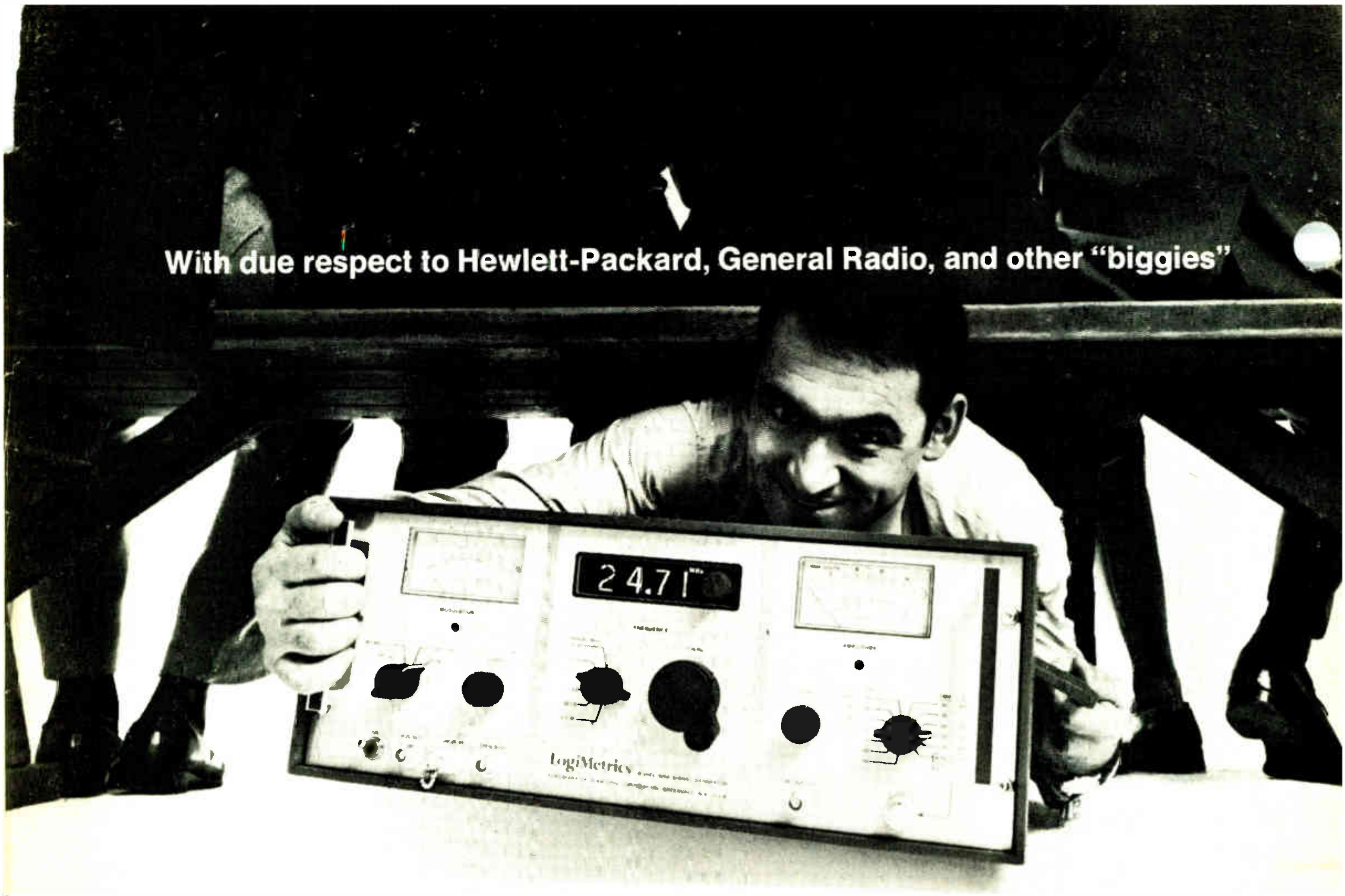
IEEE spectrum

features



LogiMetrics audaciously introduces the world's most accurate RF signal generator with continuously viewable (non-ambiguous) display.

● Unprecedented stability and resolution ● No separate counter or accessory equipment required ● A true combination of improved performance and reduction in cost!

Until now, a signal generator with an accuracy of 10,000 parts per million was considered acceptable. LogiMetrics' new 920A RF signal generator automatically, directly, and continuously measures the frequency to an accuracy of 60 parts per million (0.006% of full scale)! Think about that. Accuracy to 10,000 parts per million now improved to 60 parts per million, taking into account error due to all causes.

PATENTED IMPROVEMENTS AND EXCLUSIVES The 920A provides a significant improvement in accuracy, resolution, and repeatability due primarily to a patented new feature, ddc™—direct digital calibration. (There's no dial at all.) The frequency generated is displayed visually—to four significant digits. The frequency you see is the frequency you're getting—continuously! In addition, leveling is accurate to ± ½ dB across the complete spectrum of the generator and at all output levels. Modulation capability provides unmatched AM performance with both internal 400/1000 Hz modulation or external.

CLEANER AND MORE FLEXIBLE Unwanted spurious by-products are practically non-existent. For example: hum and noise, 70 dB down; residual or incidental FM, 1PPM; RFI leakage, better than MIL-I-6181D. That's performance! The unique design (primarily integrated circuit), makes calibration a snap. It also facilitates a range of options including computer and/or printer compatible read-out of frequency and narrow band frequency programming.

ASK US! If accuracy counts in your laboratory, write or call for catalog data describing LogiMetrics' "ddc" signal generators from 100 kHz to 80 MHz. Or use reader service number below.

Comparative Costs:

Conventional signal generator plus separate counter (typical)	$3,500.00
LogiMetrics Model 920A	$1,975.00

LogiMetrics INC.
A Subsidiary of Slant/Fin Corp.
100 Forest Drive, Greenvale, New York 11548
Phone: (516) 484-2222

Circle No. 2 on Reader Service Card

the cover

A color EVR film strip, in black and white, includes two sound tracks, a luminance track, a synchronization track, and a color track. When this film is run through an EVR player, the resulting picture appears in color on a standard home color television receiver. How it's done is explained in the article beginning on page 22

departments

IEEE SPECTRUM is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: 345 East 47 Street, New York, N.Y. 10017. Cable address: ITRIPLEE. Telephone: 212—752-6800. Published at 20th and Northampton Sts., Easton, Pa. 18042. **Change of address** can be made effective with any particular issue if received by IEEE Headquarters by the 1st of the preceding month. **Annual subscription:** IEEE members, first subscription \$3.00 included in dues. Single copies \$1.50. Nonmember subscriptions and additional member subscriptions available in either microfiche or printed form. Prices obtainable on request. **Editorial correspondence** should be addressed to IEEE SPECTRUM at IEEE Headquarters. **Advertising correspondence** should be addressed to IEEE Advertising Department, at IEEE Headquarters. Telephone: 212—752-6800.

OTHER IEEE PUBLICATIONS: IEEE also publishes the PROCEEDINGS OF THE IEEE, the IEEE STUDENT JOURNAL, and more than 30 Transactions for IEEE Groups with specialized interests within the electrical and electronics field. Manuscripts for any IEEE publication should be sent to the editor of that publication whose name and address are shown on page 17 of the January 1970 issue. When in doubt, send the manuscript to E. K. Gannett, Editorial Services, at IEEE Headquarters, for forwarding to the correct party.

IEEE also copublishes, with the Institution of Electrical Engineers of London, England, *Electrical & Electronics Abstracts* and *Computer & Control Abstracts*, as well as their companion titles journals, *Current Papers in Electrical & Electronics Engineering* and *Current Papers on Computers and Control.*

Forum

Readers are invited to comment in this department on material previously published in IEEE SPECTRUM; on the policies and operations of the IEEE; and on technical, economic, or social matters of interest to the electrical and electronics engineering profession.

Vale

The volunteer Editor of *IEEE Spectrum* for the past 2½ years, Dr. J. J. G. McCue, has stepped down, and I believe that our readers should become aware of their loss. Before he became Editor, a great effort was put into getting out a good journal, yet many of us were disturbed by what seemed to be a too-casual, too-dilatory, and too-disinterested readership. Jerry, as our Editor is known to his Editorial Board, also saw this and, early in 1969, decided to follow the precepts of good engineers when they want to improve: they innovate. Jerry's innovation, fully supported by his Editorial Board, was to include in *Spectrum* topics relating the engineers to society, and to treat some of these topics in the way we all know they really are, i.e., as controversial. In addition, Jerry's personal contribution in the form of his editorial, "Spectral lines," was written to be stimulating and provocative.

Obviously, not everyone likes the idea of provocation and controversy. As engineers, however, we must realize that all our great men have thrived on argument and, in the long run, society has benefited. The IRE's list of awards winners in past years is full of names who have strongly supported one side or another of controversial issues. Clearly, in 1970, one no longer finds it profitable to argue Heaviside's calculus vs. the Laplace transform, or whether FM is better than AM, or ac vs. dc power transmission. Our technology has stabilized and specialized; furthermore, the current technical controversies, such as discretionary wiring vs. standard integrated circuits, thin- or thick-film hybrid circuits, or LSI vs. MSI, are not the only subjects to broaden *Spectrum*'s readership. Recognizing that the relationship of science and engineering to today's major social problems is far and away the most significant topic, Jerry embarked on a plan to increase this type of material.

Many of us have changed our reading habits since the change in *Spectrum*. As one such person, I seldom defer reading at least a part of each issue almost upon the day of arrival. Evidence that reader interest has increased can also be found in the Forum letters. Prior to 1969, nearly every letter proposed was published. During the last half of 1969, the number of published letters increased fourfold, and space limitations now permit publication of only those with unique contributions.

Jerry McCue will rank high among what I hope will become a long list of distinguished editors of *Spectrum*. We will all be sorry to see him leave, but we can be confident that his competence and leadership will provide the success he so richly deserves in his other activities. On behalf of the *Spectrum* Editorial Board, and the IEEE membership, good luck and best wishes, Jerry!

E. W. Herold
Member
Editorial Board, IEEE Spectrum

I am pleased to add a few words to those written by Ed Herold in praise of the accomplishments of Dr. McCue as *Spectrum* Editor. Jerry has a long and distinguished record of accomplishment in working with IEEE publications. He was a guest editor of the Special Issue of the *Proceedings of the IEEE* on Ultrasonics issued in October 1965. He has served 4½ years as a member of the

DeWitt named IEEE Spectrum editor

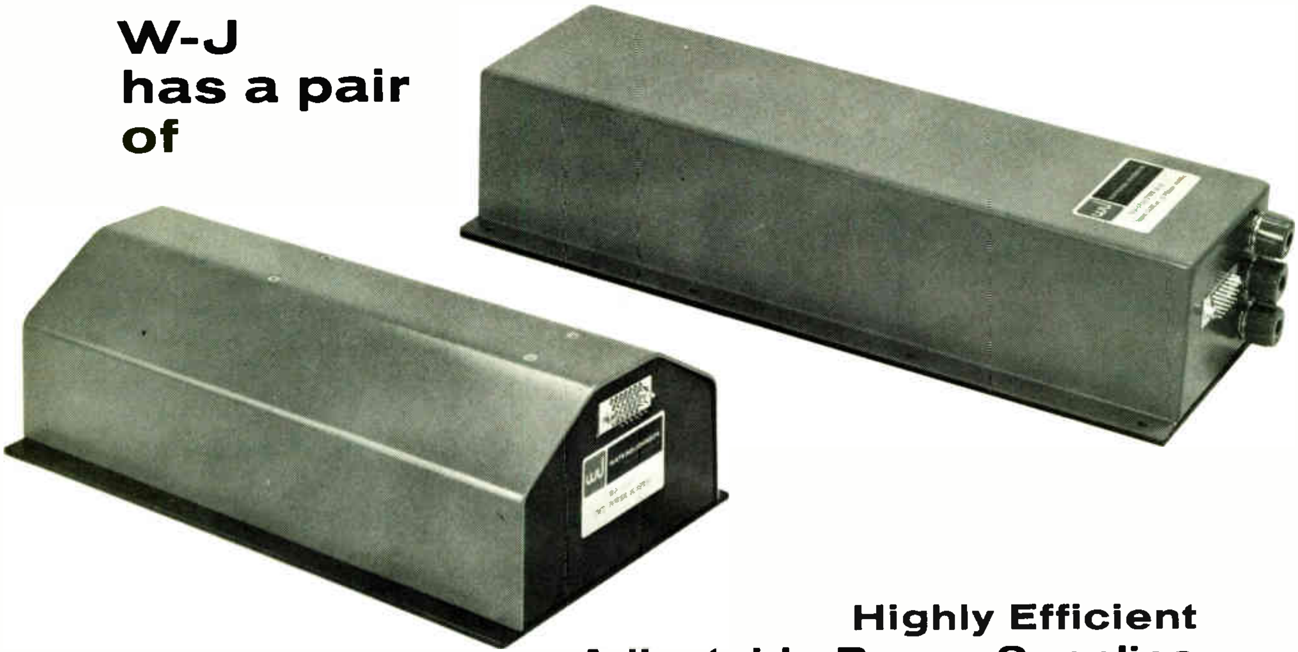
David DeWitt of International Business Machines Corporation, Hopewell Junction, N.Y., will become the new editor of *IEEE Spectrum* beginning with this issue, replacing Dr. J. J. G. McCue. Mr. DeWitt is a specialist in the development and manufacture of semiconductor devices. In addition to a long career at IBM, he was visiting professor at the University of California, Berkeley, during 1968–69. He is a member of the Semiconductor Electronics Education Committee and coauthor of two volumes issued by SEEC and published by John Wiley & Sons, Inc.

Airborne

TWT POWER SUPPLIES

For 75 to 600 Watt TWTs

W-J has a pair of



Highly Efficient Adjustable Power Supplies for High-Power TWTs

Adjustable for operation with any TWT in the RF power output range of 75 to 600 watts, the WJ-1114-1 and WJ-1169 meet all the environmental requirements of MIL-E-5400, class 2.

SPECIFICATIONS

	WJ-1114-1	WJ-1169
Power Output	75 to 300 watts	100 to 600 watts
Efficiency	89%	91%
Size	4 x 5 x 12 inches	3 x 6 x 14 inches
Weight	15 pounds	25 pounds
Collector	Regulated	Unregulated

wj WATKINS-JOHNSON

3333 HILLVIEW AVE., STANFORD INDUSTRIAL PARK, PALO ALTO, CALIF. 94304 • (415) 326-8830

Circle No. 4 on Reader Service Card

Editorial Board of the *Proceedings*, the last 2½ of these while Editor of *Spectrum*.

Jerry led the efforts of the staff of *Spectrum* to make the journal interesting, informative, and provocative, covering the technical and professional aspects of electrical and electronics engineering and including the interface of the profession and society. His tenure as Editor has truly been a time of innovation, and has set the pattern we expect that *Spectrum* will follow for years to come.

On behalf of the IEEE membership, we thank Jerry for a job well done.

M. E. Van Valkenburg, Chairman
Publications Board

Wincon revisited

I have perhaps an additional datum tending to substantiate remarks by Curtis Beach and Richard Gould in Forum for May regarding WINCON '70.

I did have a genuine need to know, and obtained the necessary clearances, but was refused admission to the classified sessions because I had not registered for WINCON. As a matter of fact, those at the gate seemed puzzled by what they obviously thought was an absurd attempt. It was pointed out that the WINCON registration desk, with cashiers paid for by WINCON, was set up right there in the hallway of the classified meeting rooms for the very purpose of registration. The only exception was that military personnel were allowed in with or without a WINCON '70 registration, since the military insists that all of its people be allowed to attend any meeting it supports. Not so, I am afraid, with IEEE. Similarly, "free" bus service (supplied by WINCON) from downtown Los Angeles to the classified meeting room was denied me although I had a clearance, and I had been accepted for attendance and am a member of IEEE, since I had not registered for WINCON.

Victor Azgapetian
Vice Chairman
Meetings and Conference Committee
IEEE Systems Science and Cybernetics Group

To code or not to code

Dr. G. David Forney's statement in his June article (pp. 47–58) that COMSAT is "hobbled" by being required to offer bandwidth "in narrow slices" gives a misleading representation of COMSAT's constraints. COMSAT is a carrier whose principal function today is to provide high-quality voice circuits. These voice circuits are required by their very nature, as well as by ITU Radio Regulations, to have a high signal-to-noise ratio at the customer interface. Since COMSAT does not share Codex's goal of providing coding per se, COMSAT cannot be hobbled by the fact that there happens to be little chance to apply coding in the provision of voice circuits.

COMSAT obtains the circuits from INTELSAT, which owns the satellites and operates them in accordance with the Radio Regulations. Unlike the voice circuits provided by COMSAT to its customers, the satellite channels are not characterized by high signal-to-noise ratio and narrow bandwidth. In fact, the reverse is true. At the present time, all INTELSAT circuits use wide-band frequency modulation at moderately low carrier-to-noise ratios.

As traffic demand grows, future generations of satellites will be required to make more efficient use of the spectrum, and carrier-to-noise ratios may be expected to increase. Coding and other means for conserving bandwidth are being investigated at COMSAT in anticipation of this requirement.

Incidentally, Dr. Forney's "simple argument" to demonstrate that in the bandwidth-limited region coding offers no dramatic gains is incorrect, as it depends on equating W to $1/2\tau$, instead of $1/\tau$ which is the case for AM. Actually, channel capacity can be approached in the bandwidth-limited region only by techniques that increase transmission rate by one bit/pulse for each doubling (not quadrupling) of the power. Amplitude modulation is not such a technique; combined amplitude modulation is not such a technique; combined amplitude and phase modulation is. The conclusion thus is still correct.

G. R. Welti, Manager
Systems Analysis Laboratory
Comsat Laboratories
Washington, D.C.

I welcome Mr. Welti's amplification of my offhand, one-sentence reference to COMSAT, which certainly was inadequate to describe its multifold services. I was concerned only with digital transmission and my point was that the regulations to which Mr. Welti refers and other inheritances from telephone practice have to date prevented COMSAT's use of satellite power and bandwidth in the most efficient way for data transmission. In fact, I believe that COMSAT now has no digital tariffs.

In turn, I object to Mr. Welti's misrepresentation of Codex's goal. We attempt to provide efficient digital transmission by the best means possible. In power-limited satellite communication, you cannot transmit efficiently without coding; this statement is due to Shannon, not Codex, and will become clear to COMSAT as it gets into digital transmission (though I am afraid that Mr. Welti will find coding of little help in bandwidth conservation). On other channels, coding may not be the answer; in Codex's 9600-b/s telephone line modem, for example, no coding is employed.

Finally, Mr. Welti has misunderstood my "simple argument." The number of samples per second and hence the bandwidth was held fixed as the power increased, hence the argument is independent of the relation between W and τ. Even so, I am sure Mr. Welti is aware of single-sideband techniques that allow AM to approach the Nyquist limit of bandwidth occupancy. This is not to say that combined amplitude and phase modulation may not be the best bandwidth-conserving approach for the Gaussian channel, but in this case each "pulse" contains two independent dimensions, and doubling the power increases transmission rate by ½ bit/dimension (sample), in accord with the capacity formula.

G. D. Forney, Jr.
Codex Corporation
Watertown, Mass.

Solid-state research engineers

The interest of electrical engineers in solid-state science has increased dramatically since the invention of the transistor. Presently a considerable number of advanced degrees are awarded by electrical engineering departments in the field of solid-state science. As a result, there are a large number of IEEE members, with electrical engineering degrees, doing basic and applied research on materials that subsequently find wide application in solid-state electronics. These people find themselves in a rather inequitable situation; i.e., they have no forum within their own professional society—the IEEE—in which to present or publish their work. These people, educated in our electrical engineering departments and supposedly represented by the IEEE, must look to physics or chemistry journals in order to publish their work.

Also, there are a vast number of electrical engineers doing development work who find it necessary to go to physics or chemistry conferences in order to gather up-to-date information on new electronic materials. It is my opinion, and that of several people with whom I have discussed the matter, that information on electronic materials research of interest to electrical engineers could be exchanged much more effectively if there were an IEEE group on electronic materials.

In this decade, we have seen electrical engineers doing research on compound semiconductors, organic semiconductors, dielectrics, amorphous semiconductors, etc. Their work, which is often of great importance to the electronics industry, is simply not publishable in any IEEE journal; for example, the IEEE Group on Electron Devices will not accept a paper unless it deals with some specific device.

In order to "rectify" this inequitable situation, I believe that it is time for the Group on Electron Devices to generalize to include electronic materials or else a new group on electronic materials should be formed.

The consequence of not taking action now will be for solid-state electrical engineers to affiliate with the American Physical Society or the Electrochemical Society rather than the IEEE.

David F. Barbe
Westinghouse Electric Corp.
Baltimore, Md.

Presentation of socio-technical material

For several months I have studied and thoughtfully considered Dr. Willenbrock's letter in *IEEE Spectrum*'s Forum of March 1970. Although I have a definite feeling that the horse has been stolen, I feel obliged to recommend a lock for the barn door.

The "socio-technical comment syndrome" has broken out in many of the technical groups. The rationale is that it is impossible to separate technical matters from their social–political impact on our society. This is a truism, but frequently it is merely an excuse for an airing of a partisan viewpoint.

I have a great admiration for the orderly quality of the engineering mind; however, objectivity frequently escapes those who discuss social and/or political problems. A technical publication is not a proper forum for social discussion; a technical journal is one of the last holdouts in this world of demagogy.

If there were no other periodical in which one could express his thoughts on social/political matters, I would heartily commend the use of *Spectrum* for this purpose. However, there is a plethora of publications in which one's socio-political thoughts may be aired.

More important, when one writes a "Letter to the Editor" of *Time*, *Newsweek*, *Fortune*, the *New York Times*, etc., his thoughts are generally considered to be an expression of his personality alone. However, when one writes to a professional journal, he is writing as one professional to other professionals and there is little credibility given to any disclaimer that this is merely a personal viewpoint. Anyone reading that technical journal takes the comments in a technical context and properly views them in the sense of the profession represented by the journal.

I give you an example of a technical journal that has lost its professional credibility. Several "concerned professionals" have been quoted in this publication as saying that the Safeguard ABM System is unworkable because it cannot be tested. In the same issue, these same "professionals" supported the solution of the world's problems by the simulation of the multitudinous interactions of world affairs. I don't deny the worthiness of either of these opinions, but obviously either one or the other is without accreditation, and must be motivated by more emotion than rationality.

This same policy has carried this publication further than they must have desired initially. They have even supported, overtly or tacitly, suggestions that employees who feel their efforts are being used for potentially dangerous purposes should destroy their work and, if necessary, destroy the equipment on which their work is performed. Obviously, this is a dangerous proposition.

What started out to be a desirable forum for professionals has turned out to be a platform for the more emotional (ergo the less professional) members of that profession. We have technical problems to spare; let us devote our technical journals to the seeking of solutions to those problems. At the same time, I commend to each and every responsible professional that he express his social and political stand in a constructive way, but in the more generic press.

Finally, 201 members out of 296 is certainly a majority, but of what? I suspect that it is a majority of a small number of the more progressive and vocal members of our profession, rather than a true spectrum (intended) of our membership.

If this is considered desirable, wouldn't it be more logical to submit this question as a referendum to the membership in toto?

T. H. Maguire, Jr., Weston, Conn.

The statement of "IEEE Policy on the Presentation of Socio-Technical Material" adopted by the Board at its January 1970 meeting and published in the March issue of *Spectrum* (p. 6) was the result of extended discussion by the Board. All of the letters received in response to Dr. Willenbrock's open letter of September 1969, as well as those commenting on specific published material, were reviewed. The Board also heard verbal arguments.

It would be impossible to set down all of the arguments the Board considered in arriving at this policy position. Since, in my view, Mr. Maguire's letter summarizes eloquently the principal arguments *against* such a policy, it may be appropriate for me to present briefly my personal views *in support* of the policy statement, without any implication that each member of the Board fully shares my convictions.

The business of engineering is the application of objective knowledge and specialized technical skill to the solution of problems that involve problems that involve people and the way they live. The success of any engineering effort ultimately is measurable only in social terms; that is, a successful engineering solution is one to which society assigns positive social value, through the economics of the market place or through the adoption of public policies that reflect a preference for (or even insistence on) certain engineering solutions as against other possible solutions. This is the rationale that Mr. Maguire acknowledges as a "truism."

I share wholeheartedly with Mr. Maguire the conviction that engineering should always be concerned with objective facts, and that engineering decision-making should, wherever possible, employ the proved logic of science. The laws of physics and the repeatable observations of laboratory experiment are not matters of opinion, and no poll of interested parties can affect their validity. The *relevance* of the known scientific facts to the solution of a particular problem is much more a matter of judgment; here expert opinion plays a defensible, even essential, role. Whenever expert opinion plays an important role in engineering decisions, and particularly when the decisions relate to matters of great public interest, controversy is inevitable. Controversy creates partisanship, as Mr. Maguire suggests.

Mr. Maguire urges that expressions of expert opinion on engineering questions of public interest should be confined to the general press. His argument, as I understand it, is that such statements in the general press are automatically discounted by the readers as an "expression of (the writer's) personality alone." If the same comments appear in (for example) *Spectrum*, he argues, "there is little credibility given to any disclaimer that this is merely a personal viewpoint." In my view, the actual situation is just the reverse of that Mr. Maguire describes. The general public is quite unable to evaluate an expert opinion or to judge the expertness of its author. Perhaps the most cogent reason for the publication of such material in a professional journal, when the subject is one that involves the technical and professional expertise of its readership, is only that there are such expert judgments exposed to the scrutiny of those individuals whose training and experience equip them for objective evaluation.

As a professional society, IEEE has an unavoidable responsibility to screen out from the mass of material submitted for publication any paper that misstates or misapplies demonstrable facts. IEEE also has the responsibility for giving priority in its publications to those contributions that maximize the product of technological insight and the significance of the potential application. (Discharging this responsibility entails personal judgment, and runs the risks of unwise choice.) Beyond these, IEEE's publications policy must recognize the responsibility to present conflicting views in those situations where expert opinion is divided with respect to engineering approaches to important problems. In my view, the policy statement adopted by the Board properly reflects all these considerations and represents an important step in the development of IEEE's service to its membership, our profession, and society at large.

John V. N. Granger, President, IEEE

Advance tables of contents

A number of our readers have told us that they object to our discontinuing publication of "Advance tables of contents." We will reinstate this department in the January 1971 issue if reader reaction is sufficient to warrant our doing so. Please address your comments to R. K. Jurgen, Managing Editor, IEEE SPECTRUM, 345 East 47 St., New York, N.Y. 10017.

Correction

George S. Moschytz of Bell Telephone Laboratories has brought to our attention several errors in his January article, "The Operational Amplifier in Linear Active Networks."

On page 47, in Eqs. (37) and (39), a minus sign should precede the entire expression on the right-hand side of the equality sign. As a consequence, in Table III (page 48) the second terms within the squared brackets in all four expressions should be positive instead of negative.

Spectral lines

Statement from the new Editor. With this issue the writer is taking over the editorship of IEEE SPECTRUM from Dr. J. J. G. McCue, who has worked hard and effectively for 2½ years in this service to our profession. It is a responsibility easy to step into because his well-defined intentions for the magazine make it fill a needed function and because he has maintained what appears to be a practical division of responsibility between the volunteer Editor and the very competent professional staff. The articles in this issue were all initiated or encouraged by Dr. McCue and his direct heritage will persist for some months.

The volunteer Editor takes primary responsibility for the subject and contents of the approximately seven feature articles that appear each month, and the selection of communications to appear in Forum. He is greatly assisted by volunteer reviewers in establishing the quality of the articles. The review service is of special importance for SPECTRUM articles because they span the entire range of electrical engineering arts and their supporting science.

The prime purpose of SPECTRUM is to display to all of us the wide range of electrical engineering so that we have initial access to all of the tools of our profession. Most of us work immersed in the details of narrow specialties. Few have the work scope of technical responsibility for large systems or the technical activities of large corporations. Hence it is easy, and sometimes necessary, to restrict ourselves to the literature of our current specialties. IEEE SPECTRUM is intended to carry tutorial and survey articles of sufficient pedagogic proficiency and authority to make it possible for any of us to invoke the power of arts we have not practiced, either because our current work can use them or because necessity or inclination require that we change our specialties.

The internal subtlety and the proliferation of electrical engineering specialties have made it impossible for schools of engineering to teach much more than the science and mathematics on which the current arts are based. The applications courses are often well selected in fields that will offer opportunity in the near future, but an engineering career is over 30 years long. The most we can hope for in engineering school is some practice in rational design so that new engineers understand the difference between engineering and science as professional fields and have enough personal experience with design to know if engineering really is their calling.

A second purpose of SPECTRUM is to tie our profession together by publication of matters of general interest to all of our members. Such articles include studies and proposals where the tools and insights of our profession are applied to social and international problems and the fine arts. We comprise a significant fraction of the world's educated innovators, working with some of the most sophisticated concepts and powerful aids to analysis and synthesis that the race has devised and financed. We live in a world whose major nations are engaged in apparently necessary but defensive and mutually reactive games that waste vast resources and use as their ultimate moves the infliction of suffering, death, and destruction. Those nations necessarily employ many of us in their defense. Since the demise of Hitlerism every major nation claims as its basic purpose the achievement of a world of peace, prosperity, and freedom for all peoples. In our most advanced nations, there are significant social problems and problems arising from lack of overall coordination of the many local optimizations at which we engineers have been so successful. And, finally, our ability to reproduce and broadcast the existing fine arts is probably only a beginning. Our tools, redesigned to go directly into the hands of artists, may lead to new and unimagined delights of creation and appreciation.

Can SPECTRUM really make it possible for each of us to have access to the other's specialty? It is easy to characterize this as a vain hope and a presumptuous claim. My belief that it is worth trying for is based in part on personal history. I was educated in power machinery, power systems, radio receivers, and wave filters. I worked at radio aids to navigation for 12 years and then abruptly went into semiconductor device design and fabrication, working in a small company. In large part the perhaps quixotic courage to enter this alien field came from tutorial magazine literature.* For about seven years the schools did little to prepare people for semiconductor device work and we were willing to take in inexperienced people who had the initiative to study the literature.

If you have read this far you may feel the call to share your arts and skills with all of us, or you may know of an area you would like to see covered. How do SPECTRUM articles originate? Some of them are on editorial initiative, with the Editor selecting a subject and finding an expert willing to cover it. Where an expert cannot be found or the material covers too broad a field, the job is undertaken by the professional staff basing their work on interviews with experts. Some of the articles are based on appropriate meeting presentations, either solicited by us or volunteered by the authors. About one third of the papers published are submitted without any initiative on the part of the editor or staff. Your new Editor hereby solicits your requests for subjects to be covered by review articles. Although this is no way for you to solve a pressing problem (you can count on a six- to nine-month delay), we are truly here to serve your needs. At best, editorial judgment on what to feature is weighted by a particular background. *David DeWitt*

*See "The physics of electronic semiconductors" by G. L. Pearson, *AIEE Transactions*, vol. 66, pp. 209–214, 1947.

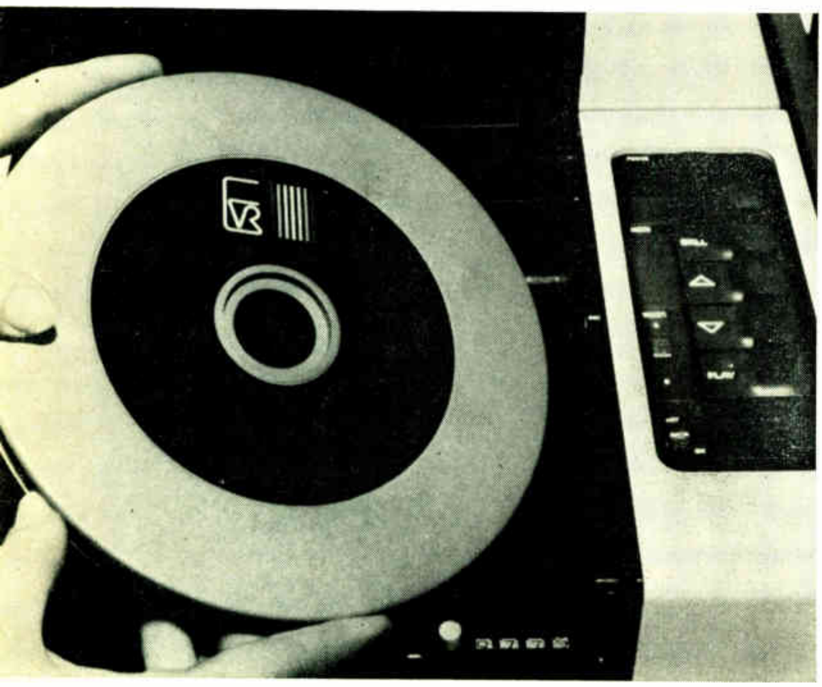
Color EVR

In color EVR, color television signals are encoded as a reel of black-and-white film images packaged in a disk. Run through a player that extracts color and luminance content, the film provides nearly a half hour's color television programming, or still pictures for study

*Peter C. Goldmark** President, CBS Laboratories

In electronic video recording (EVR),† the color television signals are recorded on a special 8.75-mm-wide, high-resolution photographic film, with two frames side by side. One frame contains luminance information; the other, coded chrominance information. Sound is recorded on two magnetic edge stripes. An 18-cm-diameter cartridge can play 25 minutes of color programming or twice that duration of black-and-white programming. Described in this article are the processes that, starting from an original program on video tape, produce a reproduction with excellent fidelity upon playback. A detailed explanation is provided of the mechanical, optical, and electronic arrangements—including the transcoding of the chrominance signal into a conventional color television signal.

FIGURE 1. The EVR player cartridge.



Through a unique combination of photography, optics, and electronics, EVR plays prerecorded video programs, including sound, through standard television receivers—truly a visual counterpart of sound reproduction from a long-playing record. The nature of the medium used lends itself to low-cost, high-volume production of monochrome or color programs.

The EVR program is contained in a cartridge measuring 17.7 cm in diameter and 1.3 cm thick. With its large center hole, the cartridge resembles a 45-rpm phonograph record. (See Fig. 1.) The EVR video signal is electronically printed on a special photographic base, 8.75 mm wide and approximately 75 μm thick. The sound is recorded on two narrow magnetic stripes at each side of the EVR film—providing either stereo or two independent audio programs. One cartridge can hold two black-and-white programs of 50 minutes' duration (60 minutes with European standards). A color program occupies both video tracks, and hence has half the playing time of the black-and-white counterpart. (With thinner film becoming available, the program duration per cartridge will be accordingly longer.)

The information contained in the EVR cartridge is translated by a player into an RF signal that encompasses the combined video and audio programs, which is then carried by wire to the antenna terminals of a standard color or monochrome television receiver. Push-button-operated controls take care of the threading, stopping, fast-forward, and rewind operations of the cartridge. With a suitably indexed film, one can stop at any particular part of a program and view still pictures for any length of time. Manual forward- or reverse-browsing facilities are provided. These facilities can also be used when stopping a motion sequence. Since each picture is exceedingly small—0.31 by 0.25 cm—

* The development of electronic video recording (EVR) has been a CBS Laboratories team effort. The following made significant contributions to the project or helped to write this article: Robert A. Castrignano, John W. Christensen, C. Russell Dupree, Bernard Erde, Dennis Gabor, William E. Glenn, Abraham A. Goldberg, Patrick F. Grosso, John M. Hollywood, Renville H. McMann, Ivan A. Purt, Robert B. Rhoades, Donald W. Ridley, Andrew A. Tarnowski, John C. Wistrand.

† EVR (a registered trademark) was developed by CBS Laboratories; it is described here in detail for the first time.

EVR has a large storage capacity with significant potential for reference-library or other types of visual-information storage.

The EVR system is optimized to provide quality equivalent to the best-received off-the-air pictures on U.S. or European color receivers. On a suitable closed-circuit monitor, the monochrome resolution of EVR can reach 500 television lines in a horizontal direction, and is limited only by the phosphor-dot structure of the color tube.

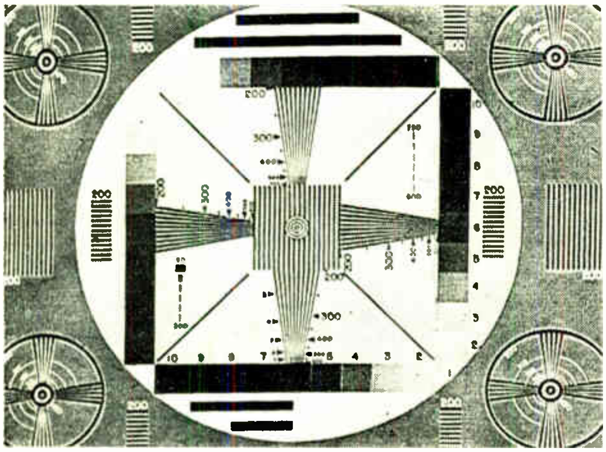


FIGURE 2. Test pattern enlarged from an EVR release-film frame.

The wear qualities of EVR cartridges are extremely good. Hundreds of plays are possible without noticeable deterioration. Neither is there significant wear in any part of the player. A three-inch flying-spot cathode-ray tube will provide several years of service. If necessary, it can be replaced simply and inexpensively.

The film is important

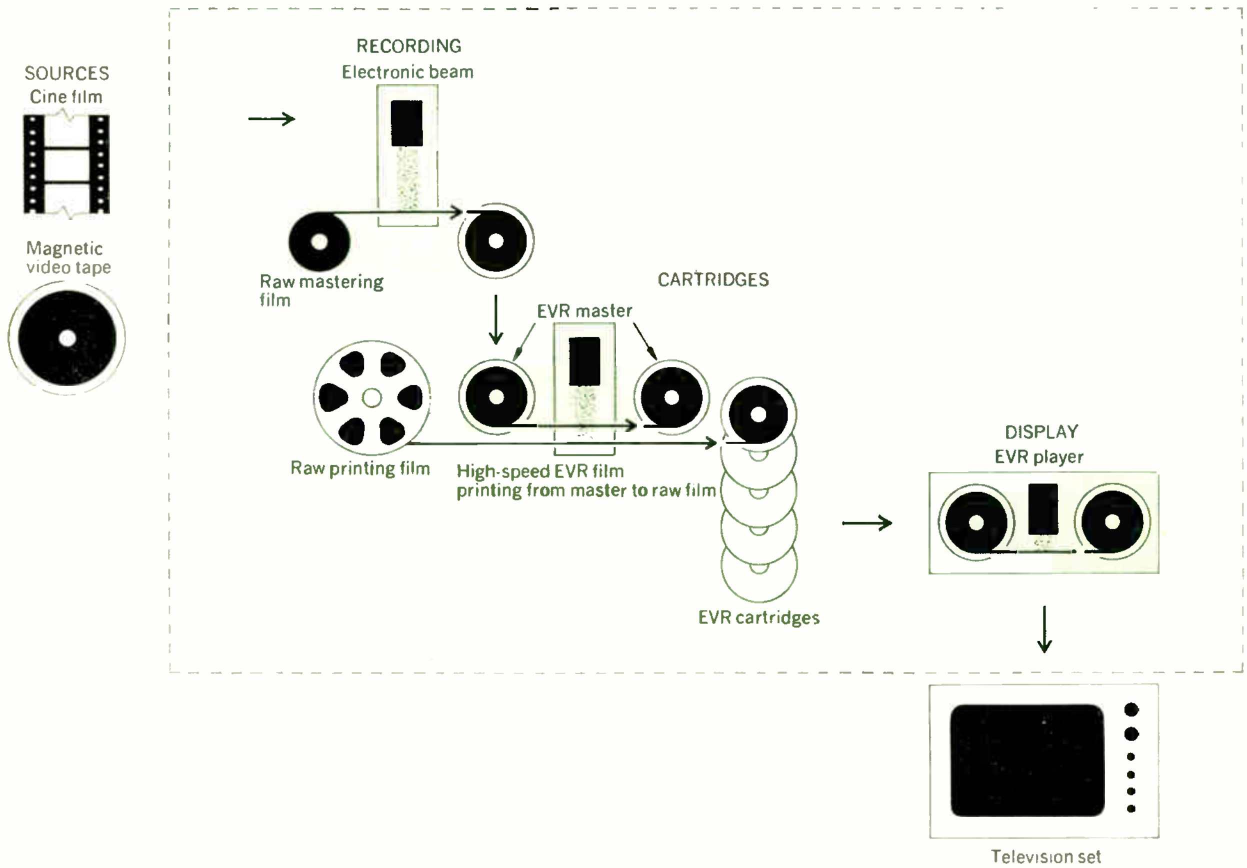
The film sensitometry in the EVR system represents an essential element in interfacing electron-beam recording, printing duplicates, and playback.

A major goal during the development of the EVR system was to devise a film-recording and -duplicating method that would permit quantity production of inexpensive film cartridges containing high-quality video programs. For high-speed production of EVR duplicates, a low-cost ultrahigh-resolution silver halide film was developed by Ilford Ltd. of England.

Modern, very-fine-grain films have a high capacity for information storage and are relatively insensitive to ordinary light, yet very responsive to a high-energy electron beam. The stock film used for making the master electron-beam recording has a crystal size less than 0.1 μm and the definition obtainable is of the order of 800 line-pairs/mm. This film too has been developed especially for electron video recording by Ilford Ltd. Figure 2 shows the enlargement of a single release-print frame.

Stringent quality and size requirements resulted in the selection of direct electronography (exposing the film

FIGURE 3. EVR system concept is shown within color tint.



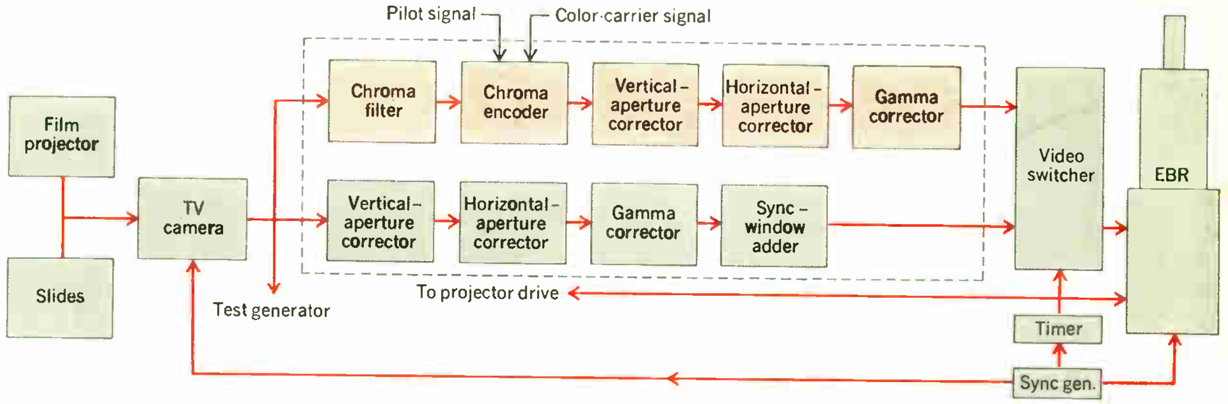


FIGURE 4. Electron-beam-recorder signal processing.

FIGURE 5. Frequency-transfer characteristics of the EVR system. Numbers along light dashed line indicate the change of the 4-MHz-frequency preemphasis along processing route.

