shunt capacitors. Figure 15 shows a two-resonator version.

If the interelectrode capacitance C' of Fig. 3(A), is negligible, the circuit corresponds to two intercoupled single-tuned circuits and consequently produces the selectivity of a double-tuned LC circuit. If C' is not negligible, insertion-loss peaks appear at finite frequencies in the upper or lower stopband. Beyond these peaks, the insertion loss declines, and rises to peaks again at zero and infinite frequency.

Ladder filters

To understand the characteristics of ceramic ladder filters, consider the basic two-resonator ladder (also referred to as L section) of Fig. 16, which includes the load R_2 and the generator with its internal resistance R_1 . Let the series and parallel resonant frequencies of resonators Z_1 and Z_2 be f_{r1} , f_{a1} , and f_{r2} , f_{a2} , respectively.

If, for example, $f_{r1} = f_{a2} = f_o$, one obtains maximum signal transmission to the load at $f = f_o$. At f_{r2} (which is smaller than f_o) and at f_{a1} (which is larger than f_o) one obtains minimum signal transmission (corresponding to insertion-loss peaks). Because of the limits on Δf , these peaks will be in the vicinity of f_o and assure good passband selectivity. At frequencies below f_{r2} and above f_{a1} , both resonators act like a capacitive voltage divider. The larger the ratio C_{o2}/C_{o1} , the larger is the insertion loss in the stopband.

Figure 17 shows the response for given resonator frequencies f_{a1} and f_{r2} and for C_{o2}/C_{o1} ratios of 100, 30, and 10 [curves (a), (b), and (c), respectively]. The stopband rejection is relatively constant over a large frequency range and rises to peaks again at very low and very high frequencies.

Ceramic ladders with n series resonators and m shunt resonators may have a response with up to n and m separate insertion-loss peaks in the lower and upper stopband, respectively.

The main advantage of ceramic ladders is their relative insensitivity to element variations. For a quantitative illustration, the nominal parameters of the filter with the response shown in Fig. 17(B) were changed as follows: (1) $R_2 \pm 50$ percent; (2) $R_1 \pm 50$ percent; (3) $R_1 + 50$ percent and $R_2 + 50$ percent; (4) $R_1 - 50$ percent; and $R_2 - 50$ percent; (5) $f_{r2} \pm 0.5$ percent; (6) $f_{a1} \pm 0.5$ percent; (7) $f_{a2} \pm 0.5$ percent; (8) $f_{r1} \pm 0.5$ percent; and (9) $Q - 50$ percent.

In cases (1) to (6), the passband response between the 20-dB limits is essentially unchanged, and the only significant variation occurs for cases (1) and (2), with changes of approximately ± 2 dB in stopband insertion loss and for cases (5) and (6) with a 0.5 percent shift of the insertion-loss peaks at f_{r2} or f_{a1} , of Fig. 17. Cases

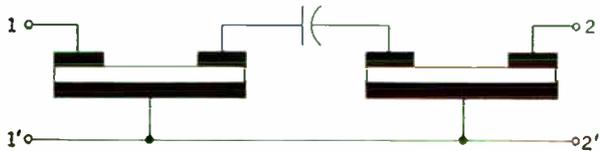


FIGURE 15. A two-resonator version of a composite filter using three-electrode resonators.

FIGURE 16. Basic two-resonator ladder filter network.

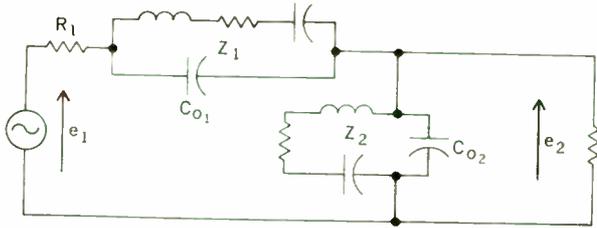


FIGURE 17. Response for given resonator frequencies f_a and f_r , with C_{o2}/C_{o1} ratios of (A) 100, (B) 30, and (C) 10.

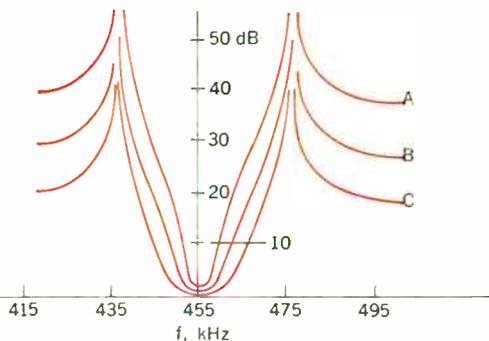


FIGURE 18. Ceramic ladder filter used in conjunction with an LC circuit.

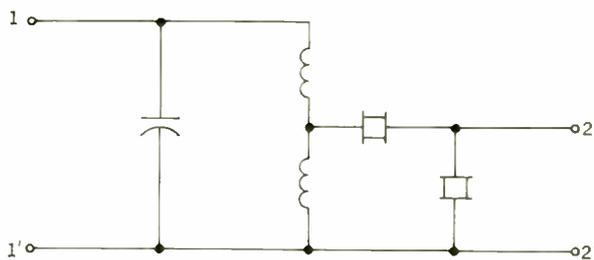
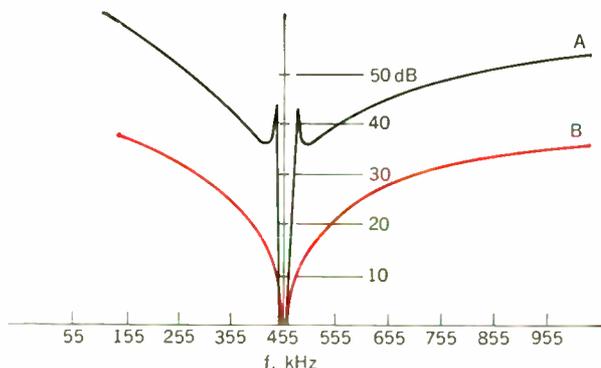


FIGURE 19. Response (A) of the LC-ceramic ladder network compared with response (B) of a single-tuned circuit of the same bandwidth.



(7) and (8) result in approximately ± 0.2 percent shift in center frequency (affecting only the response between approximately the 15-dB levels) and in case (9) the insertion loss at f_o doubles and leaves the remaining response unchanged. Variations of the shunt capacitances C_{o1} and C_{o2} have the same effect as changing the termination impedances. For example, an increase in both C_{o1} and C_{o2} by 50 percent corresponds to changing both R_1 and R_2 by -50 percent.

As mentioned before, the ceramic frequency-temperature coefficient can be controlled by processing to be positive or negative. In ladder filters, the series resonators normally have positive and the shunt resonators negative temperature dependence. This tends to improve the center-frequency stability. Further, it counteracts the band-narrowing effect of the declining Q toward higher temperatures.

The main drawback of ladders with only a few resonators is the low stopband rejection. Presently, the practical upper limit on the C_{o2}/C_{o1} ratio of Fig. 16 is about 40. For a two-resonator ladder, this corresponds to a stopband rejection of about 28 dB. There are a number of ways of alleviating this problem.

1. Increase the number of resonators. This depends on economic feasibility. For a C_{o2}/C_{o1} ratio of 40, each additional resonator would add about 14 dB to the stopband rejection. If the number of resonators is odd, the passband response will be slightly asymmetric. If, for example, the network of Fig. 16 were completed to a T section by adding another series resonator Z_1 , the insertion-loss peak at f_{a1} would be higher (due to parallel resonance in the two series resonators) than the insertion-loss peak at f_{r1} .

2. Use additional capacitors. For example, by connecting a capacitor $C = C_{o2}$ in parallel to Z_2 of Fig. 16, the parallel capacitance of the shunt resonator could effectively be doubled, thereby raising the stopband rejection. However, this procedure has limits because it reduces the effective Δf of the shunt resonator. Other resonator-capacitor ladders that are not limited in this way will be discussed later.

3. Use one or more LC circuits in conjunction with ceramic ladders. This can produce a very efficient filter since it combines the steep-skirt characteristics of the ceramic ladder with the high stopband rejection of the LC circuit.⁴

Figure 18 shows a simple version of such a network. Curve (A) of Fig. 19 is its response when designed for the same capacitance ratio, C_{o2}/C_{o1} , and the same frequencies, f_{a1} and f_{r1} , used to obtain the response shown in Fig. 17(B) of the transformerless L section. The comparison shows that the introduction of the transformer raised the minimum stopband rejection by about 16 dB. Curve (B) in Fig. 19 shows, for reference, the response of a single-tuned circuit with the same 3-dB bandwidth as curve (A).

In terms of interchangeability in conventional IF circuitry, the network of Fig. 19 offers several advantages, including dc continuity at the input terminals, flexible impedance transformation, and suppression of spurious resonances.

Lattice filters

Figure 20(A) shows a full symmetric lattice network, Fig. 20(B) its half (or hybrid) lattice equivalent whose

transformer is ideal—it has infinite inductance, unity coupling, and no losses. The only purpose of the ideal transformer is to unbalance the network and to provide a 180-degree phase inversion between the voltages v_1 and v_2 with respect to ground. The transformer could, for example, be replaced by an active phase inverter.

Most lattice filters are based on the network shown in Fig. 20(B) because it provides a common ground for input and output terminals and requires only half the number of branch impedances shown in Fig. 20(A). Frequently the transformer then is made as ideal as possible—with as large an inductance as is possible—and the finite inductance is neutralized by a shunt capacitor. In this case, the contribution of the transformer to the filter selectivity is negligible (for later reference, we call this “case A”). If, however, the transformer is made intentionally nonideal, then the filter network may be considered and designed as a tuned circuit coupled to a symmetric lattice.¹⁵⁻¹⁷ and the tuned transformer may contribute significantly to the filter selectivity (call this “case B”).

Z_1 and Z_2 of Fig. 20(B) may be reactances of any complexity. With economical filters in mind, we shall only consider the two following cases:

1. Z_1 corresponds to a ceramic resonator with shunt capacitance C_{o1} , and Z_2 represents a capacitor $C_a = C_{o1}$. Figure 21 shows the response of this circuit for case A and case B with both curves designed for the same 3-dB bandwidth.

Using a well-known method of lattice decomposition, it can be shown that case A is equivalent to the network of Fig. 22, where the *LRC* branch is identical to the series branch of Z_1 . In other words, the case A filter response will be that of a single-tuned circuit; case B approximates that of a double-tuned circuit.

If $C_2 > C_{o1}$ or $C_2 < C_{o1}$, the response becomes asymmetric due to the introduction of insertion-loss peaks at finite (rather than at zero or infinite) frequencies in the lower or upper stopband, respectively.

2. Z_1 and Z_2 each correspond to ceramic resonators with shunt capacitances C_{o1} and C_{o2} respectively. The response has insertion-loss peaks at zero and infinite frequencies and may have an additional peak each in the lower and upper stopband whose positions depend on the ratio C_{o2}/C_{o1} .

Figure 23 illustrates typical responses for case A and case B filters designed for the same 3-dB bandwidth and arbitrary skirt steepness.

The increased insertion loss for case B in both Figs. 21 and 23 is due to the reduction of the transformer inductance to a small value such that its equivalent parallel resistance (for an unloaded transformer Q of 80) absorbs a substantial portion of the signal. As in Fig. 21, case B provides significantly more selectivity than that provided by case A.

The ceramic lattice avoids the previously mentioned handicaps of the ladder, such as the extreme C_{o2}/C_{o1} ratios and the limited mobility of the insertion-loss peaks. Further, for a given Δf of the resonators, lattice filters can be designed with wider bandwidths than ladders. For this reason, for example, the lattice filters of Figs. 21 and 23 are based on using the high- Q , narrow Δf material of Table I, whereas the previously shown ladder filter responses all depended upon using the lower- Q , wide- Δf material.

The main drawback of the lattice is its sensitivity to network parameter variations. The networks of both Fig. 20(A) and 20(B) are bridge circuits. Their performance depends on the balance of the impedances Z_1 and Z_2 and on the symmetry of the transformer. As a result, the filter response is especially sensitive to variations in the transformer balance or the capacitance ratio C_{o1}/C_{o2} . The latter is illustrated in Fig. 24, where curve (A) is identical to the class A response of Fig. 23 and curves (B) to (E) were obtained from curve (A) by changing C_{o1}/C_{o2} from 0.9 to 0.96, 1, 1.04, and 1.1, respectively [this presumes $f_{r2} > f_{r1}$ for the resonant frequencies of resonators Z_1 and Z_2 of Fig. 20(B)].

The results of Fig. 24 could be duplicated by maintaining the nominal C_{o1}/C_{o2} ratio and upsetting the balance of the center-tapped transformer of Fig. 20(B). Since the ratio C_{o1}/C_{o2} is critical, the temperature dependence of the filter does not suffer as long as C_{o1} and C_{o2} have the same temperature coefficient.

Figure 24 demonstrates that even if the transformer is perfectly balanced, changes of ± 2 percent in the branch impedances produce response variations that are normally unacceptable in IF circuitry. Generally, the branch resonators will have to be sorted and matched in pairs. This characteristic represents a significant economic handicap.

Another limitation of the lattice in its conventional form, Fig. 20(B), is the requirement for a wire-wound and balanced inductance. Several ways of using active phase inverters instead of transformers have been suggested.^{18,19} This type of approach is especially attractive for integrated-circuit applications.

Inductorless IF circuits

The strong modern trend toward eliminating wound inductors from discrete, and especially integrated, IF circuitry was discussed previously; also mentioned were the reasons why, in some cases, a combination of *LC* circuits with ceramic filters may be desirable or (as in lattice networks) necessary. However, the overriding reason for using *LC*-ceramic combinations is the need for the suppression of spurious responses.

Every resonating body has harmonic or nonharmonic overtones of its desired and undesired modes of vibration. For example, the resonator of Fig. 1 has radial-mode overtones around k_i times the fundamental frequency f_r , where $k_i \approx 2.5, 4, 5.5,$ and $7,$ respectively, for the first to fourth radial overtones. Figure 25 shows the impedance vs. frequency response of a 430-kHz radial-mode resonator with unsuppressed overtones.

Further, the same resonator has thickness-mode responses. The fundamental thickness-mode frequency depends on the resonator and electrode geometry. For the materials of Table I this value is about $f_{t1} = 254/T$, where f_{t1} is measured in kilohertz and the thickness T is measured in centimeters. Figure 25 refers to a resonator with $T = 0.0254$ cm whose thickness resonance lies beyond the plotted frequency range.

The undesired resonances are called spurious responses. To account for them, the approximate equivalent circuit shown in Fig. 2 should be expanded to that illustrated by Fig. 26.