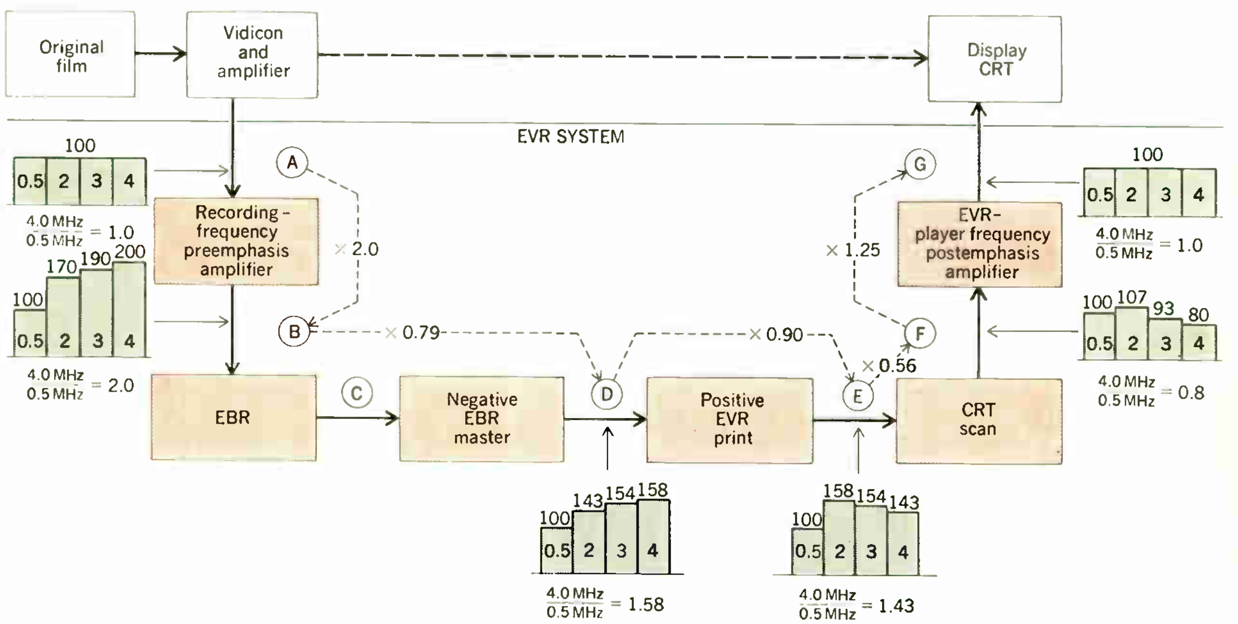
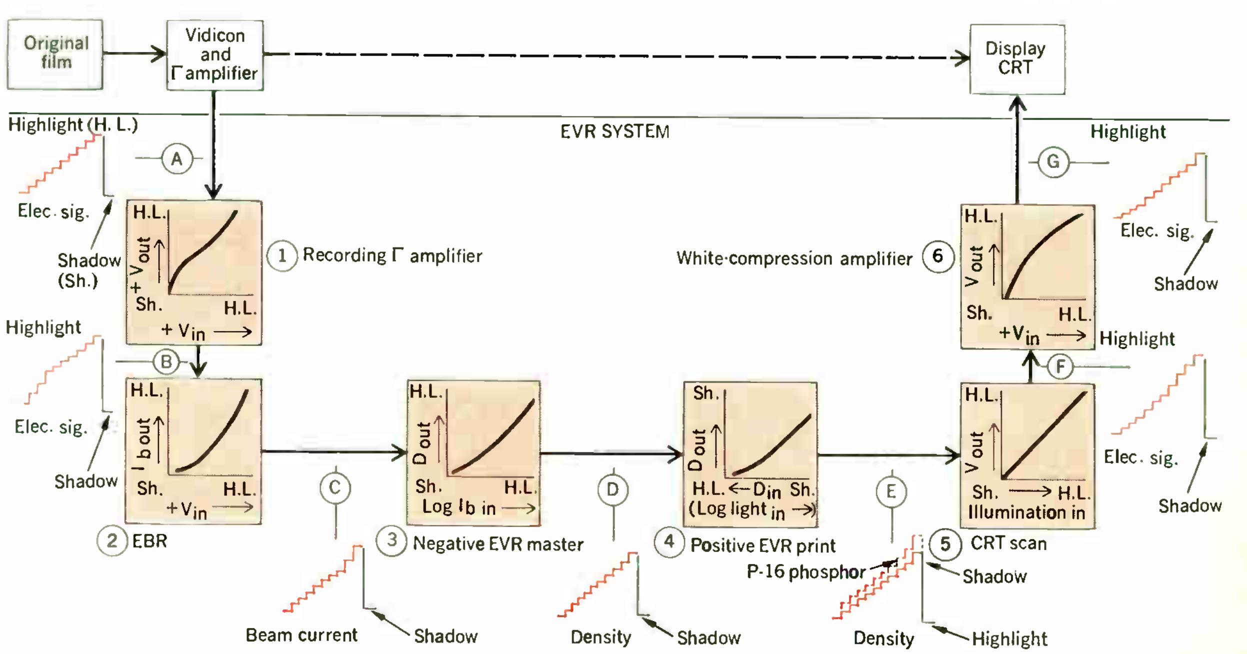


FIGURE 6. Gray-scale transfer characteristics of the EVR system. I_b is the beam current.



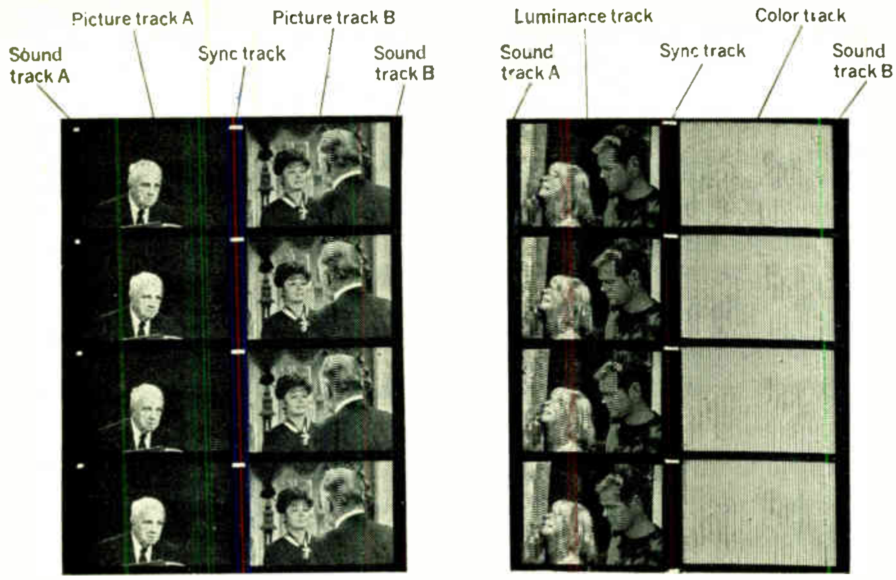


FIGURE 7. Two EVR film strips. On left, two black-and-white programs. At right, a color feature.

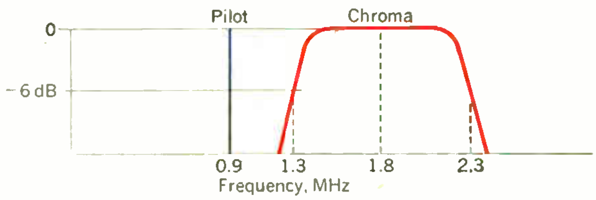


FIGURE 8. EVR pilot- and chroma-frequency spectrums.

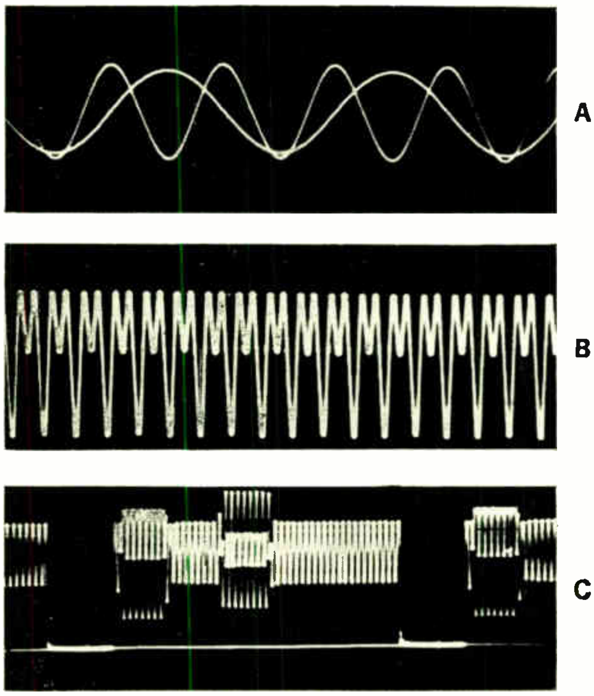


FIGURE 9. A—Chroma subcarrier has half the frequency of the pilot and leads it by 90 degrees. B—Sum of chroma and pilot carriers pictured in (A). C—Sum of chroma and pilot subcarriers during negative I (see box) test pulse.

in vacuum by a finely focused electron beam) as the method for creating the master record.

Processing the master through electronic recording makes it possible to produce a predetermined frequency preemphasis (of the television signal as recorded onto the film) and to introduce gray-scale correction while keeping the film density range limited to the desired values. All of these controls are not collaterally available in a purely optical-transfer system.

The picture on the film is different in appearance from that of normal cinematographic film, especially since color pictures appear in monochrome. (The chroma content is coded.) A further difference is in the number of pictures recorded per second onto the master and replayed per second by the recorder. Through the use of extremely fine-grain print material, very small images can be produced, making it economical to print 60 frames per second (50 in Europe). This rate leads to simplifications in the playback machine and tends to provide a high degree of visual integration (that is, smearing) of grain or other imperfections during replay.

The three basic components of the EVR system (see Fig. 3) are the recording apparatus for conversion of programs from films or tape to a master film; the printing process to mass-produce the cartridges; and the player, which is attached to the antenna terminals of one or more television sets.

The principal steps involved in producing the EVR cartridges are

1. Program preparation.
2. Generation of the master negative by electron-beam photography.
3. Printing, processing, and slitting of the films and loading of the cartridges.

The program-preparation process electronically precorrects the video signal for any losses that will occur throughout the entire system, including the player. As shown in Fig. 4, the original NTSC color television signal is separated into its luminance and chroma components. Both signals are enhanced through vertical and horizontal detailed adjustments and both are gamma (contrast) corrected. In Europe the same process takes place using the PAL system requiring a luminance bandwidth somewhere between 5 and 5.5 MHz. France and the U.S.S.R. have an 8-MHz channel with a luminance bandwidth of 6 MHz. The EVR system can easily satisfy these values.

Figure 5 indicates the amplitude-versus-frequency characteristics in the various parts of the system. Careful gamma control is employed throughout, from recording to playback. Figure 6 illustrates the typical gray-scale characteristics at various points of the process.

A synchronization mark (or window), indicating the start of each frame, is recorded on the master film during the horizontal-blanking interval. This line of windows provides synchronization in the player between the film transport and the CRT scanner in the player. Figure 7 shows a section from a monochrome- and a color-release print. Each frame is 2.34 mm high by 3.12 mm wide. The synchronizing window next to each frame is visible. Color information is recorded in a coded form on the frame adjacent to the luminance-signal frame.

It is essential to record the chroma signal in such a way that reproduction is independent of the scanning linearity in the record and playback systems. The chroma frame therefore contains a color-carrier frequency that is an integral multiple of the line-scan frequency.

In order to provide a reference carrier for the color signal, an unmodulated pilot signal having exactly half the color-carrier frequency is combined with the latter for recording across the chroma portion of the EVR film. Scanning nonlinearity, raster-size changes, and film shrinkage—among other possible influencing factors—will not interfere with the proper demodulation of the chroma carrier since the phase relationship between it and the pilot carrier automatically is maintained within the required accuracy. Figure 8 shows the chroma- and pilot-carrier relationship together with the chroma sidebands and Fig. 9 contains typical oscillograms of the two carriers.

All synchronizing signals and the pilot carrier are divided down from the color-carrier frequency. Because of the integral relationship between the color and pilot carriers and the electron-beam recorder (EBR) horizontal-scan frequency, the pilot and chroma signals appear on the master film as a series of vertical bars, deviating in a horizontal direction only where color changes occur in the picture. Actually, the recorded chroma picture is an intensity rendition of a composite signal like that of Fig. 9(C).

In the EBR color encoder the red–yellow and blue–yellow signals are arranged to modulate the phase and amplitude of the color carrier, which is then mixed with the pilot carrier. Preceding the chroma information, four cycles of the pilot carrier are recorded, appearing at the left side of the chroma area. This recorded information causes the pilot extraction filter in the EVR player to "prering," thus rendering its transient invisible in the reproduced picture.

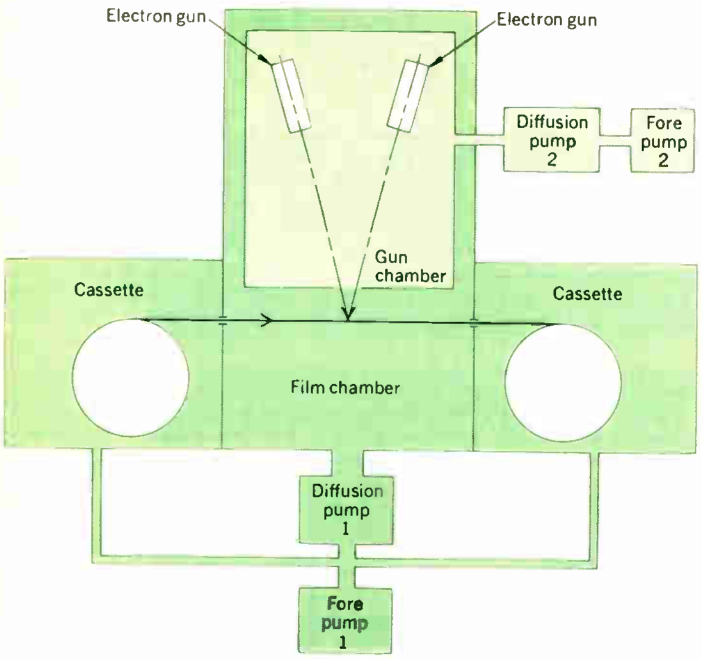
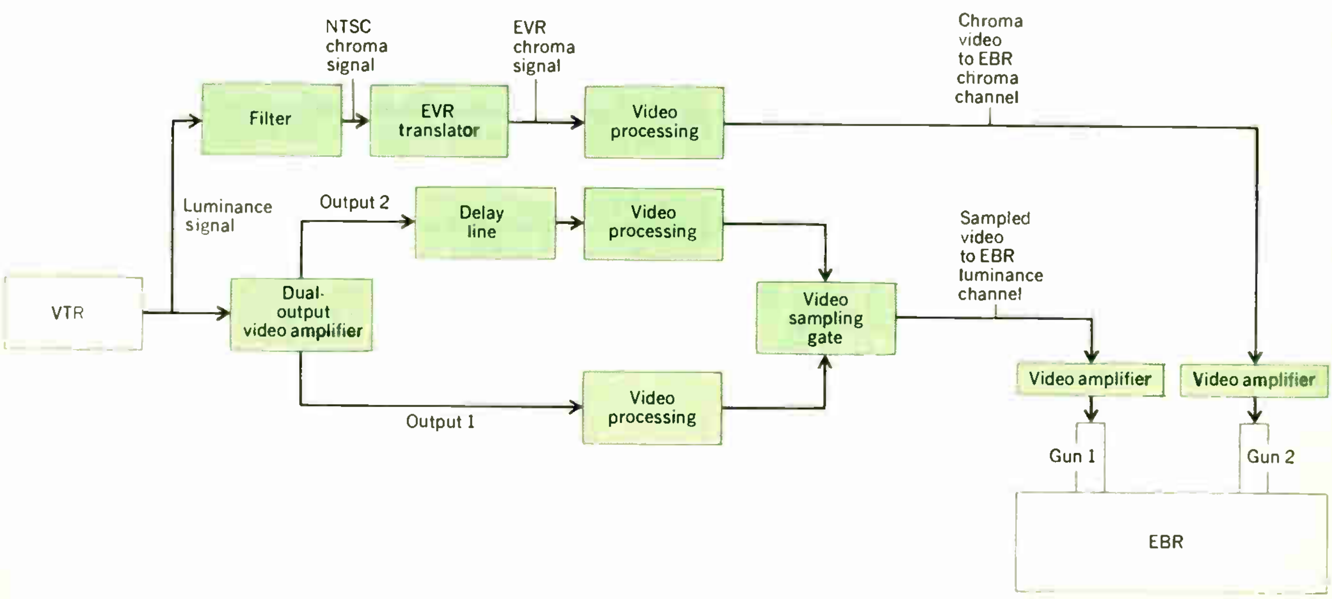


FIGURE 10. Electron-beam-recorder vacuum system.

FIGURE 11. Use of delay and sampling for real-time recording of 525 lines per frame.



The electron-beam recorder

The diagram in Fig. 10 shows the various vacuum chambers. In the space where the video-modulated electron beam exposes the film, the electron path from cathode to film is maintained at a pressure of approximately 10^{-8} atmosphere in order to insure satisfactory beam focus and cathode life. To achieve a pump-down time of a few minutes, a double-vacuum pumping system is used. A small diffusion pump maintains the electron-gun area at the required low pressure and another vacuum system maintains the film-transport mechanism and the film reels at approximately 10^{-3} atmosphere.

There is virtually no loss of resolution in the actual process of electron-beam recording because the diameter of the electron beam is 2.5 μm and the recorder is capable of a resolution of 400 line-pairs/mm. Electron-beam recorders for commercial production of EVR masters have recently been completed for use in Europe and the U.S. For color EVR, the same type of machine carries a dual-beam gun for side-by-side recording of the luminance and chroma signals. A 40-mm-wide film is used for the master and 35-mm film is used for duplicate prints. The 35-mm format accommodates four 8.75-mm EVR films, which can hold eight monochrome or four color programs. The four strips are printed simultaneously and subsequently are slit.

Although the recording system can operate directly with real-time signals from television cameras or film, an operational advantage can be derived by prerecording all programs on video tape.

The EVR master-making system is illustrated in Fig. 11. The video signal generated from the video-tape recorder (VTR) is separated into chroma and luminance components. The luminance signal is then divided into a direct signal and a signal that is delayed by one field. The delayed and the undelayed luminance signals are independently passed through vertical- and horizontal-aperture correctors, and gamma correctors. Then both the delayed and the undelayed luminance signals are applied to a sampling gate operating at a 14-MHz rate in such a way that both fields are synchronously recombined during $^1/_{60}$ second ($^1/_{50}$ in Europe). Thus information corresponding to all 525 lines (625 in Europe) is recorded in each EVR film frame. The objective is to produce a luminance picture on the EVR film that, for all practical purposes, is a homogenous photograph with full vertical resolution. When this photograph is scanned in the EVR player there is no need for the flying-spot scanner to track any particular line on the film.

The chroma portion of the video-tape-recorder playback signal is extracted from the luminance information by a comb filter and then translated from NTSC frequency standards to EVR-system values. The recombined luminance signal and the chroma signal are applied through video amplifiers to the guns of the electron-beam recorder. Figure 12 shows the enlarged structure of such a recorded release print together with a highly magnified section of that print.

Two types of field-delay methods are now available: magnetic-storage disks and quartz line. One magnetic-storage system employs a recording and a playback head located 180 degrees apart; in the other system, the heads are 360 degrees apart. The quartz line contains eight delay sections connected in series. Temperature control insures maintenance of precise transit times.

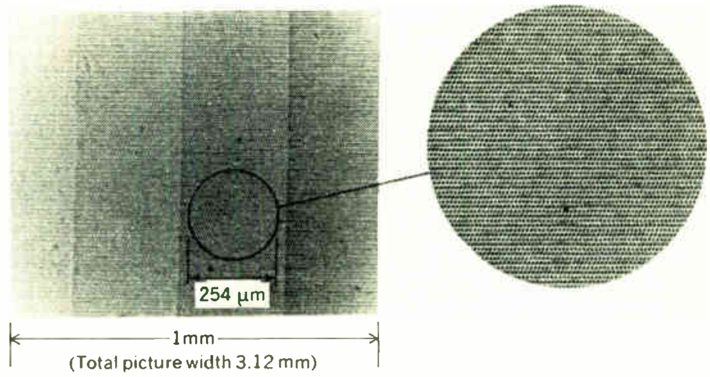


FIGURE 12. Sampling structure in real-time electron-beam recorder.

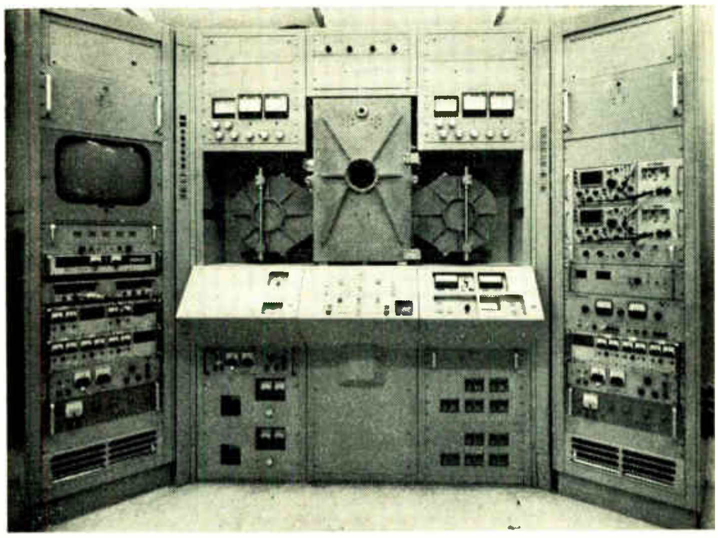
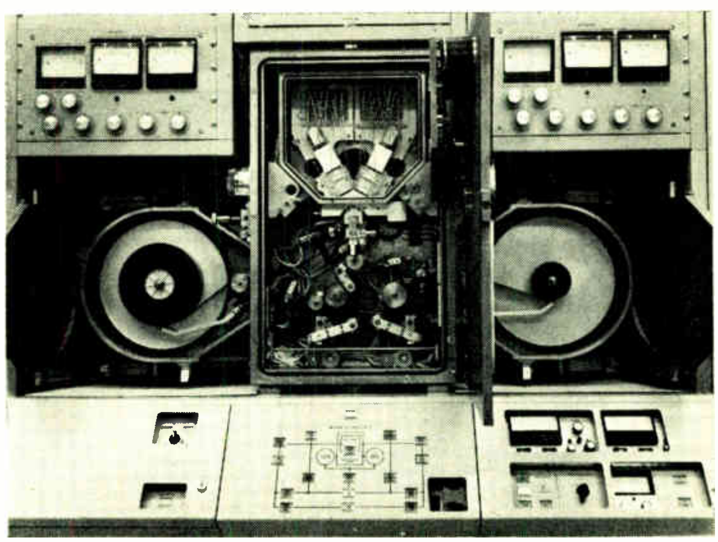


FIGURE 13. Electron-beam recorder.

FIGURE 14. Electron-beam-recorder vacuum chambers, electron guns, and film-transport mechanism.



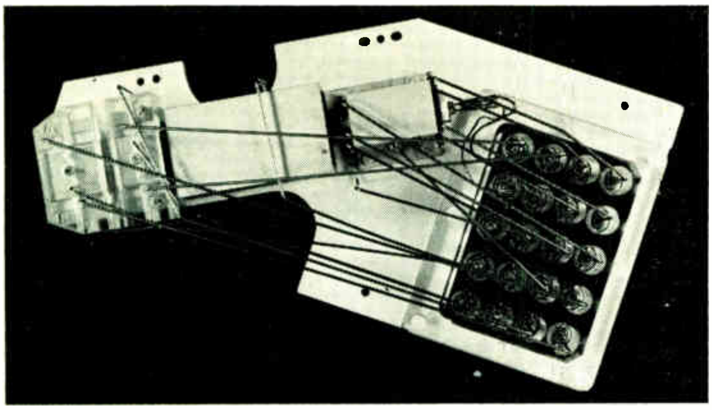


FIGURE 15. Electron gun of electron-beam recorder.

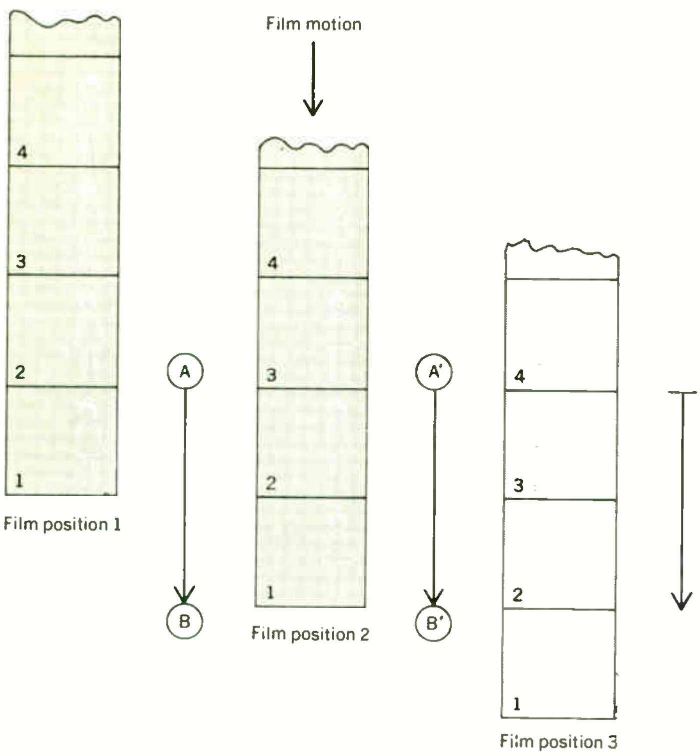
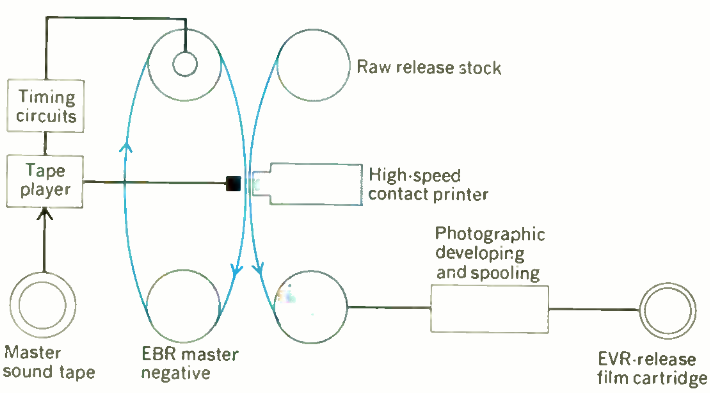


FIGURE 16. EVR following-raster, scan-timing diagram. Vertical scan starts at (A), corresponding to the top of frame 1. Film and vertical scan travel downward with the scan traveling at twice the velocity of the film. See text for complete details.

FIGURE 17. High-speed printing method for EVR release film.



The electron-beam recorder is shown in Fig. 13. The recorder contains all the circuitry necessary for operating the electron gun as well as the vacuum system. As shown earlier, the latter consists of three chambers: one housing the film magazine, the second incorporating the film drive, and the third enclosing the electron gun. Figure 14 is a close-up of the vacuum chambers and the film transport. The film magazine capacity is 630 meters of 40-mm-wide film, sufficient for recording 4 hours of color or 8 hours of monochrome programs without breaking vacuum.

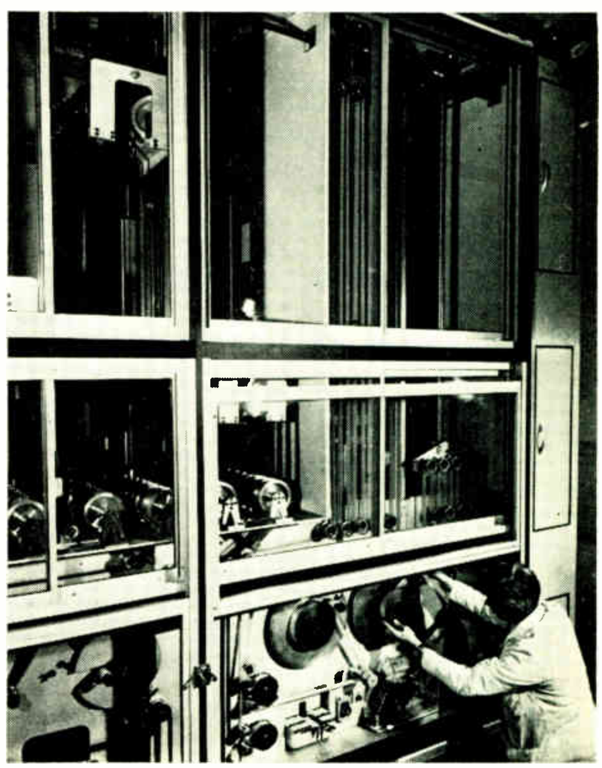
Electron-beam recording on the film is accomplished with two electrostatically focused and deflected electron guns located in the gun chamber. The latter can be indexed to four discrete horizontal positions, thus making it possible to record sequentially four dual tracks across the width of the 40-mm master film.

The special electron gun designed for the EVR master recorder is shown in Fig. 15. The vertical deflection of the modulated electron beam follows the direction of film motion but at twice the film's velocity. As shown in Fig. 16 the vertical scan starts at the top (*A*) of the film image and at the end of $^{1}/_{60}$ second reaches the bottom of the frame while the film has moved from position 1 to 2. During vertical blanking the electron beam returns to start the process over again on the next film frame.

The film drive provides accurately controlled, continuous motion at 15 cm/s for the U.S. version and 12.5 cm/s for the European version. An electronic servo drives the film at constant velocity, locked to the vertical scan and synchronized with the video-tape recorder.

A special multihead, wet-gate duplicating printer of entirely new design has been developed by Ilford Ltd. in England—a member of the EVR partnership in

FIGURE 18. High-speed EVR film production printer.



Europe. Figure 17 is a diagrammatic view of the high-speed printing operation showing one printing stage only and Fig. 18 shows a portion of the printer itself. The wet-gate EVR printer minimizes light dispersion and also protects the master film. The EVR printing and processing equipment is capable of running at speeds of up to 60 meters per minute. Through the use of multiple heads, 16 color programs, together with sound, are generated each time the master loop passes through the printing machine, and the rate at which the printer produces EVR copies is approximately 100 times (in Europe 125 times) faster than the actual playing time of the original program. Thus, a half-hour color cartridge can be produced every 18 seconds (14.5 seconds in Europe).

Experiments have been carried out at CBS Laboratories over a period of several years to investigate the use of Diazo film for EVR duplication. Diazo film is less expensive than silver halide film, is virtually grainless, and requires only two steps for the duplication process. The first step is the exposure to ultraviolet light through the master film, and the second involves the conversion of the latent image into a visual one in a hydrous-ammonia atmosphere. The Diazo film, which has been specifically adapted for EVR use, is positive working, has unity gamma, and is coated on a 71-μm-thick triacetate base. Figure 19 represents typical sensitometric data for Diazo film especially made for EVR duplication.

The EVR player

The laboratory-prototype color EVR player, on which the Motorola production players are based, is shown in Fig. 20. In Fig. 21, the cover has been removed to show the cartridge deck, cathode-ray tube (CRT), and associate circuitry, all mounted on an internal metal frame.

A cartridge is played by opening the small door over the well, placing the cartridge on the hub in the well, and closing the top. To the right of the well are six push buttons for controlling the film-transport functions. Pressing the *Play* button causes the cartridge leader to thread through the deck, and the player automatically to start. The large knob on the front permits the film to be moved backward or forward while one is viewing still pictures.

FIGURE 19. Transfer characteristics of Diazo print film.

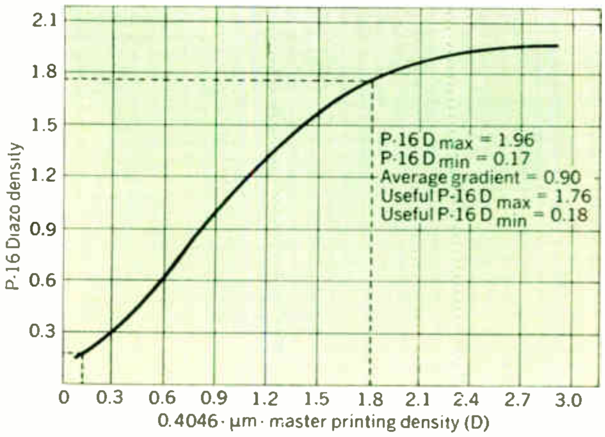


Figure 22 shows a unique CRT, also developed in the CBS Laboratories, that scans the film image by having its flying spot projected through suitable optics. The light is modulated by the film and converted by photomultiplier tubes into luminance and color signals. The CRT has a resolution equivalent to 50 percent contrast at 10 line-pairs/mm across the entire raster. Magnetic focus and deflection are employed. The tube has a diameter of 3 inches, is 19.8 cm long, and operates at 20 kV.

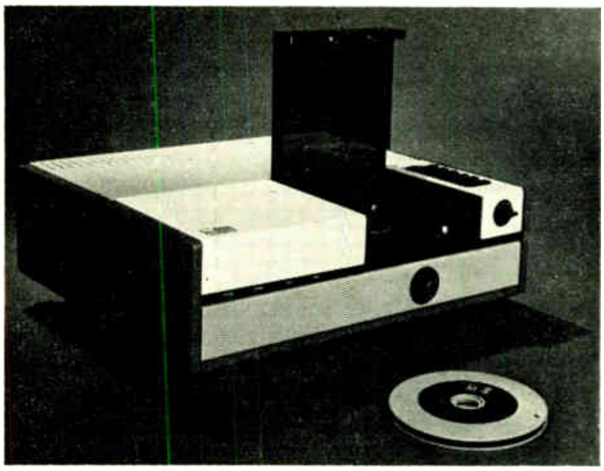


FIGURE 20. Prototype EVR player.

FIGURE 21. Internal view of EVR production player.

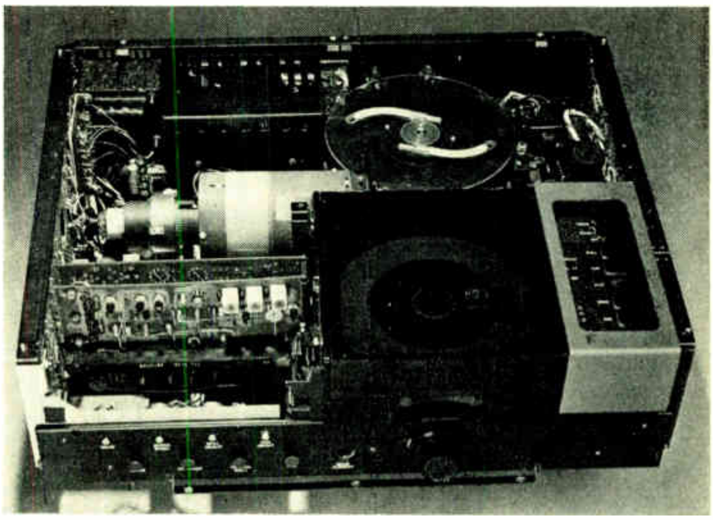
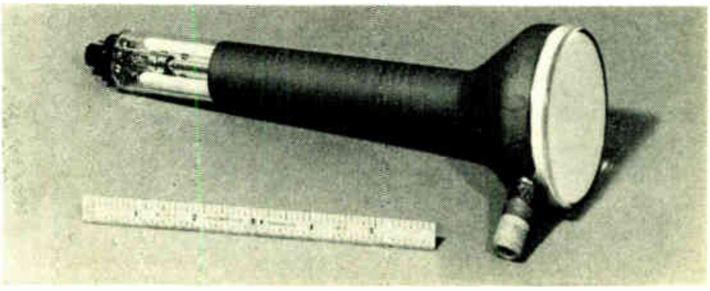


FIGURE 22. Flying-spot CRT for EVR player.



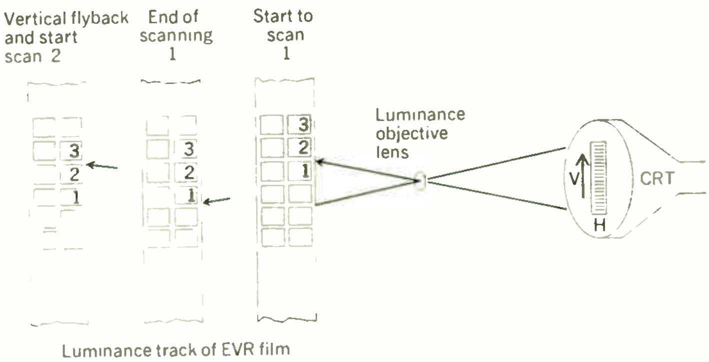


FIGURE 23. EVR player forward-scan raster.

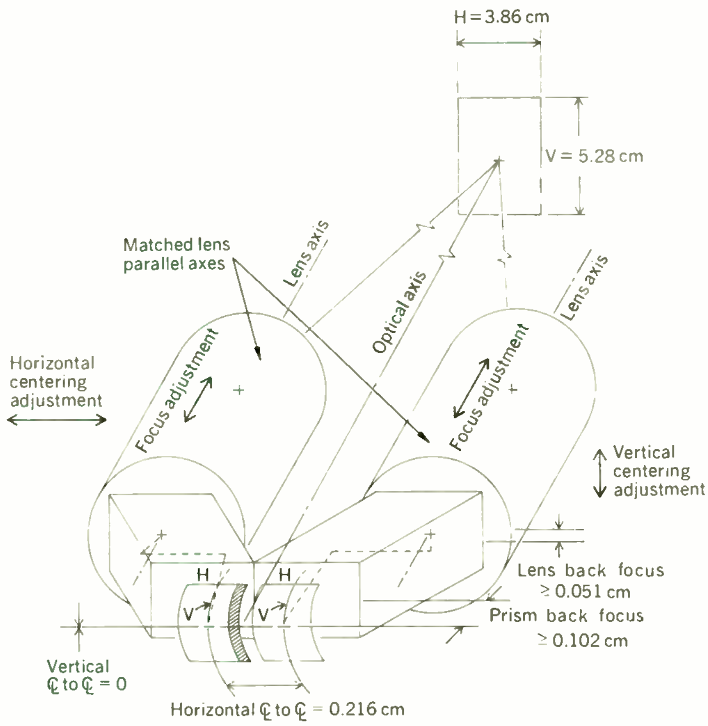
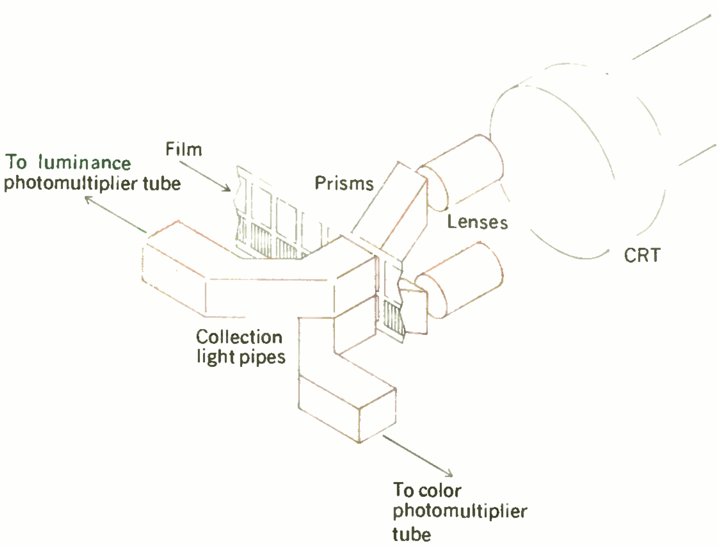


FIGURE 24. EVR color-player lens and prism assembly.

FIGURE 25. EVR color-player optical system.

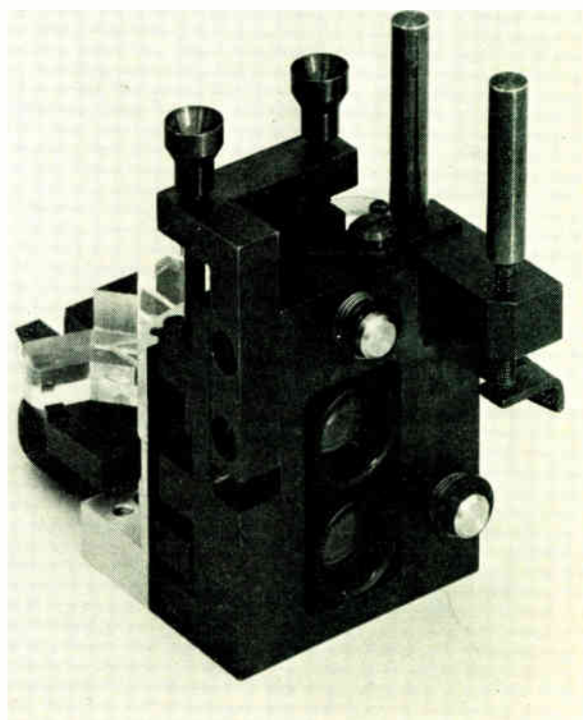


The CRT light output is kept constant throughout the life of the tube by an automatic brightness control. The P-16 phosphor decreases its light output as it ages so that the initial light-beam current of 10 μA can be increased gradually to about 80 μA without affecting the system resolution—thereby appreciably extending useful tube life. This amperage readjustment is accomplished by use of a photoresistor positioned to view the raster and control the bias of the CRT through a closed-loop, automatic brightness circuit.

Optical scanning of the luminance and chroma tracks of the EVR film, shown in Fig. 23, also employs the forward-raster-scan technique. At the start of the field, the beam from the CRT scans the top of the first picture. As the film moves at a constant speed of 15 cm/s, the beam also moves in the same direction but at twice the velocity, or 30 cm/s. Thus, by the time picture 1 has moved to the position shown in the second strip of film, the light beam has completely scanned it and rests at the foot of picture 1. At this instant, a vertical-sync pulse, derived from light flashing through the clear window on the film, initiates vertical flyback. The CRT beam returns to the top of picture 2, ready to start the next scanning period.

Since the timing of the CRT beam is controlled by the film velocity, the film speed can vary within a certain range without affecting the vertical steadiness of the reproduced television picture. The film drive is servo-locked to the 60-Hz main by comparing the speed of the recorded field pulses with the mains frequency. The horizontal oscillator of the player is free-running at a frequency of 15 734 Hz, thus the picture has random interlace. A fully interlaced picture can be obtained at the cost of more circuit complexity but subjective tests

FIGURE 26. EVR color-player gate, lens, and light-collection assemblies. Lenses and light pipes are visible.



show that random interlace is quite acceptable.

The luminance and color frames of the color EVR film or the two program tracks of a monochrome film are scanned through a dual optical system. Figure 24 is a schematic view of the total optical assembly. The imaging optics consist of two lenses, two rhomboidal prisms, a lens mount, and a film gate that holds the film in a cylindrically curved image plane. The schematic arrangement of these components is shown in Fig. 25. Each lens images the 3.8- by 3.1-cm CRT raster onto the film gate, forming two identical phosphor-dot images with centers 0.36 cm apart and reduced by a factor of 11.3.

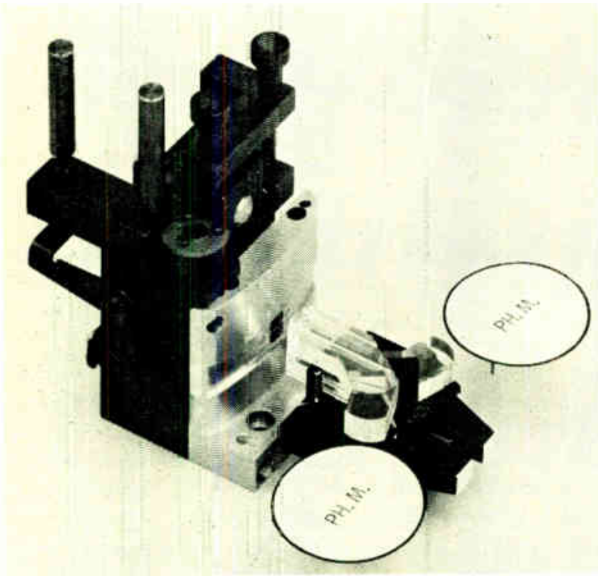


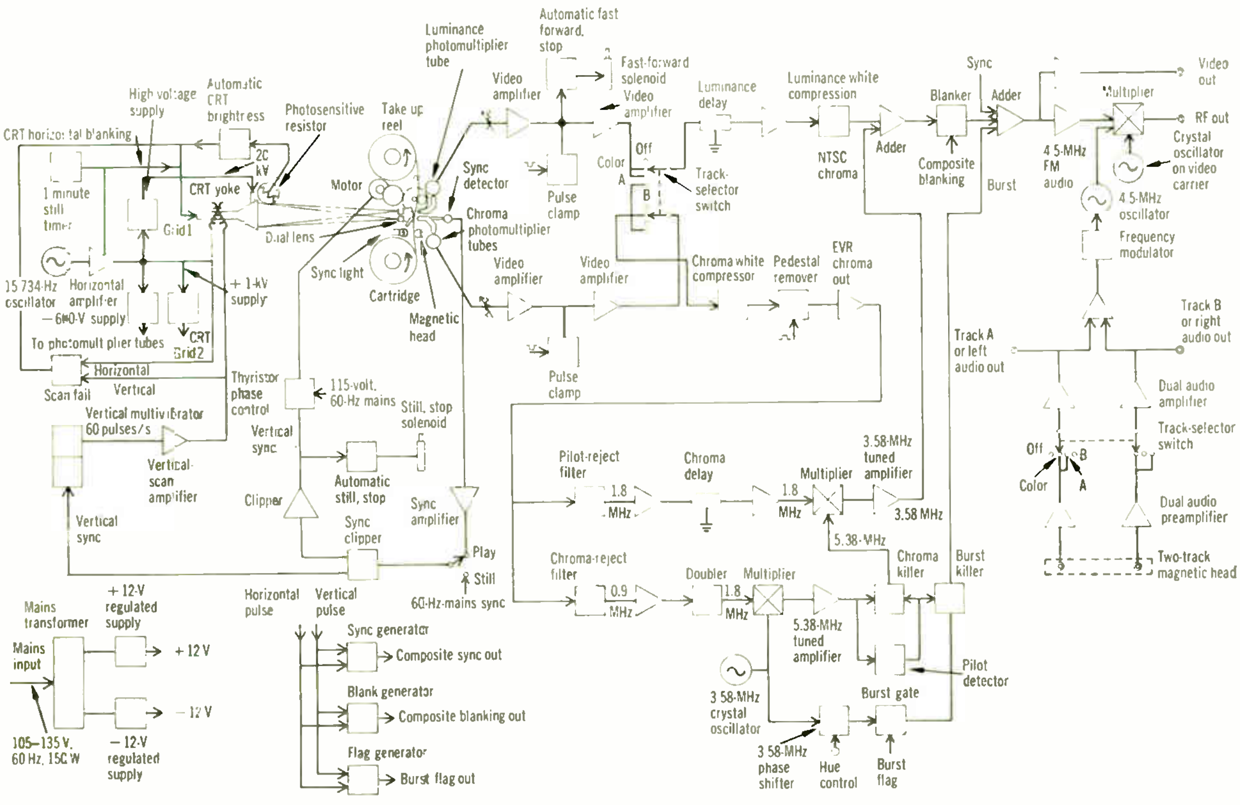
FIGURE 27. Opposite view of assembly shown in Fig. 26. Gate, light pipes, and photomultiplier tubes are exposed.

Both lenses and prisms are designed to have optimum performance for the spectrum of the P-16/S-4 (CRT/photomultiplier tube) combination with peak energy at 0.385 μm. Each lens-and-prism combination is designed for a flat object field and for an image field spherically curved with a radius of about 15 cm. This setup provides the best practical match to the cylindrically curved gate having a radius of curvature of some 7.5 cm. Adjustments are made at the factory for centering and focusing the lenses. The two small prisms, which provide greater separation of the lenses, require no adjustment. The mechanical assembly of the imaging optics is shown from the CRT side in Fig. 26. The other side, showing the two apertures in the curved gate, two lucite light pipes, and the positions occupied by the photomultipliers, is shown in Fig. 27.

Each lens has a focal length of 15 mm and will resolve in the film plane 60 line-pairs/mm with 60 percent response relative to 5 line-pairs/mm at an effective aperture of *f*/2. Registration of the luminance and chroma images is relatively easy because the low-resolution color image, containing all primaries, is superimposed on the high-resolution luminance picture. Registration is factory adjusted by varying the *X*-axis position of the luminance lens against the *Y*-axis position of the chroma lens while observing reproduced pictures on a television monitor.

The basic circuit elements of the color EVR player are shown in Fig. 28 and include:

FIGURE 28. Color EVR player block diagram.



Characteristics of the EVR-encoded color signal

The EVR color signal E_m consists of the linear sum of a pilot signal E_p, a chrominance signal E_c, and a color-difference video signal E_v.

$$E_m = E_p + E_c + E_v \qquad (1)$$

The pilot carrier frequency f_p is the 56th harmonic of the line-scanning frequency f_h.

$$f_p = 56 f_h \qquad (2)$$

The chrominance carrier frequency f_c is the second harmonic of the pilot carrier frequency.

$$f_c = 2 f_p \qquad (3)$$

The chrominance signal consists of the sidebands of two suppressed carriers in quadrature.

$$E_c = E_Q' \sin(2\pi f_c t) - E_I' \cos(2\pi f_c t) \qquad (4)$$

The amplitudes of the quadrature carriers are generated by mixing the red, green, and blue video signals.

$$E_I' = 0.60\, E_R' - 0.28\, E_G' - 0.32\, E_B' \qquad (5)$$
$$E_Q' = 0.21\, E_R' - 0.52\, E_G' - 0.31\, E_B'$$

The pilot signal is given by

$$E_p = A_p \sin(2\pi f_p t) \qquad (6)$$

A bandwidth-limited, color-difference video signal E_v, corresponding to $-E_I'$ of amplitude k relative to E_I' max, is added to the pilot and chrominance signals to achieve minimum peak-to-peak excursion of the composite signal envelope.

$$E_v = -k\, E_I' \qquad (7)$$

- CRT-deflection and high-voltage supply.
- Dual photomultipliers (PMT) and video amplifiers.
- Chroma translator for converting EVR signals to NTSC.
- Pulse generation for blanking and composite sync.
- Magnetic heads, audio amplifiers, intercarrier sound generator.
- RF links: video and sound modulators, carrier generators.
- Power supplies and motor-control circuit.

Chroma translator

Color EVR, as noted, divides the picture information into two frames: one for the luminance, and the other for color. The EVR film is scanned without line tracking and each luminance frame is composed of 525 lines (625 in Europe). It was mentioned, in relation to electron-beam recording, that EVR chroma is composed of two color-difference signals, each modulating quadrature-phased variants of the same suppressed carrier. The color-carrier frequency is 1.8 MHz and the width of the color-difference-signal sidebands is ±0.5 MHz. (The color equations are given in the accompanying box.)

Again it should be emphasized that the width and linearity of the recorded images, and of the CRT raster, cannot be maintained sufficiently uniform in all parts of the picture to assure constant frequency and phase values for the chroma subcarrier. A continuous 0.9-MHz pilot signal is added to the chroma carrier during recording to make the system self-correcting on playback.

The maximum EVR color-difference-signal bandwidth is the same as the Q bandwidth in the NTSC system: 0.5 MHz at −6 dB. In EVR, the I and Q bandwidths (see box) are equal because nearly all color television receivers are designed for the same bandwidth-color-difference demodulators. It is simple to convert EVR color to NTSC directly, by frequency translation, without demodulation to baseband color.

As shown in the block diagram, Fig. 29, the combined chroma and pilot signal from the film is first separated into its two components by frequency filters. The 0.9-MHz pilot carrier is doubled to 1.8 MHz and applied to multiplier A together with a locally generated 3.58-MHz signal. The 5.38-MHz sum-signal output of multiplier A is selected by a bandpass filter and applied to multiplier B together with the EVR chroma signal centered around the color carrier at 1.8 MHz. The difference frequency of 3.58 MHz from multiplier B is extracted by a bandpass filter and becomes the NTSC chroma signal. An analysis will show that regardless of a shift in EVR chroma frequency, the frequency of the translator output signal remains constant at 3.58 MHz.

Blanking, synchronizing, and burst signals are added to generate the composite NTSC signal. The color burst is obtained by gating the 3.58-MHz locally generated

FIGURE 29. EVR-to-NTSC chroma translator used in player.

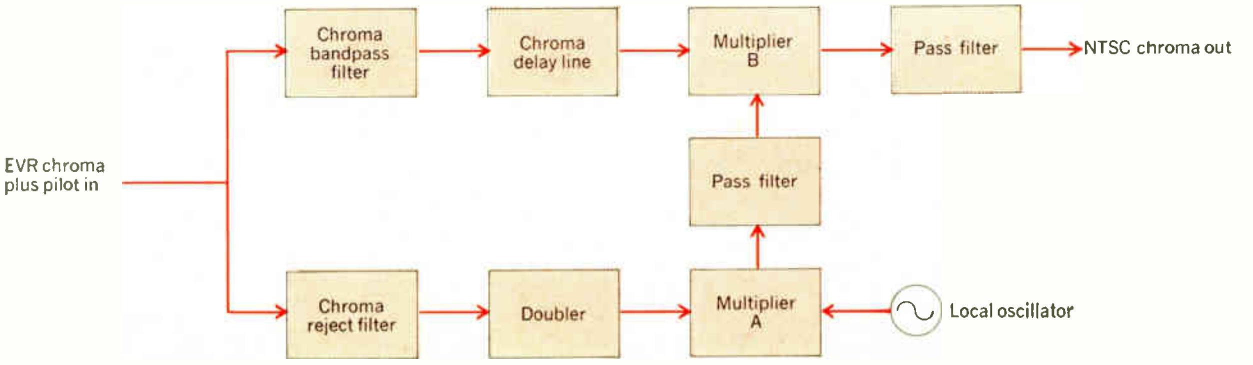


FIGURE 30. EVR cartridge components.

signal with a burst-flag pulse. Prior to this, the 3.58-MHz frequency is set to the correct phase by the player's hue control.

A color killer is employed in the player to disable the color circuit whenever a monochrome cartridge is played so as to avoid spurious beats in the picture. During "monochrome filming" the absence of the pilot signal is sensed and both the chroma and color burst are removed from the outgoing video signal. This procedure activates the color killer in the color television receiver—important when color and monochrome images are intermixed on both tracks for certain educational applications.

Audio

Magnetic recording has been chosen for EVR in order to provide good-quality sound from the player at low cost. The edges of the film carry a magnetic track on each side so that, for monochrome cartridges, each picture track utilizes one magnetic stripe. Color cartridges can use the two magnetic tracks for stereophonic sound or for multilingual presentations. The audio-frequency response is essentially flat from 60 Hz to 10 kHz.

Direct audio outputs at 600-ohm impedance are available at the rear of the player. A single-channel audio signal also is applied to a 4.5-MHz FM oscillator to generate the intercarrier sound for the RF link.

The cartridge itself—the most important item as far as the user is concerned—features simple construction, as shown in Fig. 30. The two half shells are held together by a coil spring and a plastic lock washer. When not in use, the cartridge remains automatically closed and the film is totally protected by its plastic leader.

Potentials

Because electronic video recording enables the user to play video programs of his choosing over an ordinary television set at his own convenience, it has enormous potential for the "knowledge industry" and as entertainment in the home, classroom, office, and factory.

EVR should give new scope to television's immense potential in education. As the Carnegie Commission for Educational Television noted, a more versatile playback technology in educational television is one thing needed to return to the classroom the flexibility that the present uses of broadcasting deny it. By providing this technology, EVR can help educational television make the "massive contribution to formal education" that the Carnegie Commission feels not only is possible, but is imperative.

EVR is also a logical key to home study. The most inspiring educators and other leading authorities in their fields will now be able to appear on EVR and give the student the next best thing to face-to-face lectures. And the student will be able to study at his own learning place—in the privacy of his own home.

On the job, EVR can become a major management and training tool because of its low dollar-per-minute cost for creating, disseminating, and displaying training programs. Trainees, whether studying to be salesmen, dental technicians, lathe operators, engineers, or computer programmers, should learn more and learn it faster with EVR. Hospitals will be able to exchange staff films on medical hygiene or new-patient-care techniques; at his leisure, the busy surgeon will be able to watch on his color television set close-ups of new surgical techniques. Scientists and engineers will see and hear about the latest developments in their areas of specialization at home and not just read about them.

Peter C. Goldmark (F) has been with the Columbia Broadcasting System, Inc., since 1936, and has been president and director of research of its CBS Laboratories Division since 1954. Besides conceiving and then directing the development of EVR, he has been actively advancing high-resolution recording and readout of photographic imagery for space reconnaissance applications. Earlier at the laboratory, Dr. Goldmark headed the development of the long-playing record and field-sequential color television. Dr. Goldmark is a Fellow of the Society of Motion Picture and Television Engineers, the Audio Engineering Society, the British Television Society, and the American Physical Society. He is a member of six other societies, including the American Association for the Advancement of Science and the Optical Society of America.

The impressive list of honors bestowed upon him includes two honorary doctorate degrees, the Franklin Institute's Elliot Cresson Medal, and the Achievement Award by the IRE Professional Group on Audio. Dr. Goldmark holds the B.S. and the Ph.D. degrees in physics from the University of Vienna, Austria.

Laid off!

The laid-off engineer is possibly in greater need of advice about what he should do to get a job interview than what he should do once the interview is granted

Arthur R. Pell Harper Associates Agency, Inc.

Not in recent times has so much of the engineering force been on the unemployment rolls. Job seeking is never easy, and for many engineers the experience may be novel. This article covers the steps that the engineer job seeker should follow initially, and cites some of the dos and don'ts to help him sell himself.

Laid off! After eight years with Electro-Space Industries, there was no work for Jim Roberts. What next? He hadn't had to seek a job since college. How would he find one now?

Finding a job—the right job—requires much care and preparation. As a job seeker, you must know just what to do, and how to do it. Luck plays an important part, but you have to be prepared to make the breaks by properly exposing yourself to all possible job opportunities and handling the leads that you get in such a way that you have a good chance of landing the job.

The first step is to prepare a good résumé. The résumé is your ad. It is the come-on piece designed to make the potential employer interested enough in you to call you for an interview. It also will serve later to remind him who you are. If your ad doesn't draw, you'll never be able to show the company how good you are. The situation is the same as having an excellent product that nobody will buy because it is poorly marketed.

A proper start

The most common type of résumé is a simple chronological history of your background. This can be done very well, or so poorly that it spoils any chance of getting an interview.

Jim Roberts prepared a chronological résumé. He started with his schooling, and then listed all his positions—covering every month of his life from graduation to his last job. He worked long and hard at it. It covered six pages of single-space typing, including a full-page description of his duties as a shipping clerk's assistant while he was in college.

Did you ever take the trouble to read a six-page, closely typed ad some hopeful vendor sent to you? Very few personnel men or engineering managers will do it either.

Jim spent time and money to have his tome typed and duplicated. He mailed it to many companies, box-number ads, and employment agencies—with no result. It was not until weeks later—irretrievably lost weeks—that he decided to discuss his résumé with a personnel consultant, who put him on the right track.

At the other extreme, Jim's associate, Bill, knew brevity was important. But, after noting his name and address and personal data, he simply listed his current company's name and his title, assuming this would be enough to get him an interview.

Bill, too, received no replies to his inquiries. In discussing his problem with the counselor at an employment agency in his town, he was advised to give more details about his accomplishments. Just listing a job title doesn't convey a whole picture—especially for an engineer.

The consultant who saw Jim and the agency counselor who spoke to Bill both pointed out that an effective résumé should cover the highlights of the job-applicant's background in enough detail to make it clear that what he has done will make him valuable to an employer.

Start with your most recent (or most responsible) position and, after giving title and company, break down your major functions briefly and concisely. The sample résumé shown in the box is the most effective type in most cases. If you have had two jobs of equal importance, break both down as shown for the first job in the sample. Naturally, each person's background is different, so you must construct your résumé to show your background to best advantage. It might be possible in some instances to tailor your résumé to fit the requirements of a very special job. Don't prevaricate, but emphasize phases of your background most pertinent to the job requirements.

In most cases, however, it is necessary to have a general résumé. You should have a good supply duplicated so that one is always available when you need it. Multilith®, offset—not ditto or blurred carbons—and a clean Xerox®-type copy are the neatest and most professional-looking methods of duplication.

Don't use a photo

Do not include a photograph with your résumé. No matter how good a photo is—and most photos enclosed with résumés are the passport type and far from good—it shows you as you were for one fraction of a second of your life. It is only a two-dimensional sketch of you that may give the viewer a completely wrong impression. Either consciously or subconsciously, he may make a firm and final judgment about you from his impression of the photo: "This man is too serious; too flippant; cruel looking; smug; stupid; shifty-eyed; insincere." These are some of the comments made by prospective employers who have seen photos of job applicants. Often they refuse to grant an interview just because they don't like the picture.

A human being is a four-dimensional complex, not a two-dimensional photo. In addition to the two dimensions shown in a picture, he manifests a third dimension in the constant changes of his expression and actions. His fourth dimension is his personality. A photo can never

James Smith
3 Hemlock Drive
Silver Spring, Md. 20904
301-223-5944

Offering a broad background in the design of communications systems, utilizing digital techniques and various types of modulation methods.

BOOK COMMUNICATIONS CORP., Washington, D.C. (1967-date)
Project Engineer

DESIGN & DEVELOPMENT: Responsible for design, development, and project engineering in communications technology. Performed studies for application of digital techniques for the telephone networks, an analysis of a digital multiple-access discrete-address system for use in satellite communications, the effect of intentional interference on a binary communication receiver, and atmospheric effects on line-of-sight wave propagation. Also analyzed the effectiveness of modulation and error-encoding techniques against atmospheric-type noise. Other contributions were in the design of digital communications systems using different types of modulation techniques. Also developed the methodology and flow charts for a computer program to evaluate system-transmission parameters.

ENGINEERING ADMINISTRATION: Coordinated engineering activities between engineering department and the military in specifying a shipboard satellite terminal. Wrote proposals and participated in the design and development of secure communications systems. Supervised 8 engineers. Planned and scheduled assignments. Responsible for final screening and making hiring decision for staff. Maintained all engineering records for reimbursement in CPFF contract.

TELE, INC., Falls Church, Va. (1964-1967)
Senior Engineer
Designed and developed a VHF-UHF frequency-hopping system using thin-film techniques. Performed theoretical investigations in the field of ultrasonic propagation.

SPECIFIC PRECISION, New York (1958-1964)
Junior Engineer
Participated in design, development, and flight testing of airborne Doppler radar.

EDUCATION: MSEE - George Washington University, 1969
BSEE - Columbia University, 1958
Information Theory & Coding, Univ. of Maryland, College Park, Md., 1969
RCA Digital Communication Course, Wash., D.C., 1968

PERSONAL: Age: 33, two children, SECRET clearance, fluent in German and French

MEMBERSHIPS: IEEE
Communication Technology Group - IEEE

show this. Do not let a potential employer judge you without seeing you. If your background fits his job as shown in your résumé, he will ask you to come in for an interview in order to see what you are like.

Should you indicate salary requirements? Generally it is wise to omit salary in the résumé; but include it in your letter of transmittal. This will give you flexibility since it is quite possible that your salary requirements will vary: Most people will work for less money if the opportunity for growth exists or the position is in a part of the country where the cost of living is lower.

The next step

Once you have your résumé, the next step is to get it into the hands of potential employers. The best sources:

1. Personal contacts. Send your résumé with a brief note of transmittal to companies you know and that know you, advising them of your availability. Write to trade-association officers. They usually know of openings.

2. Related industries. Engineers have a great deal of know-how valuable to companies engaged in similar projects. Research these companies and send each of them your résumé with a letter suggesting that your special knowledge of their needs will be of value to them. Such a letter might read:

> Dear Mr. Wilson:
> Has your company the need for a well-rounded design engineer with specific experience in ——— ?
> The enclosed résumé will give you a summary of my background and experience. which I believe would make me a valuable asset to your purchasing staff.
> May I have the opportunity of discussing this with you?

3. Employment agencies. These professionals are geared to help you find a job. They not only have definite job openings listed, but know many firms that may be interested in your background. Select the agencies in your area that are known to handle engineers. Also, if you are willing to relocate, mail your résumé to a few leading agencies in major cities where you might be interested in working. Many agencies are national in scope and can assist you in locating jobs in other areas. If jobs in your field or specialty are scarce in your locality, they can refer your material to affiliates in areas where there is a demand for your services.

Working with an employment agency has many advantages. They can advise you about the effectiveness of your résumé, your presentation when participating in an interview, and the details of any job to which they refer you. There usually is no charge for this service, their fee being absorbed by the employer.

4. Ads. Watch the ads in your local papers and the major national-coverage papers (*New York Times, Wall Street Journal, Chicago Tribune*). And don't overlook the ads in trade and professional journals.

5. Personnel department of present company. Some companies will make every effort to relocate a dismissed employee.

The personal approach

From the five sources cited previously will come leads that will produce interviews. Now your program requires a study of selling techniques. Remember the approaches of the best salesmen with whom you have dealt. How did they sell their product? Remember now, you are selling a product—and the product is you.

- Check your appearance. You cannot help it if you are tall or short; fat or thin; blond, brunet, or bald; but you can make sure that you are well groomed, are shaved, have shined shoes and clean fingernails. Dress neatly and conservatively. Wearing a flashy tie or a sports ensemble at an interview has cost many a man a good job.
- Be prepared to answer questions about yourself and your background, work experience, and education. Some psychologically oriented personnel men ask questions about your home life or ask you to evaluate your own strengths and weaknesses. Prepare yourself for the interview by asking yourself all the questions you would want answered if you were doing the hiring.
- There are certain questions that almost certainly will be asked. The interviewer will usually want to know just what kind of work you have been doing. You should be prepared to give him a concise, yet complete, account of this. He will want to know why you left your previous job or why you want to leave your present one. He will ask you about your past earnings. He will inquire about your education, military service, and other activities.
- Answer each question frankly and in as few words as possible. Do not digress and become involved in unrelated details. The interviewer is a busy man. He wants precise answers.

Just as a salesman must size up his prospect, so must an applicant analyze his interviewer. The key to this can be summed up in one word: *Listen!* Do not become so involved with yourself that you do not listen to the questions or remarks made by the interviewer. You can learn about the other man and his company this way.

If he is the type who likes to talk, listen to him and do as little talking as possible. Many men have obtained their job by remaining almost completely silent. These men knew the trick of listening carefully to a verbose person and just interjecting remarks at pertinent places in the conversation. When the interview is over, the interviewer, who has done all the talking, is likely to have sold himself on you because you listened to his many words. This may be poor interviewing, but if you can size up your man as this type, it might get you the job.

If the interviewer is the silent type, he will expect you to carry the conversation. In this case it gives you the chance to steer the interview to stress your strongest points and rapidly pass over your weak points. The danger, again, is overselling yourself by too much talk.

Most interviewers will be somewhere between these extremes. You will have to be alert to the interviewer's reactions and analyze his intent and his methods.

If the interviewer has not told you anything about the job and the company, you should inquire. It is as important for you to know about the job as it is for the interviewer to know about you. Few men will resent questions about their company; most will be happy to talk to you about it and respect you because you asked.

Show enthusiasm during the interview. Smile. Be animated in your conversation. You want this job. Do not be afraid that you will appear overanxious. It is best to show that you are interested in the job and company.

Closing the interview is the prerogative of the interviewer. He usually will end it when he feels he has enough information about you. When he does, accept immediately and do not try to extend the interview. Many a job has been lost because the applicant talked too much and would not let the interview come to a close at the right time.

Don't stop here

Following up an interview is almost as important as the résumé and the interview itself. Surprisingly, very few applicants do this. The interviewer sees a number of applicants for each job. It is important for you to do something to make him remember you. It may make the difference between you and a competitor. One good method is to write a short note thanking him for the courtesies extended during the interview. This should be written as soon as possible after your meeting. It should be brief. The following is typical:

> Dear Mr. Wilson:
> Thank you for the courtesies extended to me at our meeting yesterday afternoon.
> I am certain that my background and experience in designing electronic components will qualify me for the position in your company.
> Sincerely,

Looking for a job is never a pleasant experience. Most people are tense and nervous during this period. As time goes on and one does not "connect," there is a tendency to lose confidence. It might help to remember that the usual time lapse between jobs for engineers in the electronics market is between two and four months. Do not get panicky. It reflects in your job interviews and makes you less desirable to a potential employer.

Using the right techniques when writing your résumé, aligning your sources, and handling the job interview will help you get the job.

Arthur Pell has been a personnel executive for 20 years; during 17 he has been vice president of Harper Associates (personnel agency), managing its headquarters in New York and its 26 branch offices. He is also a professor at New York University, where he originated and now coordinates a course for training employment-agency personnel. Professor Pell received the A.B. and the A.M. degrees from New York University and has a professional diploma in personnel psychology from Cornell University. He is a member of several personnel associations, and has been appointed consultant on education for the National Employment Assn. He has written seven books.

Optical transmission utilizing injection light sources

Injection light sources give optical transmission systems such versatility that the promised potential of these systems might be realized sooner than anyone expected

K. L. Konnerth IBM Thomas J. Watson Research Center

B. R. Shah IBM Federal Systems Division

Although a large amount of work in optical transmission has been accomplished through the use of lasers, modern developments have enabled noncoherent light sources also to be used in this area with great success. In particular, injection light sources—in the form of the coherent injection laser and the noncoherent light-emitting diode (LED)—offer designers of optical transmission systems characteristics that are in many ways superior to those of other methods. The present article deals essentially with how injection light sources may be employed to transmit or "communicate" information optically. Such optical techniques can be used not only in communication systems, in the classic RF sense, but in computer, video, telemetry, and detection systems as well.

Optical transmission systems, in various forms, have been in existence for several decades. The first demonstration of a laser in 1960, however, inspired a renewed interest in this area that has never waned. Most work has been directed toward the two highly publicized properties of such laser systems: the potentially very-high bandwidth, and the use of heterodyne detection techniques made possible by the coherence of laser radiation.

For the most part, research on optical transmission (optical "communication" is also an appropriate terminology, in the sense of the transfer of information) has concentrated on the use of gas and solid-state lasers because of their stability and very small beam angles. However, there are numerous applications in which the special properties of the injection laser make it extremely more useful. Moreover, at the present state of development, the spontaneously emitting semiconductor injection light source—the light-emitting diode (LED)—is far more suitable for many types of optical transmission than any existing laser. Such interconnections, although limited in range and bandwidth, are numerous and certainly within the present technology.

This article discusses the relevant properties of these two injection light sources and explores their application in optical interconnections for various types of electronic systems.*

Optical transmission—potential advantages

The reasons for using optical rather than other methods of transmission are varied, and depend upon the specific circumstances of application. In those instances where optics have the advantage over cable systems, it is usually assumed that transmission would also be possible at radio frequencies, However, cables generally can be replaced much more readily with an optical link (over the distance for which they are suitable) than with an RF link because of restrictions on the use of the electromagnetic spectrum, size requirements, environment, and others.

The first and perhaps most important advantage of optical transmission systems lies in the fact that they possess a much broader bandwidth than radio transmission systems because of the inherently higher carrier frequency. Proper exploitation cannot be made of this property at the present time, however, because of the lack of adequate external modulators at extremely high frequencies and the poor response time of the light sources when directly modulated.

Another advantage, although obtainable only with laser sources, is that of coherence, which permits the use of heterodyne detection principles that have been applied to radio frequencies with great success for many years. This approach was originally expected to yield considerable improvement for optical transmission over simple photon-counting schemes (i.e., direct detection of the incoming light with some type of photodiode), since in a heterodyne system the background and detector noise are not amplified with the signal. But further comparison[1] has shown that for some wavelength regions photon counting gives superior results. Even in the spectral region where heterodyne detection could be better, the potential advantages are usually outweighed by atmospheric path fluctuations[2] on earth or by such things as Doppler shift in space communications.[3] Heterodyne detection is of value primarily where very stringent requirements can be maintained, as in enclosed "pipeline" systems.[4]

Thus the main application of heterodyning is in a controlled environment where high background discrimination is desired but it is not worth the extra complexity imposed upon most optical transmission systems.[5] For numerous applications, the coherence advantage is not of great value and, consequently, we can consider using the incoherent LED as well as the injection laser.

A third advantage of optical transmission is that complete electrical ground isolation is achievable between transmitter and receiver. Hence, in many applications, cables can be replaced by optical links to eliminate ground loops, to eliminate the common "ground shift" problem in data circuits, for high-voltage isolation, etc.

The fact that an optical system offers transmission

* A special issue on optical communication is planned for the October 1970 *Proceedings of the IEEE.*

security that is frequently higher than that of a radio system (because of the difficulty of finding and intercepting the much narrower signal beams) is an obvious advantage. In comparison with cable transmission, there is less stray radiation that can easily be detected from the optical system, particularly since, in general, the system's location is not as obvious as that of a cable. This is not meant to imply, however, that an optical link is completely secure against interception. Some of the signal unavoidably will be scattered in unwanted directions by water droplets, dust particles, or the Rayleigh scattering from air molecules.[6]

Other advantages may be stated as follows:

An optical system does not generate and is not susceptible to outside electromagnetic interference. The receiving and transmitting circuitry can be completely shielded, except for small openings to let light in or out, and this can be done through a small tube that acts as a waveguide beyond cutoff to possible interfering signals.

Transmission in an optical system is unilateral; thus receiver and transmitter can be designed independently.

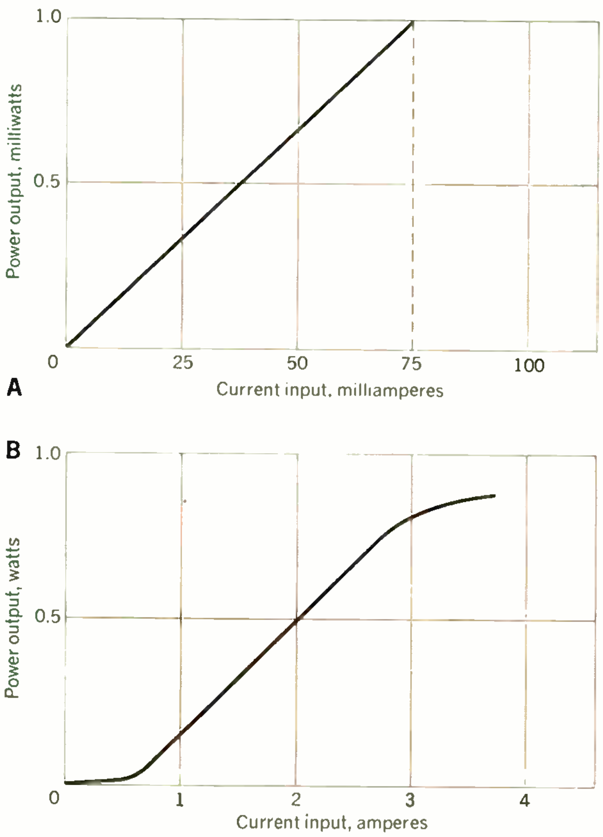
Signals are easily picked off from an optical beam, if its location is known, without the concern for impedance matching that is absolutely necessary in cable systems.

A given optical link can be made to operate over a fairly wide range of distances in contrast to the fixed length of a cable.

Optical interconnections can generally be established comparatively rapidly across a river, highway, runway, etc., whereas a cable system would require extensive tunneling or bridging.

Generally, optical transmission equipment can be made smaller than comparable radio transmission equipment. In particular, antennas (viz., lenses and reflectors) can be considerably smaller than that required for RF communications because of the very short wavelengths involved (of the order of a micrometer). In comparison with cable systems, fiber optics for light transmission are frequently much smaller than cables producing the same degradation of pulse shape.

FIGURE 1. Typical curves of power output versus drive current for injection light sources. A—GaAs light-emitting diode operated CW at room temperature. B—GaAs injection laser operated CW at 77°K.



Present limitations of optical transmission

Of the various factors limiting the use of optical transmission systems, the following are predominant:

Optical transmission is limited to "line of sight." Any opaque object intersecting the light beam will interrupt optical transmission, whereas RF signals will be diffracted around it to a rather large extent (dependent upon frequency). This problem, however, may be circumvented in many cases through the use of light-guiding structures[7] employing mirrors, periodic lenses,[8] gas lenses,[9] fiber optics,[10] and single-mode guides.[11]

Atmospheric characteristics can impose grave limitations on optical transmission, with signal strength and directionality being severely affected by the absorptive or scattering effects of rain, smoke, turbulence, etc. Whenever these problems become too severe, they can be overcome by enclosure of the light beam in a controlled environment or pipeline[12] or by the use of one of the guiding structures just discussed.

Optical transmission can be very susceptible to vibration problems. In unguided systems, because of the relatively small beam angles involved, vibration of the transmitter can cause variations in the received signal strength due to the amplified motion of the receiving end of the optical beam. Receiver motion can cause similar problems, but in many cases they are not as restricting since other practical considerations frequently result in receiver acceptance angles that are larger than transmitter beam angles.

The present state of the art in devices imposes some additional limitations upon optical transmission. At the present time, the maximum rate of data transmission is limited by transient characteristics of the radiation source or external modulator as well as detector response time. In addition, the present range of transmission is limited by the low efficiency and power-handling capabilities of the emitters, the poor sensitivity of the detectors, and losses that occur in the coupling medium.

Injection light sources

Injection light sources are small semiconductor p-n junction diodes fabricated of materials that produce radiation through carrier recombination processes when a forward-biased current is passed through them. The wavelength of this radiation is controlled by the energy gap through which the recombination occurs and is thus dependent upon device material and dopant (impurity added to produce both n and p regions).

Although injection light sources made of many different materials are available, this article will be limited to those fabricated from gallium arsenide (GaAs). Such devices are generally the most practical for optical transmission purposes because their emission wavelength

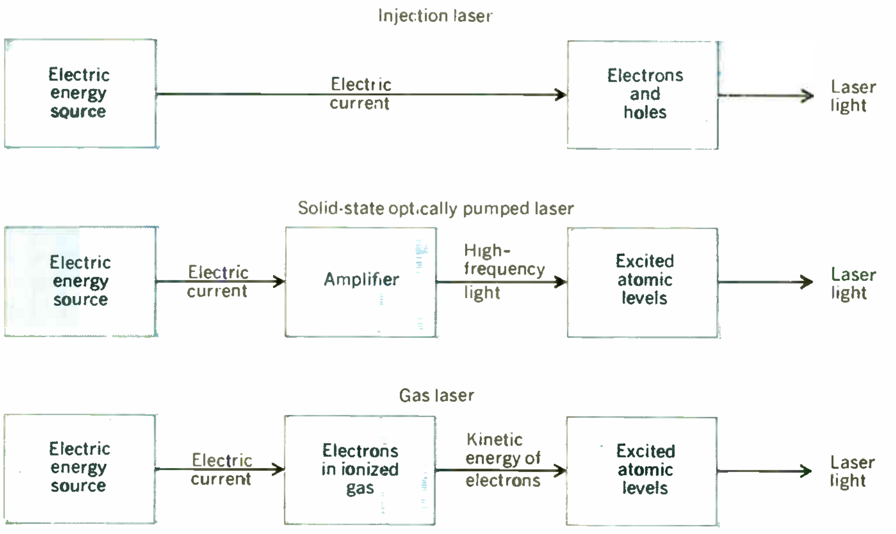


FIGURE 2. Comparison of the direct transfer of energy in the injection laser with the less-efficient intermediary forms of energy transfer employed in optically pumped and gas lasers.

matches the sensitivity peak of a silicon detector (≈ 9000 Å) and their speed, efficiency, and life characteristics are as good as, or even better than, devices fabricated from other materials.

At low current densities, GaAs injection light sources produce radiation through spontaneous recombination, which increases approximately proportional to the drive current. Device characteristics at higher current densities are dependent upon detailed material properties and physical structure. These properties are adjusted to maximize the spontaneous emission in a light-emitting diode. This device maintains an essentially linear light–current relationship [Fig. 1(A)] until heating effects are encountered. A laser, on the other hand, may not have very efficient spontaneous emission, but will produce stimulated emission that increases rapidly with current [Fig. 1(B)] once a threshold level is passed (again until heating effects become important). Both lasers and LEDs have typical semiconductor diode current–voltage curves.

The advantages of using an injection laser over other types of lasers in optical transmission are many. For one, the electric energy that is used to pump directly across the p-n junction is employed more efficiently than in other lasers, since there is no need for an intermediate stage in the transformation of electric energy to light (see Fig. 2).[13] This elimination of auxiliary apparatus results in another desirable characteristic—that of compactness. Another advantage over other lasers occurs in communication, where direct modification of the current source precludes the use of external modulation devices.

Unfortunately, the region of the injection semiconductor in which the lasing action takes place is extremely thin (a few micrometers thick). This narrow aperture of the effective lasing region causes the angular divergence of the emitted light beam to be wider than in other lasers. Another disadvantage of the injection laser lies in its inability to dissipate the heat created by the high currents that are required to produce lasing action. This latter problem, however, has been somewhat alleviated by the techniques of pulsed (rather than continuous-wave) operation, application of special heat sinks,[14] and operation in a cryogenic environment.

The radiation source properties that are most pertinent to optical transmission applications are summarized for injection light sources in Tables I, II, and III, with specific emphasis on GaAs, which will be discussed in somewhat more detail in a later section. The significant parameters of injection light sources will now be considered.

Radiation wavelength. The only essential difference in spectral output between GaAs lasers and LEDs is in line width. For continuous-wave (CW) lasers operated under carefully controlled conditions such that only one mode exists, the line width can represent less than 50 MHz.[15] For practical use, however, many modes are usually present.[16] Moreover, when a laser is pulsed, the wavelength changes radically with time as the laser begins to heat.[17] On the other hand, LED line width is of the order of several hundred angstroms.

Efficiency and power output. Light-emitting-diode efficiency is very dependent upon device geometry and the presence of any transparent coating, since the high index of refraction of the semiconductor causes much of the generated light to be internally reflected. The LED efficiencies quoted in Table I are for uncoated rectangular units and would be several times that value for coated or appropriately shaped devices.[18]

The maximum power output from an LED is limited by the power-dissipation capability of the structure and the mount used. For example, a typical dissipation limit of a transistor header not mounted on a heat sink is of the order of 100 mW. Therefore, a zinc-doped GaAs LED on such a header can produce approximately 1 mW of output radiation.

Continuous-wave outputs of the order of 1 watt can readily be achieved from GaAs injection lasers operated at liquid-nitrogen temperatures (77°K). At room temperature, however, the high series resistance and threshold current of the lasers produce severe heating, which limits pulse duration to about 1 μs and necessitates low-duty-cycle (i.e., pulsed-current) operation.* Therefore, the

*Continuous-wave operation at room temperature was announced by Bell Telephone Laboratories during the publication of this article. Full details may be obtained from an article published by I. Hayashi *et al.* in the August 1970 *Appl. Phys. Letters* (vol. 17, p. 109), entitled "Junction Lasers Which Operate Continuously at Room Temperatures."

I. Typical approximate efficiencies of various injection light sources

		Approximate Efficiency, %
Injection lasers (GaAs)		
77°K	CW	40
300°K	Pulsed	25
Light-emitting diodes (rectangular, unpotted)		
GaAs		
77°K	Si doped	10
	Others	10
300°K	Si doped	3
	Others	1
GaAsP		
300°K	6500 Å	1
GaP		
300°K	Red	2
	Green	0.05
GaAlAs		
300°K	7300 Å	2
	6450 Å	0.1
SiC		
300°K	5800 Å	0.001

II. Emission wavelength of various injection light sources

	Degrees Kelvin	Wavelength, Å
GaAs	77	8400
	300	9000
GaAsP	300	6300–9000
SiC	300	4500–5800
GaP	300	7100
		5600
GaAlAs	300	6450–9000

Peak sensitivity of silicon detector = 9000 Å
Peak eye sensitivity = 5550 Å

present characteristics of injection lasers impose rather severe constraints on their use (viz., either cryogenic operation or very high drive currents and low data rates).

Speed. The turn-on characteristics of an injection light source can be considered to consist of two parts: first, a delay between the application of a current pulse and the start of spontaneous or stimulated emission (for an LED and laser, respectively); second, the rise time of the emission once it begins. Typical values of these parameters are presented in Table III.

For LEDs, the delay is merely the time required to charge the junction capacitance and is negligible under most operating conditions. On the other hand, carrier-lifetime-dependent rise time can be anywhere from about 2 ns to 0.2 μs, depending upon fabrication and operating parameters. The diodes most useful for optical interconnections (on the basis of a general consideration of all pertinent parameters) have rise times of approximately 15 ns.

For injection lasers, a current-dependent delay exists that is caused by the time necessary to inject sufficient carriers to reach a threshold level. For 77°K operation of most lasers[19] and room-temperature operation of those fabricated in certain ways,[20] this delay is of the order of of a few nanoseconds or less and can be minimized, in many applications, by biasing just below threshold. In an equipment-limited observation, the low-temperature rise time has been measured at 0.1 ns[19] and it is probably considerably less as indicated by experiments in which a laser was modulated at 11 GHz.[21] Room-temperature rise times are generally of the order of 1 ns.

Life. Very good aging characteristics can be obtained for injection light sources by using specific fabrication processes. The only possible exception is in the room-temperature operation of injection lasers, where there is considerable variability in this characteristic although qualitative relationships have been established with other parameters.[22] Lasers have been operated CW at 77°K with about 1/3 watt of output power for over 1000 hours with negligible change in characteristics.[23]

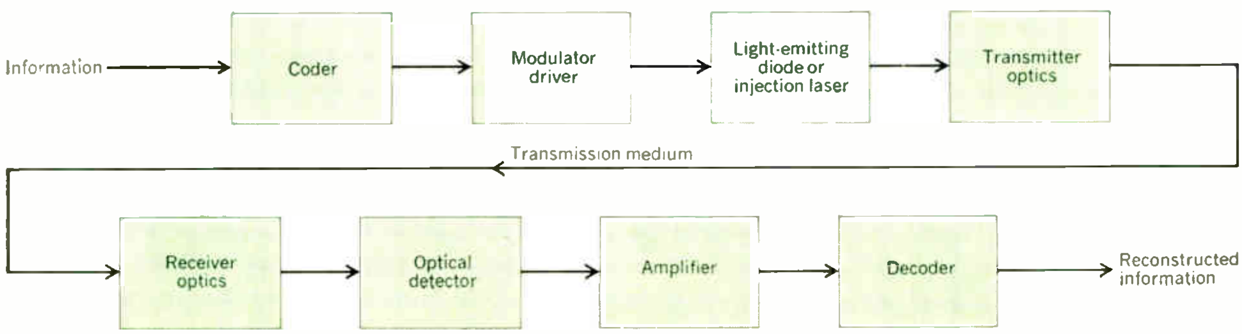
The life of LEDs is dependent upon the parameters of fabrication, but results are reproducible and diodes have been produced that degraded about 5 or 10 percent during the first few hundred hours and then changed little, if any, during the following 23 000 hours they were on test. Recent devices look even more promising.[24]

Considering these various parameters, GaAs injection light sources are seen to be an excellent choice for the type of optical transmission that is under discussion.

System considerations

General. A block diagram of an optical transmission system utilizing an injection light source is shown in Fig. 3. An information signal is coded to the modulation format (pulse-code modulation, pulse-position modulation, etc.) and applied to a driver that provides the proper excitation for the injection light source. The modulated light beam is then collimated with the help of beam-tailoring devices (lenses and/or mirrors, etc.). Signal energy col-

FIGURE 3. Block diagram of a generalized optical transmission system utilizing an injection light source.



III. Speed characteristics of various injection light sources

Injection lasers (GaAs)		
	77°K	Delay—Dependent upon current; less than 1 ns for I greater than $2 \cdot I_{threshold}$
		Rise time: less than 0.1 ns
	300°K	Delay—Dependent upon current; fabrication may be as long as 500 ns but best devices are comparable to 77°K
		Rise time: several ns
Light-emitting diodes (GaAs)*		
	77°K	Si doped—Rise time: about 100 ns
		Others—Rise time: less than 2 ns
	300°K	Si doped—Rise time: 0.2 to 1 μs
		Good Zn diffused—Rise time: about 10 ns

*LED delays are only significant at very-low current densities where the junction capacitance must be charged.

lected by the receiving optics is focused onto a photodetector and its output is then amplified and restored to a suitable form. Frequency conversion is performed at the receiver for subcarrier modulation systems.

The performance of the system is determined by the signal-to-noise ratio (SNR) at the receiver. For a given photodetector and operating range, the SNR (or error probability in case of digital systems) depends on the receiver aperture, receiver field of view, receiver temperature, transmission-medium characteristics, and transmitted power.