Due to spurious responses, ceramic filters have spurious passbands. With radial resonators operating at the fundamental radial frequency f_r and with $D > T$, all

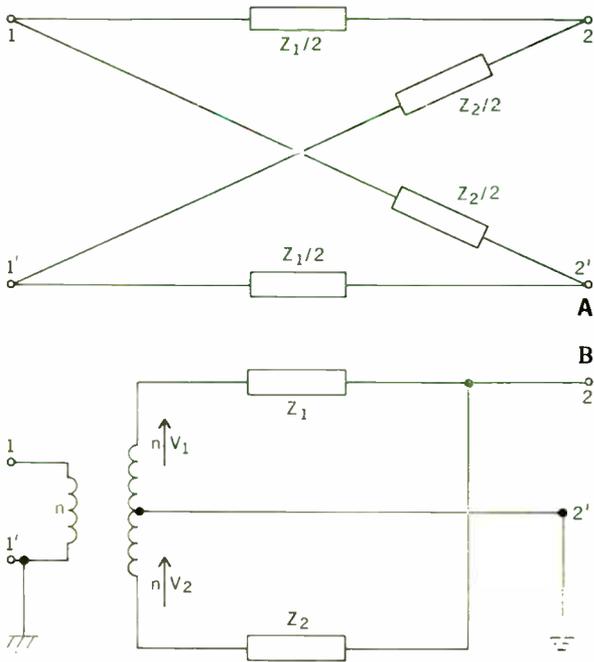


FIGURE 20. Full symmetric lattice network (A), with its half-lattice (or hybrid-lattice) equivalent (B) whose transformer is ideal.

FIGURE 21. Response of the hybrid lattice circuit with ideal (black curve) and intentionally nonideal (colored curve) transformers.

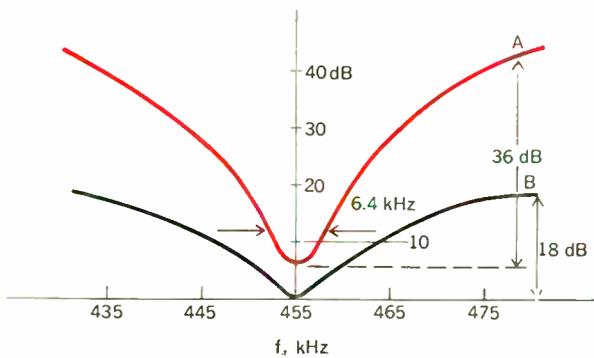


FIGURE 22. Schematic equivalent of hybrid lattice network with ideal transformer using a resonator in one branch and a capacitor in the other.

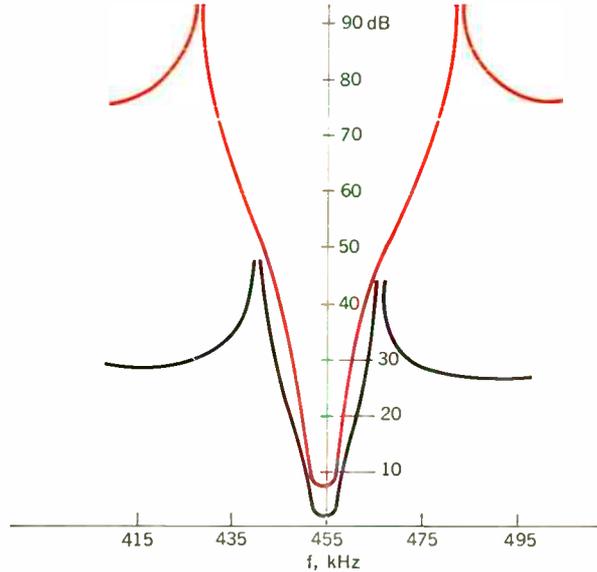
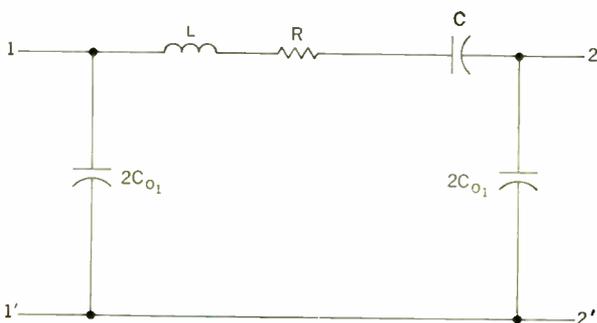


FIGURE 23. Typical responses for two-resonator hybrid lattice filters, using ideal (black curve) and nonideal (colored curve) transformers at identical 3-dB bandwidths.

FIGURE 24. Response of a two-resonator hybrid lattice filter with ideal transformer, compared with responses derived from the curve by varying C_o/C_{o1} from (A) 0.9 (original value) with (B) 0.96, (C) 1.0, (D) 1.04, and (E) 1.10.

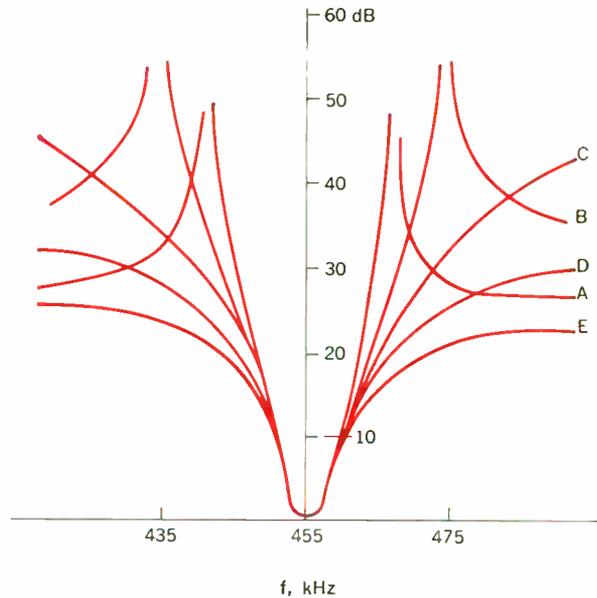
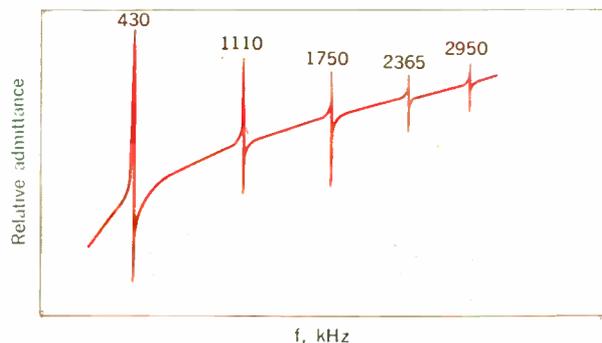


FIGURE 25. Relative admittance vs. frequency response of a 430-kHz radial-mode resonator with unsuppressed overtones.



spurious passbands lie at frequencies $f > f_r$. Figure 27 shows the wide-spectrum response of a commercial ceramic ladder filter. Note the spurious peak at $f = 2.1$ MHz, which is due to thickness resonance of the series resonators whose T is 0.12 cm.

There are a number of ways of dealing with responses due to undesired vibrational modes. We mentioned, for example, that radial-mode resonators are spring-mounted at their node of radial vibration. This mechanically dampens the thickness vibrations. Further, by making $T \ll D$, the thickness responses can be located at frequencies far removed from the desired passband.

Until recently, little had been done about the overtones of the desired vibrational mode. The reason was that a single-tuned circuit with a typical loaded Q of 35 and tuned to the center frequency f_0 of a ceramic band-pass filter contributes almost 40-dB stopband rejection at the frequency $f \approx 2.5 f_0$ of the first radial overtone. In conjunction with the filter shown in Fig. 27, for example, this simple LC circuit would be sufficient to suppress all spurious responses to about 60 dB below the passband level.

Spurious responses within the individual resonator may be suppressed by optimizing both the processing and the geometry of the resonator. However, some of these approaches are not compatible with economy. For instance, one known method completely eliminates all radial overtones, but results in a more costly resonator and a

bulkier, less rugged package. On the other hand, complete elimination of one radial overtone, or partial suppression of two neighboring radial overtones, may be achieved by controlling the resonator geometry without sacrificing economy, size, or ruggedness.

Whatever the method, not all spurious responses within the resonator can be eliminated. However, the remaining ones can be confined to frequencies $f \gg f_r$ where both their amplitude and the amplifier gain are reduced.

Additional spurious suppression may be obtained by proper design of the filter network. As was mentioned previously, efficient spurious suppression within the resonator requires a specific resonator geometry, in particular, a limited range for the ratios D/T and D/D' of the device shown in Fig. 1, or a limited range for the resonator impedances. This condition can easily be achieved in lattice filters, but it excludes the previously described ladder shown in Fig. 16 with its extreme C_{02}/C_{01} ratio.

Note that this extreme capacitance ratio was needed to raise the stopband rejection. The same purpose can be served by combining resonators of equal or approximately equal impedance levels (i.e., $C_{02}/C_{01} \approx 1$) with capacitors in a ladder network. Figures 28(A), (B), and (C) show a few examples of basic sections, which may also be cascaded to obtain more selective filters.

Networks illustrated by Fig. 28(B) and (C) are especially interesting for "economy" filters since they may be

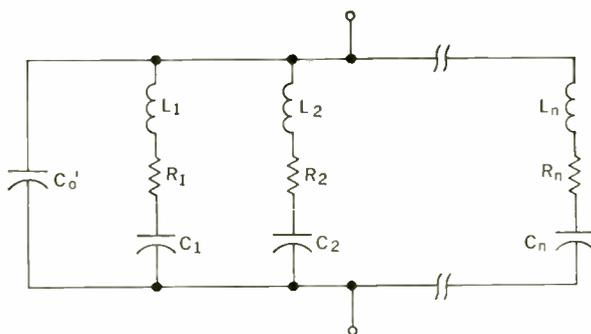


FIGURE 26. An expansion of the approximate equivalent circuit shown in Fig. 2. This circuit explains undesired, or spurious, responses.

FIGURE 27. Wide-spectrum response of a commercial ceramic ladder filter. Spurious peak at 2.1 MHz is due to thickness resonance ($T = 0.092$ cm).

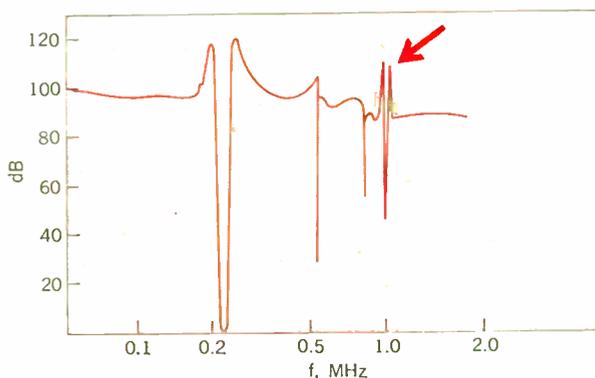
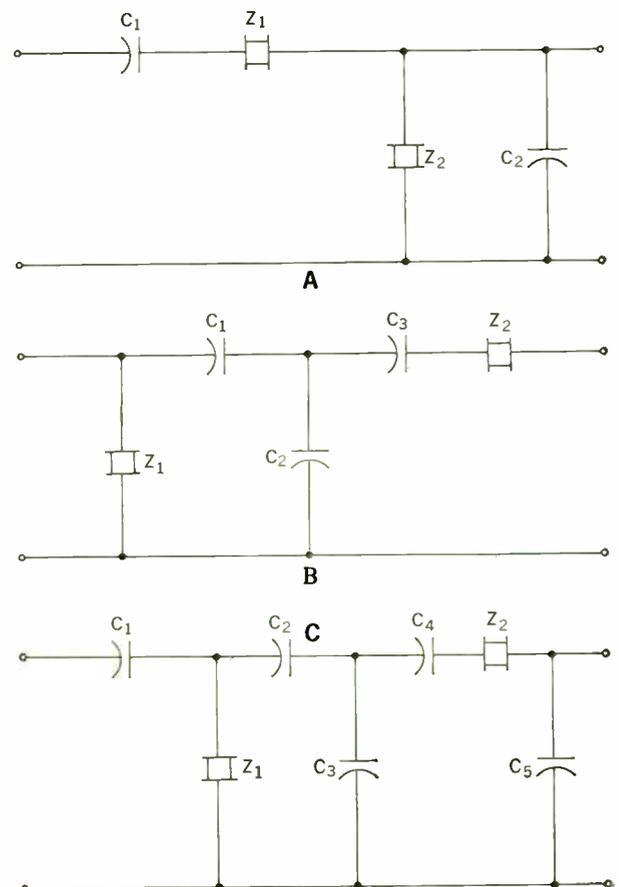


FIGURE 28. Three examples of basic resonator-capacitor circuits.



designed to employ identical resonators $Z_1 = Z_2 = Z$ in both the series and shunt branches and allowed to cover a continuous range of bandwidths and stopband rejections by merely changing the capacitors.²⁰ Each of the response curves in Fig. 29 corresponds to a different set of capacitors C_1 to C_5 and an identical resonator Z . For a quantitative example assume a shunt capacitance of 400 pF for resonator Z . Then one obtains the response shown in Fig. 29(B) by adjusting capacitors C_1 to C_5 to the values 220, 450, 1300, 800, and 750 pF, respectively. The resulting network has an input impedance of 2.1 kilohms and an output impedance of 0.36 kilohm.

Aside from the spurious suppression within the resonators, the networks shown in Fig. 28(B) and (C) contribute additional suppression due to capacitive voltage division. In conjunction with resonators having partially suppressed first and second radial overtones, this type of filter has no spurious passband up to the thickness-mode responses, which appear around 7 MHz and are partially suppressed. If necessary, their effect may be further reduced or eliminated by low-pass filtering in the associated active circuitry.

Circuit applications

So far, application work has lagged, in that it has either concentrated on customer applications with specific constraints (for example, direct exchangeability with conventional IF transformers) or lacked sufficient collaboration and communication between the circuit designer and the filter designer.

A few applications of single two-electrode resonators were discussed previously, and a number of modifications and refinements have been suggested. For example, Fig. 11 indicates that in order to make full use of the resonator selectivity, the resistor R_c in the emitter-bypass

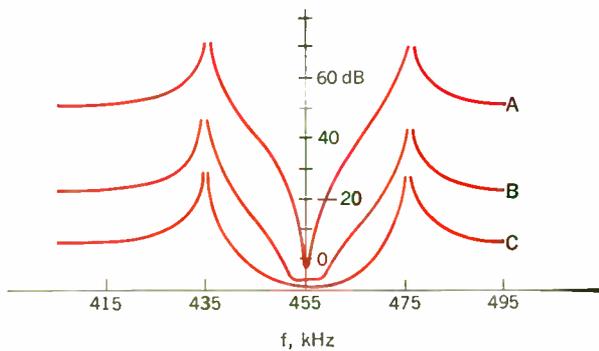
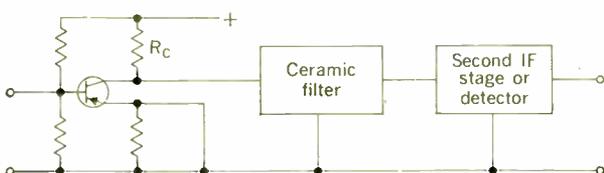


FIGURE 29. A set of response curves derived for the network shown in Fig. 28(C).

FIGURE 30. An example of applying ceramic filters to transistor IF circuits.



configuration shown in Fig. 8(B) should be as large as possible. This may be done by replacing R_c by the constant-current-source emitter-collector path of a second transistor to provide an adjustable, and potentially very large, resistance in parallel to the resonator.²¹

The use of three-electrode resonators for selective and impedance transforming intercoupling of IF stages has been described previously.⁹ The basic circuit is shown in Fig. 30 and is principally applicable to ceramic filters. As explained in conjunction with the circuits shown in Figs. 8 and 9, the compromise in choosing the value for R_c reduces the gain of the stage unless the collector direct current is supplied through a choke or IF transformer.