Detection. As previously described, the use of injection light sources limits us to direct detection systems (DDS). The detectors for direct detection are basically photon counters and include photomultipliers, photodiodes, photoconductors, phototransistors, and others. Considerations of spectral match, gain, and speed of response generally limit one to photodiodes and photomultipliers. Photomultipliers are most useful in cases where noise considerations necessitate large postdetection gains (power gains, for example, over 100 dB). Ordinary photomultipliers have a flat frequency response up to about 100 MHz. Beyond this point, time delay and spread in transit time contribute to sharp frequency drop-off. The disadvantages of photomultipliers include their large size, need for a high-voltage power supply, and poor discrimination against high-level background radiation because of saturation effects.

Recent improvements in sensitivity and frequency response of silicon photodiodes make them the most useful photodetectors in many receiving applications. They have a peak response at about 8500–9000 Å, thus serving as sensitive solid-state detectors for GaAs injection sources. Their small size, ease of fabrication (discrete as well as arrays), low operating voltages, and reliability make them ideal for use in optical transmission.

Transmission medium. The transmission medium for optical systems can be open (free-space propagation) or closed (optical fibers, or evacuated or dry-gas-filled pipes). Unfortunately, the open links on the earth are subject to interruptions due to the atmospheric effects of molecular and aerosol scattering. In addition, there are many other important effects (e.g., beam bending, "image dancing," beam spreading, and scintillation) that are caused by localized variations in the atmospheric index of refraction and density, primarily due to temperature inhomogeneity and turbulence. These effects are intimately related and contribute toward degrading system performance.[25-27]

The problem of losing too much received signal strength because of beam bending and spreading can be overcome satisfactorily through proper design choice of transmitter beam width, transmitter and receiver apertures, and receiver acceptance angle. The worst case of beam bending, for example, is expected to be less than 5×10^{-4} radian[27] and thus a transmitter beam width of 10^{-3} radian or more should handle this problem adequately.

The primary result of scattering also has the effect of reducing received signal strength. The variability of this effect, however, makes it much more difficult to overcome. It is, therefore, a main factor in determining error rates, hence system availability. System availability of greater than 98 percent has been observed for ranges of up to 3 km.[28] Thus, optical links are practical for many short-haul applications.

In general, fluctuations in the atmospheric density and index of refraction that result from turbulence tend to induce multiplicative noise on the signal component. These scintillation noises vary in direct proportion to the optical signal strength and therefore cannot be overcome by increasing signal strength, as is possible with background and receiver noises. The effect of any slowly varying multiplicative component (or modulation), however, can usually be removed through the use of an automatic gain control (AGC) circuit in the receiver to vary the receiver gain inversely to this interference signal.

As previously discussed, the atmospheric limitations can be overcome by the use of a closed transmission medium such as incoherent and coherent optical fibers or gas-filled pipes. A variety of configurations is possible with evacuated or sealed pipes and considerable research is being carried out on this type of device.[29] However, the use of sealed pipes for the relatively low data rates implied by the use of injection light sources (less than 10^9 bits/s) would probably be very expensive.

The flexibility, relatively small bend radius, and ease of handling of optical fibers make them attractive candidates for short (up to a few hundred meters) and tortuous paths. Incoherent fibers 0.5 to 1.25 cm in diameter and consisting of bundles of discrete fibers 5 to 50 μm in size are readily available and are already cost-competitive with coaxial cables. Their rather severe attenuation (3 dB per 3 meters) is acceptable over short distances. Coherent fibers are more expensive, but offer attractive possibilities for array transmission.

Transmitter and receiver optics. Injection lasers have a narrow cone of emission (5×20 degrees) and collection and collimation of radiation is relatively straightforward. Both flat and domed LEDs, however, have a 180-degree cone of radiation. This constitutes a problem since the light-gathering efficiency of lenses and mirrors is limited. It can be shown that, for a given operating range, LED, and receiver diameter, the received power of such a system varies as the square of the transmitter aperture (for small f/number optics, this relationship is only partially true; however, it holds to within 20 percent). As discussed in the previous section, transmitter beam angles of the order of 10^{-3} radian are usually appropriate for open links through the atmosphere.

Receiver design is concerned primarily with limiting the field of view to reject excessive background noise. The field of view must be large enough to cover the transmitter aperture effectively during the maximum expected displacement of the beam due to vibrations or atmospheric effects. In most situations, a field of view of 10^{-3} radian (the same as that previously cited for the transmitter beam angle) is completely satisfactory and readily achievable. The use of an optical filter at the field stop reduces incident background noise and thus improves the received signal-to-noise ratio.

Types of multiplexing

Three types of multiplexing can be considered for an optical transmission system:

Frequency multiplexing. Two types of operation can be considered under the heading of frequency multiplexing. The first would be the modulation of various frequency subcarriers, which in turn would be combined together to intensity-modulate the injection light source. The second would involve the use of various sources with different frequencies (wavelengths) of emission that would be transmitted along the same optical path.

In the first case, the subcarriers can be amplitude- or frequency-modulated, with the choice of subcarrier frequency and modulation technique dictated by the bandwidth of the light source and the data to be transmitted. A disadvantage of frequency multiplexing is the requirement for a controlled oscillator for each channel and numerous expensive filters to minimize crosstalk effects. The FSK video link described later, however, is an example of an application where frequency multiplexing offers a number of advantages because both binary data and video signals are being transmitted simultaneously.

Since injection light sources covering a broad spectral range (from 4500 to 9000 Å and beyond) are now available, the impetus for using the second form of frequency multiplexing—color multiplexing—is strong. In this case, the outputs of various LEDs emitting at appropriate wavelengths can be optically combined to form a single beam and transmitted over the same path. At the present state of the art, however, this technique does not lend itself to many practical uses because of the limited power outputs of injection sources in many spectral ranges, as well as the high cost of optics that must be corrected for aberrations over a wide band (rather than the relatively low-quality optics needed for a quasimonochromatic source).

Time multiplexing. Time multiplexing is particularly attractive for transmitting binary signals, since the signal is already in the proper form and the cost of logic circuits (required for time multiplexing) is decreasing rapidly. The state of the art in GaAs LEDs allows a maximum data-handling capability of 50 megabits per second, sufficient to time-multiplex literally hundreds of telemetering signals. Injection lasers operated at room temperature are limited to data rates of a few hundred kilobits per second because of present speed characteristics. However, because of their much higher peak power, injection lasers allow transmission over longer distances.

Time multiplexing implies a basic delay of one multiplex period. This is a serious disadvantage for some applications. For example, in a processor-to-storage transfer system, it adds two multiplex periods to the memory access time. For most other applications, however, this additional delay is of no basic consequence and the simplicity makes time multiplexing a very attractive possibility.

Space multiplexing (array transmission). Parallel optical transmission is essentially a one-for-one replacement of cables with light beams. Ordinarily, an array of LEDs or injection lasers is imaged onto a similar detector array. Three coupling techniques are possible with this system: (1) collimating and collecting optics, (2) periodic lens system, and (3) coherent fiber bundles.

1. In the case of collimating and collecting optics (Fig. 4), the off-axis sources cause varying degrees of beam divergence. The overall divergence is dependent upon the size of the LED array and the focal length of the transmitting lens. The amount of divergence that can be tolerated is determined by the allowable variations in signal strength at the detector end.

It is presently possible to make a 10 by 10 array of LEDs on 5/1000-inch (125-μm) centers. For a transmission path of 3 meters, a system can be designed with two 5-cm-diameter $f/2$ lenses. In this case, the intensity of the outermost spots on the detector array would be about 40 per cent of those on the optical axis. This system has two basic problems: (a) The $f/2$ lens collects only 3 per cent of the LED radiation and contributes to poor overall efficiency of the system. One possible improvement would be through the use of injection lasers in place of the LEDs; however, this probably is not practical at the present state of the art. A more practical possibility is the use of more closely spaced LEDs (producing less beam divergence with the same lens or the same divergence with a lens of the same diameter but shorter focal length and thus higher collection efficiency). Another improvement would be the use of LEDs fabricated so as to provide higher directionality. (b) The angular system vibrations would result in misalignment and crosstalk. For the system just considered, an angular movement of 1.25×10^{-3} radian would cause complete misalignment.

2. A periodic lens system[8] (Fig. 5) is a more efficient arrangement than a collimating and collecting optics system. The total number of lenses would be determined by path length and a consideration of the tradeoff between overall efficiency and number of lenses. The obvious disadvantage of this method, for many applications, is the need for optical elements in the space between the transmitter and the receiver. Also, alignment problems turn out to be very critical.

3. The use of coherent fiber bundles appears attractive. Although in a sense a "cable," the fiber bundle allows a much higher data flux density. For example, it is possible to transmit data practically over a distance of 30 meters using a 10 by 10 array of emitters and detectors with a 1-mm fiber bundle; a cable bundle to transmit 10 by 10 bits over a comparable distance would be about 2.5 cm in diameter. The primary problems in this case are cost, registration of the fiber bundle, and coupling from source to fiber and from fiber to detector array.

Applications

Computer interconnections. Most present-day computer systems consist of large numbers of independent modules interconnected with coaxial cables. These modules perform specialized functions such as arithmetic processing, storage, input/output control, switching, etc. Thus, this modular form offers great flexibility in designing a computer system.

However, for a large system, interconnection problems tend to offset the advantages of the modular assemblies. Here, the large number of cables and connectors that are needed occupies an enormous amount of space and panel area, thereby complicating hardware design, installation, and maintenance. The interconnection requirements of parallel systems and those brought about by the application of large-scale integration (LSI) to digital systems force the realization that interconnections are more important in determining performance than all other hardware factors. In fact, the problems of physical size and bulk, dc shift over long cables, reflections, crosstalk, radio-frequency interference, and skin-effect degradation are limiting computer-system interconnection capabilities.

Three classes of computer transmission requirements may be distinguished: (1) high-bit-density, high-speed parallel interconnection networks among processors, memories, and channel controllers (the "inboard" communication problem); (2) low-bit-density, fairly high-speed "longer"-distance communication between I/O devices and channel controllers (the "outboard" communication problem); and (3) low-bit-density communication to remote terminals (the remote communication problem).

Inboard connections. Inboard connections are characterized by short distances (up to 15 meters) and high transfer rates (up to 50 million transfers per second). The data and control signals are transferred in parallel and the major emphasis is on total transmission time, which must be as short as possible. The only suitable optical-transmission technique available, therefore, is space multiplexing. Anders and Callahan[30] have described an experimental link using a 5 by 5 matrix of 1-mm-diameter LEDs with 2-mm center-to-center spacing (only the five center LEDs actually were used in the system). The optics used for the demonstration were $f/1$, 23-cm parabolic mirrors and the transmission range was 25 meters. A 4 by 4 array of LEDs on 0.25-mm centers has also been reported.[31] It is reasonable to assume that newer technologies will allow even better packing density. Such higher packing density, of course, requires better-quality optics, and large arrays require correspondingly larger lenses or mirrors. Use of coherent optical fibers or a periodic lens structure should alleviate some of the problems.

Outboard connections. The demands placed upon outboard connections are for high data rates (up to 20×10^6 transfers per second for on-line bulk storage devices), flexibility, and reconfigurability. Since the distances are somewhat long (up to a hundred meters), array transmission is not very practical. Time multiplexing and, to a lesser extent, frequency multiplexing, on the other hand, are potentially useful. In the case of time multiplexing, the basic delay of one multiplex period does not constitute a problem for most outboard applications. An interesting example of a possible application consists of an experimental full-duplex optical link designed by one of the authors. Called adapter logic for optical transmission (ALOT), this system is an attempt to replace electrical connections from a channel to an I/O control unit with a full duplex optical link (Fig. 6).

System ALOT consists of two identical transceiver units. The ALOT I is connected to the channel (selector) and communicates with ALOT II at the control unit. All outbound lines are handled by the transmitter portion of ALOT I and the receiver portion of ALOT II, with the opposite being true of the inbound lines. This arrangement enables independent communication in each direction. Both serialization and deserialization of the normally parallel data are accomplished through time-multiplexing techniques controlled by two independent clocks (one at each end). The first bit in each byte of transmitted serial data is a signal to initiate clock operation in the receiver. The serialized information being transmitted is gated to the LED driver at a 20-Mb rate. The GaAs LED is located at the focal point of a 2.54-cm-diameter $f/2.5$ doublet lens and is driven by 250-mA pulses. The transmitted power is 45 μW in a beam angle of 3.2×10^{-3} radian; the received signal at 15 meters (with a 2.54-cm-diameter lens) is 5 μW. The minimum detectable power is 0.5 μW, thereby giving a fade margin of 10 dB.

The entire system employs one board of logic at each end in addition to the optical system. The GaAs diode driver is housed in the optical assembly, as is the receiver circuitry. Because of the very low dynamic impedance of the GaAs LED, it is essential to have the driver in close proximity to the diode. Actual optical transmission occurs near the room ceiling level to reduce interference from personnel and equipment.

Remote connections. Short-haul optical links (less than 3 km) offer advantages with respect to electromagnetic spectrum crowding, multipath interference, ease of installation, isolation, etc., that are ideal for low-bit-density

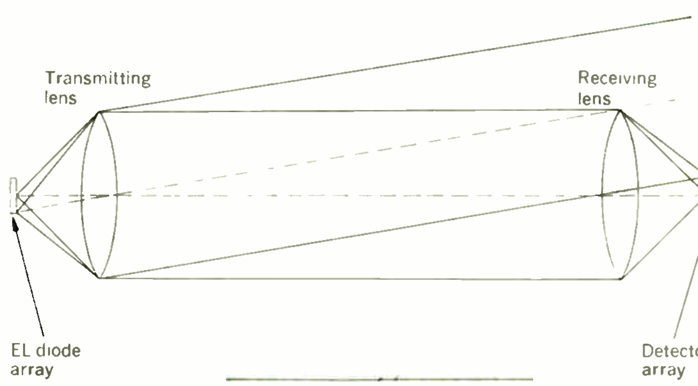
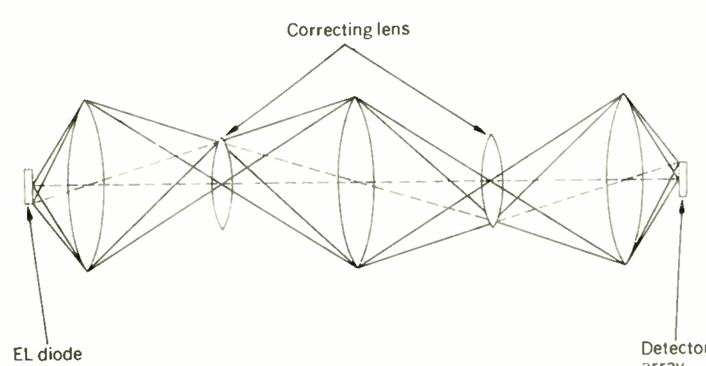


FIGURE 4. Space multiplexing (array transmission) utilizing a transmitting lens to collimate the output of the various emitters and a receiving lens to form an image of the emitted array upon the detector array.

FIGURE 5. Space multiplexing (array transmission) utilizing a periodic lens system. Only two "periods" are shown here, whereas in practice many repetitions of this structure would probably be used.



transmission to remote terminals.[32] The experimental optical link designed by IBM to connect an IBM System/360 Model 50 computer to an IBM 2250 graphic display unit fully illustrates the application's potential (Fig. 7, link *B*). This optical link was the second of two experimental links exhibited at Expo '67 in Montreal, Canada (Fig. 8).

The link was designed to transmit binary data and video information simultaneously over a distance of 700 meters, with the data rate of the system at 1.3 MHz. In this system, the modulation technique employed for data

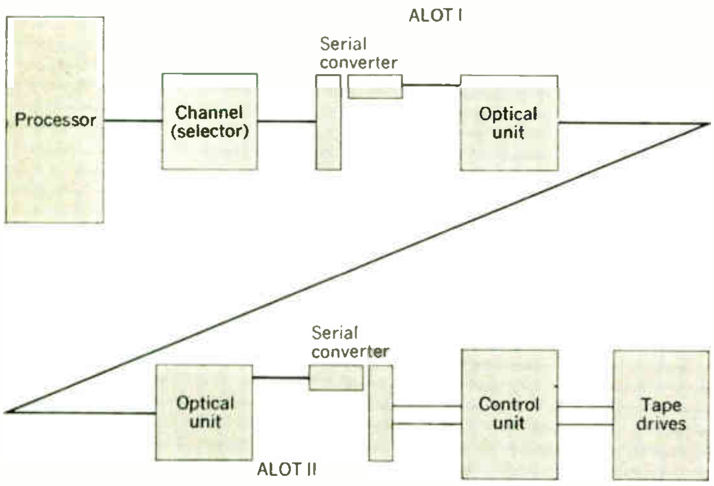
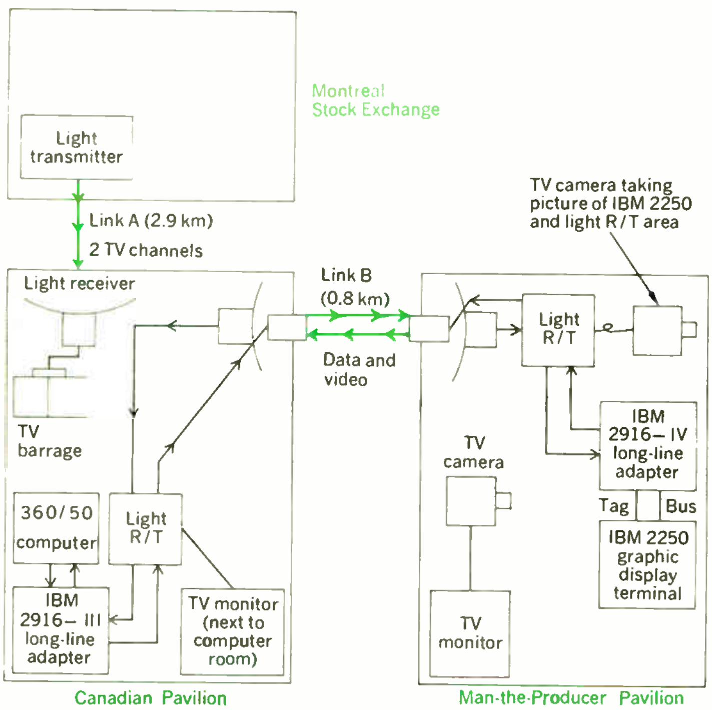


FIGURE 6. Block diagram of a system in which a full duplex optical link—ALOT (adapter logic for optical transmission)—was used.

FIGURE 7. Block diagram of the combination video and data transmission optical links used at Expo '67 in Montreal. GaAs light-emitting diodes serve as the light source for this IBM system. The power transmitted in link A is about 8 mW; the shorter link B about 1.5 mW. Transmission wavelength is 9000 Å (infrared).



transmission is frequency-shift keying (FSK/AM), where a "one" state is represented by a 16-MHz sinusoidal burst, and a "zero" is represented by a 10-MHz sinusoidal burst. The video signal is transmitted as a vestigial sideband modulation on a 5-MHz subcarrier (upper sideband removed to reduce crosstalk); the combined signal is used to intensity-modulate a GaAs LED. An $f/2.5$ achromat lens serves as the transmitter optics as well as the objective of a viewing telescope with the same field of view as the transmitter in order to assure easy alignment. The receiver optics consist of a 30-cm parabola with a 23-cm focal length, and a glass filter (Schott RC10) is used at the field stop to reduce the background radiation. The signal is demodulated by an SGD100 (EG&G) silicon photodiode. For the video signal, the information from 0 to 0.5 MHz is suppressed to avoid low-frequency interference (120-Hz background and atmospheric modulation).

Video applications. The possible video applications for optical transmission include links to remote displays or closed-circuit television systems. With injection light sources, it is possible to transmit baseband or modulated subcarrier video as well as digital television.

GaAs LEDs have been used to transmit television signals over paths ranging in length from 15 meters[33] to over 48 km.[34] Boerschig[35] has described a 12-MHz video-signal transmission system using a GaAs LED and a photomultiplier. Another interesting example of video application is the IBM experimental link that was used at Expo '67 in Montreal to transmit two television channels simultaneously over a distance of 2.9 km (Fig. 7, link *A*). This link uses subcarrier vestigial sideband–intensity modulation (SCVSB–IM). That is, the subcarriers (5 MHz and 16 MHz) are each modulated with one of the television channels using amplitude-modulation/vestigial-sideband techniques. The resulting two signals are then linearly combined to intensity-modulate the GaAs LED output. At the receiver, a dual photodiode (EG&G SGD 444-2-C) is used to separate the two channels; the optical system focuses the incoming radiation on both halves of this diode. Each half is independently connected to a tuned receiver to detect one or the other of the subcarrier signals. These are then amplified, filtered, and detected in separate channels to provide two individual composite video signals. Under normal conditions, the received power is 5 μW per channel. Minimum detectable power for this application is 0.25 μW per channel. This provides a 13-dB margin, which was quite satisfactory for the 6-month period of operation. During constant use, only very heavy rain or rain accompanied by fog was able to put the system out of communication.

Telemetry. Injection light sources are potentially very useful for telemetry, whether for computer- or noncomputer-related applications, primarily because of one important property they possess—isolation. Remote sensors, which are connected to a computer in a process control system, frequently have stringent electrical isolation requirements placed on them. These may arise from various safety considerations, common-mode rejection limitations, crosstalk problems, ground loops, or a combination of these. The near-perfect electrical isolation that can be achieved between the two terminals of an optical link permits independent design of both ends of the system and minimizes the problems that could be caused by such drastic requirements.

A related situation occurs when a sensor must be maintained in some type of extreme environment where it is not practical to couple conductively but an optical path is available. This includes, for example, data transmission into and out of very-high-voltage environments. This approach has been used for the computer control and monitoring of a linear accelerator ion source.[36] In this particular case, the signals were optically transmitted across a 750-kV interface using GaAs LEDs and silicon photodetectors.

Optical interconnections for telemetry purposes in a noncomputer-oriented system have been applied to situations closely related to that just described; specifically, high-voltage power system instrumentation. In one such system,[37] a GaAs laser and associated coding and driving circuitry were inductively coupled to a high-voltage transmission line. The light output was transmitted through a $3^1/_2$-meter-long fiber optic to ground potential. The unit was successfully demonstrated on 115- and 230-kV transmission lines, permitting current monitoring with an accuracy of better than 0.3 percent. In a similar system,[38] a design was proposed for current and voltage transformer instrumentation on a 400-kV line in which a GaAs LED was used to transmit the data through a fiber optic mounted inside a hollow, oil-filled, porcelain insulator.

Miscellaneous applications. The three areas just discussed are those in which a majority of the work involving optical interconnections using injection light sources has been done, but there are also numerous other areas for which systems have been built or at least proposed. These include portable long- and short-range voice communication systems[39, 40]—so-called "optical connectors"—in which a matrix of emitters would be placed adjacent to a similar matrix of detectors, photon-coupled isolators (commercially available devices in which an LED and detector are mounted in the same package for the purpose of providing electrical isolation), etc.

Work has also been accomplished in areas that, although not strictly within the limits of definition, are closely related to optical transmission. For example, intrusion-detection systems utilizing injection lasers[41] or GaAs light-emitting diodes[42] and fog-detection systems containing injection lasers[43] have already been built, as well as some specialized military systems. This is certainly not meant to be an all-inclusive list of achievements, but merely a peek at some of the diversified applications that have been considered thus far.

Conclusions

An attempt has been made to describe the potential advantages that are available from optical transmission through the use of injection light sources and to show representative examples that illustrate the present application of such systems. Because of the numerous factors that must be considered in the various possible application areas, no single statement can be made as to when optical techniques are practical and when they are not. In some instances, a particularly desirable property (for example, isolation) may be so important that it will outweigh the disadvantages that may also be present. More generally, however, all of the factors discussed in this article must be weighed carefully in order to determine whether a proposed application is practical.

Although some of the properties described are basic to the nature of optical transmission, there are numerous others that must be modified as the state of the art in devices changes. It is to be expected that more efficient and more highly directional light-emitting diodes will decrease the cost of the required optics in such systems or, alternatively, permit either longer-distance operation or operation with higher signal-to-noise ratios. Development of LEDs with faster turn-on characteristics or room-temperature lasers that can operate at nearly CW will certainly open up new application areas where very-high-speed modulation is a necessity. Moreover, production of lower cost, oriented-fiber optics would greatly increase the possible utilization of spatial multiplexing techniques.

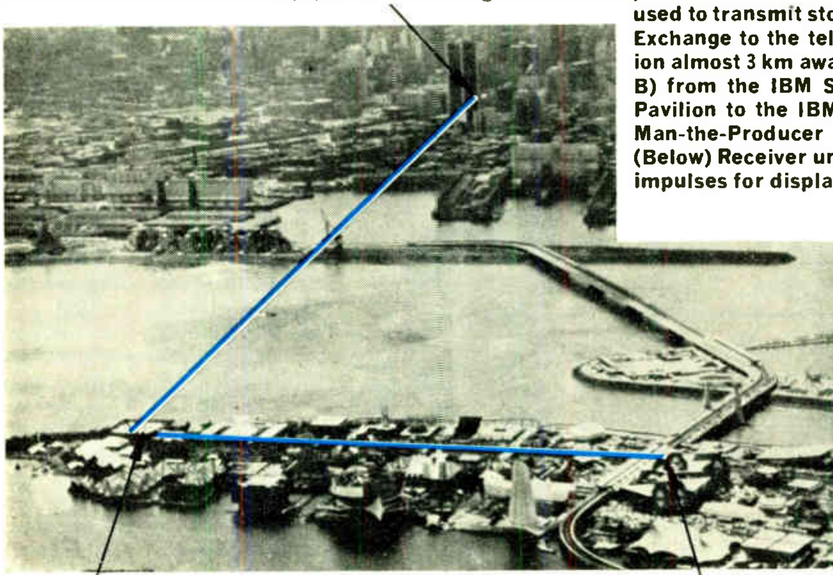


FIGURE 8. The IBM experimental optical transmission system at Expo '67. (Left) Aerial view of the communication path of the invisible infrared light beam (color overlay) used to transmit stock quotations from the Montreal Stock Exchange to the television display in the Canadian Pavilion almost 3 km away (link A) and two-way data/video (link B) from the IBM System/360 computer in the Canadian Pavilion to the IBM 2250 graphic display terminal at the Man-the-Producer Pavilion almost a kilometer away. (Below) Receiver unit reconverts light signals into electric impulses for display on the IBM 2250 terminal at the right.

On the other hand, looking at things in a more negative way, the very overcrowded electromagnetic spectrum may some day force the wider acceptance of optical transmission.

REFERENCES

1. Fried, D. L., and Seidman, J. B., "Heterodyne and photon-counting receivers for optical communications," *Appl. Opt.*, vol. 6, pp. 245–250, Feb. 1967.

2. Cooper, B., "Optical communications in the earth's atmosphere," *IEEE Spectrum*, vol. 3, pp. 83–88, July 1966.

3. Brookner, E., Kolker, M., and Wilmotte, R. M., "Deep-space optical communications," *IEEE Spectrum*, vol. 4, pp. 75–82, Jan. 1967.

4. DeLange, O. E., "Optical heterodyne detection," *IEEE Spectrum*, vol. 5, pp. 77–85, Oct. 1968.

5. Ross, M., *Laser Receivers*. New York: Wiley, 1966, chap. 4 and p. 327.

6. Kainer, S., "Laser beam 'security,'" *Proc. IEEE*, vol. 53, pp. 1752–1753, Nov. 1965.

7. Karbowiak, A. E., "A close look at optical waveguides," *Microwaves*, vol. 5, pp. 37–46, July 1966.

8. Goubau, G., "Lenses guide optical frequencies to low-loss transmission," *Electronics*, vol. 39, pp. 83–89, May 1966.

9. Marcuse, D., and Miller, S. E., "Analysis of a tubular gas lens," *Bell System Tech. J.*, vol. 43, pp. 1759–1782, July 1964.

10. Gilleo, M. A., "Radiation transfer by a light pipe between media with high indices of refraction," *Appl. Opt.*, vol. 3, pp. 765–767, June 1964.

11. Schineller, E. R., "Single-mode-guide laser components," *Microwaves*, vol. 7, pp. 77–85, Jan. 1968.

12. Prochazka, M., Pachman, J., and Muzik, J., "Experimental investigation of a pipeline for optical communication," *Electron. Letters*, vol. 3, pp. 73–74, Feb. 1967.

13. Keyes, R. W., "Injection lasers," *Industrial Research*, pp. 46–53, Oct. 1964.

14. Marinace, J. C., "High power CW operation of GaAs injection lasers at 77°K," *IBM J. Res. Develop.*, vol. 8, pp. 543–544, Nov. 1964.

15. Armstrong, J. A., and Smith, A. W., "Interferometric measurement of linewidth and noise in GaAs lasers," *Appl. Phys. Letters*, vol. 4, pp. 196–198, June 1964.

16. Lavine, J. M., and Iannini, A. A., "Temperature dependence of the multimode behavior of GaAs lasers," *J. Appl. Phys.*, vol. 36, pp. 402–405, Feb. 1965.

17. Konnerth, K. L., "Junction heating in GaAs injection lasers," *Proc. IEEE*, vol. 53, pp. 397–398, Apr. 1965.

18. Carr, W. N., "Photometric figures of merit for semiconductor luminescent sources operating in spontaneous mode," *Infrared Phys.*, vol. 6, no. 1, pp. 1–19, 1966.

19. Konnerth, K. L., and Lanza, C., "Delay between current pulse and light emission of a gallium arsenide injection laser," *Appl. Phys. Letters*, vol. 4, pp. 120–122, Apr. 1964.

20. Marinace, J. C., and Konnerth, K. L., "Room temperature delay times in diffused junction GaAs injection lasers," *IBM. J. Res. Develop.*, vol. 12, pp. 272–274, May 1968.

21. Goldstein, B. S., and Weigand, R. M., "X-band modulation of GaAs lasers," *Proc. IEEE*, vol, 53, p. 195, Feb. 1965.

22. Kressel, H,, and Byer, N. E., "Physical basis of noncatastrophic degradation in GaAs injection lasers," *Proc. IEEE*, vol. 57, pp. 25–33, Jan. 1969.

23. Konnerth, K. L., and Marinace, J. C., "Noncatastrophic degradation of GaAs lasers under CW operation," *IEEE J. Quantum Electronics*, vol. QE-4, pp. 173–175, Apr. 1968.

24. Marinace, J. C., Konnerth, K. L., and Topalian, J. C., "Zinc-diffused GaAs electroluminescent diodes with long operating life," *J. Appl. Phys.*, vol. 41, pp. 2060–2063, Apr. 1970.

25. Chatterton, E. J., M.I.T. Lincoln Laboratory Technology Report 392, June 1965.

26. Haversten, E. V., "The atmosphere as an optical commonwealth channel," *1967 IEEE Internat'l Conv. Rec.*, pt. 2.

27. Shah, B. R., "Optical communication atmospheric effects," 15th annual Opt. Soc. Am. meeting, Mar. 1966.

28. Shah, B. R., "High speed optical data links," Internat'l Commun. Conf., Philadelphia, Pa., 1966.

29. Gaubau, G., "Optical wave-guides," 1965 IEEE NEREM Conf.

30. Anders, R. A., and Callahan, D. E., "Multi-channel electro-optical data link utilizing arrays of non-coherent GaAs light sources," *Nat'l Telemetering Conf. Proc.*, May 1967.

31. Lynch, W. T., and Furnanage, R. A., "Planar beam lead GaAs electroluminescent arrays," *1966 Internat'l Electron Devices Meeting Abstracts.*

32. Baird, J. R., "An optical data link for remote computer terminals," *Datamation*, vol. 16, pp. 125–126, Jan. 1970.

33. Keyes, R. J., *et al.*, "Infrared communications using GaAs diodes," *Electronics*, Oct. 1962.

34. Keyes, R. J., *et al.*, "Modulated infrared diode spans 30 miles," *Electronics*, Apr. 1963.

35. Boerschig, B. A., "A light-modulated data link," *IEEE Trans. Broadcasting*, vol. BC-10, pp. 4–7, Feb. 1964.

36. Budge, E. C., "Data transmission across high voltage interfaces via light links," *IEEE Trans. Nuclear Science*, vol. NS-14, pp. 1074–1077, June 1967.

37. "Traser reaches production stage," *Laser Focus*, pp. 4–5, Nov. 1965.

38. "Light link CT's for 400 kV," *Electrical Rev.*, vol. 183, pp. 768–769, Nov. 1968.

39. Vollmer, J., "Applied lasers," *IEEE Spectrum*, vol. 4, pp. 66–70, June 1967.

40. Eliseev, P. G., Ismailov, I., and Fedorov, Y. F., "Injection lasers for multichannel optical communication," *IEEE J. Quantum Electronics*, vol. QE-6, pp. 38–41, Jan. 1970.

41. "Laser system serves as invisible sentry," *Laser Focus*, pp. 10–11, May 1966.

42. Lytollis, J., Kao, C. K., and Turner, G. I., "Infra-red optical communication systems," *Infrared Phys.*, vol. 8, pp. 123–129, 1968.

43. "Fog: diode laser signals warning," *Electron. Design*, Aug. 1966.

Karl L. Konnerth (M) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Carnegie-Mellon University in 1954, 1955, and 1961, respectively. An assistant professor of electrical engineering at Carnegie-Mellon since 1961, he became a member of the research engineering staff of the IBM Thomas J. Watson Research Center in 1964. Although his experience has been varied, his primary field of investigation has been, until recently, the properties and applications of semiconductor injection light sources. This work has included studies of high-speed transient properties and life characteristics of both injection lasers and light-emitting diodes, as well as their application to optical transmission and displays. At present, however, he is investigating the area of computer applications. Dr. Konnerth is active in the IEEE Resource Committee on Material Sciences in addition to the New York Section's New Technical and Scientific Activities Committee. He is also on the EIA JEDEC Task Force to standardize light-emitting-diode measurement techniques, and is a member of the Society for Information Display, Sigma Xi, and Tau Beta Pi.

Bankim R. Shah (M) received the B.E.E. degree from Poona University, India, and the M.S. degree from Rensselaer Polytechnic Institute. He joined the IBM Corporation in 1962 and began working on transmission aspects of broadband systems, both cable and optical. Other areas of application that he has investigated range from electrooptics in computer systems to system interconnections. His work in optical communication has included the two operational optical links demonstrated by IBM at Expo '67 in Montreal, Canada, and mentioned in the present article. Mr. Shah is currently with the Communications and Engineering Sciences Center of IBM's Federal Systems Division.

Applications of a dc constant-current source

Although most sources of electric energy, including those found in nature, are essentially constant-voltage power supplies, there are many applications in which the availability of a source of constant current would be more desirable

Geoffrey F. Walker Hewlett-Packard Company

Measuring resistance, performing a coulometric titration, measuring semiconductor breakdown voltages, operating IMPATT diodes, testing electrolytic capacitors, measuring magnetic field intensity—these represent only a few of the applications of constant-current sources. Some of the features of constant-current supplies, as well as the requirements imposed on these sources and techniques for using them in a variety of applications, are described in this article.

A constant-current source may be defined as a current generator that will supply any output voltage necessary to keep its output current constant, regardless of the resistance of the load connected to it. It will supply this same constant current at zero voltage to a short circuit and will try to supply the same current at infinite voltage to an open circuit. This characteristic can be contrasted with that of a constant-voltage source, in which the output *current* changes to keep the output *voltage* constant regardless of the size of the load connected. Sources of the latter type are more familiar, perhaps because most common power sources (batteries, the ac power line, laboratory power supplies, etc.) as well as most sources of electric energy found in nature all approximate constant-voltage sources.

The key word here is *approximate*. In practice, a voltage source cannot supply infinite current, nor can a current source supply infinite voltage. In fact, most common laboratory-type constant-voltage supplies have some sort of protective device that limits the current supplied to the load. Similarly, a constant-current source should have some means (preferably adjustable) for limiting the voltage appearing across the load.

Desirable characteristics

The voltage-limit requirement is only one of many characteristics that a practical current source should possess; others include excellent regulation, a means for measuring the output voltage without degrading the regulation, low output capacitance (with corresponding high output impedance at high frequencies), and rapid programming ability.

Regulation. Regulation is a measure of the "constancy" of a constant-current source. It is defined as the change in output current that occurs when the load is changed from a short circuit to the value that requires the current source to supply the full output voltage. Regulation is thus the most basic specification for a constant-current source. As an illustration, consider making accurate resistance measurements from several ohms to several hundred thousand ohms. As resistance changes, current must remain constant if the resistance measurements are to be accurate. Generally, the regulation of the constant-current source should be a factor of ten better than the desired measurement accuracy.

Load-voltage measurement capability. One problem that can arise when constant-current sources are used is that of measuring the output voltage. If a voltmeter is connected directly across the load, the impedance of the meter shunts the load impedance, and the current supplied from the constant-current source divides between the load and the voltmeter. The load then does not receive the previously set value of constant current, and the voltmeter indicates a lower output voltage than was present before the meter was connected.

It is thus necessary to provide another method for measuring the load voltage. One technique is to supply (from inside the current source) a voltage source that is held equal to the load voltage by means of feedback action. If the meter is connected to this voltage source, any current drawn by the meter is supplied by the voltage source and not by the main current supply, thus effectively isolating the meter from the load circuit and eliminating any error resulting from the introduction of the meter into the circuit.

Low output capacitance. In a current source, low output capacitance is essential since excessive output capacitance would cause the high dc output impedance to decrease with increasing frequency, resulting in current transients in rapidly changing loads. For example, if a diode being supplied by the constant-current source goes into avalanche breakdown, its impedance suddenly changes. If the current source has an energy-storage element such as a capacitor in the output, the sudden change of impedance will cause a large amount of energy to be dumped into the diode, possibly destroying it.

Rapid programming ability. The ability to rapidly program the current output opens up two large areas of measurement applications for constant-current sources—those of incremental measurements and high-speed

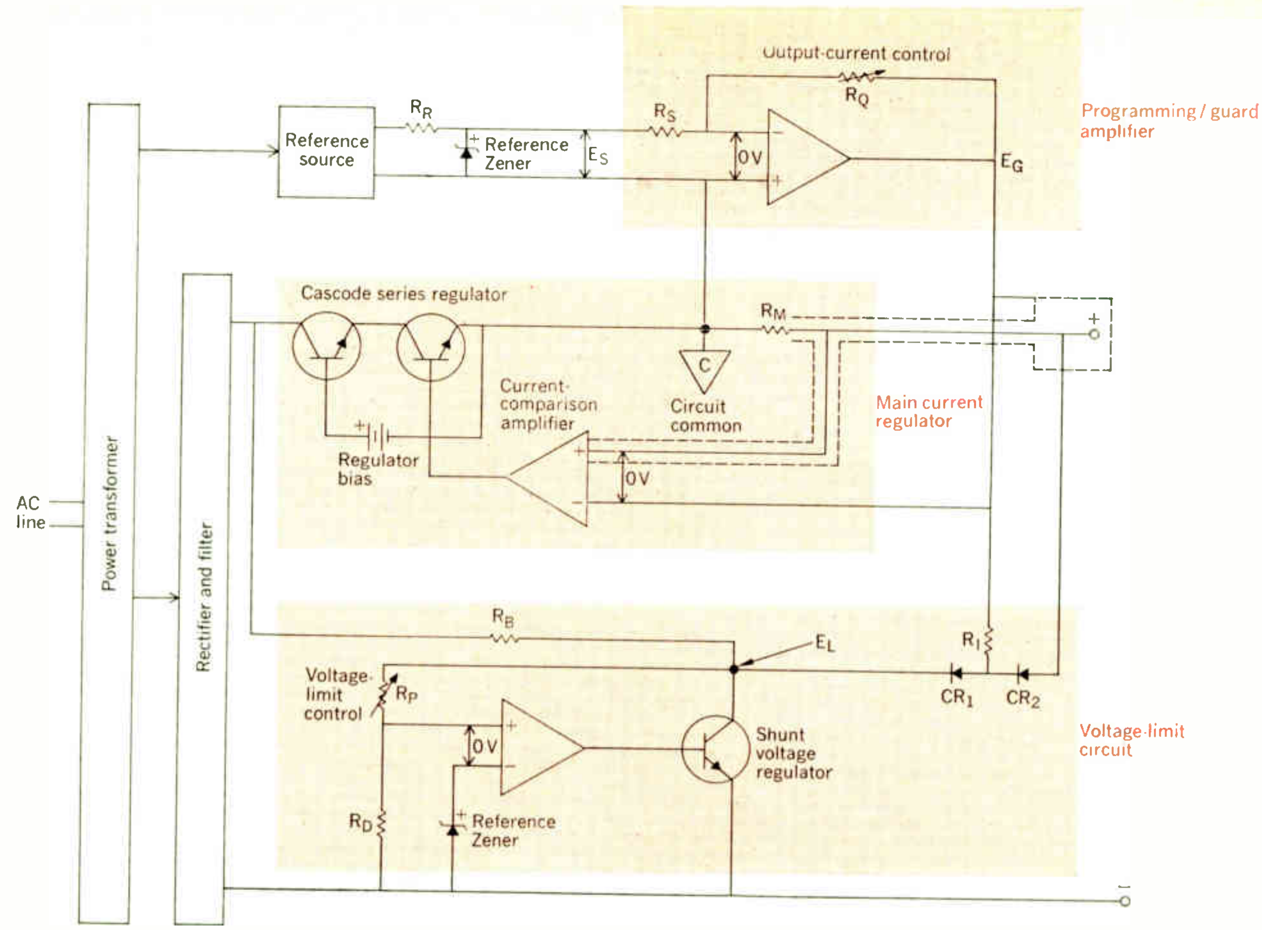


FIGURE 1. Block diagram of typical constant-current source.

(automated) testing. For example, the incremental h-parameters of transistors are commonly measured by supplying the device with two currents: a dc bias current, and a small-signal alternating current that causes incremental changes around the bias point. If a current source can be programmed rapidly enough to supply the latter current, one instrument can supply both currents, thus greatly simplifying the measurement.

Theory of operation

As an example of one possible method of achieving the desirable characteristics described in the preceding section, consider the approach used in Hewlett-Packard's constant-current sources. As shown in Fig. 1, the constant-current source is comprised of three primary sections: the programming/guard amplifier, the main current regulator, and the voltage-limit circuit.

The programming/guard amplifier is an independent, variable-voltage source whose output E_G is linearly dependent upon the setting of the output current control R_Q. In its programming function, the programming/guard amplifier provides the programming voltage E_G for the main current regulator. This dc voltage, which is negative with respect to the circuit common, is applied to one of the inputs of the differential current-comparison amplifier. The other input of this differential amplifier is connected to the current-monitoring resistor R_M. The current-comparison amplifier continuously compares the voltage drop across the current-monitoring resistor ($I_O R_M$) with the programming voltage E_G. If these voltages are momentarily unequal, due to a load disturbance or a change in the output current, the difference is amplified and applied to the main current-regulator transistors, altering the current conducted through them and forcing $I_O R_M$ once again to equal E_G. The output current I_O is thus related to the programming voltage E_G and the reference voltage E_S by the following relationship:

$$I_O = \frac{E_G}{R_M} = \left[\frac{E_S}{R_M} \cdot \frac{R_Q}{R_S}\right] \qquad (1)$$

Returning to the guard function of the programming/guard amplifier, note that the output of this amplifier (E_G) is connected to a guard conductor (shown by the dashed lines in Fig. 1) surrounding the positive output terminal, the current-sampling resistor, and the non-inverting input to the current-comparison amplifier. Since E_G is held at the same potential as the positive output voltage by the main current regulator, no leakage current flows from the positive output terminal or any of the internal circuit elements connected to it. Leakage currents that would normally flow from the positive output circuitry flow instead from the guard conductor. Note that since the programming/guard amplifier is a low-impedance source referenced to circuit common, any leakage current supplied by the guard originates from circuit common via this amplifier and bypasses the current-monitoring resistor. Thus *only* the output current flows through the current-monitoring resistor. In this manner, leakage current flowing directly between the supply's two output terminals is eliminated, and precise load regulation is obtained.

Voltage-limiting action is provided by the circuit shown in the bottom portion of Fig. 1. Normally, when voltage-limiting action is not occurring, the setting of the voltage-limit control R_P establishes across the shunt voltage regulator a preset voltage limit E_L, which is higher than the positive output voltage and its twin, the guard voltage. Since there is zero voltage across the series combination of isolation diode CR_2 and resistor R_I, no current flows through them. Potential E_G is thus present at their junction, and isolation diode CR_1 is reverse-biased. If the output voltage exceeds the preset limit value (E_L), CR_1 and CR_2 conduct, allowing the shunt regulator to conduct a portion of the current that would otherwise flow to the load, thereby clamping the output voltage to the preset limit value.

Even during voltage-limiting action, the guard voltage E_G continues to be maintained at a value equal to the potential at the positive output terminal, and both guarding action and the normal control action of the main current regulator continue, thus minimizing any output current transients that might occur during transfer from voltage limit to the normal output current mode. In addition, note that the shunt regulator conducts a "standby" current through bias resistor R_B; this current insures that the shunt regulator is operating in its linear region, ready to react quickly when voltage-limiting action is required, thus preventing output-voltage crossover transients.

Resistance measurements

Silicon-wafer resistivity (four-point measurement). Measurement of the resistivity of silicon slices used in the manufacture of semiconductors is well suited to the constant-current method. Resistivity (ρ) is defined as the resistance in ohms between opposite faces of a cube of material 1 cm on a side, and its unit is the ohm-centimeter. The resistivity of a semiconductor wafer is a direct measure of the doping impurity in it.

By far the most common method used in making this measurement is the four-point probe method. As shown in Fig. 2, the silicon-wafer sample is contacted by a probe containing four in-line contact points. The outer two contacts supply current through the sample, and the inner two contacts measure the voltage drop across a portion of the sample. The use of four probes rather than two avoids the error caused by contact resistance. If the current and voltage were to use the same contact, there would be an IR drop across the probe-to-sample contact resistance and in the body of the probe itself between the surface of the sample and the connection to the voltmeter lead. This could cause a significant error when measuring low resistivities (0.005 to 1.0 ohm·cm), in which case the IR drop may begin to approach the magnitude of the sample voltage.

By picking an appropriate value of current supplied to the sample, the resistivity measurement can be made very easily. Selecting the exact value of current requires that the resistivity be given in terms of the sample size and four-point probe geometry as follows[1]:

$$\rho = \frac{Vt}{I}\left[\frac{\pi}{\ln 2 + \ln \dfrac{(d/s)^2 + 3}{(d/s)^2 - 3}}\right] \qquad (2)$$

where ρ is the resistivity in ohm-centimeters, V is the measured voltage in millivolts, I is the applied current in milliamperes, t is the thickness of the wafer in centimeters, d is the diameter of the wafer in centimeters, and s is the voltage-probe-point spacing in centimeters. This expression is valid only for circular, thin slices, where the thickness is less than one half the voltage-probe spacing. Comparable expressions for other sample geometries can be found in the extensive literature on silicon-wafer resistivity measurements. For shallow diffusions into the surface of the silicon, forming a p-n junction below the surface, the doping value changes from a high value at the surface to progressively lower values in the interior until the doping level equals that of the opposite conductivity type of the original wafer. In such a case, the resistivity changes rapidly with depth and it is convenient to

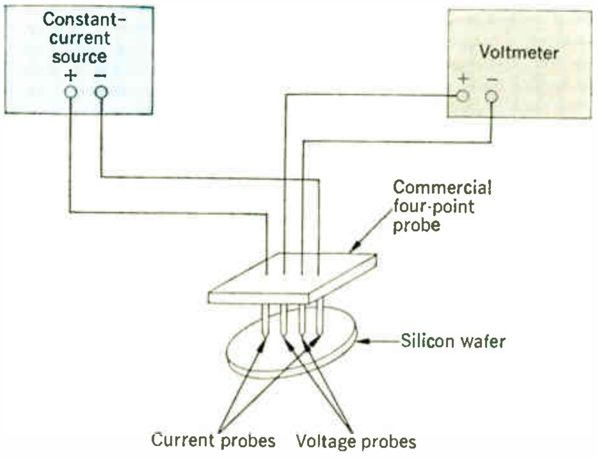
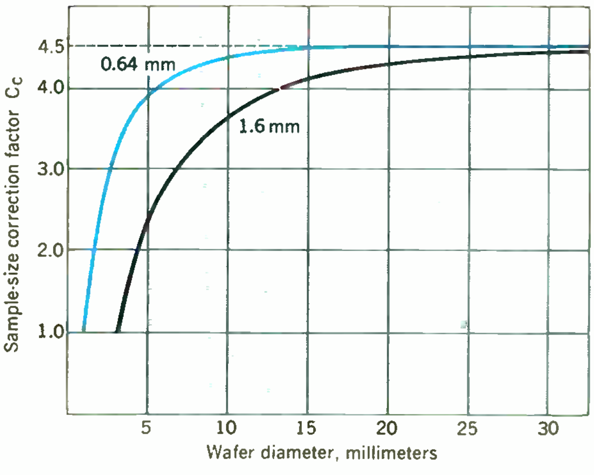


FIGURE 2. Four-point probe method for measuring silicon-wafer resistivity. Constant current is supplied to outer two probes and the voltage drop is measured across the inner two probes. This method eliminates the effect of contact resistance.

FIGURE 3. Resistivity correction factor C_c vs. sample size diameter, see Eqs. (2) and (3), for two common voltage-probe spacings. Correction factor is 4.5 for large sample diameters and decreases rapidly with diameter.



characterize the diffusion by a sheet resistivity, in ohms per square, given by

$$\rho\ (\text{per square}) = \frac{V}{I}\left[\frac{\pi}{\ln 2 + \ln \dfrac{(d/s)^2 + 3}{(d/s)^2 - 3}}\right] \qquad (2a)$$

If the entire expression within the brackets is called C_c, a correction factor, Eq. (2) can be written as

$$\rho = C_c t \frac{V}{I} \qquad (3)$$

If the applied current in milliamperes is set equal to C_c times the thickness, it can be seen that the number of millivolts measured is equal to the resisitivity in ohm-centimeters, and no calculations are required once the value of C_c has been determined for the particular probe-point spacing and sample size in use.

C_c is plotted as a function of the wafer diameter in Fig. 3 for standard probe spacings of 25 and 62.5 mils (0.64 and 1.6 mm). It can be seen that ignoring the effect of small sample size can result in quite a large error—with the latter probe spacing and a sample size of 8 mm, the error in the measured value of resisitivity is +25 percent.

In addition to the sample-size error, there are three other factors affecting the accuracy of the measurement. First, the foregoing equations were derived for a measurement made in the center of the wafer; other probe locations require greater corrections. Figure 4 shows the measurement error as the probe is moved away from the center of a 20-mm-diameter wafer.[2] Second, the heating of the wafer by the applied current must be considered, since typical temperature coefficients of resistivity for zone-refined silicon rods vary between 0.5 to 1.5 percent per degree C. It is difficult to specify a maximum current that can be used; in general, however, currents over 10 mA will begin to cause heating difficulties. The relatively high temperature coefficient also points out the need for control of the ambient temperature if accurate results are to be achieved. The third factor is the mechanical construction of the four-point probe itself. Four-point probes are commercially available at prices ranging from several hundred to several thousand dollars. Most of these models are rigidly constructed with very close tolerances on the probe assembly, and thus errors resulting from probe wander are minimized.

In addition, most probes are equipped with mechanisms that automatically adjust the contact pressure, thereby eliminating another error-causing variable.

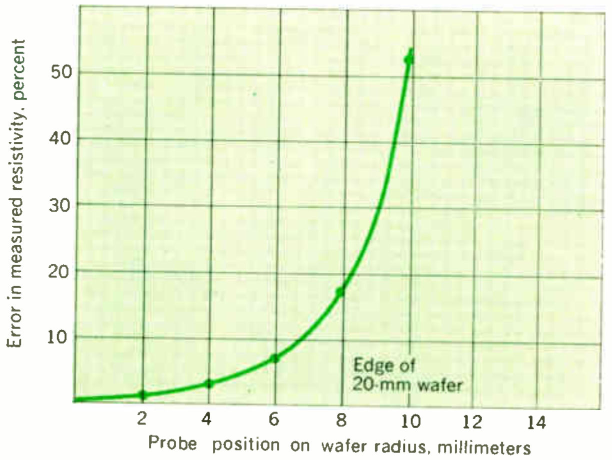
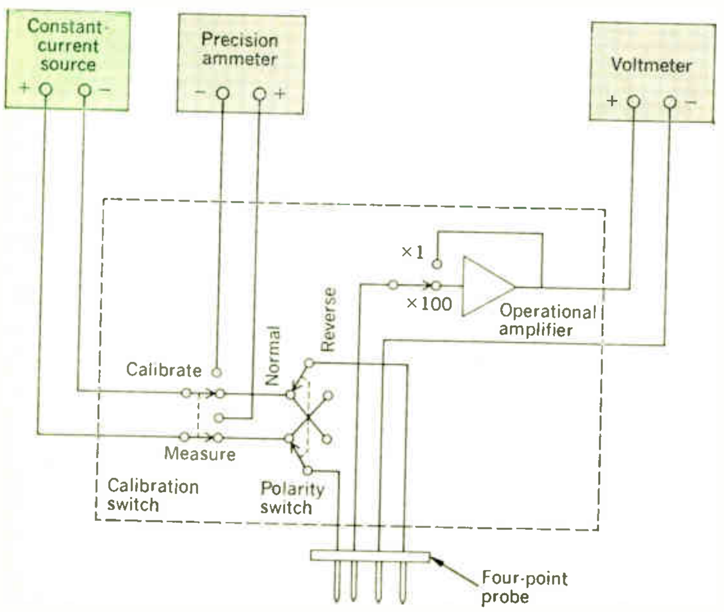


FIGURE 4. Percent error in measured resistivity as the four-point probe is moved away from the center of a 20-mm-diameter wafer.

FIGURE 5. Four-point probe setup with added sophistications: "calibrate–measure" switch, precision calibrating ammeter, polarity-reversing switch, and operational amplifier for measuring low resistivities.



There are several minor refinements that can be included in the measurement. The first of these is a "calibrate–measure" switch. As shown in Fig. 5, an external precision ammeter can be connected to the current source in the "calibrate" position to allow the current to be set accurately to the desired value. The second refinement is the use of the voltage-limit control on the constant-current source. Applying too high an initial contact voltage to very thin wafers or wafers having very thin surface films can quite easily break down the wafer or film. With the voltage limit set to a value slightly above the maximum voltage that is being measured, the samples are protected from the damage that otherwise might occur when the probe is first touched to the sample. A third refinement is the use of a polarity-reversing switch, which is placed in the current leads of the probe in such a manner as to reverse the direction of current flow through the sample. Reversing the current allows the contact between the probe and the wafer surface to be checked for undesired diode action. An alternative technique involves the use of the range switch on the current source to check the contact condition. If poor contact due to heavy surface oxide or other causes is suspected, the range switch can be switched to the next highest (or lowest) range, increasing or decreasing the current by a factor of 10. If the measured voltage does not change by the same factor (within the range switching tolerance), the probe-to-surface contact is behaving in a nonlinear fashion and should be adjusted. Caution should be exercised in using this checking method at currents over approximately 10 mA in order to avoid undesirable heating effects. A final refinement is the use of an operational amplifier to amplify the probe voltage when measuring very low resistivities. For example, with a sample of silicon with a resistivity of 0.005 ohm·cm and an applied current of 1 mA, the probe voltage is only 5 μV; measurement accuracy at this low level is difficult to achieve. Using a stabilized $\times 100$ operational amplifier, the measured voltage would be 500 μV—a level somewhat easier to measure accurately.

Production-line resistance testing, grading, and trimming. A constant-current source can serve as an integral part of a resistor production-line test setup. A typical production-line application is in the grading process. A production line may be set up to produce a resistor of a given value ± 10 percent. The resulting resistors are then measured and graded according to how close they are to the nominal value. The current source is set to provide a fixed current in multiples of milliamperes such that the voltmeter reads directly in kilohms or any convenient multiple thereof (for example, at 0.1 mA, 1 volt equals 10 kΩ). The voltmeter is connected to the meter terminal of the current source as shown in Fig. 6 in order to avoid the measurement error associated with connecting the meter across the resistance being measured. However, since the voltmeter is sensing at the current-source output terminals, the resistance of the leads connecting the test terminals to the current source and the contact resistance of the calibration switch must be as low as possible. To set the current accurately, a precision resistor (0.01 or 0.001 percent) can be connected across the current source and the current can be adjusted for an exact voltmeter reading. The voltage-limit control can be set to eliminate high voltages appearing across the test points when one resistor is removed and another is inserted.

In testing thin-film chip resistors, the main problem is one of heating. When an ohmmeter (digital or analog) is used as the measuring instrument, the current supplied to the resistor is usually of the order of 1 mA. This is too much current to be passed through a typical 28- by 38-mil (0.71- by 0.97-mm) micropower chip resistor rated for 25 mW maximum dissipation. However, controlled currents of 1 to 10 μA are ideal for this purpose, since heating problems are eliminated. On the other hand, if the heating effect *is* desired (e.g., in value trimming by oxidation with pulsed direct current), it is readily available. The voltage-limit control is particularly useful in working with thin-film chip resistors, since application of too high a voltage can very easily cause arcing and breakdown on chips where the resistance tracks are only 25 μm apart and film thickness is typically 500 Å.

The constant-current resistance measurement technique is also useful in trimming operations based on abrasive or laser methods. For instance, a thin-film resistor may be vacuum-deposited in a matrix pattern on a silicon substrate. Connection branches are provided in the matrix pattern, so by removing one branch with a laser, the value is changed by a small percentage. The resistance can be continuously monitored with constant current while this operation is taking place.

Still another use is in the analysis of integrated circuits containing deposited tantalum nitride resistors. Often these resistors are in parallel with an emitter–base or collector–emitter junction. If it is discovered that leakage is present in the device, constant current can be applied to the junction and by observing the relationship between the voltage and the current, it can be determined whether the leakage is in the active device or in the resistor.

Resistance measurements such as these lend themselves particularly well to computer automation. The current output and voltage limit of the constant-current source can be controlled by a computer through a D/A interface; output resistance readings can be punched on paper tape or another output medium. In the resistance-trimming operation, a feedback system can be set up in which resistance measurements are fed back to the controlling computer to determine when the resistance is close to the desired value.

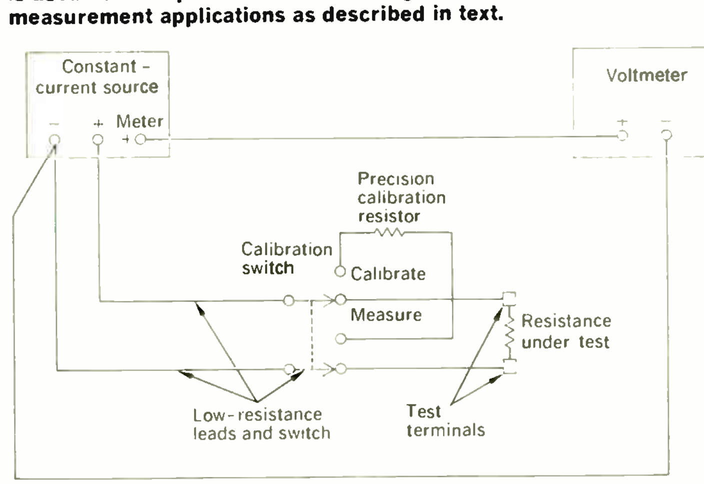
Semiconductor device measurements and related applications

What can you do with constant current? An application area that comes to mind as frequently as that of resistance measurements is that of semiconductor device measurements. Intuitively, one expects constant current to be useful with current-controlled devices, particularly in such measurements as breakdown voltage of transistor junctions where the breakdown voltage is specified at a constant current, and in small-signal *h*-parameter measurements where the transistor must be supplied with a constant dc bias current on which ac modulation is superimposed.

However, not all semiconductor parameters are suited to measurement by constant-current methods. Such parameters as collector leakage current (I_{CEO}) and emitter–base cutoff current (I_{EBO}) in which the current is specified with a constant *voltage* applied to the device are not easily measured with a constant-current source. In addition, even if a method is devised to measure these parameters with a constant-current source, the magnitudes of the currents are below the settable output of most current sources (for small-signal transistors, I_{CEO} is usually less than a microampere, and I_{CBO} is often in nanoamperes).

Measurement methods for those parameters that are well suited to constant-current measurement techniques are discussed in the following paragraphs. As a first example, consider the measurements that can be made on a diode. A typical diode (p-n junction) characteristic is

FIGURE 6. Basic resistance-measuring setup. Constant current is supplied to unknown resistance, and measured voltage drop is directly proportional to resistance. Setup is used both for production-line testing and temperature-measurement applications as described in text.



shown in Fig. 7. Measurements of two parameters shown on the characteristic (forward voltage drop and reverse breakdown voltage) can be made by simply reading values on a voltmeter; measurement of a third parameter not shown on the characteristic (temperature coefficient) can be made with the addition of a temperature-controlled oven.

Diode forward voltage drop. The maximum value of the forward voltage drop V_F is specified on data sheets at one or more values of forward current I_F (see Fig. 7). For controlled-conductance diodes (used in logic circuits to permit greater design margins or additional logic states), both the minimum and maximum values of forward voltage drop are specified at several different values of forward current. Measurement of this parameter entails simply putting a constant current through the diode in the forward direction, and measuring the voltage drop across it, as shown in Fig. 8. Selecting different values of forward current will move the measuring point along the characteristic curve: points *A*, *B*, and *C* in Fig. 7 represent operating points at increasingly higher values of forward current.

Diode reverse breakdown voltage. The minimum value of the reverse breakdown voltage is specified on data sheets at a fixed value of reverse current. In the same manner as forward voltage drop, this parameter is measured simply by applying a constant current through the diode in the reverse direction and measuring the voltage across it. The setup shown in Fig. 8 can be utilized by reversing the diode.

Note that although the parameter being measured is a "breakdown" voltage, it is being measured nondestructively by virtue of the controlled current. The breakdown voltage as shown on the characteristic in Fig. 7 is located at the beginning of the avalanche region. In this region, very large changes in reverse current result in only very small changes in reverse voltage. This means that if a constant-voltage source were used to make this measurement, a very small change in output voltage could increase the reverse current (and thus the power dissipation) to the point where the diode fails. To avoid this possibility, the variable that must be controlled is the current rather than the voltage—thus the necessity for a constant-current source.

When actually performing the measurement, varying the current supplied to the diode will move the operating point along the characteristic as in the forward-voltage measurement, except in the reverse direction. However, because the magnitude of the leakage current shown in Fig. 7 is so small (often less than 1 μA for silicon signal diodes), the almost-horizontal portion of the characteristic will be traversed rapidly as the output of the current source is increased from zero. Increasing the current output to a value greater than several microamperes will result in the measured reverse voltage increasing extremely slowly. When this takes place, it is definite indication that the diode is operating in the avalanche breakdown mode.

"Zener" voltage can be measured using this same procedure, since "Zener" voltage is simply the reverse breakdown voltage of a diode *designed* to be operated either in the Zener or the avalanche region. "Zener" voltage is usually specified on data sheets at the current resulting in a power dissipation of one quarter the maximum rated value for uncompensated diodes, or at the

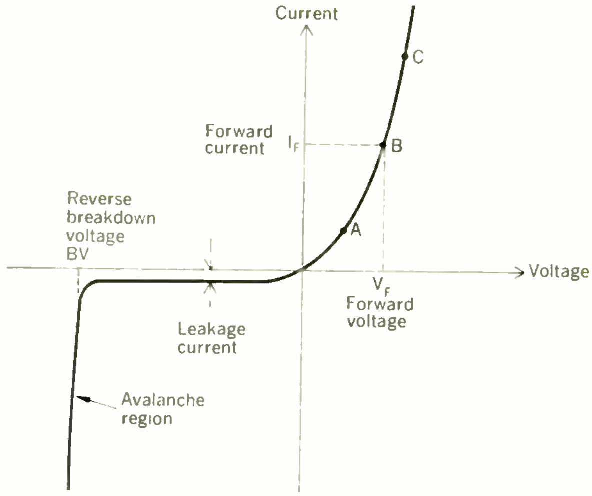


FIGURE 7. Characteristic V-I curve of diode (p-n) junction). Area to right of current axis is forward-biased region; area to left is reverse-biased region.

current resulting in the minimum temperature coefficient for reference (temperature-compensated) diodes.

Temperature coefficient. With the addition of a temperature-controlled oven, the foregoing measurement procedure can be used to determine the forward- or reverse-voltage temperature coefficient of a diode. Connected as shown in Fig. 8, the diode is placed in the oven (the current source remains outside the oven), the temperature is varied over the desired range, and the voltage is recorded at each desired temperature setting. Sufficient time should be allowed for the diode junction temperature to stabilize before each voltage reading is taken. The forward-voltage temperature coefficient of small-signal silicon diodes is typically 2 mV/°C. Typical temperature coefficients of "Zener" voltage vary from 25 mV/°C for uncompensated diodes to essentially zero for reference diodes.

Voltage–current characteristic. In essence, measuring the forward voltage drop and reverse breakdown voltage of diodes as has been described amounts to determining points on the characteristic. Instead of measuring only two specific points on the characteristic of one specific device, it is perfectly possible to determine the entire *V-I* characteristic of *any* device, linear or nonlinear, by using just the same method of supplying a known current and measuring the voltage drop across the device at many points on the characteristic. In fact, by remote-programming the current source and connecting an *X-Y* plotter to the programming source and the current-source meter terminal, completely automated characteristic plotting may be achieved.

Transistor junction reverse breakdown. Measurement of diode parameters leads naturally into measurement of transistor parameters, and perhaps the most obvious and easily measured transistor parameter is junction reverse breakdown voltage.

For a transistor junction (for example, the base–emitter junction), the breakdown voltage measurement can be made in exactly the same manner as described for a diode. BV_{EBO}, the emitter-to-base breakdown voltage with the collector open, is specified on transistor data sheets at a constant current (typically 100 μA). This current can be set on the current source, and the breakdown voltage (typically less than 10 volts for low-power silicon transistors) read directly from the voltmeter connected to the meter terminal on the current source, as shown in Fig. 9. In a similar manner, a measurement of BV_{CBO}, the collector-to-base breakdown voltage with the emitter open, can be made (see Fig. 9); typical values vary with devices selected for specific applications.

The most meaningful and commonly used breakdown voltage is the collector-to-emitter value. There are four different collector–emitter breakdown voltages that can be measured, depending on the base connection. In order of increasing magnitude, they are: BV_{CEO} (base open), BV_{CER} (base connected to the emitter through a resistor of value R), BV_{CES} (base short-circuited to the emitter), and BV_{CEV} (base reverse-biased with respect to the emitter by voltage V). A simple setup for determining these voltages is shown in Fig. 10. These breakdown voltages are usually specified at a higher collector current (typically 1 mA) than the BV_{EBO} and BV_{CBO} specifications in order to avoid difficulty with leakage-current multiplication.

DC (static) current transfer ratio. The single parameter most frequently used in describing a transistor's performance is the forward-current transfer ratio. The ratio, a measure of the gain (amplification factor) of a transistor, is commonly specified for two different transistor connections, common emitter and common base. Both ratios are easily measured using a constant-current stimulus. The following paragraphs describe measurement procedures for the common-emitter transfer ratio (h_{FE} or B); the common-base transfer ratio (h_{FB} or A) may be measured using similar techniques.

The measurement can be performed on either a qualitative or "go/no-go" basis; both use essentially the same test setup, that of Fig. 11. In both cases, constant-current sources supply the base current and collector current for the transistor under test. On transistor data sheets, B is usually given at a specified collector current and collector-to-emitter voltage. Thus, to perform the measurement, the collector-current source is set for the specified I_C, and the base-current source is adjusted until V_{CE} (displayed by the voltmeter connected to the meter terminal of the collector-current source) is the specified value. The current supplied from the base-current source is then measured, and B is calculated by dividing the set collector current by the measured base current.

For small-signal transistors, I_C is typically 1 or 2 mA and B ranges from 20 to 400. It can therefore be seen that for a typical B of 100, the current supplied by the base-current source will be 10 to 20 μA—much too small a value to be read accurately from the front-panel meter on the current source. Instead, the current must be measured with either a series ammeter, as shown in Fig. 11, or a small current-monitoring resistor and a voltmeter.

"Go/no-go" measurements, more suited to production or quality-control applications, can be made by utilizing the same procedure except that the base-current source is set to a fixed, known current value. The collector-to-emitter voltage then becomes the measured variable. If V_{CE} as read on the meter is less than that given in the test specification, B for the transistor is greater than that required; if V_{CE} is greater than the value specified, B is less than the required value.

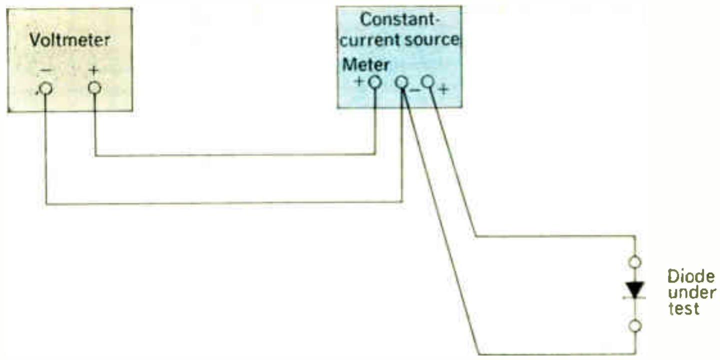
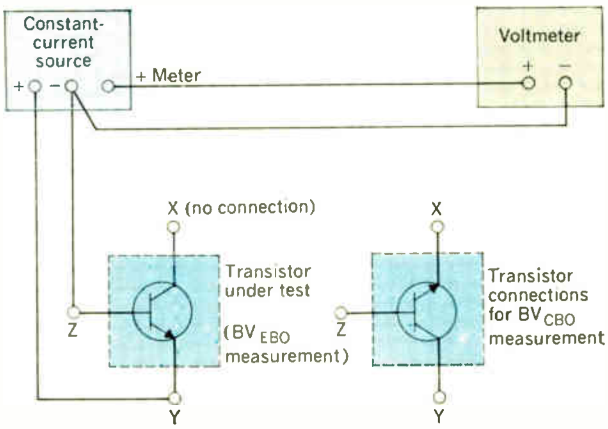


FIGURE 8. Basic test setup for measuring forward voltage drop of diode. Reversing diode allows measurement of reverse breakdown voltage.

FIGURE 9. Test setup for measuring transistor reverse breakdown voltages BV_{EBO} and BV_{CBO}. Constant-current source reverse-biases transistor junction; voltmeter connected to meter terminal on current source measures voltage drop across junction.



Junction saturation voltage. Another frequently used transistor parameter is $V_{CE(SAT)}$, the voltage from the collector to the emitter for a given I_C and I_B while biased in the collector saturation region. The value of $V_{CE(SAT)}$ is often of particular importance in computer applications where minimum and maximum values of the saturation voltage are frequently relied upon in the design. The measurement utilizes the setup of Fig. 11. The specified base and collector currents are set on the constant-current sources, and $V_{CE(SAT)}$ is read directly from the voltmeter connected to the meter terminal of the collector-current source. The currents specified are such that operation in the collector saturation region is guaranteed: if the transistor has a B of 100 specified at a collector current of 1 mA, any base current greater than 10 μA will saturate the transistor when the collector current is 1 mA. Usually a margin of safety is built into such measurements; in the foregoing example, the actual base current specified for the test might be 100 μA in order to insure that the transistor will be saturated. Typical values of $V_{CE(SAT)}$ for small-signal silicon transistors vary from 0.1 to 0.5 volt.

AC modulation method. An ac-modulated constant-current source is the ideal power supply for measuring incremental parameters of semiconductor devices, since it can simultaneously supply both a fixed direct current that biases the device into operation in its active region, and an alternating current that produces incremental changes around this bias point. However, before the actual procedures for these measurements is discussed, an explanation of the method used in modulating a current source in order.

As shown in Fig. 12, the programming circuitry of the current source is arranged so that a constant current (I_P, the programming current) flows through a reference resistor R_R and the front panel current control R_P. Because R_P has a constant current flowing through it, varying the resistance varies the voltage V_P (the programming voltage) appearing at point *P*. The output current of the instrument is held proportional to this potential. It can thus be seen that if the programming current is increased, the dc output current will increase because the potential at point *P* has increased, even though the position of the front-panel current control has not been changed.

The programming current, typically 1 mA, is constant

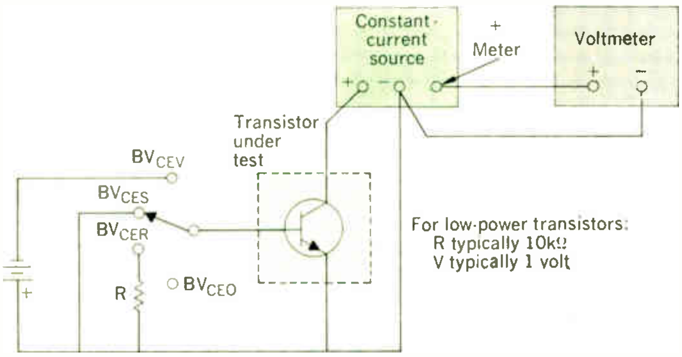
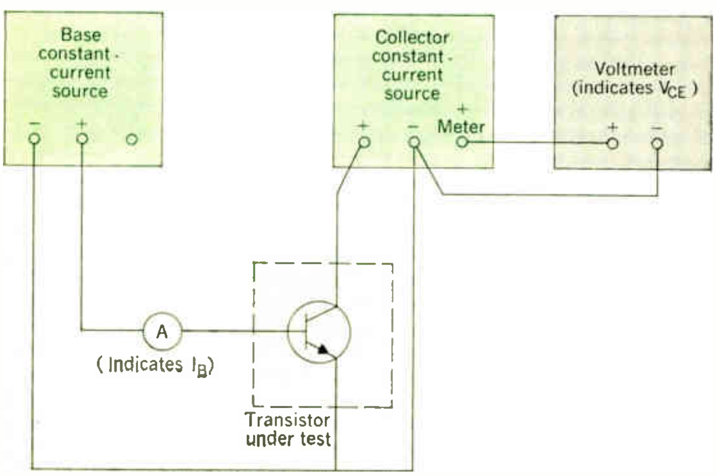


FIGURE 10. Circuit for measurement of four collector–emitter breakdown voltages: (in ascending order) BV_{CEO}, BV_{CER}, BV_{CES}, and BV_{CEV}. Final letter in voltage subscript indicates connection to base of transistor (O = open, R = resistance R between base and emitter, S = short circuit, V = voltage reverse-biasing base–emitter junction).

FIGURE 11. Measurement setup for dc common-emitter current-transfer ratio and junction saturation voltage. Constant-current sources are used to supply both the base and collector currents.



for *any* dc output current. If an additional current I_s is added to the programming current, the dc output current will increase by a proportional amount. For example, if the 1-mA programming current is increased by 0.05 mA, the dc output current will increase by 5 percent; if the dc output control has been set at 50 mA, it would increase to 52.5 mA. Note that this same 5 percent increase in output current will occur at any output current setting. If the dc output control is set at 5 mA, the output current will increase to 5.25 mA, because the percentage variation in output current is determined only by the increase in the programming current, and not by the level at which the dc output current is set.

This additional current (I_s) can be supplied from a signal generator as shown in Fig. 12. Given the peak-to-peak value of the signal-generator output voltage and the desired magnitude of current I_s, the value of resistor R_X can be determined from Ohm's law.

For example, if it is desired to vary the dc output current by ±10 percent around a dc value of 100 mA, the 1-mA programming current must vary ±10 percent, or I_S must be 0.2 mA peak-to-peak. If the oscillator output voltage (V_s) is 5 volts peak-to-peak, resistor R_X must be 5 V/0.2 mA, or 25 kΩ. Coupling capacitor C_C is necessary to prevent any dc voltage that is present in the output of the oscillator from reaching point *P*. The value of C_C should be such that its reactance at the operating frequency is less than $^1/_{10}$ the magnitude of R_X. For example, if the oscillator were set for 100 Hz in the foregoing example, the reactance $X_C = 1/(2\pi fC)$ would have to be less than 2.5 kΩ. This would require a capacitor of at least 0.7 μF.

There are two inherent limitations on the modulation process:

1. The dc output current can never swing below zero or above the rated output current of the current source if the output is to be undistorted. As an example, assume that an output current varying ±25 percent of its dc value ("25 percent modulation") is desired from a 250-mA current source. Since the +25 percent portion of the swing must not cause the dc output to exceed 250 mA, the dc output of the current source cannot be set higher than 200 mA.

2. The amplifiers in the current source place a definite limit on the maximum modulation frequency. This limitation is usually expressed as the maximum percentage modulation (±*x* percent of the dc output) that may be used at a given frequency. Typically, a current source may be specified for 100 percent modulation up to 100 Hz, decreasing linearly thereafter to a maximum of 10 percent at 1000 Hz.

Zener-diode dynamic impedance. Zener-diode incremental impedance (defined as the slope of the *V-I* curve, usually of interest only in the breakdown region) can be easily measured using the preceding method for modulating a current source. Zener impedance measurements are commonly used as a measure of the voltage-regulating capability of the device.

Zener impedance is usually specified at two current values: the quarter-power point (I_T), and a low value in the knee region (I_K) chosen so that a plot of impedance versus current on a log–log scale is essentially a straight line (see Fig. 13).[3] Two additional parameters, the measurement frequency and the percentage modulation, are usually specified on diode data sheets; common values are frequencies of 60 or 1000 Hz and modulations of 5 or 10 percent.

The test setup of Fig. 14 is used to perform the measurement. The constant-current source supplies the specified value of dc bias current (I_T or I_K), the oscillator provides incremental output current changes (ΔI) around this bias value, and the vertical channel of the oscilloscope displays the ac voltage (ΔV) appearing across the diode. Series resistor R_X is selected as described under "AC Modulation Method" to provide the desired modulation percentage (and thus determine the value of ΔI). Calculating $\Delta V/\Delta I$ yields the Zener dynamic impedance.

Note that each time a new bias current is selected, the magnitude of ΔI changes, since it is a fixed percentage of the bias current. A more convenient method of performing this measurement at many different values of bias current utilizes the "optional connections" shown in Fig. 14. These connections allow the horizontal channel of the oscilloscope to sense the current-source programming voltage. This voltage appears at the current-source meter terminal, referenced to *circuit common* (note that the voltage at the same terminal referenced to the *negative output terminal* is equal to the output voltage of the current source). Because the oscilloscope common is already connected to the current-source negative output terminal, the programming voltage must be measured *differentially* as shown. The constant of proportionality relating the programming voltage to the dc output current is simply the voltage programming coefficient of the current source. The value depends on the particular current source and output range used; typical values vary from 2 mV/mA to 10 V/mA. Since the oscilloscope horizontal channel is sensing the ac diode current (ΔI) and the vertical channel is sensing the ac diode voltage (ΔV), the resulting display is a line whose slope (vertical deflection per unit horizontal deflection) is the diode incremental impedance.

When very low incremental impedances (less than 1 ohm) are measured at relatively high frequencies (above 500 Hz), the line displayed on the oscilloscope may be an ellipse rather than a straight line. This effect, caused by the phase shift inherent in the current-source amplifier at higher frequencies, can be eliminated by connecting the oscilloscope vertical amplifier directly across the diode under test rather than to the meter terminal as shown in Fig. 14. Since the impedance of the diode is much less than the input impedance of the oscilloscope (1 ohm vs. 1 megohm), the current-source regulation will not suffer any measurable degradation.

Transistor h-parameters. Another application of an ac-modulated current source is found in the measurement of transistor hybrid parameters. The following descriptions discuss measurement of the common-emitter parameters (indicated by the *e* in the subscript of each parameter); common-base or common-collector *h*-parameters can be measured using similar techniques. Table I gives definitions, alternate symbols, and typical values of the common-emitter *h*-parameters for small-signal silicon transistors.

Since the exact values of these parameters are dependent upon the quiescent operating point of the transistor, they are usually given on transistor data sheets at a specified value of dc collector current and dc collector-to-emitter voltage (typical values for small-signal transistors are I_C = 1 mA and V_{CE} = 10 volts).

The *h*-parameters can be separated into two pairs according to the measurement conditions: (1) h_{re} and h_{oe} (measured with the transistor input open-circuited), and (2) h_{ie} and h_{fe} (measured with the transistor output short-circuited). The first pair can be measured utilizing the setup of Fig. 15 in combination with the modulation method of Fig. 12. As shown in Fig. 15, the base constant-current source supplies the dc base current to the transistor (I_B) and the very high output impedance of the current source acts as an open circuit to any ac component of the base current (i_b). The collector-current source supplies both the dc (I_C) and ac (i_c) collector currents. When the measurements are being performed, the collector-current source is set for the specified value of I_C, and the base-current source is adjusted until V_{CE} is the specified value.

The modulation percentage used in these two measurements is generally less than 10 percent in order to maintain "small-signal" conditions; note that whatever percentage is actually used, the value of i_c is determined

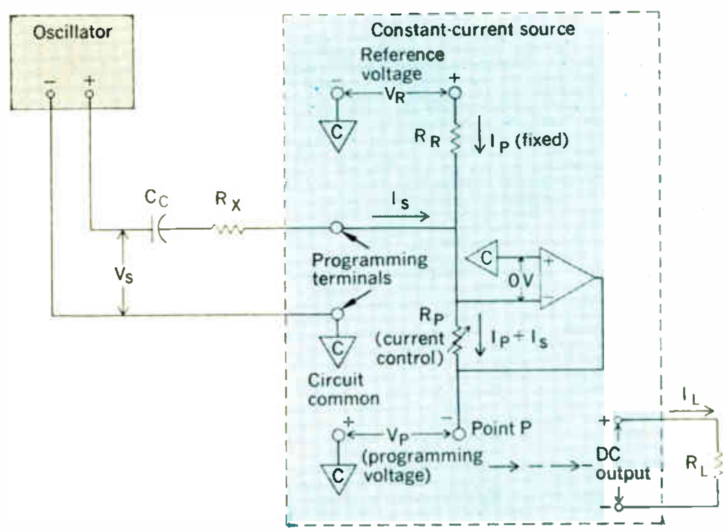
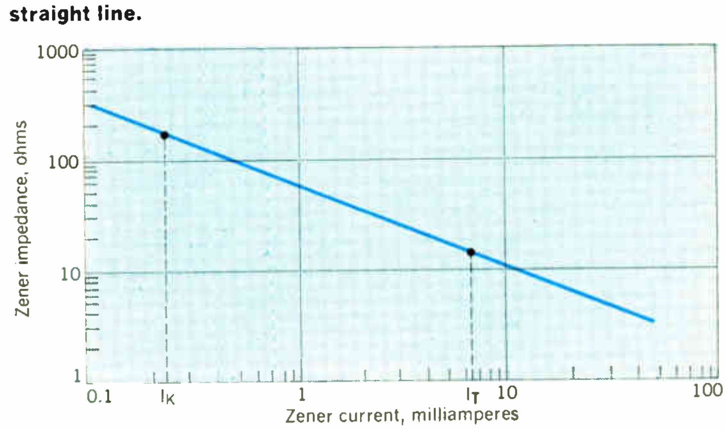


FIGURE 12. Representation of modulation scheme of constant-current source. Oscillator supplies additional (alternating) current I_S, adding to fixed (direct) current I_P. Additional current causes potential at point P to vary; output current is held proportional to potential.

FIGURE 13. Zener-diode impedance vs. diode current for a 20-volt, 400-mW IN968 diode. Two test currents (I_K and I_T) are shown on plot. I_T is quarter-power point; I_K is value in knee region selected so that a plot of impedance between I_T and I_K on a log-log scale is approximately a straight line.



Other applications

Constant-current sources are also suitable for a number of applications not specifically discussed in this article. These include:

Component testing

1. Electrolytic capacitors: Value, build factor
2. Relays: Production test, pull-in and dropout currents
3. Meters: Accuracy, movement freedom, mid-scale linearity, temperature coefficient
4. Potentiometers: Effective running resistance
5. Operating and testing IMPATT diodes

Electrochemical applications

1. Coulometric titration
2. Electrogravimetry
3. Precision electroplating
4. Chronopotentiometry
5. Electrode kinetics

Miscellaneous applications

1. Measuring contact resistance
2. Temperature measurement (ΔR method)
3. Precision specific-heat measurement

(Information on these applications is contained in Applications Note AN-128, available by letterhead request to: Richard Gooding, Marketing Services Department, Hewlett-Packard Co., 100 Locust Ave., Berkeley Heights, N.J. 07922.)

without measurement because it is a known percentage of I_C. An oscilloscope connected to the meter terminal of the base-current source will indicate v_{be}; connecting the oscilloscope to the meter terminal on the collector-current source will indicate v_{ce}. Knowing these three ac values, $h_{re} = (v_{be}/v_{ce})$ and $h_{oe} = (i_c/v_{ce})$ can be readily calculated.

Figure 16 shows a method for measuring the second pair of parameters, the principal feature of which is the use of a constant-voltage source to provide the required ac short circuit on the transistor output.

The dc and ac collector currents are measured by connecting a differential oscilloscope across resistor R_2; the value of R_2 should be the smallest possible value that will permit measurement of i_c. For example, with an oscilloscope having a sensitivity of 100 μV/cm and an i_c of 100 μA, R_2 would be 1 ohm. The value of R_2 detracts directly from the accuracy of the h_{fe} and h_{ie} measurements, since as R_2 becomes larger, the transistor output is no longer an ac short circuit. The exact value of the error will vary depending upon the magnitudes of R_2, v_{ce}, i_b, and the parameters themselves, but for the typical values given above (R_2 = 1 ohm) and in Table I, the error will be less than one percent.

The ac base current is measured by connecting a differential oscilloscope across R_1. However, since R_1 only adds to the already high output impedance of the current source, its presence has no effect on the accuracy of the parameter measurements. The value of R_1 should be selected for convenience in measuring i_b; with an oscilloscope of 100-μV/cm sensitivity and an i_b of 1 μA, R_1 is required to be at least 100 ohms. In the actual setup for the second pair of parameter measurements, the constant-voltage source is set for the specified value of V_{CE}, and the constant-current source is adjusted to obtain the specified value of I_C. The modulation percentage used in this pair of measurements is, as in the other pair, usually less than 10 percent. Once the three required ac values (i_b, i_c, and v_{be}) are determined, $h_{fe} = (i_c/i_b)$ and $h_{ie} = (v_{be}/i_b)$ can be readily calculated.

Operating Hall-effect devices. Another application in the semiconductor area is that of operating Hall-effect devices. Simply stated, the Hall effect is the generation of a voltage across opposite edges of an electrical conductor carrying current and placed in a magnetic field (see Fig. 17). The basic Hall-effect equation is

$$V_H = wR_H\,(\mathbf{j} \times \mathbf{B}) \qquad (4)$$

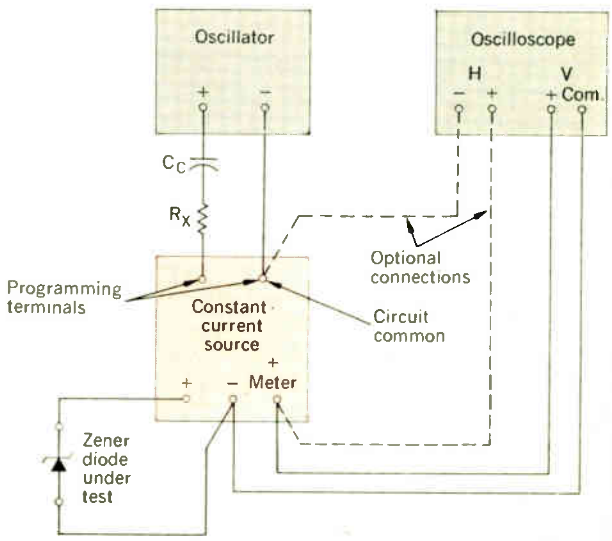
where V_H is the Hall output voltage, w is the width of the Hall plate, R_H is the Hall coefficient, $\mathbf{j}$ is the current density through the Hall plate, and $\mathbf{B}$ is the magnetic field strength. A more useful equation for practical Hall generators is

$$V_H = Y_{IB}\,(\mathbf{I}_C \times \mathbf{B}) \qquad (5)$$

where $\mathbf{I}_C$ is the control current and Y_{IB} is the open-circuit sensitivity, a constant taking into account the effects of geometry and other factors.[4] This equation indicates that the product of two inputs results in a voltage output. If either input is zero, the output is zero. If one input is held constant, the output is proportional to the other input.