The terminal impedances of ceramic filters normally range from 0.5 to 5 kilohms. The filter output is well adaptable to the input of common-emitter stages. Since the transistor's collector impedance generally is rather high, proper match at the filter input may be obtained if R_c equals the filter input impedance.

The filter response is relatively insensitive to impedance mismatch, as described earlier for ladder filters. However, if large circuit impedance variations—due to strong AVC, for example—are anticipated, they may have to be counteracted by special means, such as resistive divider networks.

Figures 31 and 32 illustrate methods of applying the ladder filter shown in Fig. 18, or lattice filters, to common-emitter circuits. The presence of the LC circuit at the filter input facilitates direct replacement in conventional IF transformer networks.

Ceramic filters offer some features that are not easily

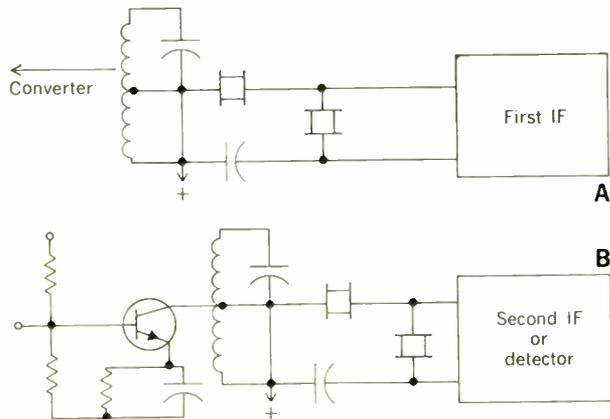
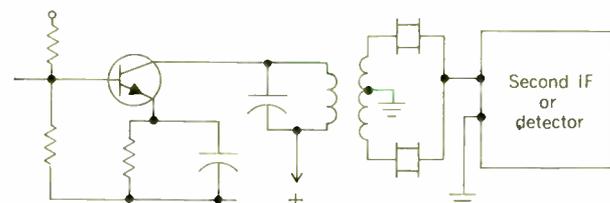


FIGURE 31. Two examples of applying ladder filters to IF circuits.

FIGURE 32. Application of a ceramic lattice filter to an IF circuit.



obtained in the conventional (*LC*-coupled stages) manner. For example, the response of ceramic ladders suggests a way to eliminate adjacent-channel interference by placing the insertion-loss peaks at the adjacent-channel frequencies. Some high-fidelity applications specify 60-dB adjacent-channel (at 455 ± 10 kHz) rejection and a 6-dB bandwidth of 8 kHz. Four double-tuned circuits would be needed to match these requirements, which could also be satisfied with a four-resonator ladder filter of the type shown in Fig. 18.

Typical United States automobile-radio IF filters have two double-tuned transformers, and their selectivity may be duplicated by a number of ceramic filter configurations. Most circuits devised so far maintain a single-tuned *LC* circuit following the converter. The remaining selectivity can be provided by a filter of the type illustrated in Fig. 18, by a T or pi section of two-electrode resonators, or by three two-electrode resonators, combining an L section similar to that shown in Fig. 16 with one of the approaches shown in Figs. 8 and 9. Incidentally, ceramic IF filters for auto radios are preferably centered around $f_o = 445$ kHz instead of the traditional 262.5 kHz because of the considerable reduction in resonator size. One may recall that the historic reason for the lower IF in automobile radios was the limit for the unloaded *Q* of *LC* components—a reason that is no longer valid for high-*Q* ceramics.

A similar variety of choices is available for the relaxed selectivity specifications, but stringent cost requirements, of portable-radio IF amplifiers. For example, the circuit of Fig. 31(A) can produce an IF response of the form shown in Fig. 19(A) with a minimum stopband rejection of 46 dB, and bandwidths of 8, 20, and 31 kHz at the respective attenuation levels of 6, 23, and 40 dB. Other alternatives are filters like those shown in Figs. 15, 16, or 28, used in conjunction with a single-tuned circuit. Intermediate-frequency amplifiers without *LC* circuits are feasible when used with spurious-free ceramic networks similar to those described earlier.

Table II lists some selectivity specifications of various economical 455-kHz filters. B_6 and B_{40} are the bandwidths at the 6- and 40-dB level, respectively. $R_{\pm 10}$ is the rejection at the adjacent channel frequencies ± 10 kHz off the center frequency, and R_s is the minimum stopband rejection. The table refers to nominal design values, which do not take into account parameter tolerances, and the filter types correspond to configurations such as (A) identical-resonator structure illustrated by Fig. 28(B); (B) identical-resonator structure obtained by

II. Selectivity specifications for various 455-kHz ceramic filters

Filter type	B_6 , kHz	$R_{\pm 10}$, dB	B_{40} , kHz	R_s , dB
A	4.7	29	31	39
A	4.7	27	36	46
B	8.5	46	17	49
B	8.4	43	19	65
C	8.0	23	31	46
D	8.0	60	17	60
E	5.5	33	25	55
E	7.0	29	27	52
E	5.5	31	27	67
E	7.0	27	31	62

cascading two sections [Fig. 28(C)]; (C) ladder structure shown in Fig. 18; (D) ladder structure similar to that shown in Fig. 18, but with four resonators; and (E) identical-resonator structure illustrated in Fig. 28(B), combined with an *LC* transformer at the filter input.

Ceramic filters in integrated circuits

The ultimate goal of IC designers presumably is to combine all circuit functions, including the IF section, into one monolithic structure. As was mentioned previously, there is at present no economically feasible means to this end.

Ceramic filters, in their present form, cannot be integrated monolithically. However, hybrid integrated circuits have been made using thickness-mode ceramic IF filters. Also, radial-mode resonators become small enough at higher frequencies (> 1 MHz) to be included in IC packages. They are not economically competitive with conventional IF transformers at present; however, some experimental circuits are being developed using this technique, and pilot quantities are being evaluated for high-volume commercial systems.

The reason for distinguishing between discrete-component and integrated-circuit applications is the assumption that ICs will sooner or later require inductorless filters. The requirements for this type of filter were discussed earlier, and some inductorless ladder filters were mentioned. This section concerns preliminary results for inductorless lattice filters, where differential transformers are replaced by differential amplifiers¹⁹—structures that are common in ICs but would not be economical for discrete-component IF circuits.

The hybrid lattice filter shown in Fig. 20(B) is equivalent to the full lattice shown in Fig. 20(A) only if the differential transformer is ideal. The lattice shown in Fig. 20 has the transfer function

$$r(s) = \frac{e_o}{e_{in}} = \frac{R(Z_b - Z_a)}{2[R^2 + R(Z_a + Z_b) + Z_a Z_b]} \quad (14)$$

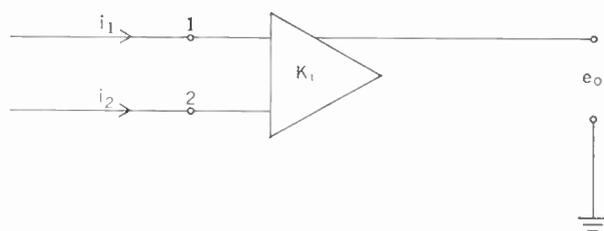
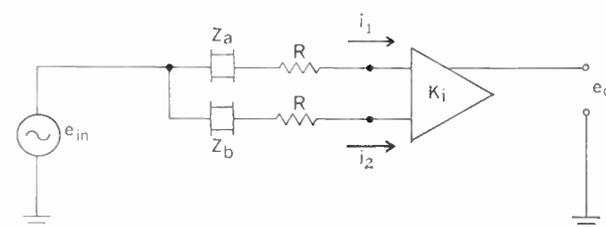


FIGURE 33. An ideal differential current amplifier.

FIGURE 34. An active filter that is equivalent to a full-lattice filter.



where the generator and load resistance is R .

An ideal differential current amplifier is shown in Fig. 33. Pins 1 and 2 are both assumed to be at ground potential, and the bandwidth is assumed to be infinite. The output voltage e_o is given by

$$e_o = K_i(i_1 - i_2) \quad (15)$$

If resistors and reactive elements (later to be ceramic resonators) are combined with the ideal differential current amplifier as illustrated in Fig. 34, and it is analyzed using Eq. (15) to find the transfer function, the result is

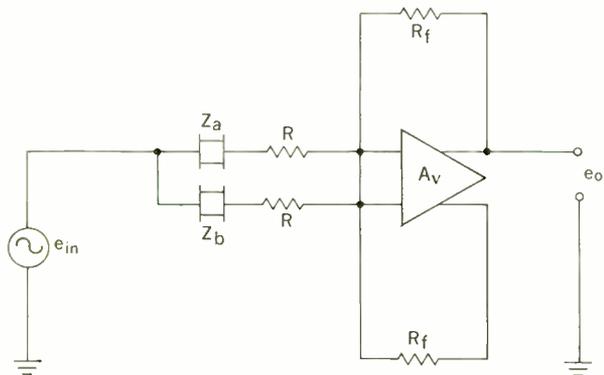


FIGURE 35. A practical utilization of Fig. 34, using a differential-input, differential-output operational amplifier.

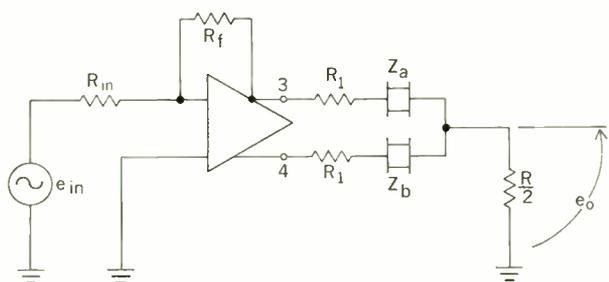
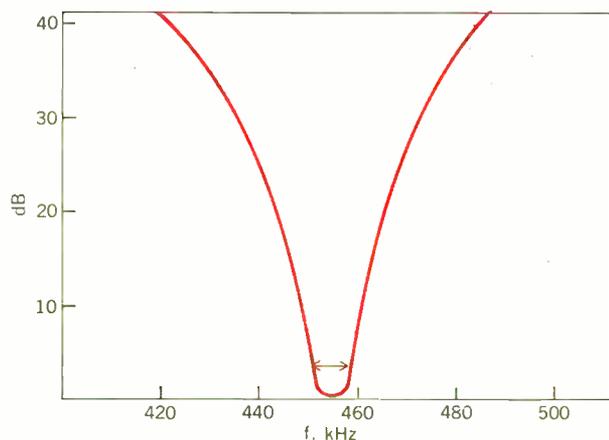


FIGURE 36. Operational amplifier with possibility for higher gain due to the fact that ratio R_f/R_{in} sets voltage gain.

FIGURE 37. Response curve of an experimental filter similar to that shown in Fig. 36.



$$\frac{e_o}{e_{in}} = \frac{K_i(Z_b - Z_a)}{R^2 + R(Z_a + Z_b) + Z_a Z_b} \quad (16)$$

Note that except for the gain constant, Eqs. (16) and (14) are exactly the same. This shows that an ideal differential current amplifier can replace the ideal differential transformer of a hybrid lattice circuit.

The ideal differential current amplifier can be approximated using the differential-input, differential-output operational amplifier shown in Fig. 35.²²

If the operational amplifier is ideal (has infinite gain, infinite bandwidth, infinite input impedance, and zero output impedance), the circuit shown in Fig. 35 is exactly equivalent to Fig. 34. Real operational amplifiers, with high input impedance and low output impedance, will be nearly ideal if the ratio R_f/R is small compared with the open-loop voltage gain.

The equivalence between the symmetrical lattice loaded at both input and output with resistance R and the circuit shown in Fig. 35 holds only if the voltage source in Fig. 35 has zero internal impedance. The assumption of a zero source impedance may not be practical. In addition, the circuit shown in Fig. 35 has the disadvantage of relatively small gain, since the ratio R_f/R must be small in comparison with the open-circuit voltage gain of the amplifier. The circuit illustrated in Fig. 36 does not suffer from this disadvantage. The ratio R_f/R_{in} sets the voltage gain, which must be less than, or equal to, the open-circuit voltage gain.

The voltage at pin 3 is $-(R_f/R_{in})e_{in}$, and that at pin 4 is $(R_f/R_{in})e_{in}$.

Analysis of the circuit of Fig. 36, assuming $R_f = R/2$, gives the transfer function

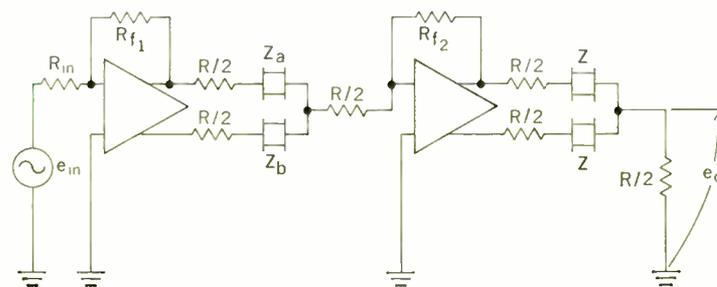
$$t(s) = \frac{e_o}{e_{in}} = -\frac{R_f}{R_{in}} \frac{Z_b - Z_a}{\frac{3}{4}[R^2 + R(Z_a + Z_b) + Z_a Z_b]} \quad (17)$$

Except for gain constants, the only difference between (14) and (17) is the factor $\frac{3}{4}$ for R^2 .

Figure 37 shows the response of an experimental filter of the type illustrated by Fig. 36, designed for a 3-dB bandwidth of 8 kHz. It corresponds approximately to the selectivity of a critically coupled double-tuned LC circuit.

Equation (17) is used in a procedure developed to design ceramic active bandpass filters with either Chebyshev or Butterworth polynomial characteristics or filters with insertion-loss peaks at finite frequencies. Each basic sec-

FIGURE 38. Filters with greater selectivity can be designed by cascading sections.



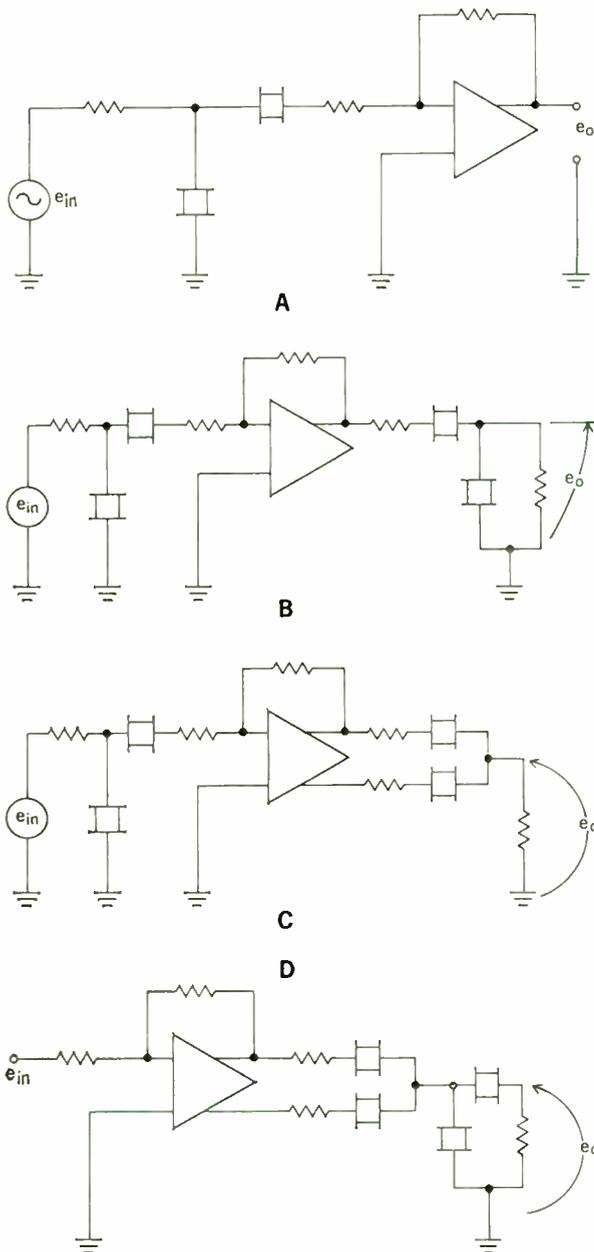
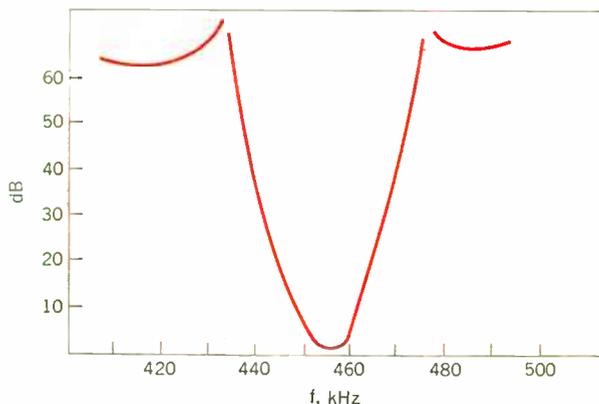


FIGURE 39. Alternate methods of combining ceramic filters with operational amplifiers.

FIGURE 40. Measured response of a circuit of the type shown in Fig. 39(D).



tion of Fig. 36 corresponds to a second-order response, and more selective filters are designed by cascading the sections as in Fig. 38. Further details of this design method will be made available at a later date.²³

The previous circuits are by no means the only configurations for applying ceramic resonators in conjunction with linear integrated circuits. We have thus far only shown lattice structures with integrated circuits. Figure 39 shows several ways of using ceramic ladder filters alone and with lattice filters.

Combining a ladder and lattice with a linear IC combines the desirable properties of both. That is, the ladder can supply stable insertion-loss peaks near the passband edges, and the lattice can give increasing rejection far away from the passband. The measured response of a circuit with the structure represented by Fig. 39(D) is shown in Fig. 40.

The active lattice of Fig. 35 or of Fig. 36 may also be combined with equal-resonator ladder filters (discussed previously). Such a structure has the advantage that all resonators may be types with reduced spurious responses. The effects of spurious responses could also be reduced by using active RC filtering in the feedback loop of the differential amplifier.

REFERENCES

1. Kulesar, F., U.S. Patent No. 3 006 857.
2. Curran, D., and Gerber, W., "Piezoelectric ceramic IF filters," *Proc. IEEE Electron. Components Conf.*, pp. 160-165, May 1959.
3. Macario, R. C., "Design data for bandpass ladder filters employing ceramic resonators," *Electron. Eng.*, Mar. 1961.
4. Sauerland, F., "Design of piezoelectric ladder filters," *IEEE Internat'l Conc. Record*, vol. 13, no. 10, pp. 111-115, 1965.
5. Waren, A., and Schoeffler, J., "Approximation problem for resonator ladder filters," *IEEE Trans. Circuit Theory*, vol. CT-12, pp. 215-222, June 1965.
6. Curran, D., and Gerber, W., "Low-frequency ceramic bandpass filters," *Proc. IEEE Electron. Components Conf.*, pp. 108-113, May 1963.
7. Curran, D., and Koneval, D., "Miniature ceramic bandpass filters," *Proc. Nat'l Electron. Conf.*, vol. 17, pp. 514-520, Oct. 1961.
8. Lungo, A., and Henderson, K., "Application of piezoelectric resonators to modern bandpass amplifiers," *IRE Internat'l Conc. Record*, no. 6, pp. 235-242, Mar. 1958.
9. Lungo, A., and Sauerland, F., "A ceramic bandpass transformer and filter element," *IRE Internat'l Conc. Record*, pp. 189-203, Mar. 1961.
10. Jaffé, H., "Piezoelectricity," in *Encyclopedia Britannica*, vol. 19, 1965, pp. 910-916.
11. Sykes, R., Smith, W., and Spencer, W., "Monolithic crystal filters," *IEEE Internat'l Conc. Record*, vol. 11, pp. 78-93, Mar. 1967.
12. Terman, F., *Electronic and Radio Engineering*. New York: McGraw-Hill, 1955.
13. Munk, E. C., "The equivalent circuit for radial modes of a piezoelectric ceramic disk with concentric electrodes," *Philips Research Reports*, vol. 20, pp. 170-189, Apr. 1965.
14. Lungo, A., U.S. Patent No. 3 217 265.
15. Sauerland, F., U.S. Patent No. 3 349 347.
16. "Ceramic filters edge out costly IF transformers," *Electronics*, vol. 39, pp. 160-163, Nov. 14, 1966.
17. Hoke, L., et al., "Monolithic circuits for radio receivers," *Appliance Eng.*, vol. 1, no. 4, 1967.
18. Fogle et al., U.S. Patent No. 3 206 692.
19. Humphreys, D. S., "Active crystal filters," *Electro-Technol.*, vol. 78, July 1966.
20. This type of filter and its design will be described in *Electronic Design* in the near future.
21. Balfour, D., internal report, Brush-Clevite.
22. *Handbook of Operational Amplifier Applications*, Burr Brown Research Corp., Tucson, Ariz., 1963.
23. Ph.D thesis to be submitted to Case Western Reserve University.

Spark breakdown in air at a positive point

A technique originally developed in 1777 is still being employed in the study of electrical discharge in air. Photographic results now form a basis for improved understanding of spark breakdown phenomena

Essam Nasser Iowa State University

When a voltage across an insulating gas is increased beyond a certain limit, the gas breaks down, becoming a conductor of electricity. A photographic method, the Lichtenberg figure technique, is used to study the process. Above a certain voltage an ionization wave, called a streamer, proceeds from the highly stressed electrode, branching out along the way and extending far into low-field regions. At elevated voltages the vigorous streamers reach the cathode with a high-potential front. Owing to the front, the short-lived local field at the cathode triggers electron emission, also producing negative streamers. The negative streamers greatly increase the density of the ionized particles in the channel, yielding what is known as a backstroke. For long air gaps, streamers are incapable of reaching the cathode; intensive secondary channels therefore develop at the anode and proceed toward the cathode. Either the backstroke or the secondary channel paves the way for the full ionization of the channel and spark breakdown.

Unlike metals, gases under normal conditions have very few free electrons, making them almost perfect dielectrics or electric insulators. When subjected to high electric fields they suddenly cease to behave as insulators, becoming instead good conductors of electricity. This process is manifested by nature in the dramatic and awesome occurrence of lightning, where the air suddenly exhibits the features of a good conductor. The transition occurs by the separation of a majority of electrons from their parent molecules that produces a new medium containing free electrons and negative and positive ions. The conductive medium, known as a plasma, is utilized in such well-known components as spark plugs, lasers, fluorescent tubes, arc furnaces, and welders, and in controlled thermonuclear energy conversion.¹

In nature, plasmas exist in the upper atmosphere of the earth, on the sun, and in interplanetary space.² The understanding of this medium is therefore vital to

many areas of science and technology.

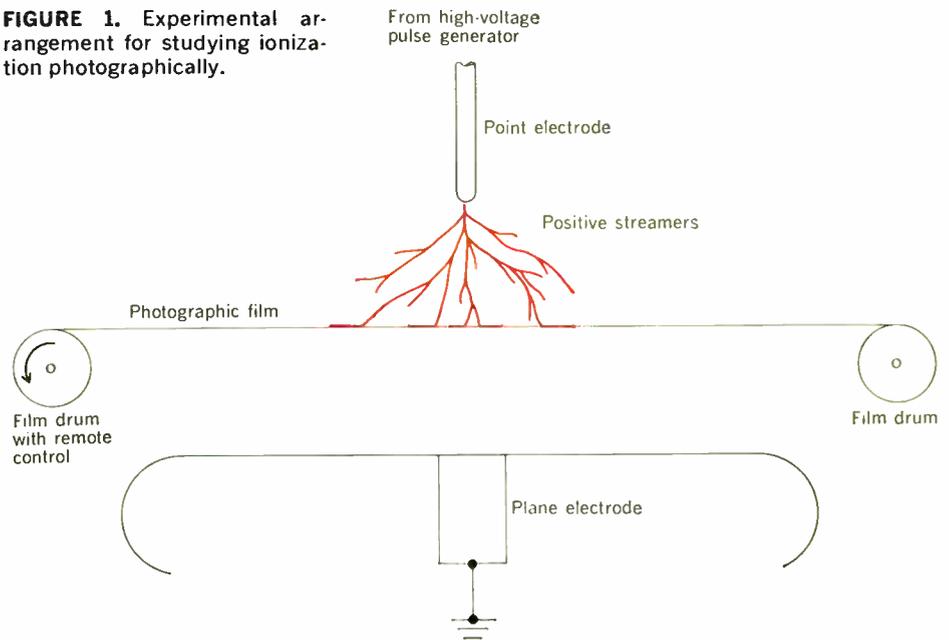
Experimenting with static electrification in 1777, Lichtenberg³ found that when rods were charged by friction and then placed on dusty surfaces, figures of peculiar patterns were produced in the dust. The patterns, which he called *Gleitfiguren* (glide figures), differ in character depending upon whether the rod was positively or negatively charged. Today, through the use of photographic plates instead of dust, more permanent figures can be produced that exhibit extremely fine detail, which, according to Merrill and von Hippel,⁴ "set a riddle akin to hieroglyphics of the Egyptian script." Without even a clear understanding of their nature, the figures were used in the 1930s to record high-voltage transients on power lines. In 1939 Merrill and von Hippel undertook the first extensive investigation to decipher the many fine details of Lichtenberg figures, providing an atomistic interpretation of what they called one of the most beautiful spectacles in science. The correct physical interpretation of some of the features of Lichtenberg figures made their use as a detecting tool possible.

When a point electrode is made positive with respect to another electrode a few centimeters away, waves of ionization or plasma trails, called *streamers*, are found to proceed from the point source to the electrode. These streamers are the first manifestation of ionization; they are accompanied by radiation so weak that it is almost undetectable with the naked eye or with ordinary photocells.⁵ Streamers may develop as a result of static potentials or rectangular voltage pulses a few tenths of a microsecond in length. Under uniform fields, a streamer is extremely difficult to observe because in a few nanoseconds it changes to a bright spark, which bridges the electrodes.

Experimental techniques

Ionization detection and recording have always been important obstacles in the study of ionization phenomena in gases. This problem was solved in these investigations

FIGURE 1. Experimental arrangement for studying ionization photographically.



by use of the Lichtenberg figure technique. In simple terms, a sensitive photographic emulsion is placed in any location at which ionization may occur. Because ionization, even when weak, is always accompanied by excitation, radiation is emitted. The radiation is difficult to detect because it is quickly attenuated by absorption before reaching the detecting device. With sensitive film placed in the path of ionization, however, distances are too short for effective absorption, and traces are produced on the film. Even radiation from an avalanche produced by a single electron is recorded. Because ionization and the accompanying excitation take place very close to the photographic emulsion, sharp and well-defined traces are obtained.⁶ When ionization is slightly off the surface of the film, vague patterns are recorded; if the distance is more than 1 cm, a general haze is produced on the film.

To study ionization leading to spark breakdown in asymmetrical gaps, the photographic film is placed perpendicular to the point axis of a point-to-plane electrode arrangement, as shown in Fig. 1. The film can be moved to any distance from the point; it also can be arranged so that the sensitive side faces either the point or the plane electrode. In some cases double-emulsion film is used to record simultaneously the ionization processes occurring at both sides of the film for correlating the involved processes. For other experiments, the film is placed parallel and close to the point axis, perpendicular to the plane. This arrangement, called the coplanar mode, serves to measure the axial and radial propagation of the streamers.

Most investigations were carried out using pulse voltages of short duration in order to avoid the formation of surface charges on the film and distortion of the initial field. Generally, the point electrode was held positive (positive point) with respect to the plane. Some results were obtained, however, with the point electrode held negative (negative point) for which the development of corona to spark is still less understood. The positive

point provides more consistent data with little spread in values; the negative point is associated with an erratic spread of data. Although the technique is applicable to static voltages, care must be exercised to prevent the charging of film and subsequent field distortion. There is less chance of this happening when pulse voltages are used.