An obvious application of the Hall effect is to measure magnetic field strength by holding $\mathbf{I}_C$ constant and measuring V_H. Since the Hall output voltage is directly proportional to the current times the field strength, the current must be held absolutely constant if high accuracy in the measurement is to be obtained—thus the necessity for a precision constant-current source. The direction and polarity of the field can be determined as part of this measurement: the output voltage polarity shown in Fig. 17 will reverse if the direction of the magnetic field reverses, and the output voltage will be a maximum when the lines of force are perpendicular to the plane of the element. A dc control current may be used for measuring

FIGURE 14. Test setup for measuring Zener-diode incremental impedance. An optional connection allows the oscilloscope to display line whose slope represents the incremental impedance.



ac fields, where the Hall output is an ac voltage at the frequency of the field with its magnitude proportional to the instantaneous value of flux density. Applying this output to an oscilloscope will allow observation of the true waveshape of the ac field.

In actual measurements of the magnetic field strength, the Hall output voltage is usually amplified with a differential amplifier in order to eliminate the zero-field residual voltage (typically between 150 μV and 1.5 mV). Hall generators are available with open-circuit sensitivity constants (Y_{IB}) from 20 mV/G to 300 mV/kG, allowing measurement of a wide range of field intensities.

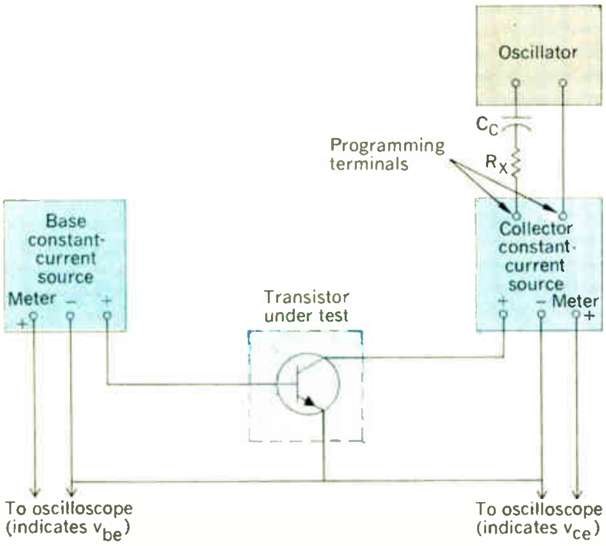
Additional Hall-generator characteristics affecting the accuracy and resolution of magnetic field intensity measurements are the temperature coefficient, the linearity, and the physical dimensions. Since a typical temperature coefficient of the Hall output voltage is 0.06 percent per degree C, a limitation is placed upon the accuracy that can be achieved without control of the ambient temperature. Hall generators with high basic linearity and individual linearity deviation curves are available to combat the second source of error (typical linearity specifications are 0.25 percent of reading from -10 kG to $+10$ kG, and 1.0 percent of reading from -30 kG to $+30$ kG). To achieve the highest possible resolution, the Hall element should be as small as possible in order to minimize output variation due to unequal sensitivity over the surface area. Active areas as small as 10 by 20 mils (50 by 100 μm) are available.

Other applications of a Hall generator in combination with a constant-current source are illustrated in Fig. 18. Figures 18(A), (B), and (C) illustrate the use of the Hall generator as a linear or angular displacement transducer. Slight movement in the direction of the heavy arrows in (A) will produce a large positive or negative output; movement in the direction of the heavy arrow in (B) will produce approximately a linear output versus displacement; and movement in the direction indicated

I. Common-emitter h-parameters

Parameter	Alternate Symbols	Definition	Typical Value
h_{ie}	h_{11}, r_{n}	Input impedance (v_{be}/i_{b}), output short-circuited	2 kΩ
h_{re}	h_{12}, μ	Reverse-voltage amplification factor (v_{be}/v_{ce}), input open-circuited	5×10^{-4}
h_{fe}	h_{21}, β	Forward current gain (i_{c}/i_{b}), output short-circuited	100
h_{oe}	h_{22}, g_{o}	Output admittance (i_{ce}/v_{ce}), input open-circuited	10 μmho

FIGURE 15. Measurement setup for determining open-circuit h-parameters h_{re} and h_{oe}. Base constant-current source supplies dc base current and acts as open circuit to ac component of base current. Collector constant-current source supplies both dc and ac collector currents.



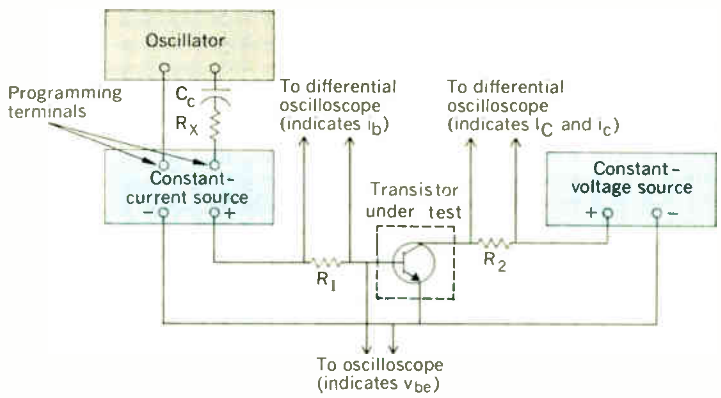
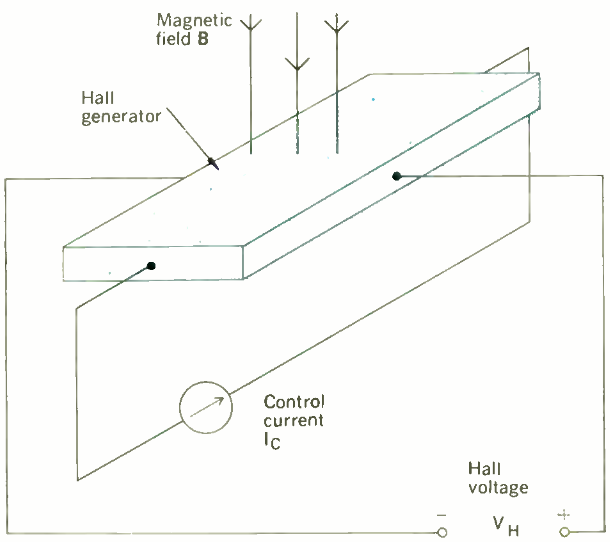


FIGURE 16. Measurement setup for determining open-circuit h-parameters h_{ie} and h_{fe}. Constant-voltage source supplies dc collector–emitter voltage and acts as short circuit to ac component of collector current; constant source supplies dc base current.

FIGURE 17. The Hall effect. The generation of a voltage across opposite edges of an electric conductor carrying current and placed in a magnetic field. Effect can be used to measure magnetic field intensity by holding input current constant and measuring Hall voltage.



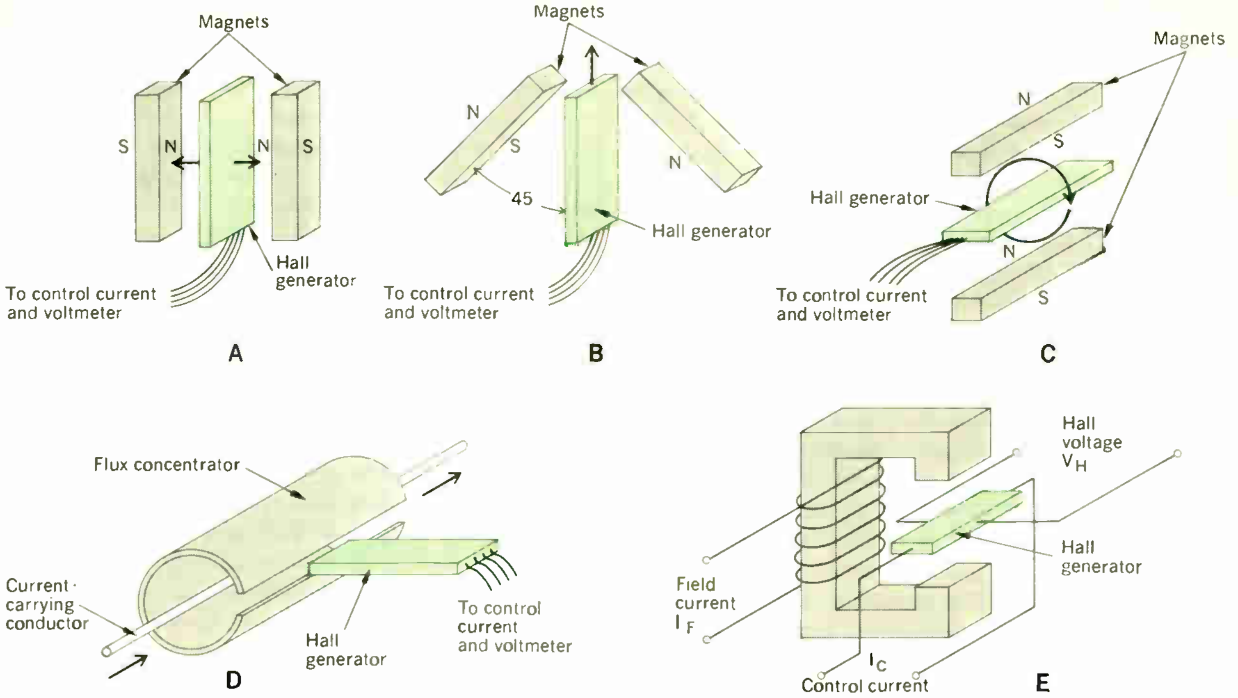


FIGURE 18. A—C: Linear and angular displacement transducer based on Hall effect. Movement in direction of heavy arrows produces output voltage. D: Current measurement utilizing Hall effect. Currents from milliamperes to kiloamperes can be measured with no electrical connection. E: Multiplier using Hall effect. Output voltage is product of two inputs: control current and field current. Either or both inputs may be ac, dc, or combination of ac and dc.

by the heavy arrow in (C) will produce an output voltage that is a sine function of the angle between the plane of the element and the direction of the magnetic lines of force.

Figure 18(D) illustrates the use of the Hall generator to measure current (any conductor carrying current has an associated magnetic field around it whose magnitude is proportional to the current). Very small currents may be measured with the use of a flux concentrator as shown; for larger currents, the Hall generator need only be placed adjacent to the conductor. Currents ranging from milliamperes to kiloamperes can be measured in this manner with essentially zero losses and no electrical connection.

A final application of the combination of a Hall generator and a constant-current source is the multiplier shown in Fig. 18(E). The device has two inputs: the control current to the Hall generator (I_C) and the current through the field coil (I_F). Either of these inputs may be ac, dc, or any combination of ac and dc (the constant-current source can be ac-modulated, as described earlier, to produce such a combination signal with minimum effort). The output voltage, as shown in Eq. (5), is the product of the two inputs. The multiplier may be used in such applications as modulators, analog multipliers, power transducers, variable attenuators, frequency doublers, and square-law detectors.

The author wishes to express his gratitude to the following companies and institutions for providing the basic application ideas presented in this article: Burroughs Corp., Carborundum Co., Bell Telephone Laboratories, F. W. Bell Inc., General Electric Co., CBS Laboratories, IBM Corp., Northrop Nortronics, Polaroid Corp., Princeton University, Rensselaer Polytechnic Institute, RCA Corp., Applied Nuclear Corp., Hamco Machine and Electronics Corp., and Solid State Scientific, Inc.

REFERENCES

1. Smits, F. M., "Measurement of sheet resistivities with the four point probe," *Bell System Tech. J.*, pp. 711–718, May 1958.

2. Combs, J. F., and Albert, M. P., "Diameter correction factors for the resistivity measurement of semiconductor slices," *Semiconductor Products*, pp. 26–27, Feb. 1963.

3. "Zener impedance," in *Zener Diode Handbook*, Motorola Inc., 1967, pp. 4–5.

4. "The Hall effect and its applications," 2nd ed., booklet published by F. W. Bell Inc., Columbus, Ohio.

BIBLIOGRAPHY

Perkinson, J. C., and Pierce, W. C., Jr., "Precision dc current sources," *Hewlett-Packard J.*, vol. 21, pp. 15–20, Sept. 1969.

DC Power Supply Handbook, HP Application Note AN-90A, Hewlett-Packard Co., Berkeley Heights, N.J., 1970.

Geoffrey F. Walker (M) received the B.S. degree in English from the Polytechnic Institute of Brooklyn in 1969. He is currently working toward the B.S.E.E. degree, also from Brooklyn Polytechnic, which he expects to receive in 1971. He has been employed since graduation as a technical writer by the New Jersey Division of Hewlett-Packard. In his present position Mr. Walker is responsible for creating operating and service manuals and other technical support information for newly developed regulated dc power supplies and other products originating from the division. In gathering the basic applications information for this article he conducted interviews with more than a dozen companies and institutions using Hewlett-Packard's CCB series constant-current source.

Inductorless filters: a survey

II. Linear active and digital filters

It is not possible to recommend particular types of inductorless filters, many of which have not yet been proved in actual practice. The choice, of course, will depend upon the application

George S. Moschytz *Bell Telephone Laboratories, Inc.*

In this two-part survey the author has attempted to review the main methods of inductorless filter design being pursued at the present time. These filters are of particular interest because they may provide the replacement for the LC filters that are being eliminated from electronic equipment as a result of the trend toward microminiaturization. This month's installment discusses inductorless filters of both the linear active and digital types.

Linear active filters

Less than five years ago, the prospects for economically feasible filters in any integrated form looked dim.[29] A turning point came in 1965 with the introduction of the first commercially available monolithic, silicon operational amplifier.[30] This amplifier, and the many improved amplifiers that followed in rapid succession at drastically reduced costs, owed much to the ingenuity of R. J. Widlar. Widlar considered integrated circuit design a new discipline that obeyed rules differing significantly from those used for discrete-component circuit design.[31] Thus, today there are many electronic systems with highly sophisticated linear microcircuits operating in industrial, military, and commercial fields that exceed even the most optimistic predictions of less than a decade ago.

Nevertheless, high-Q linear filters have continued to defy monolithic silicon implementation and, in all likelihood, will continue to do so for some time, because of the wide variations of silicon integrated circuit components with temperature and aging, and because the required component stability of active frequency-selective networks is very much higher than that of passive (e.g., LC) networks. It is no wonder then that most active filters considered for or now in production use a hybrid integrated-circuit approach—i.e., the combination of monolithic silicon gain elements with thick- or thin-film RC components.[32-34] These circuits are designed so that the thick- or thin-film passive components determine the Q and frequency stability of the overall networks. In this respect it can be argued that these active filter schemes are based on the characteristics of a component technology such as those discussed in Part I of this article, rather than on circuit design techniques. It takes specialized circuit design techniques, however, to realize active filter networks whose Q and frequency stability are determined primarily by the passive components.

Although the first active filter in the form of a frequency-selective amplifier goes back to the late 1930s,[35] the present widespread interest was triggered some 15 years ago by Linvill, Sallen, and Key.[36,37] The numerous possibilities for active filter design that have been suggested since 1954[38,39] are based on the use of four basically different circuits or devices:

1. The negative impedance converter (NIC).
2. The gyrator.
3. The operational amplifier (op amp).
4. The automatic phase-locked loop (APLL).

The NIC has not been actively pursued recently because of the practical problems encountered in its usage, and so it will not be discussed here. Each of the other circuits offers certain advantages to the filter designer, and they will be described briefly in the following.

Inductance simulation using the gyrator. The gyrator is a two-port device with an impedance-inversion property that converts a capacitive load into a simulated inductive reactance at its input terminals.[40] This is shown qualitatively in Fig. 13. However, only the simulation of a grounded inductor is provided. A floating inductor is obtained with two gyrators as shown in Fig. 14. Inductor simulation using a silicon integrated gyrator combined with a chip or thin-film capacitor comes closest to achieving a microminiaturized inductor and, consequently, microminiaturized versions of equivalent LC filter networks. Herein lies one of the main attractions of this approach; that is, the well-established and tabulated data on the design of LC filters can be directly adapted to gyrator filter design. To convert an LC to a gyrator filter, each grounded or floating inductor simply is replaced by a single- or double-gyrator–capacitor combination, respectively. An alternative to using two gyrators per floating inductor is to use a gyrator operating from a floating power supply. Clearly, this is not a very useful alternative from a practical point of view. Another advantage of the inductor simulation approach is that the inherently low component sensitivity of an LC filter is carried over directly to its gyrator–capacitor equivalent.[41] Furthermore, in contrast to most other active filter methods, gyrator–capacitor filters permit the design of bidirectional filters, just as LC filters do.

Typically, the LC notch filter shown in Fig. 15(A) can be replaced by the equivalent gyrator–capacitor network shown in Fig. 15(B). This equivalence assumes ideal gyrators—which, of course, do not exist physically. Thus, in practice, gyrators have parasitics that effect the desired high input and output impedances. There is also an inherent phase shift within the gyrator loop that can cause instability unless proper compensation is provided. The parasitics limit the Q as well as the frequency range obtainable with gyrator–capacitor filters[42] to the values listed in Table II, which are based on experimental gyrator circuits fabricated in monolithic integrated circuit form as reported in the literature.[43,44] Since integrated gyrators are presently unavailable commercially (although reports of their imminent appearance on the market are not infrequent), it is difficult to relate these numbers to the performance of microminiaturized gyrator–capacitor filters built with mass-produced commercial units, if such exist.

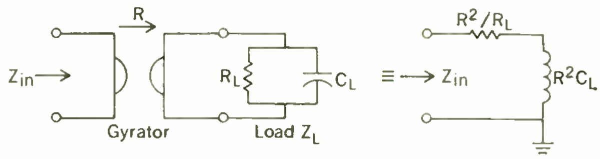
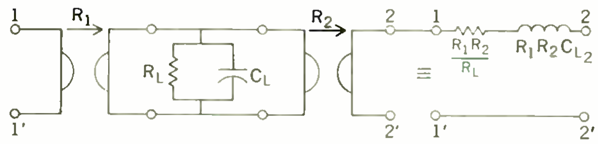


FIGURE 13. Simulation of a grounded inductor by a gyrator–capacitor combination; R = gyration resistance.

FIGURE 14. Simulation of a floating inductor by a double-gyrator–capacitor combination.



Numerous publications[45–49] that describe gyrators built with discrete components or multiple chips indicate that the performance ultimately to be expected from gyrator–capacitor filters may exceed the frequency and Q ranges in Table II. Frequency ranges up to tens of megahertz and Q values in the thousands have been reported for these circuits, and it very well may be that once an integrated gyrator is finally available commercially, it will live up to expectations.

On the other hand, the problems involved in designing monolithic gyrators are not trivial. The circuits suggested so far either have required matching p-n-p and n-p-n transistor pairs with high current gain (neither are easy to make, particularly for high frequencies) or combinations of n-p-n transistors and FETs (the latter also being incompatible with operation in the megahertz range). To

II. Characteristics of hybrid integrated active filters*

Type	Frequency Range	Maximum Pole Q	$\frac{\Delta f}{f}$, %	$\frac{\Delta Q}{Q}$, %	Functional Versatility	Complexity	Power Consumption	Signal Dynamic Range, 20 dB S/N	Compatibility with HIC Technology
Gyrator[a]	0–350 kHz	500	0.2–0.8	10–25	Excellent (includes bidirectional circuits)	Fair (two gyrators or floating gyrator per floating inductor)	(Not reported)	(Not reported)	Fair (may require FETs or high-beta p-n-p transistors)
Single-loop[b] op amp	0–100 kHz	20[c]	0.36–0.5[d]	5	Good	Good	Good (one op amp)[e]	50 dB[e]	Good
Multiloop[b]	0–100 kHz	500	0.36–0.5[d]	5	Good	Fair	Fair[e]	50 dB[e]	Good
APLL[a]	0.1 Hz–25 MHz	400	<0.1%/°C		Poor (special-purpose FM filter and discriminator)	Fair	Fair	40–60 dB	Good

* The characteristics mentioned here may be mutually exclusive.
Notes:
[a] Experimental silicon monolithic.
[b] Op amp: SIC: RC network: tantalum integrated circuit (TIC).
[c] Below 10 kHz Q_{max} for $\Delta Q/Q$ = 5 percent is 50; from 100 kHz to 1 MHz, Q_{max} is 5.
[d] 0.36 percent: 5-year life, 10° to 60°C, 95 percent humidity.
0.5 percent: 20-year life, 0° to 60°C, 95 percent humidity.
[e] Depends on op amp.

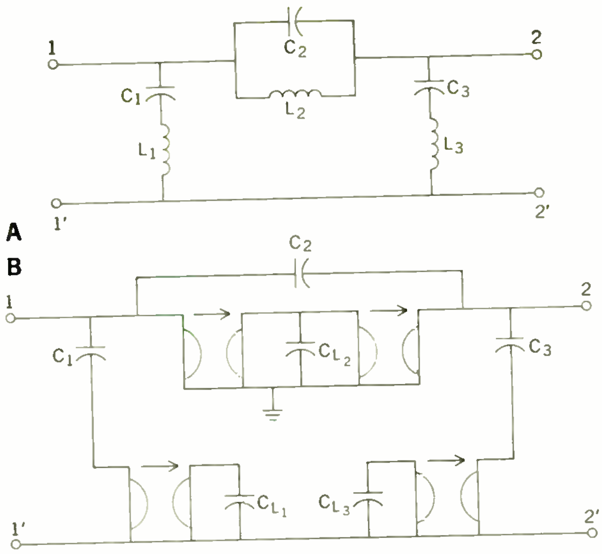


FIGURE 15. LC notch filter (A) and its gyrator-capacitor equivalent (B).

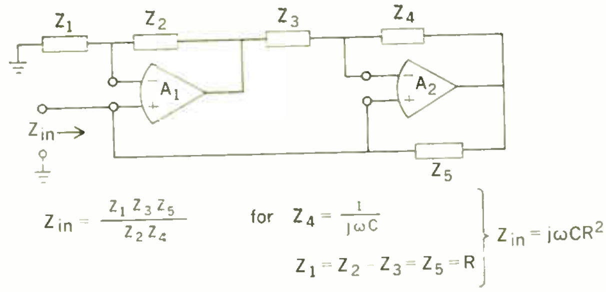


FIGURE 16. Gyrator simulation using two differential amplifiers.

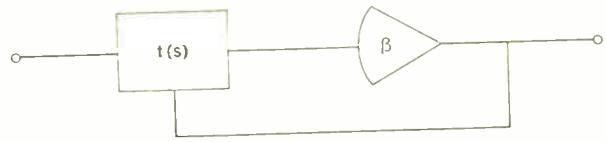


FIGURE 17. Basic single-loop feedback structure using one operational amplifier.

operate up to these frequencies, the gyrator must be designed entirely with n-p-n transistors, which results in biasing problems, or with two chips, one of which contains n-p-n, the other p-n-p transistors. This is uneconomical. Thus, there seems to be an increasing tendency to look to other methods of designing gyrator filters using available linear monolithic integrated circuits.

The most common of the aforesaid circuits is the operational amplifier; two are required to simulate a gyrator. Excellent results have been reported with the configuration shown in Fig. 16.[50] It has been argued that with cheap, monolithic operational amplifiers available, the need for monolithic gyrators has diminished. However, since two amplifiers are required for every gyrator, the power consumption for gyrator filters is doubled. The filter shown in Fig. 15, for example, would require eight operational amplifiers; this scheme does not seem economical, either on the basis of device count or of power consumption. Nevertheless, the availability of low-cost monolithic operational amplifiers makes a strong case for their use in active filters. It is not surprising that the greatest efforts and the most progress in active filter design have included the operational amplifier in one way or another, with the most common methods involving configurations consisting of either positive or negative, single or multiple feedback loops. These will be described in the next section.

Active filters using the operational amplifier. As discussed previously, the high quality and low cost of monolithic integrated operational amplifiers provide an enormous incentive for the development of the active filters incorporating them. It seems unlikely that any other linear integrated devices, such as gyrators, ever will attain the cost advantage of operational amplifiers, since the latter can be used in so many other applications in addition to active filters. Indeed, operational amplifiers had reached their present low cost long before they were incorporated into active filters on any large scale. By contrast, the demand for integrated gyrators, when they are available, will be limited to linear filter applications—a much more restricted market.

In general, active filters using the operational amplifier are realized as a cascade of second-order filter sections. These sections are designed to have high input and low output impedances so that they can be cascaded without additional buffer stages. Cascade design of this type leads to lower sensitivities and permits independent tuning of individual sections because of the lack of interaction between sections. It also has been found that the order of a filter as compared with its *LC* equivalent can thereby be reduced. Consequently, most methods considered for practical implementation today are based on the realization of cascadable second-order filter sections.[51,52] The most commonly used methods require either low- or very-high-gain voltage amplifiers in the inverting or noninverting mode; either can be realized by commercially available monolithic operational amplifiers. To provide filter sections with a low output impedance, the configurations are chosen so that the filter output coincides with the output terminal of one of the amplifiers.

To be economically competitive, active filters must employ integrated-circuit-processing techniques and, for the sake of filter stability, *hybrid* integrated circuit techniques must be used. A useful figure of merit for the bandwidth or Q stability of hybrid integrated circuits using operational amplifiers has been shown to be the product of gain and Q sensitivity; the frequency stability is determined by the tracking capabilities and the temperature-coefficient matching of the resistors and capacitors.[53] The active filter methods considered in the following have been designed using hybrid integrated circuit (HIC) techniques consisting of either thick-film resistors and ceramic-chip capacitors combined with monolithic silicon operational amplifiers or thin-film tantalum integrated-circuit (TIC) resistors and capacitors combined with beam-leaded monolithic silicon operational amplifier chips. Both the single-loop and multiple-loop feedback schemes are well suited to either variation of HIC technology.

Single-loop feedback schemes using the operational amplifier. The method of inductorless filtering described in this and the following sections differs fundamentally from

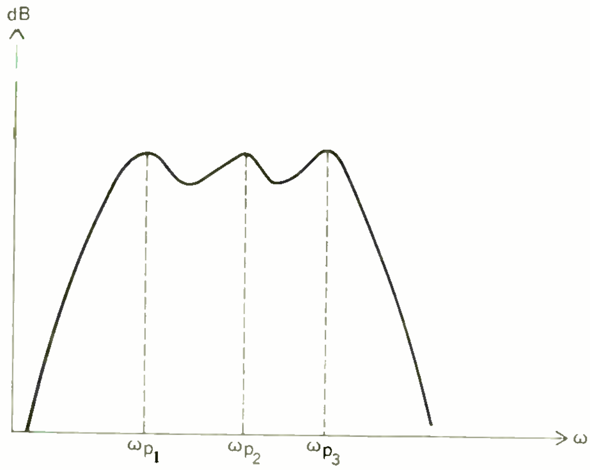


FIGURE 18. Frequency response of sixth-order stagger-tuned bandpass filter.

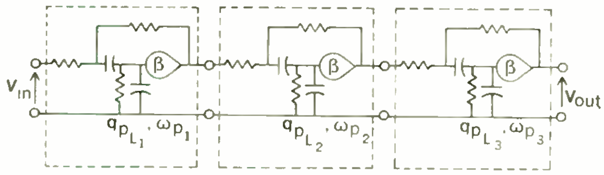


FIGURE 19. Single-amplifier realization of low-Q bandpass filter shown in Fig. 18.

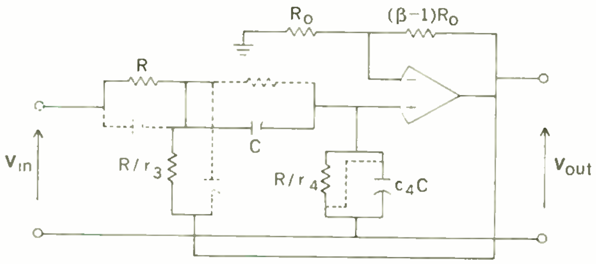
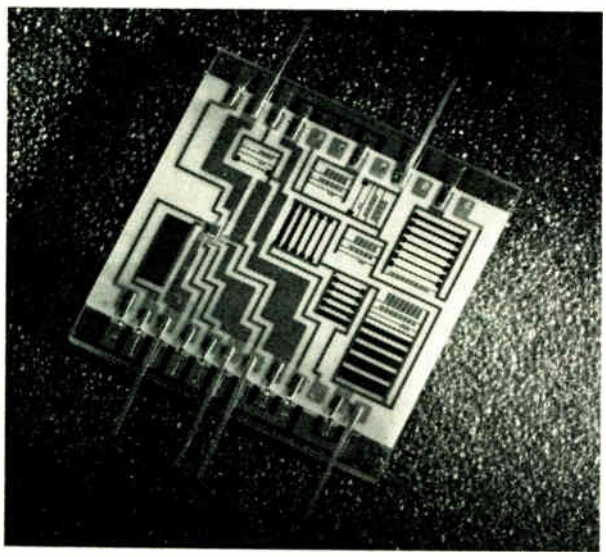


FIGURE 20. Circuit diagram of universal second-order filter building block.

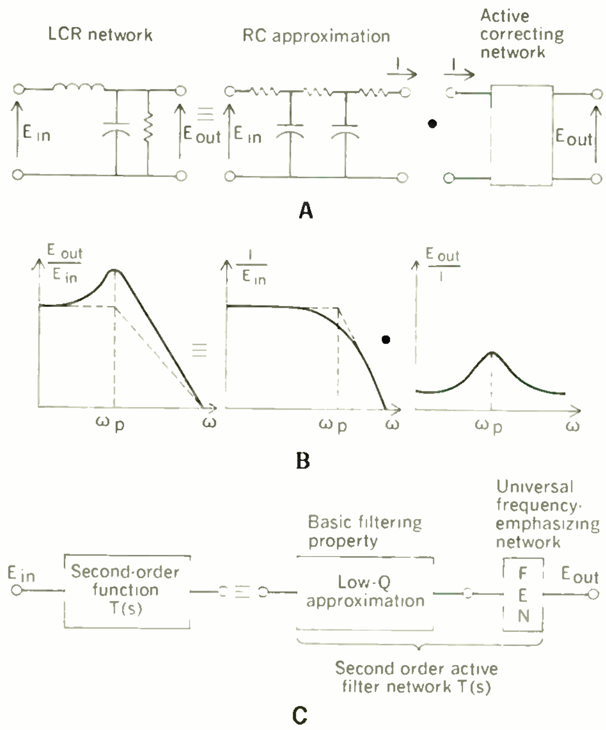
FIGURE 21. Hybrid integrated realization of filter building block whose circuit is shown in Fig. 20.



any of the preceding schemes in that a network *transfer function*, rather than a *simulated LC network* capable of realizing this transfer function, is to be designed. Thus, any one-to-one relationship between inductors and capacitors and their active equivalents disappears. Instead, the objective is to find feedback networks in which operational amplifiers are combined with passive *RC* networks that will provide the transfer functions desired.

The single-loop feedback configurations are essentially single-amplifier structures of the kind shown in Fig. 17, consisting of an amplifier (β), which is either in the positive or negative mode, and a passive *RC* network $t(s)$. By appropriate use of positive or negative feedback, the negative real poles of $t(s)$ are shifted into the s-plane to provide the desired conjugate complex pole pair. To achieve this pole shift, the feedback loop must have a prescribed frequency response, thus the passive *RC* network $t(s)$ must provide a low-pass, high-pass, or band-rejection frequency response within the feedback loop when β is negative and a bandpass response when β is positive. The band-rejection response can be obtained by a bridged- or twin-T network; the other response by *RC* ladder networks. The desired input–output transfer function is then obtained by inserting the input (voltage) signal source into the feedback network in such a way as not to change the basic frequency response in the feedback loop while at the same time providing the desired overall response. In general, the positive-feedback types require low gains and are useful up to reasonably high frequencies, but the Q sensitivities

FIGURE 22. Decomposition of second-order network function into low-Q approximation and active correcting network. A—LCR low-pass network decomposed into passive RC and active correcting network. B—Corresponding frequency responses. C—Generalized decomposition into low-Q approximation and universal frequency-emphasizing network (FEN).



are high (namely, proportional to Q). By contrast, the negative-feedback schemes require high gain and are more frequency-limited but exhibit low Q sensitivity. Either way, the gain–sensitivity product limits single-amplifier second-order filter sections to low-Q applications.

Of numerous available single-amplifier schemes, it has been shown in Ref. 53 that the positive-feedback Sallen–Key[37] circuits in hybrid integrated form are superior to various high-gain negative-feedback equivalent circuits, despite their high-Q sensitivity. These circuits also can be used for higher frequencies than the negative-feedback versions because of the low gain required by the amplifier.

The performance capabilities of hybrid integrated single-amplifier filters using thin-film RC components and silicon integrated operational amplifiers are summarized in Table II. In spite of the fact that these circuits can be recommended only for low and medium Q, they are very useful for many communications systems in which pole Q values no larger than 10 or 15 ever occur.

The power consumption of active filters is a very critical parameter that must be minimized, both for economy and maximum equipment packaging density. It is, therefore, important to use filtering schemes in which the number of amplifiers can be proportionately scaled to the pole Q required. Typically, a sixth-order stagger-tuned bandpass filter with the response shown in Fig. 18 (where the pole Q values are assumed to be low) might have the form shown in Fig. 19. Single-amplifier low-Q filter building blocks have been described in the literature[54,55] in which the components necessary for any second-order function are deposited in thin film on a substrate; all interconnections necessary for any given second-order function are shown in Fig. 20.

To obtain the bandpass sections required in the filter shown in Fig. 19, the components shown by dashed lines in Fig. 20 are simply open-circuited. This approach has been used to design an all-purpose hybrid integrated filter building block, as shown in Fig. 21, that can be modified to any given requirement. The approach is very similar to that taken in the field of digital integrated circuits, where complex systems are broken down into small families of logic building blocks for high production quantities and resulting low costs. In fact, such a building-block technique in integrated active filter design is typical of the methods in current use. It is also utilized in the multiloop feedback design approaches described in the next section.

Multiloop feedback networks using operational amplifiers. It can be shown that the gain–sensitivity product of a single-amplifier second-order filter section increases with increasing Q. Thus, as higher Q is required, more than one amplifier must be incorporated in the feedback loop. One method is based on the decomposition of a second-order network into a low-Q approximation of the desired second-order function and, in cascade with it, an active correcting network. The two networks in tandem provide the required network function. An example of this decomposition is illustrated in Fig. 22. Figure 22(A) shows a second-order LCR network decomposed into a passive RC and an active correcting network. The corresponding frequency responses are shown in Fig. 22(B). Note that the RC approximation accurately provides the asymptotic slope of the required function (in this case, -12 dB/octave), but is greatly overdamped in the vicinity of the natural frequency ω_p. The active correcting network reduces the damping in the vicinity of ω_p as required by the specified network function. It is called a frequency-emphasizing network (FEN) and can be designed as a universal building block, as summarized in Fig. 22(C).

For higher Q values of the order of a few hundred, a low-Q active circuit, such as that shown in Fig. 20, is used as the input network to provide the asymptotic characteristics of the required network function. This is then cascaded with an FEN. The latter uses two operational amplifiers and has been built in hybrid integrated form using integrated passive components combined with silicon-integrated operational amplifier chips, as shown in Fig. 23. Using these building blocks to design medium-Q and high-Q versions of the sixth-order tuned bandpass filter shown in Fig. 18 would require three sections of the kind shown in Figs. 24 and 25 respectively.

The performance capabilities of active filter building blocks using tantalum integrated circuits (TICs) and silicon

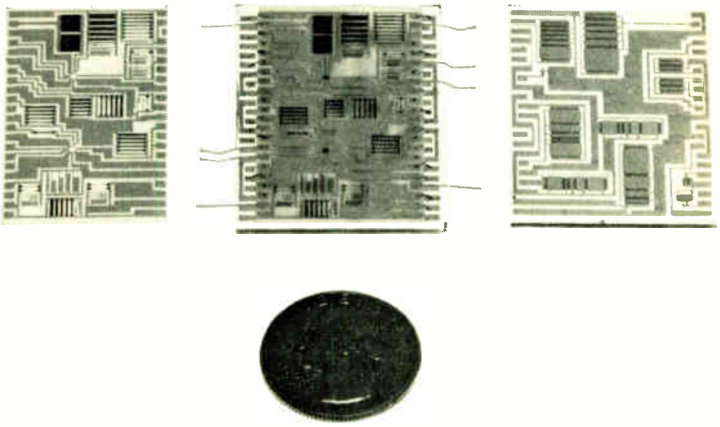


FIGURE 23. Hybrid integrated filter building block using FEN for high-Q applications, showing assembled unit (center), resistor substrate (left), and capacitor substrate (right).

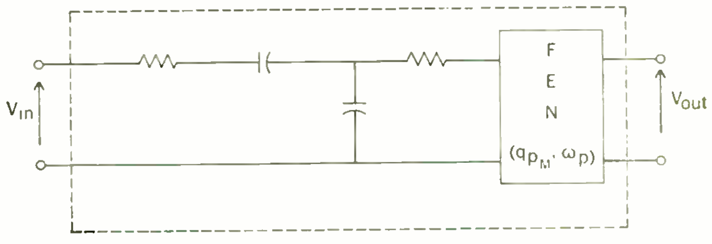
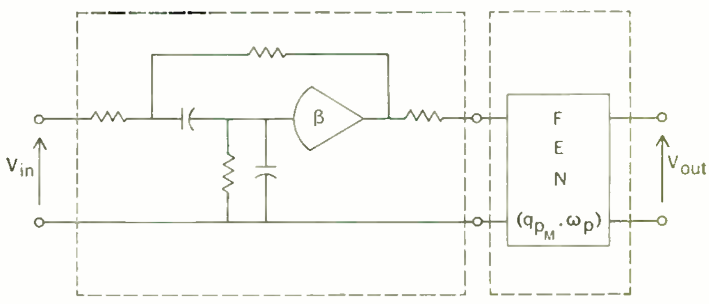


FIGURE 24. Medium-Q bandpass section using FEN.

FIGURE 25. High-Q bandpass section using FEN.



integrated circuits (SICs) are summarized in Table II. Note that apart from the higher Q, the characteristics of the FEN or multiloop networks are the same as those of the single-loop amplifier circuits, since the same technology has been used. The complexity and power consumption of the high-Q circuits (i.e., FENs) are greater. However, by combining the high- and low-Q circuits appropriately, the FEN method has the advantage that it requires increased complexity and power consumption only for high-Q circuits; it can resort to single-amplifier building blocks wherever low pole Q values occur. Excellent network versatility and efficiency result with only one pair of filter building blocks.[55]

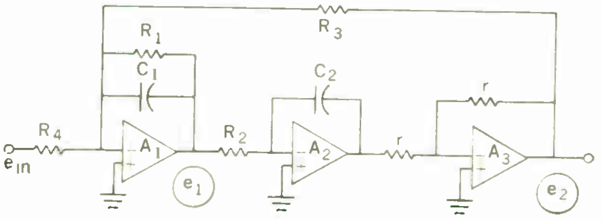


FIGURE 26. Generation of complex pole pair using two integrators and an inverter in the feedback loop.

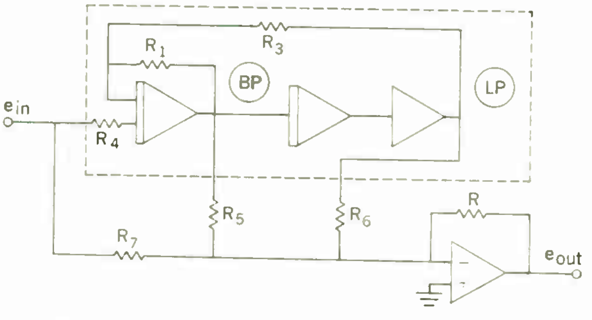
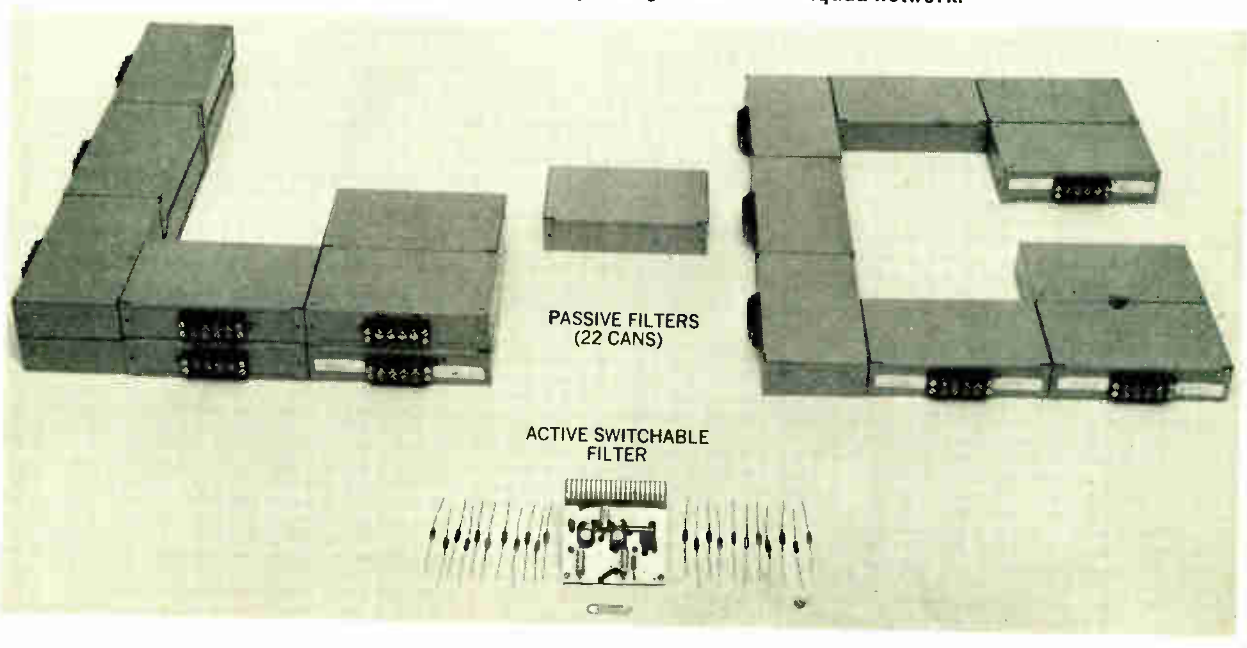


FIGURE 27. Filter building block based on state-variable-design approach.

In many communications systems, particularly those comprising pulse-shaping networks, most pole Q values will be low, say less than 20, and only one or two high-Q poles may be required for pilot-tone pick-off filtering or the like. In such applications, production of the FEN building block (see Fig. 23) for the high-Q poles in addition to the low-Q building block (see Fig. 21) may hardly be justified. Since the FEN itself consists of a low-Q active twin-T network (obtainable from the low-Q building block) in the feedback loop and an inverting summing amplifier in the forward loop, a very simple and economical solution presents itself. The FEN is designed using one of the HIC low-Q building blocks (modified to provide the active twin-T function) in the feedback loop and an external discrete summing amplifier in the forward loop. No determination of frequency stability is thereby incurred since the active twin T that determines this stability is in HIC form. It would seem that, unless the quantities of high-Q poles are very large, this method of FEN filter design would be the most economical; certainly it guarantees the use of *only* low-Q building blocks when designing filter networks requiring low- *and* high-Q poles.

Another commonly used multiloop building-block approach is based on the analog computer simulation of a second-order network—more recently referred to as the state-variable approach.[56] It is a modification of the original scheme in which complex poles were generated by simulating the reactances of a passive LC network by integrators on a one-to-one basis. In the second-order building-block method, two integrators and an inverting amplifier, as shown in Fig. 26, are required to generate a pole pair. With a fourth amplifier outside the pole-generating loop, an all-purpose filter building block is obtained that provides any second-order transfer function, depending on which output terminal or combination of output terminals is used. This building block, referred to as a biquad,[57] is shown in Fig. 27. Its versatility as a

FIGURE 28. Comparison of 22 LC filter cans replaced by a single switchable biquad network.



switching filter has been demonstrated by virtue of the fact that the pole frequency can be varied by switching resistor R_3 (see Fig. 27) while maintaining a constant bandwidth.

The luxury of using four amplifiers for a second-order function was demonstrated to be worthwhile in an application in which 22 passive filters, any one of which was to be switched into a particular system, were eliminated by a single biquad having a 22-position switch connected to the 22 corresponding values of frequency-determining resistors R_3. The comparison between the resulting active filters and the original passive filter is shown in Fig. 28. Clearly, for this particular application even a discrete version of the active filter dramatically outperforms and costs less than its *LC* equivalent. In general, where more mundane applications are required, neither four amplifiers per second-order section nor a discrete component realization can be afforded. This filter type, therefore, has been integrated in hybrid form using thick-film technology for the passive components and silicon technology to provide a three-amplifier chip (see Ref. 32).

Rather than all-purpose operational amplifiers, a custom-designed triple-amplifier chip has been used. Since the amplifiers are intended for low-frequency applications, their bandwidth or frequency capabilities are not critical and they have been specifically optimized for minimum power dissipation. As a result, the three amplifiers dissipate no more than 300 μW to 9.6 mW, depending on the power supply used. Hybrid integrated filters of this kind have been on the market for some time.[28] Because thick-film, rather than the more stable thin-film, components are employed, the obtainable Q and frequency stability are somewhat less than indicated in Table II.

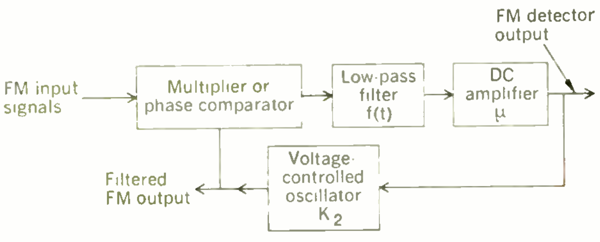
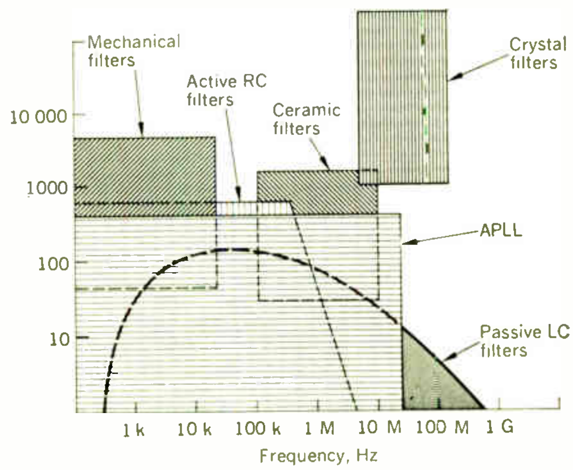


FIGURE 29. Block diagram of automatic phase-locked loop for use as a microminiaturized frequency-selective network.

FIGURE 30. Q versus frequency capabilities of electromechanical and hybrid-integrated linear-active filters.

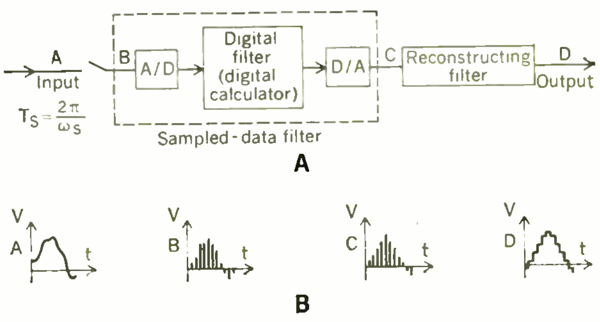


Frequency-selective networks using the phase-locked loop. Although the automatic phase-locked loop has been utilized in communications systems for many years, it was not until the early 1960s that its potential as a microminiaturized frequency-selective circuit was recognized.[59] At that time, the APLL was designed for hybrid integrated implementation. Because linear silicon integrated circuits did not exist, the individual functions required by the APLL, as shown in Fig. 29, were individually designed.

In brief, the APLL provides frequency selectivity for a band of frequency-modulated signals in the following way. Assuming a number of FM signals present at the input to the APLL, the voltage-controlled oscillator locks onto the signal closest to its own center frequency. The two signals (the incoming FM signal and the VCO signal), which are in phase to within a phase error $\Delta\phi(t)$, are passed through a phase comparator and low-pass filter. An error signal at the output, proportional to the phase difference $\Delta\phi(t)$ between the two signals, controls the frequency of the VCO. This error voltage can be considered as the output of a frequency discriminator, since it is a measure of the input frequency. At the same time, the APLL exercises a smoothing effect on the error voltage equivalent to the effect of a second-order bandpass filter with respect to noise and interference from neighboring channels. In other words, if there are a number of FM signals at the input to the APLL, a filtered (and amplified) replica of the signal whose frequency is closest to that of the VCO will appear at the output of the VCO, and a demodulated and filtered version of this signal will be available at the VCO's input.

The importance of the APLL for integrated circuit design is that, first, its individual functions can be designed in active *RC* form. Second, because of the locking property of the APLL, which means that its frequency tracks or locks on to the input signal to be filtered and detected, the stability of the overall circuit is not as critical as that of other active filters. Thus, where it is inconceivable that the filter schemes described in the previous sections could be manufactured using present-day monolithic integrated-circuit techniques, *this is no longer true of the APLL.* The work described in Ref. 59 was picked up again some years later when the era of linear silicon-integrated circuits had emerged and monolithic integrated phase-

FIGURE 31. Basic diagram of generalized digital filter. A—Block diagram including input sampler and output reconstructing filter. B—Typical waveforms at points A through D.



locked loops have been successfully designed.[60]

By nature, the APLL is limited to applications requiring the selection or filtering of one of a number of FM signals. In contrast to this limited functional versatility, it can be used over a very wide frequency range merely by changing the center frequency of the VCO. It is, and will remain, a special-purpose circuit, which, however, may find its way into some of the same consumer communications systems for which numerous electromechanical filters are being developed. One possible application is a high-frequency (1- to 25-MHz) FM amplifier–demodulator circuit, which is designed as a monolithic replacement for the IF strip and the detector section of commercial FM receivers. In another application, the APLL has been used as a monolithic FM multiplex receiver circuit for industrial applications in a frequency range of 2.1 Hz to 300 kHz. The attraction of the APLL for these applications is the fact that it can be used as a combined channel filter and FM discriminator. Tentative performance characteristics of the device as a monolithic integrated bandpass filter, based on the experimental results reported in Ref. 60, are listed in Table II.

A graphical summary of the Q versus frequency capabilities of the various filters listed in Table I (see Part I in the August issue) and Table II is shown in Fig. 30.

Digital filters

As discussed previously, the original analog computer method of active filter design uses adders, subtractors, multipliers, and integrators to simulate and solve the differential equations defining, for example, a second-order network function. Digital filters do exactly the same thing in the digital domain, and so perhaps a more accurate name for them would be *numerical* filters. This nomenclature becomes clearer if we look at the diagram of a generalized digital filter,[61] as shown in Fig. 31(A). It simultaneously covers two basic types of filters that partially or wholly rely on digital circuit techniques, the sampled-data filter and the digital filter. Both types require a signal at the input that is quantized in time. This is achieved by a sampler at the input. However, where the sampled-data filter accepts a time-quantized signal with continuously varying amplitude at its input, the digital filter requires its input signal to be quantized in amplitude as well. The latter is accomplished by an analog-to-digital converter preceding the digital filter. Clearly, in an all-digital system in which no continuous signals occur, A/D and D/A converters are not required and the sampled-data and digital filters are synonomous. Thus, it becomes a question of where in a system the A/D and D/A converters are put as to which filter is used. The closer to the actual input of a system receiving continuous signals the converters can be located, the more of this system can be implemented with digital circuit techniques—including, of course, the digital filtering. In the case of continuous signals, either sampled-data or digital filters can be applied. The former do not require A/D converters since they have the freedom of processing the incoming sampled signals by analog or digital means (in the latter case, of course, the A/D converters are present). Linear processing, in which active *RC* networks resembling those discussed under "Linear Active Filters" can be used, may be more economical when low sampling rates are sufficient for the time-quantization of the incoming signals. In this case, the capabilities of the high-speed circuitry generally employed in digital filters would not be utilized fully.

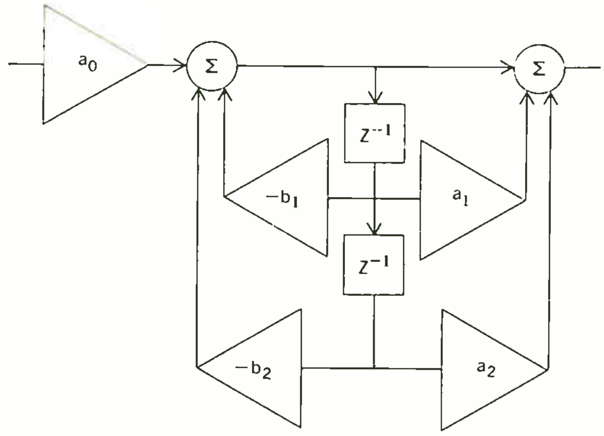
Let us return to the sampled-data filter incorporating a digital filter as its processor, which is shown in Fig. 31(A). At time $t = nT_s$, the continuous input signal is momentarily sampled, and pulse $x(nT_s)$ appears at the input to the sampled-data filter. Thus, the input (and output) signals are narrow amplitude-modulated pulses, one pulse per sampling period T_s. In the A/D converter, each pulse amplitude is converted into a *number* in the form of a digital word. This digital word is a coded sequence of binary digits (bits), which represents the amplitude $x(nT_s)$. The length of the word (i.e., its number of bits) determines the accuracy of the numerical representation. The digital calculations are performed with these words in the digital filter, which is nothing but a special-purpose digital calculator, and the calculator output word is inserted into a D/A converter to produce the output pulse $y(nT_s)$ of the sampled-data filter. The processing, which was formerly carried out continuously using analog computer techniques to simulate a given network transfer function (e.g., state-space approach), here is carried out numerically by circuits similar to those in a digital computer. A reconstructing filter follows the sampled-data filter to convert the pulse stream into a continuous output signal. A holding circuit often is used for this purpose, as shown in Fig. 31(B). Further analog filtering may be desired to remove the signal components resulting from the step approximations caused by the A/D and D/A conversions.

The operation of a digital filter is represented by a difference equation. This equation defines the output pulse amplitude $y(nT_s)$ as a function of the present input pulse $x(nT_s)$ and any number of past input and output pulses. The operations corresponding to the difference equation are performed in the digital calculator with words representing the required past input and output pulses being stored in temporary memory. These operations can be performed in either serial or parallel form. The usual practice is to simplify the notation to x_n and y_n with the understanding that n refers to the time $t = nT_s$. Thus, a general formula for the difference equation is

$$y_n = \sum_{k=0}^{m} a_k x_{n-k} - \sum_{k=1}^{n} b_k y_{n-k} \qquad (2)$$

This expression describes the algorithm performed by the

FIGURE 32. General second-order digital filter section.



digital filter within the sampled-data filter of Fig. 31(A), whereby a sampled signal or sequence of numbers $x(nT_s)$ or x_n acting as an input is transformed into a second sequence of numbers y_n termed the output signal.

Analysis of digital filters is most conveniently carried out with the use of the z-transform where z^{-1} is equal to e^{-sT}, the unit delay operator. Some examples of digital filters include discrete transversal filters ($b_k = 0$), differentiators, discrete recursion filters ($b_k \neq 0$), digital low-pass or smoothing filters, interpolating filters, digital band-pass filters, and spectrum-shaping filters. Specifically, if y_n is a function of only the present and past *input* pulses (i.e., $b_k = 0$), the filter is termed nonrecursive; if the past output pulses are also included (i.e., $b_k \neq 0$), then the filter is of the recursive type. Experimental work on both types has been reported.[62–64]

The implementation of digital filters becomes clearer if we apply the z-transform to Eq. (2):

$$y(z) = x(z) \sum_{k=0}^{m} a_k z^{-k} + y(z) \sum_{k=1}^{n} b_k z^{-k} \qquad (3)$$

The similarity between (2) and (3) is obvious. Equation (3) shows that past values of the input and output signals are simply multiplied by coefficients a_k and b_k and delayed an appropriate amount z^{-k}, where k denotes the number of sampling periods that the pulse or word has been delayed.

The choice of the scaling or weighting coefficients (a_k and b_k) determines the performance of the digital filter. In effect, these coefficients play the same role as the coefficients in the transfer function of a continuous system. In fact, they can be derived from the corresponding continuous transfer function in a number of ways. This involves various transforms that convert an analog transfer function in s into a corresponding digital transfer function in z. These transformations consist of a set of rules for converting from linear continuous operation (summing, scaling, integration) to linear discrete operation (summing, scaling, delay). Conventional analog memory (capacitance, inductance) thereby yields its digital equivalent in the form of shift registers.[65]

It has been shown[66] that coefficient accuracy requirements are least in digital filters when the order of the system is smallest. As a result, and in complete analogy to the continuous active filter case, the second-order system has been used as a basic building block for the synthesis of higher-order systems because, as in the linear case, it is the lowest-order system that can realize complex poles with real coefficients. The general realization of a second-order digital filter section is shown in Fig. 32. As seen from this configuration, the basic arithmetic operations required by the filter are those of addition, multiplication, and unit delay. Thus, in both the recursive and the nonrecursive designs, one is accumulating products of signal filter-coefficient pairs.

Digital filters by nature are characterized differently than their continuous counterparts, although they may be required to perform identical operations. In spite of the different characterization, for every problem in the continuous field there is an equivalent one in the digital field, and vice versa. For example, where in the continuous field frequency stability was shown to depend upon the accuracy and matching of passive components, it can be shown here to depend on the accuracy of the sampling frequency and the number of significant digits used to define the co-

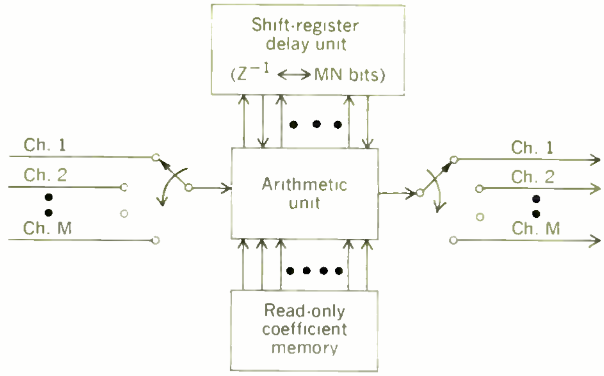


FIGURE 33. Type 1 multiplexing for M input channels.

efficients b_k and a_k. Numerous other error sources, such as those caused by the quantization errors resulting from the A/D converters, round-off errors caused by finite word lengths in the digital calculations, aliasing errors caused by signal frequencies higher than half the sampling frequency, etc., must be dealt with. In an all-digital system where A/D conversion is not required, many of these error sources are absent.

What, then, are the advantages and disadvantages of using digital as opposed to continuous or electromechanical filters?

In an all-digital signal-processing system, digital filtering should be considered first. In such systems, the digital operations required to perform the numerical calculations in the filter are similar to those carried out elsewhere in the system; these operations, in turn, are similar to the operations in any digital computer. The economical considerations for the filters are very similar to those valid for most large-scale integrated circuits. These are, briefly[67,68]: (1) The circuits used should be digital, a seemingly obvious requirement, except that analog calculators also have been used but have not proved to be feasible economically. (2) Circuits should be connected in regular arrays so that they are amenable to large-scale integration. (3) The arrays should have few external leads, and only a few types should be required. Based on these criterions, transmission terminal equipment, such as channel banks and tone-signaling units using shift registers, adders, and read-only memories as the principal circuits, have been built.

As pointed out earlier, the three basic operations to be realized in the implementation of a digital filter are delay, addition (or subtraction), and multiplication. For example, with serial arithmetic, the delays (z^{-1}) are realized simply as single-input, single-output shift registers. Realizations of serial adders (subtractors) and serial multipliers have been described in detail elsewhere.[69] The adders and multipliers, including their interconnections, can be said to comprise the arithmetic unit or the numerical calculator of the digital filter. Having realized the three basic digital filter components (delays, adders, and multipliers), the filter itself may be implemented by simply interconnecting these components in a configuration corresponding to one of the digital forms, canonical or otherwise, for the filter. However, if the input bit rate (sampling rate times bits per sample) is significantly below the capability of the digital circuits, the digital filter can be multiplexed to utilize the circuits more efficiently. This multiplexing step greatly enhances the efficiency of the digital hardware. The various multiplexing schemes are of two main types: (1) The multiplexed filter may operate upon a number of input signals simultaneously, or (2) the multiplexed filter may effect a number of (different) filter forms for a single input signal. The first type can also be interpreted as a time-sharing of the calculator part of the digital filter. A combination of the two types is possible.

To multiplex the filter to process M simultaneous inputs (type 1), the input samples from the M sources are interleaved sample by sample and fed (serially) into the filter. The bit rate in the filter is increased by a factor of M, and the shift-register delays also must be increased by a factor of M to a length of MN bits. Otherwise, the filter is identical in its construction to the single-input case. In particular, the arithmetic unit containing the adders and multipliers is the same; it just operates M times faster. The output samples emerge in the same interleaved order as the input and thus are easily separated. Type 1 multiplexing is depicted in Fig. 33.

If the M channels in Fig. 33 are to be filtered differently or if type 2 multiplexing is also employed, the filter coefficients are stored in a separate read-only coefficient memory and are read out as required by the multiplexed filter. A diode matrix provides a very fast and inexpensive form of read-only memory (ROM) for this purpose. If, however, all M channels are to be filtered identically and no type 2 multiplexing is employed, the coefficients may be wired into the multipliers and no ROM is then required. This is indicated by the dotted lines enclosing the ROM in Fig. 33. In this case, adders must be included only in those multiplier bit sections of the arithmetic unit for which the corresponding multiplier bits (a_k) equal one.

In many cases, a number of different but similar filters or subfilters are required for the same input signal. Similarly to active-filter building-block design, higher-order network functions can be obtained by connecting second-order sections in series, in parallel, or in a combination of these. The individual second-order sections are identical in form, differing only in the values for the multiplying coefficients. Type 2 multiplexing refers to the implementation of these different second-order sections with a single multiplexed second-order section, an example of which is shown in Fig. 34. As with type 1 multiplexing, the combining of M separate filters into one multiplexed filter requires that the bit rate in the filter be increased by a factor of M and that shift-register delays (z^{-1}) also be increased by a factor of M to MN bits in length. The coefficients are supplied from the read-only coefficient memory, which cycles through M values for each coefficient during every Nyquist interval. Data are routed in, around, and out of the filter by external routing switches, which are controlled from the ROM.

As an example of type 2 multiplexing, consider the implementation of a 12th-order filter in cascade form using the multiplexed second-order filter in Fig. 34. Here $M = 6$, so the bit rate in the filter must be (at least) $6N$ bits per Nyquist interval. During the first N bits of each Nyquist interval, the input sample is introduced into and processed by the arithmetic unit with the multiplying coefficients (a_1, a_2, b_1, b_2) of the first subfilter in the cascade form. This processing essentially takes only an Mth of the basic sampling period to complete. By delaying the resulting output by $1/M$ times the sampling period, it can be fed

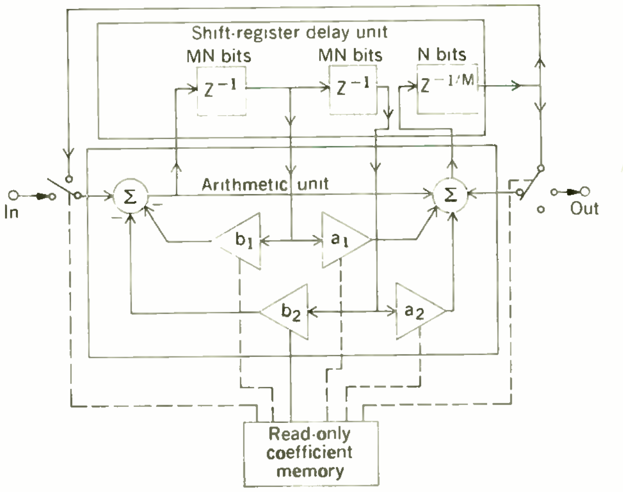


FIGURE 34. General second-order filter for type 1 and type 2 multiplexing.

back via the input routing switch to become the input to the filter during the second N-bit portion of the Nyquist interval. This feedback process is repeated four more times, with the filter coefficients from the ROM changed each time to correspond to the appropriate subfilter in the cascade form. The sixth (last) filter output during each Nyquist interval is the desired 12th-order filter output. The parallel form, or a combination of cascade and parallel filters, may be realized using the filter in Fig. 34 simply by changing the bits in the ROM that control the switching sequences of the input and output routing switches.

In audio-frequency applications, up to 100 filter sections can be multiplexed with a single section using presently available digital circuits in the medium-speed range. This assumes that ten-megabit logic is used, and that the input signal is sampled at 10 kHz and quantized at ten bits per sample, or some combination thereof. The filters are also easily modified to realize a wide range of filter forms, transfer functions, multiplexing schemes, and round-off noise levels by changing the contents of the read-only memory and/or the timing signals and the length of the shift-register delays.

At present, to summarize the capabilities of digital filters as a whole is not an easy task, but it is clear that one of their biggest advantages is the flexibility with which the filter characteristics can be changed, merely by changing the values of the stored filter coefficients (a_k and b_k). Thus, digital filters are being used extensively in laboratory experiments where variable filter characteristics are constantly required—for example, in laboratory tests dealing with optimum filtering and general studies of the human voice. It is difficult to make prognostications with respect to the economics of the integrated digital filter because its success is inextricably linked to the success (i.e., the low cost) achievable with large-scale integrated (LSI) circuit chips. Thus, any predictions for LSI can be extrapolated to digital filters. Once truly large-scale integrated chips are both readily available and inexpensive and their power dissipation is sufficiently low to permit high-density packaging and low-cost power supplies, there is no question as to the potential of digital

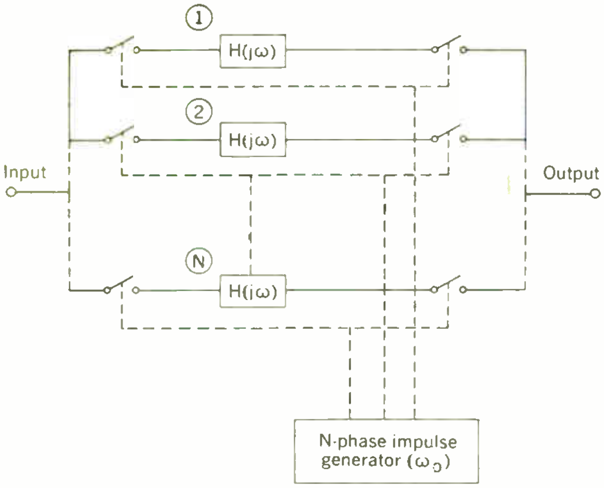


FIGURE 35. Basic form of serial-switched N-path filter.

filters. By merely extending bit lengths, filter accuracy can be improved, as can the Q of a frequency-selective circuit.

Digital filters will be particularly useful in digital systems in which analog-to-digital conversion is not required. In general, their frequency-range capabilities will depend on the capabilities of LSI chips. Certainly, in various types of voice-frequency communications equipment, the number of components will be reduced by more than an order of magnitude, with accompanying cost reductions, through the use of large-scale digital arrays.[68] Specifically, by the use of digital signal-processing techniques, equipment such as data sets, tone receivers, digital channel banks, PCM transmission systems, and even analog channel banks for FDM systems, may be constructed out of shift registers, read-only memories, and serial arithmetic circuits, and thus may be considered as potential digital-filter candidates. Depending on the arrival of true, large-scale integration, these digital versions of their present-day analog counterparts may offer many significant advantages over current implementation techniques. For example, the utilization of circuit techniques, such as multiplexed filters and digital calculator time sharing, can result in considerable additional savings in hardware.

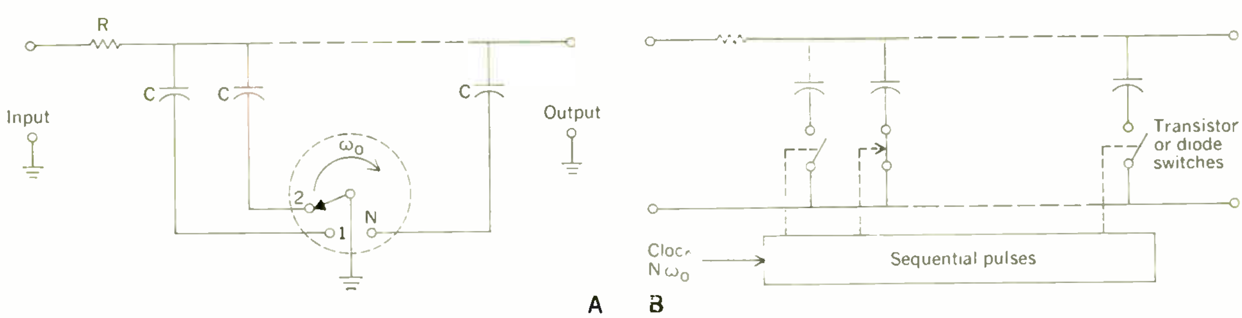
The preceding discussion has been mainly concerned with the implementation of digital filters as defined in Fig. 31. To return briefly to the sampled-data filter that only requires input signals quantized in time, refer again to Fig. 31. The sampled-data filter using a linear processor generally consists of a sampler at the input and output and, instead of the A/D and D/A converters and digital filter in the dotted box, some continuous filter circuitry resembling the active filters previously discussed. A modification of this arrangement (shown in Fig. 35), in which N parallel active or passive RC filters, each with a transfer function $H(j\omega)$, are time-division-multiplexed into the signal path, has been found to provide many useful features. Generally called an N-path filter, it was first described in 1960.[70]

Referring to Fig. 35, the N identical networks $H(j\omega)$ are cyclically switched into the signal path. Thus, the total network has time-variable characteristics. If a low-pass filter is used to provide the $H(j\omega)$ function in each of the paths, the cyclical switching process causes a low-pass to bandpass transformation. The resulting transfer characteristic is symmetrical with respect to the switching or sampling frequency ω_0. For example, if each of the low-pass elements has only a single real pole (i.e., a passive RC section), then the corresponding bandpass function will have a single pair of complex conjugate poles. Analogously, if active RC low-pass sections with conjugate complex pole pairs are used for the $H(j\omega)$ functions, the corresponding bandpass function will have twice as many pairs of conjugate complex poles. The center frequency of the resulting bandpass filter is equal to the sampling frequency ω_0, and the selectivity (Q) or bandwidth depends on the number N of parallel channels and on the poles of $H(j\omega)$. Thus, by varying the sampling frequency ω_0, time-variable bandpass characteristics with constant bandwidth can be obtained.

Various modifications of the basic N-path filter shown in Fig. 35 have been described in the literature. The replacement of the sampling switches by analog multipliers[71,72] enables all operations to be performed by sinusoidal signals and thereby reduces switching transients. It also has the advantages of extending the useful frequency range and reducing the problem of harmonic rejection. With the advent of monolithic integrated analog multipliers,[73] this should also permit a significant reduction in the complexity of implementation.

Another important modification is a rearrangment of the series-switched to a parallel- or shunt-switched N-path filter, which permits the $2N$ floating switches shown in Fig. 35 to be replaced by N switches with grounded terminals. This feature is particularly significant if semi-

FIGURE 36. Shunt-switched N-path RC filter (commutated capacitor filter). A—Basic form. B—Using electronic switching scheme.



conductor (e.g., transistor or diode) switches are to be used. If single-order RC sections are used for $H(j\omega)$, an extremely simple configuration, the so-called commutated capacitor filter shown qualitatively in Fig. 36(A), results. Notice that the N resistors of the N $H(j\omega)$ functions have been combined into one. The corresponding N capacitors, where $N \geq 3$, are sequentially connected into the circuit by a grounded commutating switch. This can be accomplished by digital integrated circuits generating an appropriate sequential pulse stream and by N electronic switches, as shown in Fig. 36(B). This result has been realized by switching transistors,[74,75] and apparently even better by a diode-bridge network.[76] The bandwidth of the resulting bandpass filter equals $2/NRC$, and therefore depends only on passive components. The selectivity or Q equals $(\omega_0 NRC)/2$, and therefore depends on the stability of the clock frequency $N\omega_0$, as does the bandpass center frequency, which is ω_0.

An inherent characteristic of *all* sampled systems is that the transfer response appears not only at the sampling frequency but also at its harmonics. Digital and sampled-data filters—including, of course, serial- and shunt-switched N-path filters—are no exceptions; responses centered at zero frequency, at the sampling frequency, and its harmonics will occur. The response at zero frequency takes place only if energy is applied to the filter at this frequency. The responses at the harmonics of the center frequency ω_0, which always occur, may be filtered out by a low-pass filter or a tuned circuit following the filter itself. In Fig. 31 this is carried out by the reconstructing filter, which must follow the N-path filters of Figs. 35 and 36 in exactly the same way. Furthermore, because the filter samples occur at a frequency $N\omega_0$, which is the clock frequency of the sequential switching network in Fig. 36(B), the input signal spectrum must not exceed $N\omega_0/2$ in order to avoid distortion (i.e., aliasing). Similar reasoning confines the main application of the filter to carrier selection in bands of less than one octave.

N-path filters have been investigated for a variety of applications, some of which would be hard to implement any other way. The multiple-response characteristic has been utilized to advantage for the realization of multiple passband or comb-filter characteristics,[77,78] which otherwise would require the use of structures containing distributed-parameter elements. Third-order shunt-switched N-path filters have been used for the IF section of an integrated AM/FM receiver,[79] where their versatility has proved a great advantage. By utilizing the capabilities of these filters appropriately, a completely new concept for a radio-receiver design permitting an unusually high degree of integration has been demonstrated. Finally, in order to alleviate a basic difficulty encountered in the design of N-path filters, namely, the requirement that the transmission characteristics of the individual paths must be closely matched, an approach resembling the type 2 multiplexing discussed previously (see Ref. 69) has been described.[80] This approach accomplishes the N-path configuration of Fig. 35 by a multiplexed version of the sampled-data filter (including the digital processor) shown in Fig. 31. Thus the time-invariant filter [i.e., $H(j\omega)$] is the digital filter, and both the input sampler and digital filter are time-division-multiplexed. In the digital multiplexed N-path system, the signals in each of the paths are processed by the *same* filter rather than by a set of *similar* filters realizing $H(j\omega)$. Since only one encoder (A/D converter) and one decoder (D/A converter) are used for all of the paths, the only components that are not shared are the output samplers.

Conclusion

It has been attempted in this two-part survey to review the main methods of inductorless filter design being pursued at the present time. Due to the extraordinary diversity of the methods, all aimed at similar objectives, it was realized at the outset that this would be a formidable task and that it would not be possible to do full justice to any one, let alone to all, of the methods discussed. The intent was, therefore, to limit the article to an outline of the basic operation and the most important features of each general filter category. Key references are included that should enable an interested reader to pursue any of the significant methods in more detail.

It will be evident to the reader that, in spite of the similarity in objectives, it is not possible to state a general preference for one or more of the described filtering methods over the others, since the choice depends to a large extent on the application involved. Moreover, many of the methods only recently have emerged from the realm of theoretical analyses to that of practical reality and much more experience with them will be necessary before they are ready to compete in the decisive area of mass production and economical feasibility.

The author wishes to acknowledge the many helpful comments from J. F. Kaiser and W. J. Spencer.

REFERENCES

1. Zverev, A. I., "The golden anniversary of electric wave filters," *IEEE Spectrum*, vol. 3, pp. 129–131, Mar. 1966.

2. Zverev, A. I., "Introduction to filters," *Electro-Technol. (New York)*, pp. 63–90, June 1964.

3. Rand, A., "Inductor size vs. *Q*: A dimensional analysis," *IEEE Trans. Component Parts*, vol. CP-10, pp. 31–35, Mar. 1963.

4. Kosowsky, D. I., "High-frequency crystal filter design techniques and applications," *Proc. IRE*, vol. 46, pp. 419–429, Feb. 1958.

5. Schoeffler, I. D., "Insertion loss design of symmetrical lattice piezoelectric resonator filters with symmetrical and unsymmetrical passbands," *IRE Trans. Circuit Theory*, vol. CT-9, pp. 251–255, Sept. 1962.

6. Humpherys, D. S., "Network synthesis of narrow-band crystal filters," *Electro-Technol. (New York)*, Nov. 1965.

7. Sykes, R. A., and Beaver, W. D., "High frequency monolithic crystal filters with application to single frequency and single sideband use," *Proc. 20th Ann. Symp. Frequency Control*, Apr. 1966.

8. Sykes, R. A., Smith, W. L., and Spencer, W. J., "Monolithic crystal filters," 1967 *IEEE Internat'l Conv. Record*, pt. 11, pp. 78–93.

9. Sykes, R. A., and Smith, W. L., "A monolithic crystal filter," *Bell Lab. Record*, pp. 52–54, Feb. 1968.

10. Spencer, W. J., "Monolithic crystal filters," presented at Third Asilomar Conf., Dec. 1969.

11. Terman, F. E., *Radio Engineers Handbook*. New York: McGraw-Hill, 1943.

12. Humpherys, D. S., "Active crystal filters," *Electro-Technol. (New York)*, July 1966.

13. Sauerland, F., and Blum, W., "Ceramic IF filters for consumer products," *IEEE Spectrum*, vol. 5, pp. 112–126, Nov. 1968.

14. Curran, D., and Gerber, W., "Low-frequency ceramic bandpass filters," *Proc. IEEE Electronic Components Conf.*, pp. 108–113, May 1963.

15. Berlincourt, D., and Sauerland, F., "Ceramic filters," presented at Third Asilomar Conf., Dec. 1969.

16. Sauerland, F., "Identical resonators cut ceramic IF filter cost," *Electronics*, pp. 102–105, May 26, 1969.

17. Tanaka, T., and Inoguchi, T., "A mechanical filter," presented at meeting of Acoustical Society of Japan, May 1957.

18. Börner, M., Kettel, E., and Ohnsorge, H., "Mechanische

Filter für die Nachrichtentechnik," *Telefunken Zeitung*, vol. 31, pp. 105–114, June 1958.

19. Johnson, R. A., "A single-sideband, disk-wire type mechanical filter," *IEEE Trans. Component Parts*, vol. CP-11, pp. 3–7, Dec. 1964.

20. Börner, M., and Schüssler, H., "Miniaturisierung Mechanischer Filter," *Telefunken Zeitung*, vol. 37, pp. 228–246, 1964.

21. Lüder, E., "Mechanische Schwinger in Filtern für Tiefe Frequenzen," *Arch Elektrotech.*, vol. 51, no. 6, pp. 351–357, 1968.

22. Johnson, R. A., Börner, M., and Konno, M., "Mechanical filters: past, present and future," presented at Third Asilomar Conf., Dec. 1969.

23. Baker, H., and Cressey, J. R., "H-shapec resonators signal upturn in tone telemetering," *Electronics*, pp. 99–106, Oct. 2, 1967.

24. Newell, W. E., "Ultrasonics in integrated electronics," *Proc. IEEE*, vol. 53, pp. 1305–1309, Oct. 1965.

25. Newell, W. E., "Miniaturization of tuning forks," *Science*, vol. 161, pp. 1320–1326, Sept. 27, 1968.

26. Nathanson, N. C., Newell, W. E., Wickstrom, R. A., and Davis, J. R., "The resonant gate transistor," *IEEE Trans. Electron Devices*, vol. ED-14, pp. 117–133, Mar. 1967.

27. Newell, W. E., "The resonant gate transistor," *Microelectronics*, vol. 1, no. 1, pp. 22–25, June 1968.

28. Newell, W. E., and Wickstrom, R. A., "The Tunistor: a mechanical resonator for microcircuits," *IEEE Trans. Electron Devices*, vol. ED-16, pp. 781–787, Sept. 1969.

29. Newell, W. E., "Tuned integrated circuits—a state-of-the-art survey," *Proc. IEEE*, vol. 52, pp. 1603–1608, Dec. 1964.

30. Widlar, R. J., "A unique circuit design for a high performance operational amplifier especially suited to monolithic construction," *Proc. Nat'l Electron. Conf.*, vol. 21, pp. 169–174, 1965.

31. Widlar, R. J., "Design techniques for monolithic operational amplifiers," *IEEE J. Solid-State Circuits*, vol. SC-4, pp. 184–191, Aug. 1969.

32. Hollenbeck, D. W., "Multiloop negative feedback active filters using thick film integrated circuit techniques," Paper 4.4, WESCON 1969.

33. Broyle, W. R., "Improved *RC* active filters using integrated circuit/thick film technology," Paper 4.6, WESCON 1969.

34. Moschytz, G. S., Medill, D. G., and Steffen, C. J., "Filter building blocks using tantalum and silicon integrated circuits," *Proc. Electron. Components Conf.*, Washington, D.C., 1969, pp. 100–110.

35. Scott, H. H., "A new type of selective circuit and some applications," *Proc. IRE*, vol. 26, pp. 226–235, Feb. 1938.

36. Linvill, J. G., "*RC* active filters," *Proc. IRE*, vol. 42, pp. 555–564, Mar. 1954.

37. Sallen, R. P., and Key, E. L., "A practical method of designing *RC* active filters," *IRE Trans. Circuit Theory*, vol. CT-2, pp. 78–85, Mar. 1955.

38. Mitra, S. K., "Synthesizing active filters," *IEEE Spectrum*, vol. 6, pp. 47–63, Jan. 1969.

39. Mitra, S. K., "Active *RC* filters," *Proc. Houston Conf. Circuits, Systems and Computers*, May 22–23, 1969.

40. Aaronson, G., "Active filters: part 10, synthetic inductors from gyrators," *Electronics*, pp. 118–125, July 7, 1969.

41. Orchard, H. J., "Inductorless filters," *Electron. Letters*, vol. 2, pp. 224–225, June 1966.

42. Rao, T. N., Gary, P., and Newcomb, R. W., "Equivalent inductance and *Q* of a capacitor-loaded gyrator," *IEEE J. Solid-State Circuits*, vol SC-2, p. 32, Mar. 1967.

43. Chua, H. T., and Newcomb, R. W., "Integrated direct-coupled gyrator," *Electron. Letters*, vol. 5, pp. 182–184, May 1967.

44. Hove, R. G., and Kleingartner, C. A., "Silicon monolithic gyrator using FET's," Paper 4.3, WESCON 1969.

45. Holmes, W. H., Gruetzmann, S., and Heinlein, W. E., "High performance direct-coupled gyrators," *Electron. Letters*, vol. 3, pp. 45–46, Feb. 1967.

46. Sheahan, D. F., and Orchard, H. J., "High-quality transistorized gyrators," *Electron. Letters*, vol. 2, pp. 274–275, July 1966.

47. Haykim, S. S., Kramer, S., Schewchun, J., and Treleaven, D. H., "Integrated-circuit implementation of direct-coupled gyrator," *IEEE J. Solid-State Circuits* (*Correspondence*), pp. 164–166, June 1969.

48. Van Looij, H, T., and Adams, K. M,, "Wideband electronic gyrator circuit," *Electron. Letters*, vol. 4, Apr. 1968.

49. Rao, T. N., "Readily biased wideband gyrator circuit for floating or earthed inductors," *Electron. Letters*, vol. 5, pp. 309–310, July 1969.

50. Riordan, R. H. S., "Simulated inductors using differential amplifiers," *Electron. Letters*, vol. 3, pp. 50–51,, Feb. 1967.

51. Moschytz, G. S., Wyndrum, R. W., "Active filters; Part 5—Applying the operational amplifier," *Electronics*, pp. 98–106, Dec. 9, 1968.

52. Mitra, S. K., "Filter design using integrated operational amplifiers," Paper 4.1, WESCON 1969.

53. Moschytz, G. S., "The gain sensitivity product: a figure of merit for hybrid integrated filters using operational amplifiers," *Proc. Midwest Symp. on Circuit Theory*, Minneapolis, Minn., May 1970.

54. Moschytz, G. S., and Thelen, W., "Design of hybrid-integrated filter building blocks," *IEEE J. Solid-State Circuits*, vol. SC-5, June 1970.

55. Moschytz, G. S., "FEN-filter design using tantalum and silicon integrated circuits," *Proc. IEEE*, vol. 58, pp. 550–556, Apr. 1970.

56. Kerwin, W. J., Huelsman, L. P., and Newcomb, R. W., "State-variable synthesis for insensitive integrated circuit transfer functions," *IEEE J. Solid-State Circuits*, vol. SC-2, pp. 87–92, Sept. 1967.

57. Thomas, L. C., "The biquad: a multipurpose active filtering system," presented at IEEE Internat'l Symp. on Circuit Theory, San Francisco, Calif., 1969.

58. Estep, G., "Universal active filter theory and application," *Applications Manual*, Kinetic Technology Inc.

59. Moschytz, G. S., "Miniaturized *RC* filters using phase-locked loop," *Bell System Tech. J.*, vol. 44, pp. 823–870, May–June 1965.

60. Grebene, A. B., and Camenzind, H. R., "Frequency-selective integrated circuits using phase-locked techniques," *IEEE J. Solid-State Circuits*, vol. SC-4, pp. 216–225, Aug. 1969.

61. Nowak, D. J., and Schmid, P. E., "Introduction to digital filters," *IEEE Trans. Electromagnetic Compatibility*, vol. EMC-10, pp. 210–220, June 1968.

62. Rader, C. M., "On digital filtering," *IEEE Trans. Audio and Electroacoustics*, vol. AU-16, pp. 303–314, Sept. 1963.

63. Kaiser, J. F., "Digital filters," in *System Analysis by Digital Computer*, J. R. Kaiser and F. F. Kuo, eds. New York: Wiley, 1966, pp. 218–285.

64. Kaiser, J. F., "The digital filter—its implementation and application," presented at Digital Filtering Seminar, Newark College of Engineering, Jan. 21, 1970.

65. White, S. A., and Mitsutomi, T., "The IC digital filter: a low-cost signal processing tool," *Control Eng.*, vol. 17, pp. 58–68, June 1970.

66. Kaiser, J. F., "Some practical considerations in the realization of linear digital filters," *Proc. Third Allerton Conf. on Circuit and System Theory*, 1965, pp. 621–633.

67. McDonald, H. S., "Digital filter capabilities," presented at Third Asilomar Conf. on Circuits and Systems, Dec. 1969.

68. McDonald, H. S., "Impact of large-scale integrated circuits on communication equipment," *Proc. Nat'l Electronics Conf.*, Dec. 1968.

69. Jackson, L. B., Kaiser, J. F., and McDonald, H. S., "An approach to the implementation of digital filters," *IEEE Trans. Audio and Electroacoustics*, vol. AU-16, pp. 413–421, Sept. 1968.

70. Franks, L. E., and Sandberg, I. W., "An alternative approach to the realization of network transfer functions: the *N*-path filter," *Bell System Tech. J.*, vol. 39, pp. 1321–1350, Sept. 1960.

71. Rigby, G. A., "An integrated selective amplifier using frequency translation," *IEEE J. Solid-State Circuits*, vol. SC-1, pp. 39–44, Sept. 1966.

72. Geffe, P. R., "Active filters," Tech. Rept. ECOM-9363-5, Westinghouse Electric Corp., Feb. 1969.

73. Bilotti, A., "Applications of a monolithic analog multiplier," *IEEE J. Solid-State Circuits*, vol. SC-3, pp. 373–380, Dec. 1968.

74. Thompson, J., "*RC* digital filters for microcircuit bandpass amplifiers," *Elec. Equipment Eng.*, pp. 45–50, Mar. 1964.

75. Harden, W. R., "Digital filters with IC's boost *Q* without inductors," *Electronics*, pp. 91–100, July 24, 1967.

76. Macario, R. C. V., and Yusuf, T., "High *Q N*-path filter using diode bridges," *Electron. Engr.*, pp. 76–79, Jan. 1969.