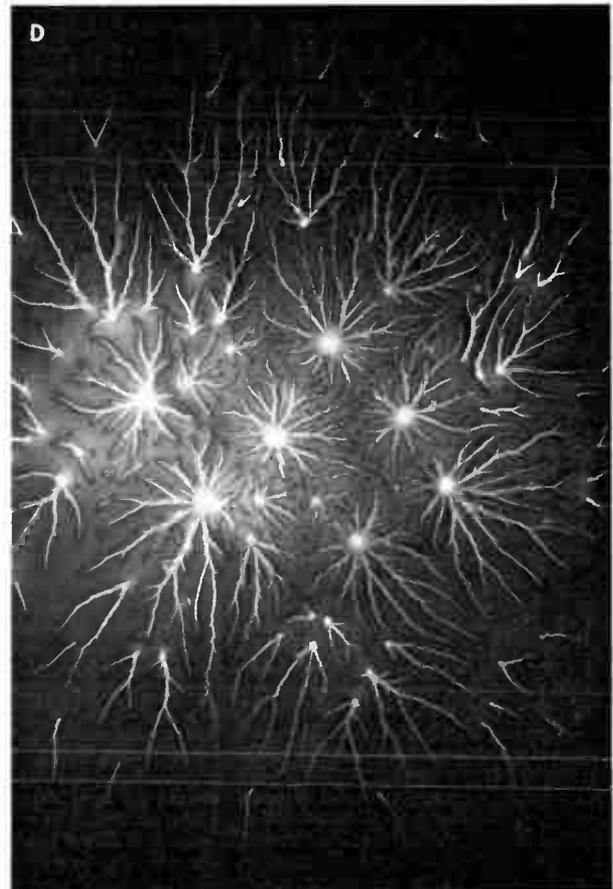
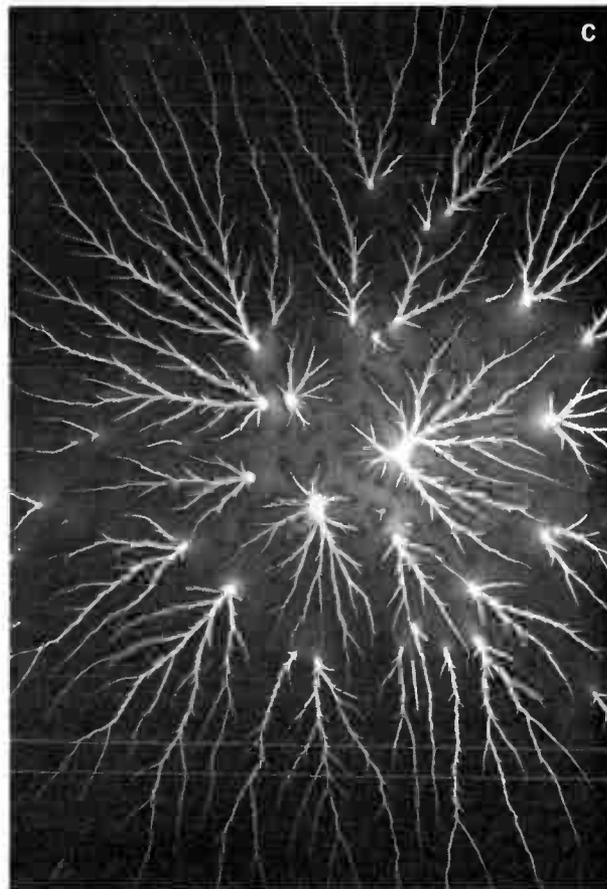
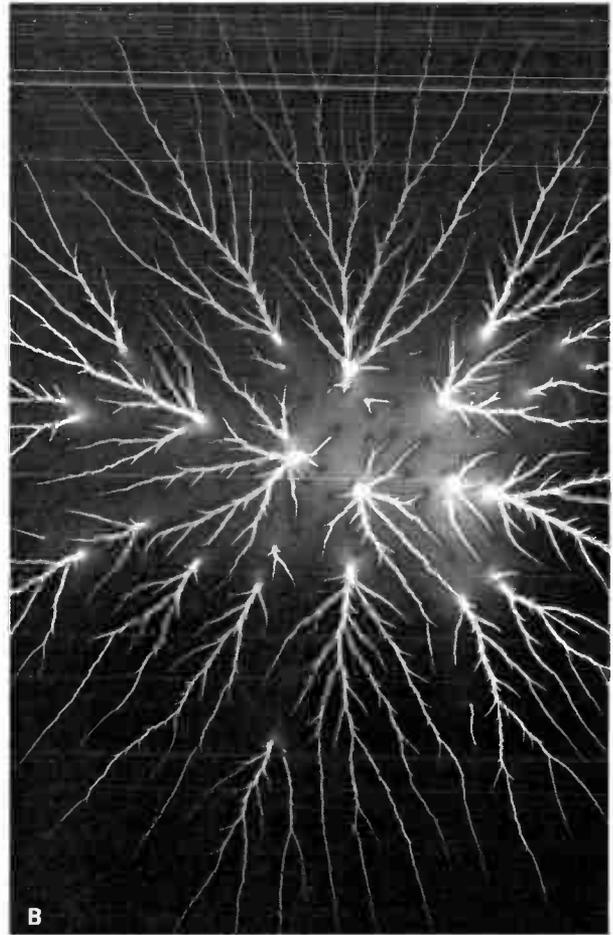
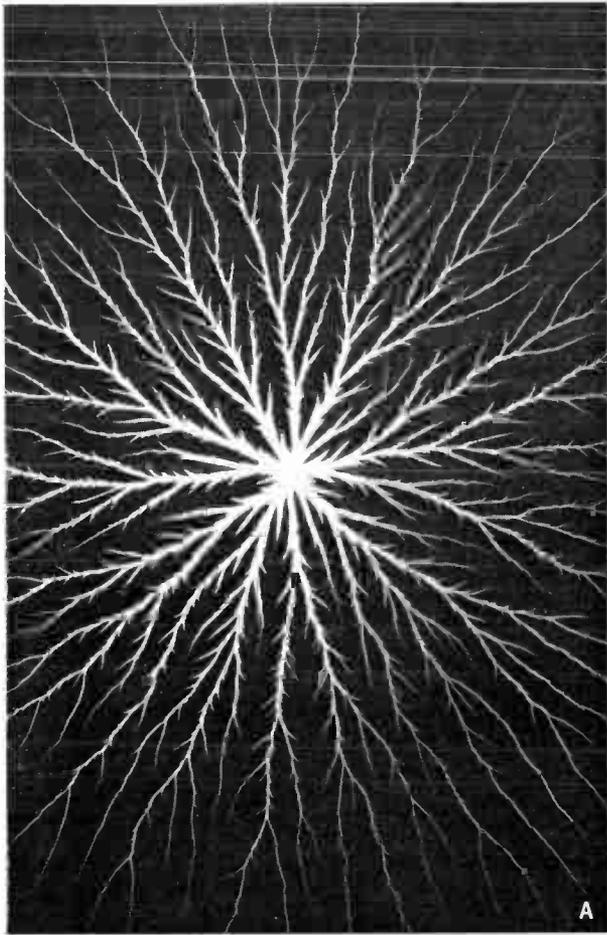
The method may be adapted to the study of various phases within the breakdown process. For example, the use of highly sensitive photographic film permits the study of primary streamers, the first manifestation of ionization at positive electrodes. Placing the film on the cathode plane makes the observation of cathode processes possible. With the use of infrared films, primary streamers go almost undetected because their emitted radiation is mainly in the blue region. Infrared film is suitable where high temperatures have developed or where infrared radiation has occurred.

Autographs

For the quantitative analysis of gaseous discharges, correct identification of the pictures obtained becomes essential. The pictures are called *autographs* because the advancing luminous front leaves or "writes" tracks as it advances along its path. Not all details in the autographs can be explained; however, many observations can be accounted for and correlated with other results.

With the point positive and facing the sensitive side of the film, the four autographs of Fig. 2 are typical of what is obtained as the film is moved from the point

FIGURE 2. Autographs illustrating streamer phenomena produced by a 30-kV pulse at an increasing distance, s , from the point electrode. Spacing between electrodes is 2.5 cm. A: $s = 0$ cm. B: $s = 1.0$ cm. C: $s = 1.5$ cm. D: $s = 2.0$ cm.



Nasser—Spark breakdown in air at a positive point

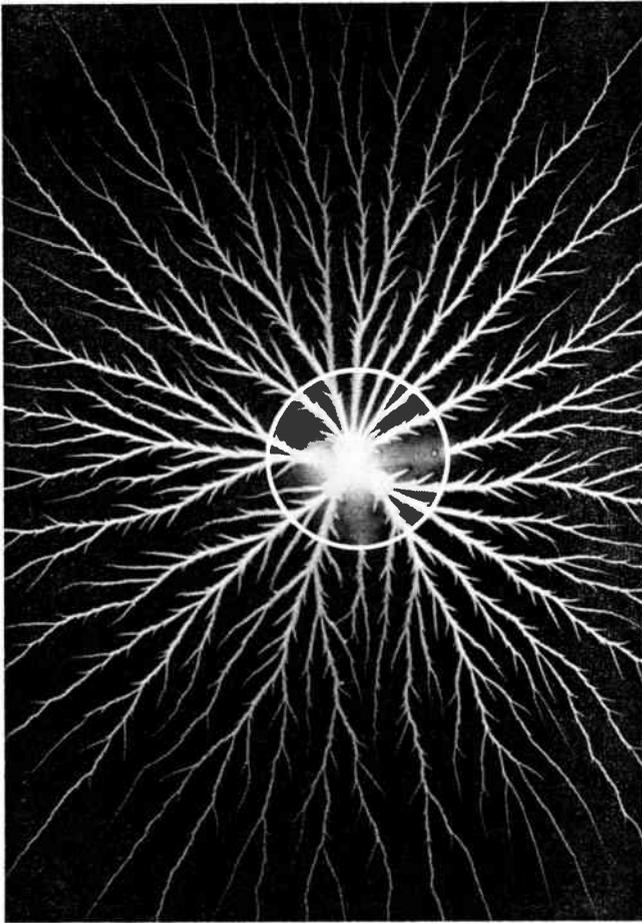
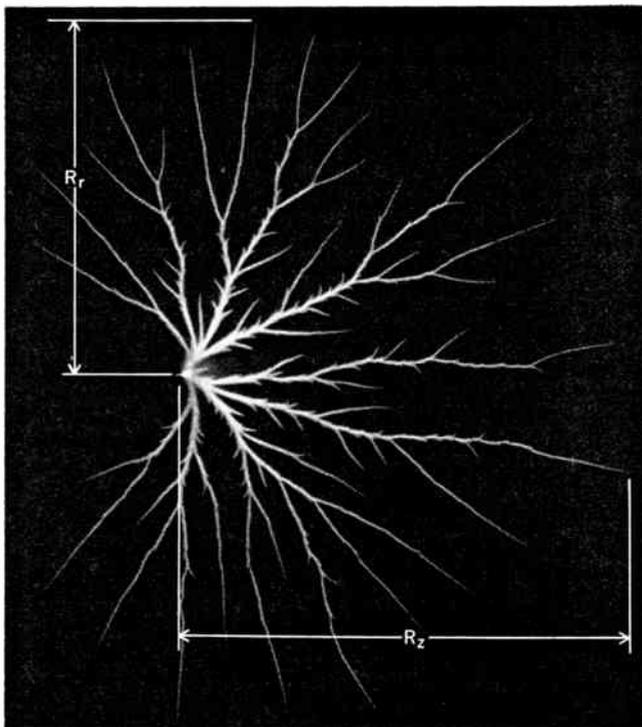


FIGURE 3. A midgap lateral autograph showing R_r at $s = 0.25$ cm. The gap spacing and voltage are the same as those given in Fig. 2.

FIGURE 4. An axial autograph illustrating R_z and R_r .



toward the plane. The autograph of Fig. 2(A) was taken with the film almost touching the point, yielding a typical positive figure resembling Lichtenberg's original dust figures. Such a figure is produced by a streamer emerging from the point, then deflected by the film. Because of the symmetry of the experimental apparatus, the streamer divides into six branches equally distributed around the circumference of the main branch. The six branches continue to branch further, resulting in the pattern shown.

Moving the film 1 cm away from the point resulted in the autograph of Fig. 2(B). Here there are a number of streamer branches, called *impact* points, out of which surface streamers, similar to those of Fig. 2(A) but having different lengths, grow radially outward. Such an autograph is produced by the several branches developed in the main streamer that strike the film and are then deflected into the field. A single streamer starting at the point may branch into as many as 40 branches. Figure 2(C), taken 1.5 cm away from the point, shows that the number of impact points reaching the film has increased tremendously. The surface tracks (glide streamers) they produce have diminished in length, however. From this fact the potential of streamer tips can be estimated.^{5,6} Finally, Fig. 2(D), taken 2 cm from the point, indicates the enormous number of branches present; most of them have short surface streamers that repel each other.

All four autographs were taken with a 30-kV voltage pulse that was well above corona onset but below spark threshold (36 kV). At reduced voltages, the manifestations shown by the autographs become continually less pronounced, finally disappearing.

Spatial characteristics

The fact that the autographs were evaluated with respect to streamer spatial propagation is perhaps one of the most significant contributions of the technique, since all streamer branches, regardless of intensity, are detected. To determine the volume enclosing all streamers and their branches, the radial and axial ranges of the primary streamers were measured. The maximum distance of an impact point of a branch from the autograph center, which is the intersection of the rod or gap axis with the plane of the film, was measured for various distances from the point; it is designated as R_r , the lateral range. The lateral range is illustrated in Fig. 3 by the radius of the circle enclosing all impact points; in Fig. 4, R_r is measured perpendicular to the gap axis and R_z is the axial range of streamers.

Results for various voltages are summarized in Fig. 5. Surprising was the fact that the streamers are capable of propagating both axially and laterally much farther than expected. The curves indicate that streamers branch in a radial direction, attaining a maximum lateral range at about 1.5 to 2 cm from the point. The maximum lateral range was also obtained from a large number of autographs taken for the same condition, plotted as a function of the applied voltage. Lateral range was found to increase with voltage until it reaches a plateau, beyond which it does not increase any further with voltage.

Axially, the streamers were found to advance toward low-field regions, striking the plane cathode at voltages much below spark threshold. The axial range of

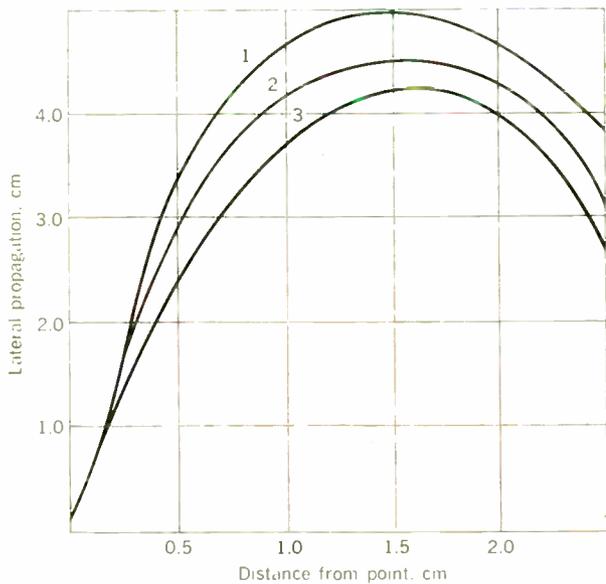


FIGURE 5. Lateral range as a function of distance for various voltages. The curves are envelopes of discharge volume at 33.5 kV (curve 1), 28.7 kV (curve 2), and 23.8 kV (curve 3).

streamers was obtained from autographs taken in both the lateral and coplanar modes, then plotted as a function of applied voltage for various distances between electrodes. A typical curve is shown in Fig. 6; note the great spread of points encountered in such studies. Recently it was found that streamers tend to stop whenever the values of the external field drop below approximately 500 V/cm. An empirical relation was found that describes the relation between axial range and applied voltage, allowing the calculation of axial ranges for many cases for which no experimental data exist.⁷

The evaluation of the number of streamer branches at various distances from the anode is also obtainable from autographs. Figure 7 shows the number of streamer branches as a function of gap distance z , for two voltages. It is interesting to note the huge number of branches that result from a single streamer. At high fields, the branching is particularly intensive, as indicated by the curves for small values of gap spacing.

Loeb and Nasser have proposed a theory that accounts for excessive branching and its decline.⁸ The theory was found to be in close agreement with data obtained so far. More work will be carried out in order to determine the degree of branching as a function of the applied field.

Temporal aspects

An important consideration is the time element in streamer development. To achieve temporal resolution, a rectangular-pulse generator was built whose pulse duration was variable from 5 to 200 nanoseconds. The use of a generator was found necessary after preliminary results demonstrated that cutting the voltage pulse after 2 to 5 microseconds did not produce any change in the character of the streamers.

Streamer ranges for various time durations t of the applied pulse are graphed in Fig. 8. It was surprising to find that streamers are capable of crossing the gap in less

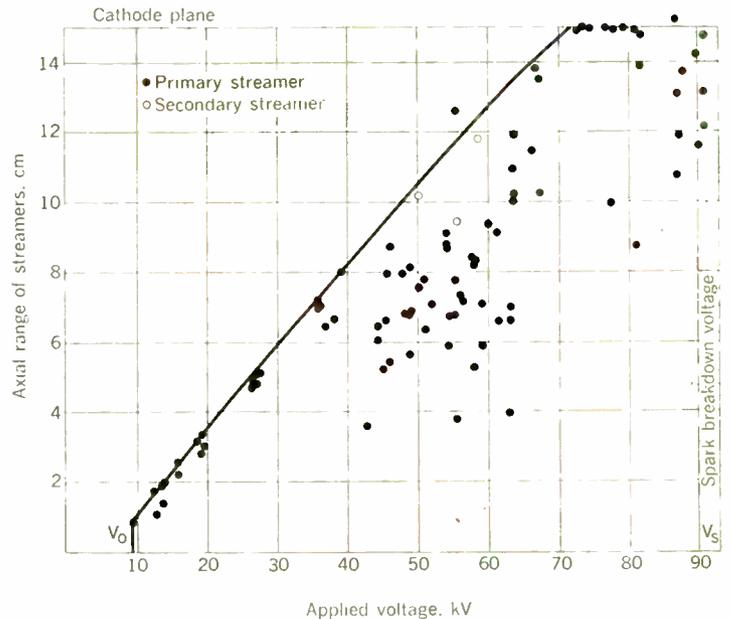
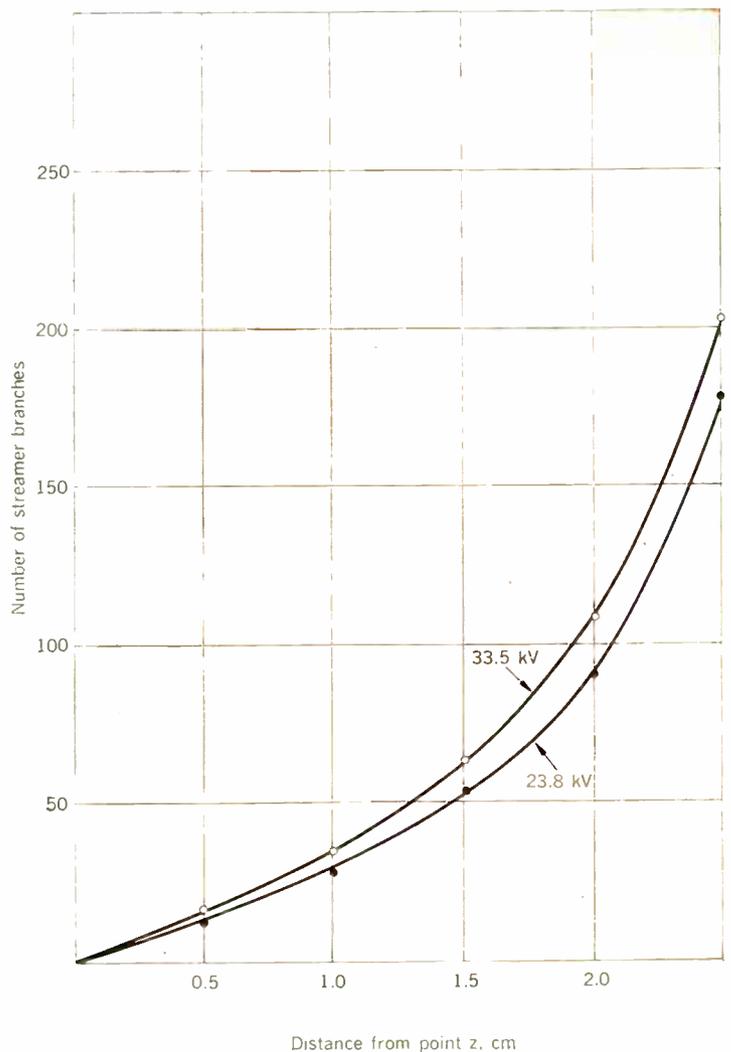


FIGURE 6. Axial range of streamers for a 15-cm gap. The concentration of data points shown is not indicative of a statistical distribution.

FIGURE 7. Number of streamer branches as a function of gap distance for two different voltages.



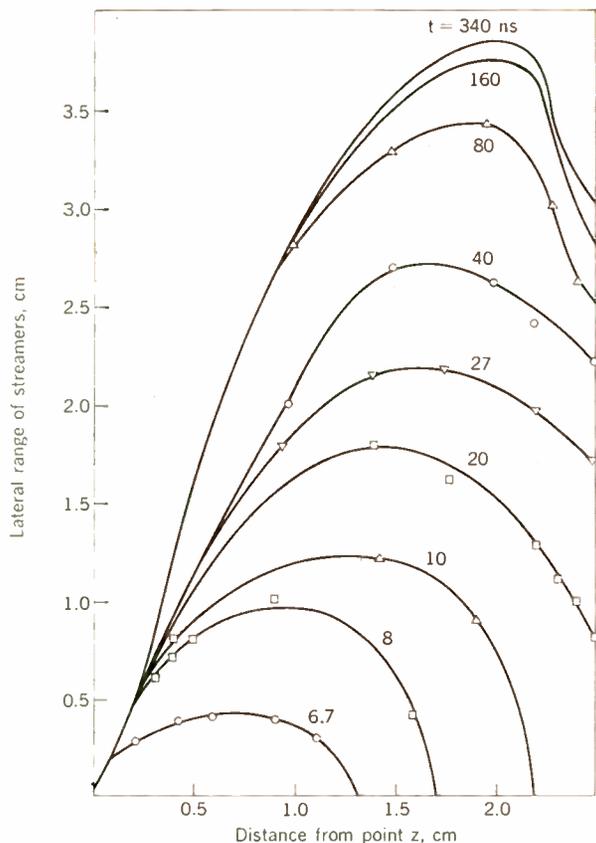


FIGURE 8. Lateral streamer range for various pulse lengths, t , of 38.5-kV amplitude; rise time ≈ 8 ns.

than 20 ns, after which the slower radial spread proceeds until the volume is filled with the advancing branches. Some weaker axial branches are relatively slow, reaching the cathode after 300 or 400 ns.

The measurements served also to determine the velocity of axial propagation of the fast and more vigorous streamers. Results obtained for a gap spacing of 2.5 cm at two different voltages are illustrated in Fig. 9. Owing to the extremely fast initial advance of the streamers, it was not possible to determine the starting propagation velocity in the anode vicinity. Velocity seems to decline very quickly, equal to 4×10^6 m/s at about 1 cm from the anode held at 38.7 kV. It then drops to 1.9×10^6 m/s, and seems to remain constant in the low-field region. Such a low speed appears to persist in the lateral propagation mode, for which no exact velocity evaluation was made.

It may be interesting to speculate about the initial speed at the anode. Extrapolating from the curves of Fig. 9 yields an initial anode velocity of 3.6×10^6 and 7.6×10^6 m/s at 25 and 38.7 kV, respectively. Such velocities are impressive if one considers that they are 1.3 and 2.5 percent of the speed of light.

Cathode processes

If very intensive streamers strike the cathode, electron emission is produced. This is due to the high electric field in front of the space charge of the streamer tips approaching the cathode; the emitted electrons are accelerated toward the positive wavefront. If there is ample

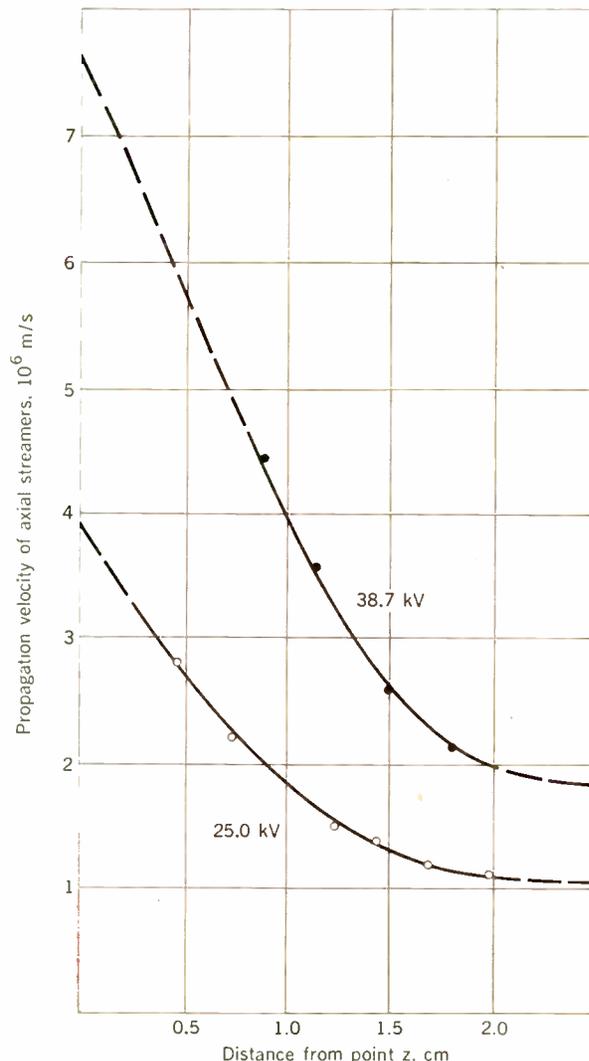


FIGURE 9. Streamer propagation velocity for two different amplitudes of an 8-ns-rise pulse; gap spacing is 2.5 cm.

time and distance, the electrons initiate avalanches and negative streamers, thereby increasing gap conductivity appreciably.⁹ When the current is observed, a pulse is produced when the streamers reach the cathode. A flash of light was also reported when the phenomenon was viewed with the aid of sensitive photomultipliers.⁹

The Lichtenberg figure technique was used to study this phase of development. The film was placed at the cathode and, depending on the orientation of the emulsion side, either the ionization processes at the cathode or those emanating from the anode, or both, were recorded. The use of color film shows both phenomena, with radiation coming through the backside exciting the red-layer emulsion and that coming from the front exciting the blue layer. Streamer impact points and tracks are seen in red, whereas cathode phenomena are seen in blue or white because of their high intensity (see Fig. 10).

Secondary channels

The action of the cathode in augmenting the charge carrier density of the original streamer channel is, how-

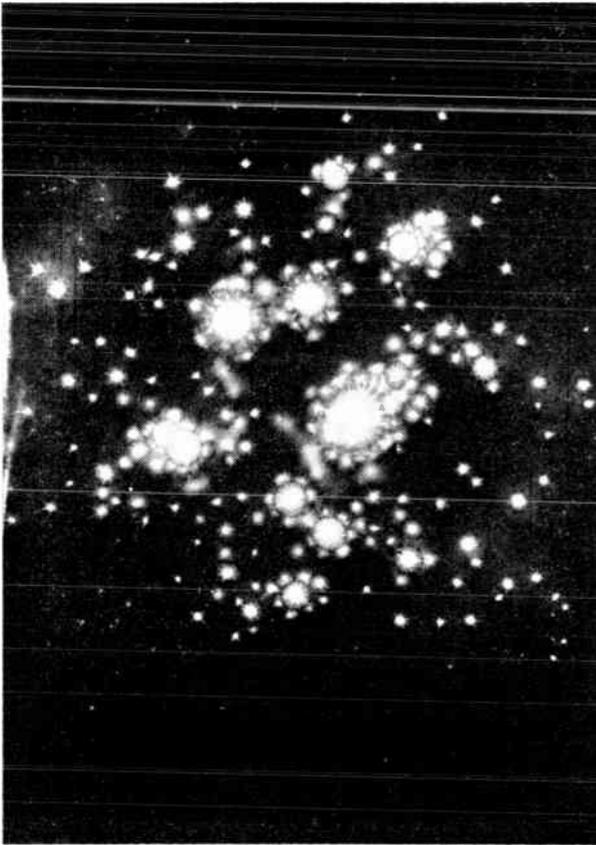


FIGURE 10. Processes observed at the cathode.

ever, restricted to short gaps, for which the streamers are capable of reaching the cathode. For greater gap distances, the primary streamers may develop such a high carrier concentration near the anode that a new ionization wave or plasma canal starts to materialize, growing from the anode toward the cathode. The new ionization wave involves *secondary* streamers, which are very different from primary streamers.

The autographs of Fig. 11 show a typical discharge in air at a pressure of 40 000 N/m² (300 mmHg) at voltages of 10 and 11.5 kV. The ionization wave has a much greater intensity than that of the primary streamers and follows channels that are narrower and probably more irregular. One of their typical characteristics is that they terminate in weak "threads" having very low intensity, then follow erratic but more or less circular paths when the film is in the lateral mode.

The bright radial portion of the secondary channels seem to follow in most cases the paths of the primary streamers. These channels have been detected and analyzed by Merrill and von Hippel in their study of Lichtenberg figures. Secondary channels also have been observed using photomultipliers, although the interpretation of such observations seems to be inconsistent.⁹ In very long air gaps and in lightning strokes, the discharges preceding full breakdown have properties very similar to secondary streamers.

It was found that secondary channels developed more readily in pure nitrogen, whereas the presence of any electronegative admixture, such as Freon, suppresses them entirely.¹⁰ This observation is in accordance with

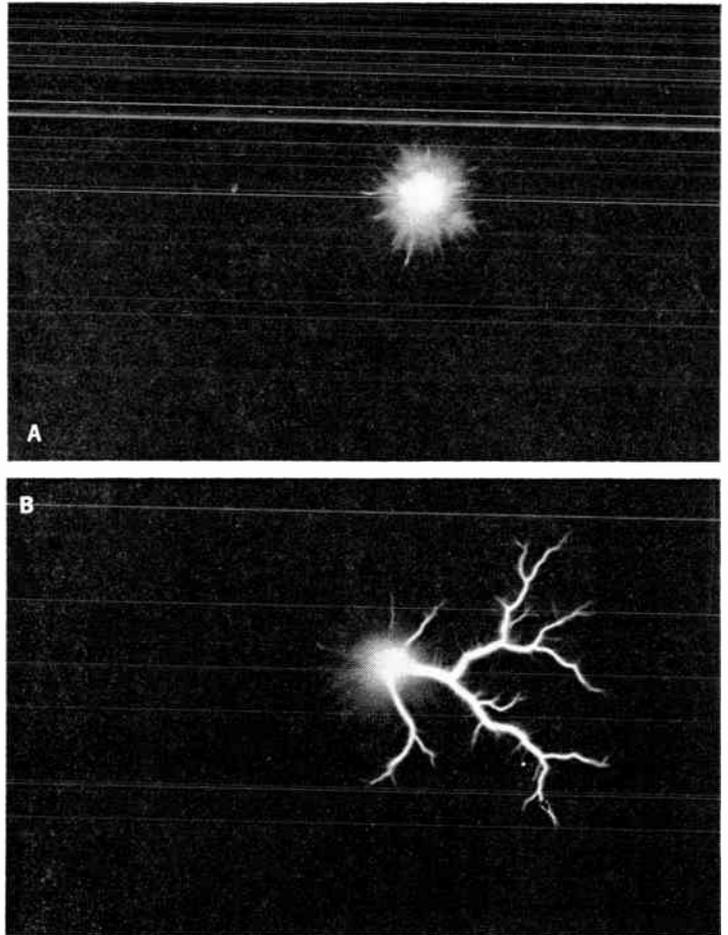


FIGURE 11. Secondary leader channels for two voltages developing from the center of a discharge after very faint primary streamers have completed their growth. Gap distance = 0.6 cm; air pressure = 40 000 N/m². A—At 10 kV. B—At 11.5 kV.

the findings of Merrill and von Hippel. Furthermore, when electrons were produced in air by a separately generated glow discharge, negative ions were formed by electron attachment to oxygen molecules. Total suppression of secondary channels was observed for this condition. This might be the reason for the high breakdown voltages of electronegative gases, although the exact mechanism still needs further study.

Conclusions

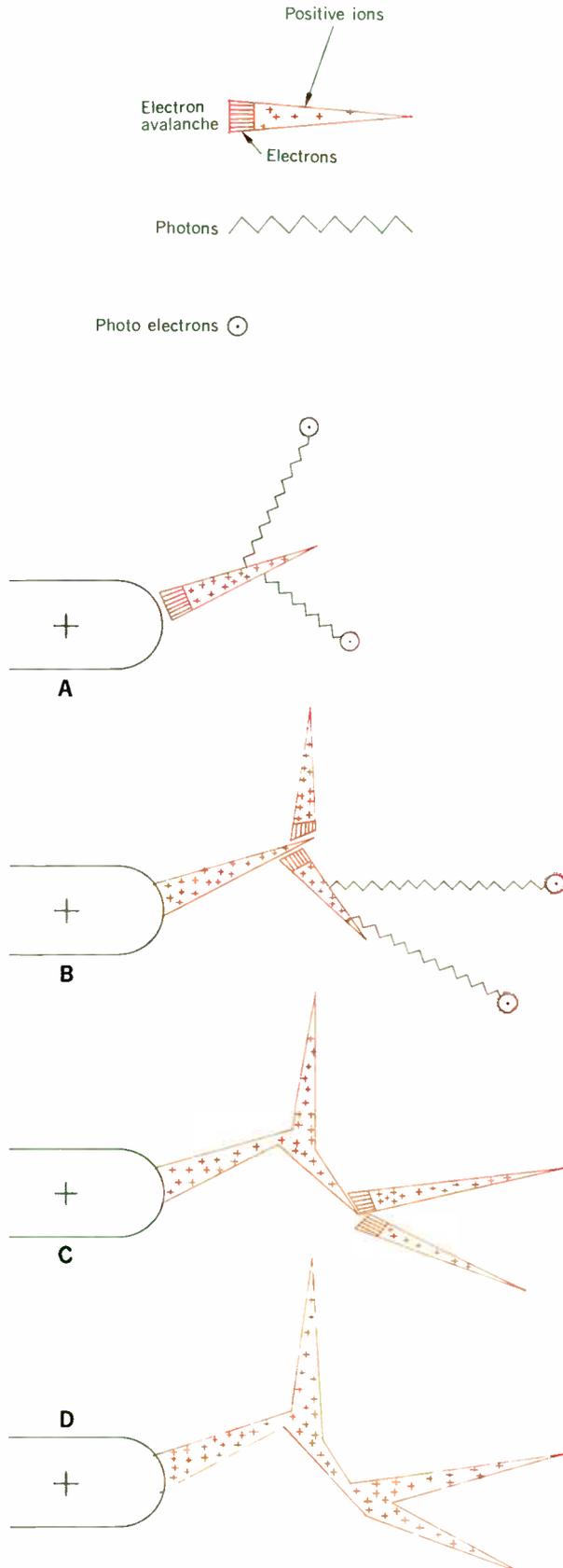
It is appropriate to offer a tentative explanation for the formation of the highly ionized channel through which an electrode charge is dissipated and a spark occurs. The streamer is initiated by a number of electrons n_0 , emitted from the cathode, leading to an avalanche similar to a Townsend discharge. The number of electron-ion pairs n produced is

$$n = n_0 \exp \left[\int_{x_1}^{x_2} \alpha dx \right]$$

where α is the ionization number (number of new electrons or ions formed per centimeter) and x_1 and x_2 are the initial and final points of measurement.

Figure 12(A) illustrates such an avalanche. A free

FIGURE 12. Streamer mechanism idealized in four steps, each of a few nanoseconds' duration. A—Initiation of an electron avalanche. B—Two photoelectrons are emitted. C—Two weak avalanches are produced. D—Positive ions drifting toward the plane cathode (not shown in figure).



electron initiates an avalanche by collision with air molecules on its way toward the positive point in a locally increasing field. Besides ionization, excited states are produced, with photons being emitted. Two such photons are shown being absorbed to ionize molecules that in turn liberate two electrons.

In Fig. 12(B), the two electrons are accelerating toward the high field at the tip of the wedge-shaped positive space left after the mobile electrons of the first avalanche were absorbed by the electrode. The electrons in front of the two second-generation avalanches proceed toward the positive electrode through the channel of the first avalanche in which a small potential gradient exists. Again, photons are emitted in all directions, two of which are shown producing photoelectrons. The two photoelectrons initiate two avalanches [Fig. 12(C)] exactly in the manner shown in Fig. 12(A), except that their magnitudes have become so small that further avalanches are not possible.

For simplicity, the preceding description was presented as though the processes occur in steps. Actually, the steps may occur at overlapping times; that is, step 2 starts before step 1 is concluded. The overlap in streamer development must occur in order to account for the measured velocities of their advance, which is of the order of 10^7 m/s.

All electrons having been absorbed, the positive ions of Fig. 12(D) begin to drift toward the plane cathode (not shown in the figure). When such a wedge-shaped streamer approaches the cathode, an extremely high field develops, producing a retroactive ionizing wave, the *backstroke*. For long gaps, however, the streamers do not reach the cathode with sufficient energy to trigger the backstroke. As explained earlier, secondary channels appear to be the mechanism for this condition.

The existence of two possible mechanisms of spark formation confirms many observations and may help explain some of the peculiar occurrences of various complete and partial ionization phenomena.

The author expresses his sincere thanks to the National Science Foundation and the Engineering Research Institute of Iowa State University for their continued financial support of this project, and to Manfred Heiszler for his devoted and able technical assistance and for obtaining the data of Fig. 6.

REFERENCES

1. Kunkel, W. B., ed., *Plasma Physics in Theory and Application*. New York: McGraw-Hill, 1966.
2. Levine, M. A., "Plasmas in space," *IEEE Spectrum*, vol. 3, pp. 43-46, Nov. 1966.
3. Lichtenberg, G. C., "Die Staubfiguren," *Noti Comment*, Göttingen, vol. 8, p. 168, 1777.
4. Merrill, F. H., and von Hippel, A., "The atomphysical interpretation of Lichtenberg figures and their application to the study of gas discharge phenomena," *J. Appl. Phys.*, vol. 10, pp. 873-887, 1939.
5. Loeb, L. B., *Electrical Coronas: Their Basic Physical Mechanisms*. Berkeley: University of California Press, 1964.
6. Nasser, E., "Räumliche Entladungsaufbau bei positiver Spitze," *Arch. Elektrotech.*, vol. 44, pp. 157-167, 1959.
7. Nasser, E., Heiszler, M., and Abou-Seada, M., "Field criterion for sustained streamer propagation," *J. Appl. Phys.*, vol. 39, pp. 3707-3714, July 1968.
8. Nasser, E., and Loeb, L. B., "Impulse streamer branching from Lichtenberg figure studies," *J. Appl. Phys.*, vol. 34, pp. 3340-3348, Nov. 1963.
9. Nasser, E., "Role of the cathode in the streamer-spark transition," *J. Appl. Phys.*, vol. 37, pp. 4712-4716, Dec. 1966.
10. Schroder, D. C., and Nasser, E., "Secondary leader channels in spark breakdown," submitted to *J. Appl. Phys.*

Scanning the issues

Integrated Microwave Circuits. An unusual, triple conjunction of interests is represented in the recent publication of a special issue on integrated microwave circuits. For the first time, three groups within the IEEE community have collaborated in the gathering and joint publication of papers on a subject that will be of wide interest. The papers appear in the IEEE TRANSACTIONS ON ELECTRON DEVICES, the IEEE JOURNAL OF SOLID-STATE CIRCUITS, and the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES. As S. Okwit, the editor of the last Transactions, points out, the joint publication demonstrates the wide interest of the technical community in the new and challenging field of microwave integrated circuits, and shows as well the accelerated evolution of a technology when broad and varied disciplines channel their efforts in a common direction. The guest editor of the special issue, William J. Edwards, defines the subject of the integration of microwave circuits quite simply—it embodies the method for the construction of microwave circuits where photolithographic and chemical processes are the dominant fabrication techniques.

Mr. Edwards pinpoints the beginnings of these efforts in 1964. At that time, there was little to suggest that microwave integration could provide the performance and cost advantages being projected—e.g., few quality components were available in compatible packages, the beam-lead technology was just emerging, etc.—but despite initial pessimism, a beginning was made with TEM and quasi-TEM transmission lines with particular attention to such adaptations as microstrip on ceramic and semiconductor substrates, which appeared to be the only practical method for making the microwave interconnections.

This special issue shows how quickly initial work spread to a variety of integrated microwave circuit developments. The papers include discussions of hybrid and monolithic circuits; ceramic, semiconductor, and ferrite substrates; distributed and lumped elements circuits; special device configurations; computer-aided design; and feasibility demonstrations of various circuit classes.

It has now become apparent, Edwards says, that microstrip is acceptable for making interconnections on a variety of substrates and that circuit losses are not prohibitive; for higher-frequency applications (above X band), where surface finish is more critical, single-crystal substrates such as semi-insulating gallium arsenide may be required. Furthermore, device packaging has become a limit to circuit experimentation and a demand is being created for characterized unpackaged devices in beam lead, or some equivalent, form. Finally, Edwards notes, integration has advanced from the realm of novel possibility to near practical reality. Much advanced systems planning is now predicated on the existence of integrated microwave circuits.

Those who wish to grasp a feeling for the problems and opportunities of applying monolithic and hybrid semiconductor technologies to microwaves might at least read the introductory paper, "Integrated Microwave Modules—A Prospectus," by William M. Webster. It contains a brief appraisal of the present situation from a semitechnical and economic viewpoint, and thus works no special hardship on the background knowledge of the general reader. (*IEEE Trans. on Electron Devices*, July 1968; *IEEE Journal of Solid-State Circuits*, June 1968; and *IEEE Trans. on Microwave Theory and Techniques*, July 1968.)

Special on Coding and Error Control. It should be noted that the current issue of the IEEE TRANSACTIONS ON INFORMATION THEORY is a special issue on the use of coding and error-control techniques in communication systems. The issue comes somewhat fittingly on the 20th anniversary of Shannon's famous 1948 paper, "The Mathematical Theory of Communication." A decade after Shannon's paper, guest editor Irwin M. Jacobs remarks, a reaction had set in to the effect that information theory would never yield results of practical value. Jacobs opines that we have now reached a more balanced position vis-à-vis information theory. Coding, he says, can provide significant if not dramatic improvements in communication over certain channels in

certain instances, some of which are detailed in papers in this special issue. On the other hand, Jacobs continues, he is struck by the fact that most applications of coding are still experimental. The widespread use of coding, it is clear, is not yet here. However, Jacobs predicts a burgeoning understanding and acceptance of now-proved coding techniques in the next two or three years. Those who are seeking understanding may find some of it in this issue. Four of the papers deal with communication over burst-noise channels, a subject that has long been of interest to applied information theorists, not only because of the commercial and military importance of telephone, HF radio, and troposcatter communication systems, but also, Jacobs says, because of the challenge presented by the algebraic rather than exponential decrease of error probability with signal-to-noise ratio. The improvement of coded over noncoded systems, although not as awesome as early studies suggested, is still regarded as significant. (*IEEE Trans. on Information Theory*, September 1968.)

Interdigital Transducer. Work that will be extremely useful in the design of ultrasonic surface-wave delay lines and resonators is reported by Chin-Chong Tseng in the current issue of the IEEE TRANSACTIONS ON ELECTRON DEVICES. His paper provides a complete mathematical treatment of the selectivity of the so-called "comb transducer" for excitation of elastic waves in piezoelectric solid material. The use of these waves has recently become the subject of considerable interest. In his well-written paper, the author shows that the interdigital transducer can respond to fundamental and odd harmonics but not to even harmonics. The author shows that the characteristics he derives analytically agree with experimental results. (C. C. Tseng, "Frequency Response of an Interdigital Transducer for Excitation of Surface Elastic Waves." *IEEE Trans. on Electron Devices*, August 1968.)

Chinese Text Analysis Via Computer. The IEEE TRANSACTIONS ON ENGINEERING WRITING AND SPEECH, which has been wondering to itself for some time whether or not its dearth of good manuscripts meant that the end was near, feels that it has suddenly struck it rich with the subject of computers. Indeed, its recent special issue on computer-aided engineering documentation merits

more than passing attention. The papers in the issue represent recent original work in the development of automated methods for engineering documentation. The hardware papers deal with photocomposition machines; the software contributions concern text editing, formatting, and other documentation programs; and the systems papers describe methods of generating documents with combinations of programmed devices. Batch processing and on-line methods are discussed, as are dedicated computer systems and time sharing.

Worthy of special note in this special issue is an unusual computer use for analyzing Chinese mathematical text with graphical inputs and outputs. The project is described by G. L. Walker and his colleagues. Although other researchers have used computers to print out nonstandard characters, the idea that is unique to this system is that of using a display scope, a RAND tablet, or a plotting table with digitizing unit, for *graphically inputting* strings of nonstandard characters.

The graphical input/output system, the authors note, has brought machine processing of Chinese text to the realm of practicability. Such projects as the compilation of a concordance of Chinese mathematical tests, the revision of a Chinese-English glossary of mathematical terms, and the production of inter-linear Chinese-English texts for the editor's easy checking of translations, etc., are aimed at the eventual implementation of an on-line, real-time system for machine aids to an editor of scientific translations of Chinese mathematical texts. (G. L. Walker *et al.*, "Chinese Mathematical Text Analysis," *IEEE Trans. on Engineering Writing and Speech*, August 1968.)

Filtering for EMC. There is conventional filtering, with which every engineer is familiar, and then there is filtering for electromagnetic compatibility, which until recently has been more of an art than a science. In an introduction to a special issue on this latter form of filtering, Heinz M. Schlicke tells us the story in a nutshell. A system may have an extremely high reliability, he says, and still be worthless in operation unless it is also electromagnetically compatible. Hence high reliability and excellent electromagnetic compatibility are twin requirements for systems effectiveness at all times and under all circumstances. Reliability has been recognized as significant and has been stipulated and implemented systematically for years.

Electromagnetic compatibility, on the other hand, although often cursed when its absence causes trouble, has only very recently been given the emphasis and support that is indispensable if an exploding communications technology is to represent progress instead of chaos.

This special issue of 25 papers assesses and surveys the state of the art of EMC filters as a key aspect of EMC work, and aims systematically at converting filtering for EMC from an art into a science. The EMC filters, in contrast to filters normally encountered, have to operate under much less defined constraints and less simple premises. They must be more universal because of the need for very-wide-band characteristics, the existence of severe bias effects, excessive power levels, very strong mismatching and detuning interfaces, spectral overlap, and indeterminacy of what has to be filtered out; these conditions may appear in various combinations. The job of the EMC engineer is to find the right balance between too little prevention (incompatibility) and too much prevention (resulting in excessive cost, size, weight, and possibly reduced reliability). Engineers who are EMC-conscious will want to look into this special issue. (H. M. Schlicke, "Survey," *IEEE Trans. on Electromagnetic Compatibility*, June 1968.)

Sea Floor Electronics. Every year, it seems, we have our new "romantic" subject in engineering, and with each new subject there seems to run a tide of hope that "here at last is the answer," the panacea for many ills. Judging from the prevalence and persistence of these high hopes, one guesses that the holding of illusions is as much a necessity to our natures as is the construction of actual works that correspond to more sober realities. (We may blame the advertising men for putting us on, but then, we must reflect, we really *want* to be put on.) Information theory, microelectronics, systems thinking, innovation, productivity, you name it . . . the amount of rhetoric that has been spilled over any of these, if taken at face value, could lead one to believe that a lot of engineers were getting more than their fair share of ecstasy. But the fact seems to be that engineers really revel in the existence of *problems*. If something is labeled difficult or impossible, they are delighted—and when systems break down, many engineers become downright gleeful. Absolute answers and final solutions, if there were such, would put an end to the whole charade. Thus we

see in every engineering specialty regular alternations, like the tides, between high hopes and sober assessments when new developments come to light.

One of the more recent glamour subjects that is now coming under more sober engineering assessment is the technological exploitation of the seas. As Victor C. Anderson remarks in an interesting paper on one aspect of this subject, the term "exploitation of the sea" has itself been exploited in the past few years in political campaigns, in popular articles, in promotion of corporate funding, and even for the justification of scientific research. But in the process, he warns, little emphasis has been given to the fact that working at sea, and, in particular, working on the sea floor, is a very expensive endeavor. Any plans for commercial exploitation, he says, where costs and values are to be carefully weighed, must be vitally concerned with the extent to which a sea-floor working capability is designed into the program. Unfortunately, there is a paucity of background data that can be brought to bear in the quantitative consideration of the cost of sea-floor work. Qualitatively, experience has shown the cost of sea-floor work to be almost unbelievably high. For example, he says, the average cost of repairing a deep ocean cable that has failed for one reason or another is in the vicinity of \$80 000 per cable repair. The cost for a comparable cable repair on land would be about one percent of this amount.

Obviously, Anderson goes on, in the face of this difference in the cost of work performance, even a qualitative look at order-of-magnitude considerations that may be used in approaching the economics of sea-floor work can be worthwhile. If large-scale sea-floor installations are to be a reality in the future, complex electric and electronic systems are inevitable, and the degree to which repairs or modifications of these systems may be carried out *in situ* will strongly influence the installation design and its cost.

Anderson's discussion of present developments in *in situ* electronics maintenance by the use of divers and by remote manipulators, and his relation of the effects of these developments to the operating cost of future sea-floor instrumentation systems, should stir the interests of any who are weighing the real prospects of the glamorous exploitation of the seas. (V. C. Anderson, "Maintenance of Sea-Floor Electronics," *IEEE Trans. on Aerospace and Electronic Systems*, September 1968.)

Computers aren't people

Dr. John Pierce writes in his article on "Men, Machines, and Languages" (July, pp. 44-49) that "progress with computers has come through employing their useful potentialities . . . But the sort of performance the artificial intelligence enthusiasts have worked toward hasn't proved useful."* He then goes on to offer some evidence to support his conjecture saying, "computers have been miserable at theorem proving, music composing (at least we don't enjoy their compositions—maybe computers do), chess playing, and general pattern recognition." Finally, he states, "In view of the tremendous usefulness of computers, all I can say about their failures in these fields is, so what?"

To my mind this is a rather strange view, since even if I were to grant his hypothesis that there have been failures in these fields, all I can say about the *tremendous usefulness* of computers is, so what? Most people already know that computers are useful. Indeed, taking a more philosophical bent, the only interesting questions to be asked about computers concern their fundamental limitations. And this leads us directly back to the field of artificial intelligence. It seems to me that in the computer field, as in any field, one tends to solve easy problems before hard ones; and in my opinion the ostensible failures in artificial intelligence derive from the fact that problems in this field are inherently more difficult than problems in other areas of computer application for which solutions have been found. It seems to me to be premature to draw any conclusions about the *potential usefulness* of theorem proving, music composing, chess playing, and pattern recognition based on the current performance of programs designed to perform these tasks. Furthermore, it is a matter of opinion as to whether current performance at these tasks has been miserable (I could argue strongly to the contrary); but even assuming this outlook, I feel it is a grave disservice to characterize efforts in these areas as failures. It seems that a more constructive attitude toward artificial intelligence is (1) hard problems take longer than

easy problems; (2) significant progress in artificial intelligence has already been made†; (3) there are no obvious barriers to continued progress; and (4) once some of these problems are solved, the benefits in terms of utility will be enormous.

There are also some misleading remarks in the section on language. Dr. Pierce correctly observes that attempts to characterize the relation between syntax and semantics in processing natural language statements has met with considerable difficulty. He then states, "No doubt, syntax comes into the understanding of speech, but meaning seems somehow to come first." My own research on natural language processing leads me to a somewhat different conclusion. Although syntax and semantics ought to remain recognizably distinct components of any natural language processing system, I believe these components should be highly interactive in the processing of any statement. It seems that any approach that argues that syntax begins where semantics leaves off, or conversely, that semantics begins where syntax ends, is ultimately doomed to failure. This interrelation between syntax and semantics is not intended as an apologetic compromise between contending proposals, but as a precise reflection of the facts of language.

Finally, we come to the section entitled, "Where do we go from here?" Instead of awakening the engineering community to some really important possibilities and their consequences (some of which will shortly be within the state of the art) this article does its greatest disservice by lulling us into complacency about future prospects. This section essentially reinforces the following simplistic argument: (1) human beings are wonderful! (we can open things we've never seen before, we can perform the "miracle" of responding to instructions phrased in

*Notice how the wording tempts one to draw the erroneous conclusion: therefore, workers in artificial intelligence can make no progress.

†Although I could document this claim by citing a large number of programs in each of the above-mentioned fields, I refer the interested reader to the ACM SIGART *Newsletters*, nos. 1-10.

Advanced Systems Specialist

An excellent opportunity offering stimulating challenges and a real chance for personal advancement in the rapidly growing Electronics division of General Dynamics. Background should include 5-7 years industry experience and familiarity with:

- Organization and direction of electronics systems synthesis and analysis projects.
- Application of systems engineering methods to avionics and precision tracking-system projects.
- Coordination of advanced systems projects with user agencies.

An MS degree is required; a PhD is preferred. Please send your detailed resume to:

Mr. V. K. Finley,
Professional Placement
P.O. Box 127
San Diego, California 92112

GENERAL DYNAMICS

Electronics Division
San Diego Operations
An Equal Opportunity Employer

ENGINEER Electric Power Distribution

Minimum of 5 years electric utility experience is desired. Duties will include: preparation of designs, layouts and specifications for electric power distribution systems, review of contract proposals and supervision of construction.

Some foreign travel involved.

In addition to job security and excellent growth potential, we offer a full range benefit package which includes; profit sharing, annual cash bonus, comprehensive insurance, and tuition refund.

Submit resume in strict confidence to Mr. Ralph Nelson.

HARZA
ENGINEERING COMPANY
400 West Madison Street
Chicago, Illinois 60606

An Equal Opportunity Employer

THE EXCEPTIONAL MAN



Wanted—for a challenging position in an American industrial or research organization . . . The Exceptional Man.

He is ready for a career opportunity with a nationally known organization offering immediate rewards and unlimited potential for personal growth.

He has a background in electronics or mechanical engineering, physics, mathematics, operations research, chemistry, materials science or industrial engineering. Opportunities exist for Bachelors, Masters and PhD candidates.

He is wanted by one or more of our prominent clients for a future in industry, university-affiliated research or non-profit research. These organizations, located throughout the United States, pay all our charges.

They want us to locate The Exceptional Man. May we hear from you?

Charles A. Binswanger ASSOCIATES
INC.

EXECUTIVE SEARCH SPECIALISTS
725 JACKSON TOWERS • BALTIMORE, MD. 21201 • (301) 728-5000

Keep abreast of progress— get the new Proceedings!

Will advanced research and development help you in your job? Would you like to know about new thinking in the electrical/electronics field—new ideas that in time would help you get ahead?

Would you like in-depth, up-to-the-minute information on new concepts—from theory to applications?

If you'd like to be "the man with all the answers," send in your subscription today!

Subscription rate for IEEE members, \$3.00 per year. For non-members, \$22.00 per year (\$23.00 if outside United States and Canada).

Address: Proceedings of the IEEE, Circulation Dept., 345 East 47th St., New York, N.Y. 10017.

We have been placing B.S., M.S. & Ph.D.

ELECTRICAL & ELECTRONIC ENGINEERS

in FEE PAID positions
throughout the U.S.
since 1959, through our
PERSONALIZED SERVICE.

Over 1000 client companies
rely on us to provide carefully
screened, technical referrals.
We are staffed by experienced
graduate engineers . . .

WORKING FULL TIME FOR YOU!

SEND RESUME TODAY
(if none, SEND COUPON
for confidential application)

Name

Address

City State

A. L. Krasnow, Member IEEE

ATOMIC PERSONNEL, INC.

Suite C, 1518 Walnut St.
Philadelphia, Pa. 19102

AN EMPLOYMENT AGENCY FOR ALL TECHNICAL FIELDS



broad concepts., etc.); (2) computers do miserably at tasks that are trivial for people (pattern recognition, resolving ambiguity, etc.); (3) substantive progress on known problems has been slow (so far, the results have been more startling than useful); (4) there are still some extremely difficult problems ahead to which little attention has been paid (such as forming a map of even a portion of the real world); *therefore*, (5) we are not likely in the near future (if ever) to "succeed very well in making computers interact usefully with natural languages."

Although the first four premises are all more or less true, their relationship to the conclusion is somewhat tenuous. Indeed, I am already prepared to offer a counterexample to this conclusion. For the last three months we have been experimenting with a deductive question-answering system in the Stanford Research Institute Information Science Laboratory capable of accepting English sentences (declarative, interrogative, and imperative) taken from a restricted subset of natural language in a conversational manner and usefully acting on them in conjunction with a "map" of the world automatically generated by various sensors, including a television camera. A fuller discussion of the capabilities of this system can be found in an article by L. S. Coles entitled "An On-Line Question-Answering System with Natural Language and Pictorial Inputs" that was presented at the ACM Conference in Las Vegas, Nev., August 27, 1968, and might form the basis for a future SPECTRUM article with the same title.

*L. Stephen Coles
Stanford Research Institute
Menlo Park, Calif.*

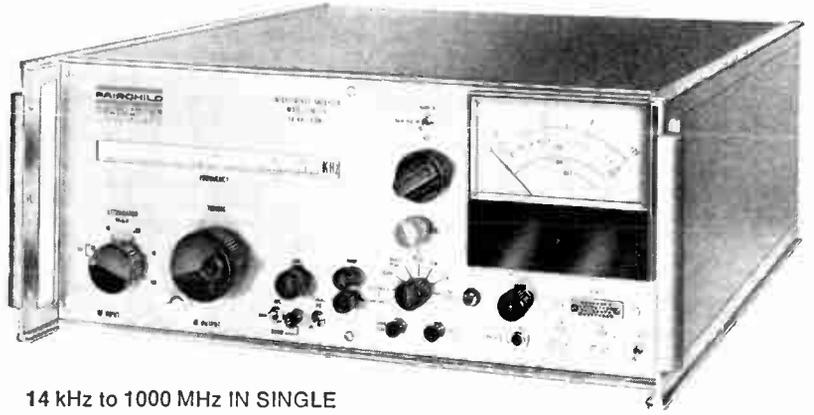
To an engineer, Dr. L. Stephen Coles' statement that "the only interesting questions to be asked about computers concern their fundamental limitations" seems rather negative. Would one say the same thing about man? What about wonderful and far from fully exploited capabilities of computers?

Certainly, sensible people (excluding some philosophers) start with easy problems and progress toward harder problems. That's how progress is made. Computers and computer applications are progressing thrillingly as well as usefully in just this way. I don't observe that this is so of what goes on under the name "artificial intelligence." Further, I just don't believe that "once



Delivery! SWEPT-RECEIVERS S-T-C-H-I-N-G MANPOWER

Get full information on a wide variety of drying substrates, wafers, and crystals. Filaments, baskets, and trays are precision-molded from a material that is almost 100% inert to extreme temperatures. Write for the "Semi-conductor Process Catalog," including information on the spin dryer.



14 kHz to 1000 MHz IN SINGLE
COMPACT PACKAGE—BATTERY POWERED

†Military Nomenclature AN/URM-178; FSN 226625-115-2775

SOLUTIONS FROM PRE-SET PROCEDURES

Whether used alone or with one of its specially-designed peripheral instruments, Interference Analyzer Model EMC-25† is the most advanced receiver system currently available. Hundreds of these varactor-swept instruments are already in field and laboratory use, speeding up tests and monitoring, and relieving manpower shortages. It has outstanding sensitivity and two bandwidths throughout the full range. Fully remotely controllable...digitally over telephone lines if required.

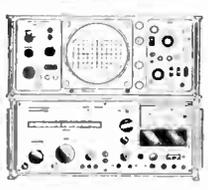
With its accessory programmer, the full range of 14 kHz to 1000 MHz can be plotted on a single sheet of pre-calibrated paper in minutes.

When used with the accessory Display Module, a full octave band dispersion can be shown on the CRT at one

time...or any selected portion... in linear or log!

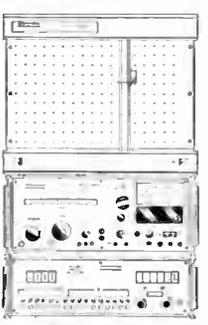
When combined into the fully automated Spectrum Surveillance System, all these functions are performed, plus many more.

Hundreds of EMC-25's, with and without peripheral equipment, are being used all over the world. Contact your Fairchild/Electro-Metrics representative for your catalog, to see how you too can s-t-r-e-t-c-h your valuable manpower.



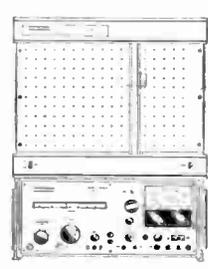
BAND-BY-BAND SPECTRUM DISPERSION

By displaying the EMC-25's 15 octave bands, one at a time, on the CRT of Display Module SPD-125, the full range of 14 kHz to 1 GHz is visually displayed.



FULL FREQUENCY RANGE AUTOMATIC PLOTTING

The ESC-125A Programmer adds the capability of plotting the EMC-25's 14 kHz to 1 GHz range on a single sheet of pre-calibrated graph paper, and with automatic antenna selection.



BAND-BY-BAND SPECTRUM PLOTTING

Built-in electronic sweep in EMC-25 provides push-button control of octave-band plots on the EXY-125A X-Y Plotter.



DUAL-SYSTEM FSS-250

For full 20 Hz to 1 GHz range:
Here is a system with both visual CRT octave dispersion and automated X-Y plotting...plus built-in calibrator control, amplifier and speaker, and many optional operator controls.



EMC-10 for VLF

Interference Analyzer EMC-10 covers the frequency range from 20 Hz to 50 kHz. It also is solid-state, battery operated and electronically scannable. Category A approved; FSN 6625-911-0840 and (for "E" model) 6625-937-6523.



100 CHURCH STREET, AMSTERDAM, N.Y. 12010