77. Smith, B. D., "Analysis of commutated networks," *IRE Trans. Aeronautical and Navigational Electronics*, vol. PGAE-10, pp. 21–26, Dec. 1953.

78. Le Page, W. R., Cahn, C. R., and Brown, J. S., "Analysis of a comb filter using synchronously commutated capacitors," *AIEE Trans. Communications and Electronics*, vol. 72, 1953.

79. Langer, E., "Tune in with a new *N*-path filter," *Electron. Engr.*, pp. 62–66, Nov. 1969.

80. Glaser, A. B., "A digital time division multiplexed *N*-path system," presented at IEEE Internat'l Symp. on Circuit Theory, 1969.

The early history of electronics

VI. Discovery of the electron

The invention of the first thermionic devices, which led to the massive development of electronic telecommunications in the 20th century, was predicated on a discovery made in a British university laboratory as the 19th century drew to its close

Charles Süsskind **University of California**

The preceding five articles in this series described how advances in electromagnetic theory, one of the two parent sciences of today's electronics technology, led to its first successes at the turn of the century: "wireless" communications.[1] But radiotelegraphy may well have remained an expensive and uncertain medium, restricted to emergency use in situations in which costs did not matter, if discoveries in the other parent science—particle physics—had not most conveniently come along at about the same time and opened up undreamed-of possibilities in the generation of high-frequency waves and the amplification of weak signals.

The concept of the electron

As with many major discoveries, some writers have speculated why the concept of an electron was not enunciated at least half a century before it was, as a result of the development of electrochemistry. In 1833, Faraday discovered the laws of electrolysis, by which the rate of decomposition of an electrolytic solution is proportional to the electric current and independent of the strength of the solution or the size of the electrodes. In a comparison of different electrolytes, a given electric current liberates one atom of any element in the time it would take to liberate n atoms of hydrogen, where n is the valency of the element in question. (A charge of 96 580 coulombs, known as a faraday, is required to liberate Z/n grams of each ion, where Z is the atomic weight. To be sure, Faraday knew nothing of "valency," a concept that was established later; he expressed his results in terms of "electrochemical equivalents.") The implication that a given quantity of electricity is associated with every atom was clear to Faraday, but he was wary of basing a new theory of matter on it "for," as he wrote, "though it is very easy to talk of atoms, it is very difficult to form a clear idea of their nature, especially when compound bodies are under consideration."[2]

As we shall see, the idea of the discreteness of charge, based on Faraday's laws, was to be reopened in the 1880s by Helmholtz and others. At the same time the question was also agitated in a new context, the experimental investigation of electric discharges in rarefied gases. This study went back to the beginning of the 18th century, to the discoveries of Francis Hauksbee (*d.* 1713?), who had observed a luminosity on glass carrying an electric charge, and of Pierre Polinière (1671–1734).[3] During the next century and a half, very little progress was made in this field, with the exception of the discovery made by Faraday in 1838 that the luminosity caused by an electric discharge between two electrodes in rarefied air exhibited a gap just in front of the negative electrode[4]; this gap is now designated as the *Faraday dark space*. The reason the subject proceeded so slowly may well have been a technological one: it was difficult to reduce pressure reliably and effectively to the extent necessary in investigating electric discharges. The dependence of the progress of science on technological innovation is a well-known phenomenon that is nowhere better exemplified than in the influence that air-pump design (accelerated by the exigencies of incandescent-lamp manufacture) exercised on electrical research.[5] Heinrich Geissler (1815–1879) provided an important new tool for the study of electric discharges when he developed the mercury air pump in 1855.

Julius Plücker (1801–1868) at the University of Bonn and his pupil, Johann Wilhelm Hittorf (1824–1914), brilliantly exploited the new Geissler high-vacuum tubes in an extensive series of researches on electric-discharge phenomenons.[6] Plücker remembered that Sir Humphry Davy (1778–1829) had shown in 1821 that an arc between two electrodes could be deflected by a magnet.[7] Plücker repeated the experiment with a gaseous discharge and obtained a similar deflection. Moreover, the luminosity from the negative electrode appeared to follow the magnetic-field lines. Plücker made two other important observations. He noticed that particles of the electrode material (which in this instance was platinum) left the cathode and were deposited on the walls of the enclosing glass bulb, and that the walls in the vicinity of this cathode exhibited a luminous glow that could be likewise moved about by a magnet. The second observation was further investigated by Hittorf, who discovered that when an object was placed in front of a point cathode, a shadow was cast in the glow.[8] From this observation, Hittorf deduced that the emission from the cathode propagated rectilinearly in "glow rays" (*Glimmstrahlen*). Moreover, he noticed that the rectilinear propagation changed to a helical path under the influence of a magnetic field.

In 1871 Cromwell Fleetwood Varley (1828–1883) performed an important experiment in which he deflected the rays by an electrostatic field.[9] He concluded that they consisted of "attenuated particles of metal, projected from their negative pole by electricity"; in other words, negatively charged corpuscles. This hypothesis also served to explain their helical path in a magnetic field.

Cathode rays

The next important discovery was made in 1876 by Eugen Goldstein (1850–1930), who found that the shadows observed by Hittorf as a result of placing a solid object in the path of the rays were cast regardless of whether the cathode was a single point or an extended surface, as long as the obstacle was placed near the cathode.[10] This experiment showed that the entire cathode surface emitted rays in a direction perpendicular to the surface. Goldstein named them *cathode rays*.

During the next two decades, scientists kept up a running argument: Were cathode rays indeed particles, or were they disturbances in the ether? By and large, the former view was held by the British physicists and the latter by the German school. In England, Crookes inclined toward the view that cathode rays were a molecular torrent.[11] He thought that molecules of the residual gas impinged on the cathode, became negative as a result, and were then repelled by the negative cathode. This hypothesis was reinforced by his observation that the width of a dark space that appeared in front of the cathode grew as the pressure in the tube was decreased, until the entire glow disappeared. (This space, whose discovery was likewise predicated on improvements in air pumps, is now known by his name and should not be confounded with the Faraday space.) Crookes described the cathode rays as an "ultragaseous" or "fourth state" of matter, by which he apparently meant a condition in which electrically charged gaseous molecules moved under pressure so low (i.e., a mean free path so long) that collisions with other particles could be disregarded. Although it is true that the theoretical explanation proposed by Crookes was rather naive (he was primarily a chemist), it cannot be gainsaid that the corpuscular hypothesis was considerably strengthened by his experiments, which were characterized by meticulous care and not a little sense of showmanship. (That was also true of the experiments that Crookes had performed with his radiometer, in which experimental technique likewise outdistanced his theoretical capabilities.)

The German physicists, notably Eilhard Ernst Gustav Wiedemann[12] (1852–1928), Goldstein,[13] and Hertz,[14] vigorously opposed the notion that the cathode rays were corpuscles. Hertz thought that the deflection of the cathode rays by a magnetic field could be explained by the action of the magnet on the medium in which the rays were propagating, much as the optical plane of polarization is rotated when the medium through which the light passes is magnetized. As for C. F. Varley's observation that cathode rays could be deflected by an electrostatic field, Hertz was unable to reproduce that result—largely because, as it proved later, his vessels were insufficiently evacuated, so that the residual gas molecules were ionized by the cathode rays and aggregated along the deflecting plates, effectively shielding the cathode rays from the deflection fields. (The correct explanation was subsequently provided by FitzGerald.[15]) In the same investigation, carried out five years before his celebrated series of experiments on electromagnetic waves described in our second installment,[1] Hertz also tried to deflect a magnetic needle by a cathode ray, and to determine whether the cathode rays carried electric charges by directing them into a metal collector (a Faraday cage) to which an electrometer was connected. Again both results were negative, owing to the employment of experimental techniques that were later shown to be open to serious objections: the first result sought is too small to be detectable by the means that were at Hertz' disposal, and the second was later contravened by a more carefully designed experiment performed by Jean Baptiste Perrin (1870–1942) and described in what was the future Nobel prizewinner's (physics, 1926) first published work.[16]

These experiments only reinforced Hertz' conviction that cathode rays could not consist of particles. His conviction was not based entirely on negative results. In what proved to be the last electrical research of his life, Hertz had confirmed a finding that seemed to him overwhelming proof that particles could not be involved: the fact that cathode rays could to some extent pass through thin metal foils, such as gold leaf.[17] This result was quite incontrovertible and was later amply confirmed by his pupil Philipp Eduard Anton Lenard (1862–1947), who managed to pass cathode rays out of the vacuum envelope through a thin aluminum window and to investigate their absorption in air.[18] The penetrating power of the rays was certainly a serious objection to the corpuscular theory. The only reply that the British physicists could muster was that in impinging on the metal window, the cathode rays might conceivably generate a new lot of particles, so that the window itself became an emitter. This suggestion was made by Joseph John Thomson (1856–1940), who had long occupied himself with these questions and who was soon to provide the correct explanation.

The electron

The end of the 19th century was an incredibly fruitful period in the history of electrical research, encompassing as it did the aforementioned researches of Hertz on electromagnetic-wave propagation, the development of a general theory of discharge and ionization in rarefied gases by Arthur Schuster (1851–1934) at the University of Manchester,[19] the discovery of the photoelectric effect by Hertz and Wilhelm Hallwachs (1859–1922) in 1887–1888, and the discovery of X rays by Wilhelm Konrad Röntgen (1845–1923) in 1895.

J. J. Thomson was deeply involved in these investigations. What impressed him more than anything was that cathode rays, as distinct from light or X rays, could be deflected by a magnetic field. He rightly surmised that the true nature of the cathode rays would be revealed by careful measurements of this deflection. He was amazed by the result. "I had for a long time been convinced that these rays were charged particles," he said afterwards, "but it was some time before I had any suspicion that they were anything but charged atoms. My first doubts as to this being the case arose when I measured the deflection of the rays by a magnet, for this was far greater than I could account for by any hypothesis which seemed at all reasonable if the particles had a mass at all approaching that of the hydrogen atom, the smallest then known."[20] Moreover, the deflection was independent of the kind of gas remaining in the tube. If one made the gratuitous but convenient assumption that the charge associated with the particles was the same as that which entered into the ratio of mass to charge of an ordinary ionized atom, the inescapable conclusion was that he was dealing with particles whose mass was smaller, by three orders of magnitude, than that of atoms.

It is hard for us to appreciate, across the decades, what courage was required to put forward the startling hypothesis that the atom of an element was not the smallest subdivision of matter. Thomson plunged in fearlessly at once. In a lecture at the Royal Institution on April 30, 1897, he astutely utilized Lenard's own results to establish the particulate nature of cathode rays. "From Lenard's experiments on the absorption of the rays outside the tube," he said, "it follows on the hypothesis that the cathode rays are charged particles moving with high velocities that the size of the carriers must be small compared with the dimensions of ordinary atoms or molecules."[21]

It should not be thought that the discovery of the electron was accompanied by immediate understanding of the structure of the atom; that came much later through the efforts of Thomson, of his pupil Ernest Rutherford (1871–1937), of Niels Bohr (1885–1962), and of others. Thomson realized that the atom was electrically neutral and that the negative charge carried by electrons had to be balanced by a positive charge, which he at first represented by an admittedly artificial model of a cloud diffused throughout the space occupied by the atom rather than concentrated at the nucleus. Moreover, it had been known for a dozen years that rays with properties corresponding to the opposite polarity existed. Goldstein had shown in 1886 that, in a discharge tube containing a perforated cathode, rays would pass through the perforations in a direction opposite to that of cathode rays; he had named them *Kanalstrahlen* (canal rays).[22] These rays were now investigated in greater detail by Wilhelm Carl Werner Otto Fritz Franz (Willy) Wien (1864–1928), who showed that these *positive ions* also behaved like particles, that (unlike the electrons) their behavior depended on the nature of the gas from which they originated, and that the ratio of mass to charge of the smallest was comparable with the ratio obtained in electrolysis—i.e., about 1000 times as large as that obtained by Thomson for the electron.[23]

Thomson had carried out the measurement of the mass-to-charge ratio before the end of 1897, having first removed the uncertainty created by Hertz' failure to deflect cathode rays electrostatically. Thomson showed that the neutralizing effect of the gas ionized by the passage of the cathode rays, which acts as a conductor to shield the rays from the deflecting plates (and which had ruined Hertz' experiment), could be minimized by insuring that the experiment was carried out in a better vacuum.[24] Hertz himself had understood that if cathode rays should prove to consist of corpuscles after all, it should be possible to make measurements on them by observing their path under the combined action of electrostatic and magnetic fields; however, since he had failed to obtain electrostatic deflection, this avenue had been closed to him. Schuster, who believed that cathode rays were made up of charged atoms, had suggested even earlier that the mass-to-charge ratio could be obtained from a knowledge of the magnetic and electric fields. But it was not until Thomson's classic experiment that these considerations were applied to the measurement of the ratio of mass to charge of an electron.

In his experiment, Thomson used an electric field E to counteract the deflecting force exerted on a particle of charge e by a magnetic field B, as indicated by zero deflection of the beam, so that

$$eE = Bev$$

Having thus determined the velocity $v = E/B$, he inferred

Sir Joseph John Thomson (1856–1940) was born near Manchester, the son of a publisher and bookseller. He showed early talent in school and at Owens College in Manchester, where he studied under Balfour Stewart and won many prizes. At Cambridge, he came in second in the mathematics examination (Joseph Larmor was Senior Wrangler that year) and continued postgraduate work at the new Cavendish Laboratory, whose directorship had passed to Lord Rayleigh in 1879 on the death of the first professor, James Clerk Maxwell. Rayleigh resigned five years later and Thomson, then 28, became the third Cavendish Professor, a position he held for 34 years and only gave up (to Ernest Rutherford) to become Master of Trinity College at Cambridge, a valuable and prestigious sinecure. During his long tenure, the Cavendish came to occupy an eminent position in the world of experimental physics (similar to that held in later years by Franck's Göttingen and Lawrence's Berkeley), owing in no small part to the important innovation of making graduates of other universities eligible to receive Cambridge research degrees and Fellowships.

Besides his epoch-making identification of the electron as a separate charged particle, Thomson's greatest contribution was a result of his work on "positive rays" (ion beams): the development of a cross-field method of identifying various atoms and molecules, including isotopes of the same element not distinguishable by chemical means. That was the beginning of mass spectroscopy.

Thomson was the first recipient of the Hughes Medal of the Royal Society in 1902 and also received its Copley Medal, as well as many other honors from scientific and engineering societies. In 1906 he got the Nobel Prize for "his theoretical and experimental investigations into the transmission of electricity through gasses." He was knighted in 1908 and received the Order of Merit in 1912. Among his pupils were Rutherford, C. T. R. Wilson, Aston, Barkla, W. H. Bragg, and Richardson, all of whom later received the Nobel Prize, as did his son G. P. Thomson.

In 1936, J. J. Thomson published his autobiography, *Recollections and Reflections* and, shortly after his death, *The Life of Sir J. J. Thomson* (1942) was published by the fourth Lord Rayleigh (a physicist like his more famous father, the third baron), who had also worked under Thomson.

the deflecting force produced by the magnetic field from measurements of the radius r of the circular path of the beam when the magnetic field acted alone:

$$\frac{mv^2}{r} = Bev$$

Hence the ratio of charge to mass was given by

$$\frac{e}{m} = \frac{v}{Br} = \frac{E}{B^2r}$$

In that connection, the Harvard historian of science, I. Bernard Cohen, has noted[25] that the description of Thomson's experiment found in many modern textbooks and college laboratory manuals is not an historic one, inasmuch as it leads the latter-day student to believe that Thomson obtained his results with a highly evacuated tube containing a thermionic cathode. In fact, the work was done with a poorly evacuated cold-cathode discharge tube, in which the ionization of the residual gas served to introduce appreciable errors, since the relationship between velocity and potential difference is substantially more complex under these circumstances than in a highly evacuated tube. In his autobiography, written nearly 40 years later, Thomson himself emphasized that point. "It is not possible to estimate from the potential difference between the cathode and anode of the discharge tube the energy possessed by a charged particle at any point in its course without knowing more about the mechanism of the discharge than we do even at the present time."[26] To which today's plasma physicist, after yet another 40 years, can only respond with a hearty "amen."

As we have pointed out elsewhere,[27] the name *electron* was not coined by the discoverer of the particle but by the Irish physicist George Johnstone Stoney (1826–1911) in 1894.[28] The view that electricity came in discrete amounts, or quantums (so that the processes of gaseous discharges could be considered to be analogous to those of electrolysis), had been advocated as early as 1881 by Stoney[29] and almost simultaneously by Helmholtz.[30] By 1891 Stoney had concluded that each atom in electrolysis carries a charge that is given up when free hydrogen is liberated. He estimated the magnitude of the charge on the basis of the number of atoms per unit volume of hydrogen gas (calculated from kinetic theory), and he proposed the name *electron* for this "natural unit of electricity."

The term "electron" for a quantum of charge achieved fairly wide acceptance and was quickly applied to the particle discovered by Thomson; FitzGerald was the first to do so when he suggested, following Thomson's Royal Institution lecture, that "we are dealing with free electrons in these cathode rays."[31] Although the new interpretation of the term was universally accepted, it is a curious fact that Thomson himself apparently had misgivings regarding the change in its meaning (from quantum to particle) and continued to use the word "corpuscles" for over 20 years. Perhaps he also did not want to prejudge the question of electromagnetic mass.

As had been the case for electromagnetic waves, the action of electrons had been observed by several inventors and scientists before Thomson came to his conclusions. One observer figured in both fields: Edison. We shall see in our next installment how his report of what came to be known as the "Edison effect" (thermionic emission), which antedated Thomson's discovery, led indirectly to the invention of the first vacuum tube, the diode.

This study was sponsored in part by Grant GS-2411 from the National Science Foundation.

REFERENCES AND NOTES

1. Süsskind, C., *IEEE Spectrum*, vol. 5, pp. 90–98, Aug. 1968; vol. 5, pp. 57–60, Dec. 1968; vol. 6, pp. 69–74, Apr. 1969; vol. 6, pp. 66–70, Aug. 1969; vol, 7, pp. 78–83, Apr. 1970.

2. Faraday, M., *Experimental Researches in Electricity*, Vol. 1. London: R. and J. E. Taylor, 1839, p. 256.

3. Hauksbee, F., *Trans. Roy. Soc.*, vol. 24, pp. 2165–2175, 1705. For a description of Polinière's simultaneous discovery of electroluminescence, see Corson, D. W., *Isis*, vol. 59, pp. 402–413, 1968.

4. Faraday, M., *Trans. Roy. Soc.*, pt. 1, pp. 125–168, 1838.

5. See, for instance, Lafferty, J. M., *Proc. IRE*, vol. 49, pp. 1136–1154, 1961.

6. Plücker, J., *Ann. Phys.*, ser. 2, vol. 103, pp. 88–106, 151–157, 1858; vol. 104, pp. 113–128, 622–630, 1858; vol. 105, pp. 77–313, 1859; *Phil. Mag.*, ser. 4, vol. 16, pp. 119–135, 408–418, 1858; vol. 18, pp. 1–20, 1859.

7. Davy, H., *Trans. Roy. Soc.*, pt. 1, pp. 425–439, 1821.

8. Hittorf, J. W., *Ann. Phys.*, ser. 2, vol. 136, pp. 1–31, 197–234, 1869. For a discussion of the contributions of Plücker, Geissler, and Hittorf, see also Espenschied, L., *IEEE Spectrum*, vol. 4, p. 140, Dec. 1967.

9. Varley, C. F., *Proc. Roy. Soc.*, vol. 19, pp. 236–242, 1871. The same volume also contains (p. 243) an important paper by Varley on the polarization of metallic surfaces in aqueous solutions. "Cromwell Varley was a man of remarkable mental vision," says Rollo Appleyard in *The History of the Institution of Electrical Engineers (1871–1931)*, IEE, London, 1939, p. 35. Although he made important scientific contributions, he was not a professional scientist but a prominent telegraph engineer, one of the 66 founder members of the IEE's predecessor, The Society of Telegraph Engineers, in 1871.

10. Hittorf, J. W., *Monatsber. Akad. Wiss. Berlin*, pp. 279–295, 1876.

11. Crookes, W., *Trans. Roy. Soc.*, vol. 170, pp. 135–164, 641–662, 1879; *Phil. Mag.*, ser. 5, vol. 7, pp. 57–64, 1879. For Crookes and the radiometer, see also Süsskind, C., *Proc. IRE*, vol. 50, pp. 326–327, 1962; Woodruff, A. E., *Isis*, vol. 57, pp. 188–198, 1966; and Brush, S. G. and Everitt, C. W. F., in *Historical Studies in the Physical Sciences*. Philadelphia: University of Pennsylvania Press, 1969, vol. 1. This volume also contains a discussion by Tetu Hirosige of the growth of the theory of the electron out of the matrix of electromagnetic-field theory and electrodynamics on the Continent, with particular emphasis on the work of Hendrik Antoon Lorentz (1853–1928).

12. Wiedemann, G., *Ann. Phys.*, ser. 3, vol. 10, pp. 202–257, 1880; *Phil. Mag.*, ser. 5, vol. 10, pp. 357–380, 407–427, 1880.

13. Goldstein, E., *Ann. Phys.*, ser. 3, vol. 10, pp. 90–109, 249–279, 1881.

14. Hertz, H., *Ann. Phys.*, ser. 3, vol. 19, pp. 782–816, 1883.

15. FitzGerald, G. F., *Nature*, vol. 55, pp. 6–9, 1896.

16. Perrin, J. B., *Compt. Rend.*, vol. 121, pp. 1130–1134, 1895.

17. Hertz, H., *Ann. Phys.*, ser. 3, vol. 45, pp. 28–32, 1892.

18. Lenard, P., *Ann. Phys.*, ser. 3, vol. 51, pp. 225–267, 1894; vol. 52, pp. 23–33, 1894.

19. Schuster, A., *Proc. Roy. Soc.*, vol. 37, pp. 317–339, 1884; vol. 42, pp. 371–379, 1887.

20. Rayleigh, R. J. S., *The Life of Sir J. J. Thomson.* London: Cambridge, 1942, p. 80.

21. Thomson, J. J., *Proc. Roy. Inst.*, vol. 15, pp. 419–432, 1897; *Electrician*, vol. 39, pp. 104–109, 1897.

22. Goldstein, E., *Monatsber. Akad. Wiss. Berlin*, pp. 691–699, 1886.

23. Wien, W., *Ann. Phys.*, ser. 3, vol. 65, pp. 440–452, 1898.

24. Thomson, J. J., *Phil. Mag.*, ser. 5, vol. 44, pp. 293–316, 1897.

25. Cohen, I. B., *Am. J. Phys.*, vol. 18, pp. 343–359, 1950.

26. Thomson, J. J., *Recollections and Reflections*. London: G. Bell, 1936, pp. 339–340.

27. Süsskind, C., *Proc. IERE*, vol. 4, pp. 95–99, 1966; *IEEE Spectrum*, vol. 3, pp. 72–79, May 1966.

28. Stoney, G. J., *Phil. Mag.*, ser. 5, vol. 38, pp. 418–420, 1894.

29. Stoney, G. J., *Trans. Roy. Dublin Soc. Sci.*, vol. 3, pp. 51–60, 1881; *Phil. Mag.*, ser. 5, vol. 11, pp. 381–390, 1881.

30. Helmholtz, H., *J. Chem. Soc.*, vol. 39, pp. 277–303, 1881.

31. FitzGerald, G. F., *Electrician*, vol. 39, pp. 103–104, 1897.

Charles Süsskind's biography appeared on page 76 of the March 1970 issue.

Circuit breakers

Physical and engineering problems

III—Arc-medium considerations

Arc extinguishment in circuit breakers can be accomplished by a number of techniques, including the use of such mediums as oil, compressed air, SF_6, and vacuum

W. Rieder **Brown, Boveri & Co., Ltd.**

The problems treated in last month's discussion are essentially independent of the type of breaker, although their relative importance may vary. More specialized problems connected with the conception of certain physical principles of current interruption—mainly the arc medium—are described in this final installment.

Magnetic-blast breakers

In the early days of electrical engineering, when the service voltages climbed up to astronomical values and threatened to approach a kilovolt, it became extremely difficult to open the contacts sufficiently wide because the arc had to be lengthened to 30 cm or more before it extinguished. Fortunately, the arc could be much longer than the contact gap if the magnetic field caused by the path of the current itself blew the arc away from the gap. Later the magnetic field was intensified by coils inserted in the current path and by iron plates concentrating the field in the volume around the arc gap. To protect the surroundings, arc chutes were built around the contacts. With increasing voltage and current, attempts were made to reduce the volume of the arc chutes by ceramic arc labyrinths inside the chutes. In this method, the arc is pressed by the magnetic field against the ceramic walls, melts their surface, and, due to this additional power loss, the electrical gradient of the arc is increased and quenching is facilitated (Fig. 22).

Another possibility is to split up the arc into many short arcs, electrically connected in series, by blowing the arc against a number of parallel metal splitter plates arranged a short distance (about 1 mm) apart. Each of the short arcs can withstand about 100 volts rms of the service voltage, provided the voltage distribution across the many gaps is uniform.

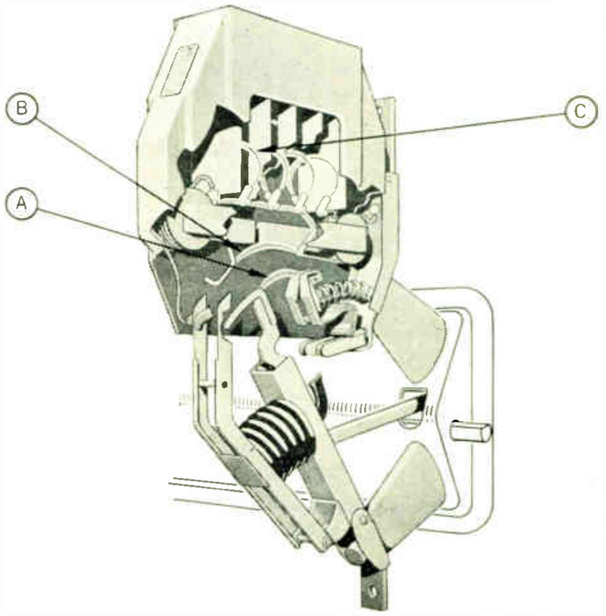
Although both kinds of magnetic-blast breakers are autonomous and current-limiting, they are also heavy, voluminous, expensive, and slow, especially in certain current ranges. At low currents the magnetic field may be too weak to move the arc[43]; therefore, in some types of breakers an additional air blast (created by a piston connected with the operating mechanism) is provided (Fig. 22).

The application of magnetic-blast breakers is limited to voltages below about 15 kV. At higher voltages, fast breakers are preferred, with minimum arc power release; these are able to interrupt the current definitely within three or two cycles, or even one cycle, after release—that is, before current limitation becomes effective in the usual current-limiting breakers.

Detonation-assisted breakers

Nevertheless, very fast current-limiting devices developed for rectifier protection do exist. The interrupter consists of a special piece of duct, which carries the current and is filled with gunpowder to be ignited by a spark released from a relay. This is sensitive to the current increase, di/dt, and explodes the duct before the maximum current occurs. Then a conventional current-limiting fuse connected in parallel takes over and interrupts the current. This apparatus, however, must not be called a circuit breaker, because it is not suitable for normal

FIGURE 22. Magnetic-blast breaker with auxiliary air blast. Arc is shown in three successive positions (A, B, and C).



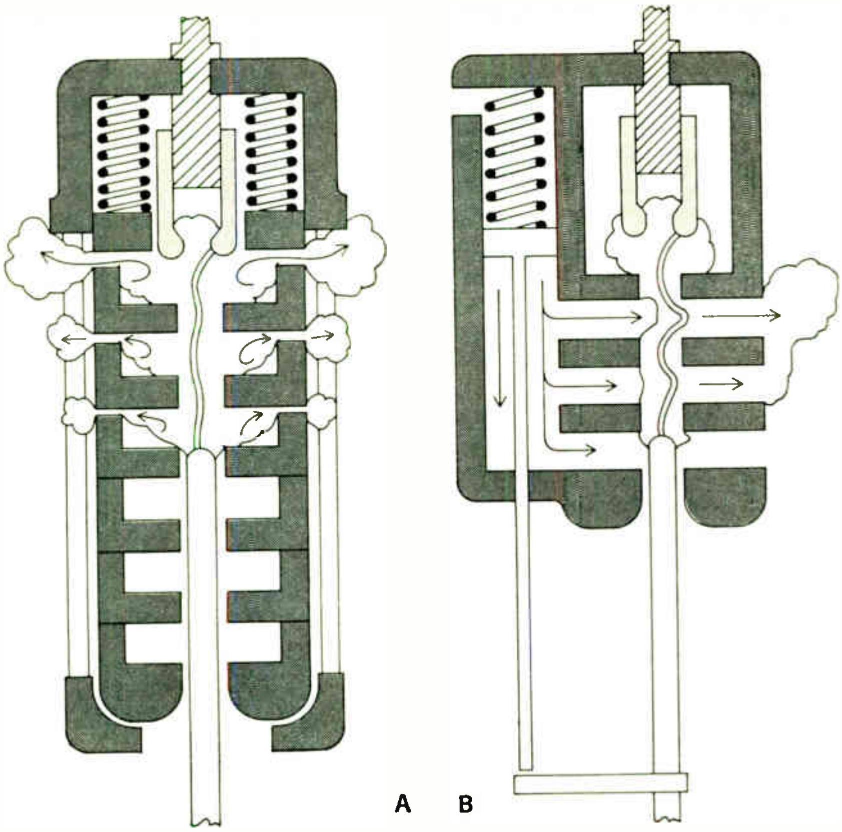


FIGURE 23. A—Elastic expulsion chamber. B—Expulsion chamber with auxiliary piston.

current interruption, nor is it able to make a current. There are other proposals for more conventional powder-driven circuit breakers, but these are not on the market or in service.

Oil breakers

That an arc is extinguished more easily in oil than in air was probably discovered by chance, because it does not seem likely that anybody would have expected highly flammable oil to be an outstanding arc-extinguishing medium. However, it worked successfully due to the high heat conductivity of the hydrogen gas released by the arc and its high pressure (see Part I)—so long as the hot gases of decomposition did not explode at the surface of the oil. Therefore, with increasing arc power, the arcing contacts had to be immersed deeper in the oil vessel. In 1908 the breaking capacity of bulk oil breakers was notably increased by the use of an insulating, rigid arc chamber surrounding the contact tulip. The chamber encloses the highly pressurized gas until the contact rod drawn through the bottom hole is replaced by the switching arc, which now is exposed to a strong flow of the expanding gas. It is an advantage of this current-dependent flow, generated by the arc itself, that it needs no gas compressor and is enhanced when the interrupted current increases. But difficulties arise at extremely high currents because of the very high pressure stresses on the chamber, and at low currents because of poor gas flow. To avoid explosion at very high currents, pressure-limiting valves may be used to protect the so-called "elastic" arc chambers; at very low currents, a mechanically driven piston guarantees sufficient flow (Fig. 23). More sophisticated devices use an auxiliary arc to generate the pressure needed, or they use a differential piston, which causes a flow by exposing two areas of different size to the same pressure. At least a cross blast was forced by the chamber geometry without any piston (Fig. 24).

Although in the United States the volume of the oil vessels increased enormously with increasing interrupting power, in Europe breakers with low oil content were developed following a major disaster in Prague after World War I. Up to 110 kV single breaks only were used in Europe, but now at higher voltages many units are employed in series—sometimes more than comparable air-blast or SF_6 breakers (Fig. 25). The units are made of porcelain and contain at most a few liters of oil.

Oil breakers have the advantage of being autonomous, in that they require no compressor or heating; moreover, the arc is enclosed in oil and thus cannot cause explosions in explosive atmospheres. If, however, the breaker should explode as a result of a mechanical failure or electrical overload, the consequences could be catastrophic.

Because of the decomposition of the oil in the arc, the oil becomes polluted by carbon particles, which reduce its dielectric strength. Therefore, the arc chambers and the total oil content must be cleaned periodically; this is a tedious job, especially in the case of bulk oil breakers.

Air-blast breakers

In air-blast breakers, compressed air for arc quenching and for succeeding insulation is stored in a vessel (and often in the arc chamber itself). In the axis of a nozzle through which the compressed air leaves the chamber,

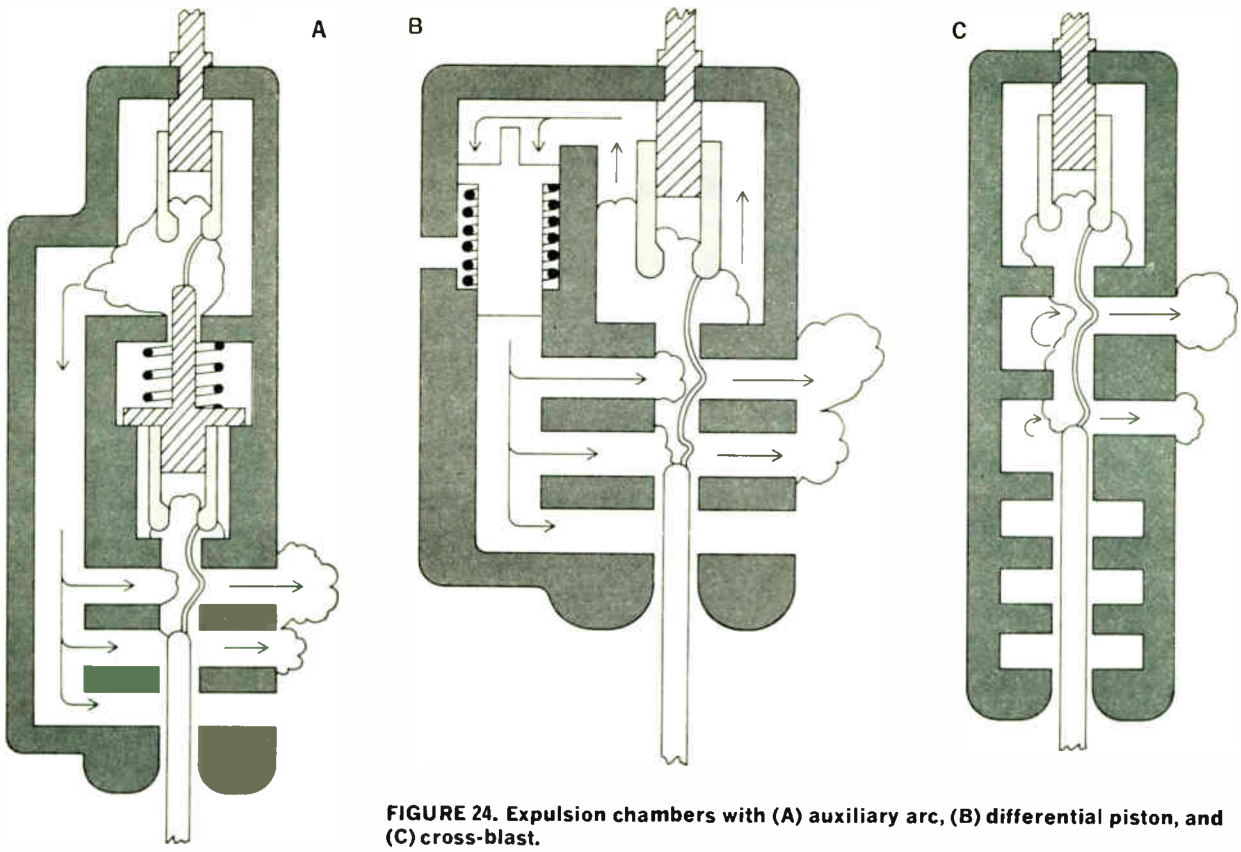
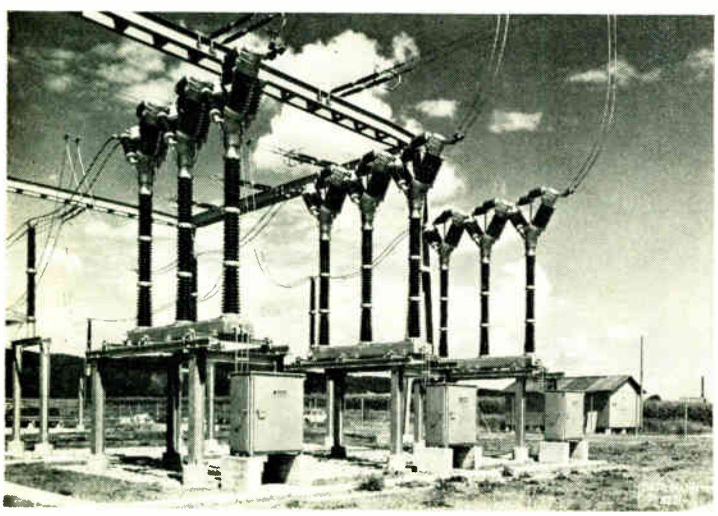


FIGURE 24. Expulsion chambers with (A) auxiliary arc, (B) differential piston, and (C) cross-blast.

the switching arc is exposed to forced convection flow. Some modern breakers use twin nozzles, which are aerodynamically in parallel but electrically in series (Figs. 16 and 17). The nozzles may be insulating (Fig. 26) or metallic (Figs. 16–18 and 20); the latter are often used as electrodes (Figs. 16 and 17). Optimum nozzle and electrode geometry depend on the type of fault to be interrupted (breaker-terminal fault, short-line fault, phase opposition, etc.).

FIGURE 25. High-voltage low-oil-content breaker, rated 245 kV, 2 kA, 17 GVA, with parallel capacitors and resistors. Hydraulic mechanism.



Compressed air is an excellent insulant, comparable to clean oil (Fig. 27), though unlikely to cause fire. Therefore, in most air-blast breakers, air at high pressure for insulation is used inside the arc chamber. But there is one type of breaker (Fig. 26) that draws the arc in open air, saving the expensive high-pressure porcelain housing of the unit. It needs a longer contact travel and more space, because of the renunciation of the high dielectric strength of the compressed air; the visible contact gap is advantageous, but the noise of the compressed air expanding directly to atmosphere is disturbing. Therefore, full air flow is released for short-circuit currents only, whereas in the other air-blast breakers the flow is independent of the current.

SF_6 breakers

Apart from its attractive features as an insulant (Fig. 27) and as an arc medium (as described in the July installment), SF_6 has two other traits that give rise to serious problems in the design and use of SF_6 breakers.

First, although SF_6 itself is nontoxic and inert, its dissociation products are rather aggressive, especially when water (in the form of vapor) is present. The sulfur and fluorine recombine to SF_6 when cooled down again, but in the interrupter some fluorine may combine with the metal vapor of the electrodes or with an insulant, and then unsaturated molecules or radicals remain. There is no serious trouble as long as no water vapor is present. However, this proviso implies cautious choice of materials and a construction that needs minimum mainte-

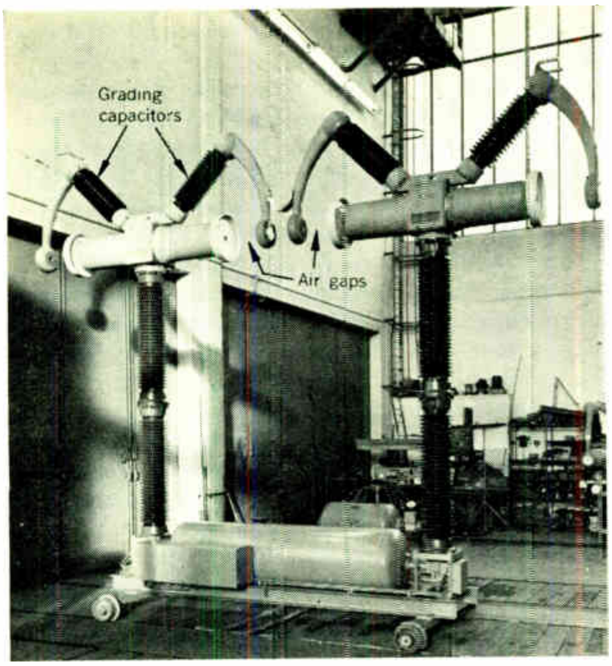


FIGURE 26. Pole of a 220-kV, 2-kA, 15-GVA, free-jet air-blast breaker with current-controlled air blast.

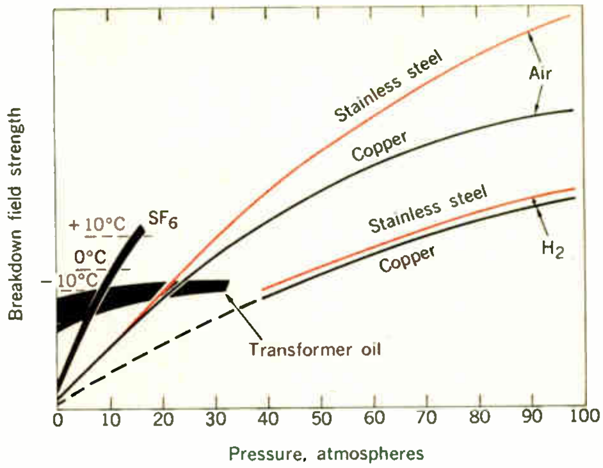


FIGURE 27. Dielectric strength of some arc mediums vs. pressure. Nearly homogeneous field and 50-Hz supply are assumed.[44]

nance during the life of the breaker.

Second, the pressure of SF_6 is limited by its triple point at 38 atm and 45°C. This does not impair its application at atmospheric pressure, but it prevents the pressure from being raised above about 18 atm and thereby limits the dielectric strength and the gas flow. At this pressure SF_6 liquidizes at 14°C; at lower ambient temperature it has to be heated. Providing heating power is no serious obstacle, but the control and supervision of this power represent an additional complication and a possible source of failure.

There are two families of SF_6 breakers in use. The single-pressure devices produce their pressure gradients needed for the forced convection either by an auxiliary arc or by a piston, as in the case of some low-oil-content breakers. Two-pressure systems store high-pressure gas from a compressor, as do air-blast breakers, although each SF_6 breaker has its own compressor.

The special advantages of SF_6 gas are exploited in SF_6-insulated high-voltage indoor stations, which make it possible to build a switchyard for 245 or even 500 kV in the cellar of a building in the center of a reasonably large city.

Vacuum breakers

Vacuum interrupters are absolutely autonomous and their short contact travel is obtainable with a cheap operating mechanism. Their main problem is their manufacture, which is a delicate and costly matter. High vacuum must be guaranteed for 20 years. For the time being, a metal bellows generally allows the movement of one contact piece (Fig. 28). Without doubt this is a weak point, because the number of mechanical operations it can perform before it breaks varies over a wide range and it is not the average but the minimum number that determines the guaranteed life. Unfortunately, the vacuum of each unit cannot be monitored economically.

The electrode material must be chosen with regard to the vapor pressure, to avoid current chopping at low currents and make vacuum recovery possible even at high currents. In addition, there are many other demands[45]: Its melting point must be high enough to allow adsorbed gas layers to bake out effectively; its electrical conductivity must be low to avoid high contact resistance; heat conductivity must be high enough to dissipate the contact heat, but low enough to prevent cathode instability at low current values; the yield strength must be low to permit opening of the welded contacts by moderate forces, but high enough to prevent vacuum breakdown due to electrostatic extraction of particles on the grain boundaries; its work function must be high, to increase breakdown field strength.

Vacuum interrupters are not in use for high-voltage applications, but there is one prototype under test. At present, these devices are used mainly in capacitor switches and reclosers.

Semiconducting breakers

Semiconductors are used nowadays in rectifiers for very high currents and voltages, and therefore the question arises as to whether they will displace the arc in the breakers as well. The main drawback is that semiconductors are neither good conductors nor good insulators. In rectifiers, the losses due to the voltage drop of the semiconducting elements in the forward direction are acceptable, because there is no other method of obtaining the same effect with less effort and loss. But it seems to be unacceptable that a utility bargaining for a tenth of a percent in the efficiency of a transformer should lose several percent in semiconductor circuit breakers, when that loss can be avoided by using the much cheaper, smaller, and more reliable metallic breakers.

On the other hand, even in the nonconducting state, semiconductors are not true insulators; therefore an isolator would be needed to interrupt the residual current and to guarantee galvanic insulation. Other disvantages of a semiconducting circuit breaker are the

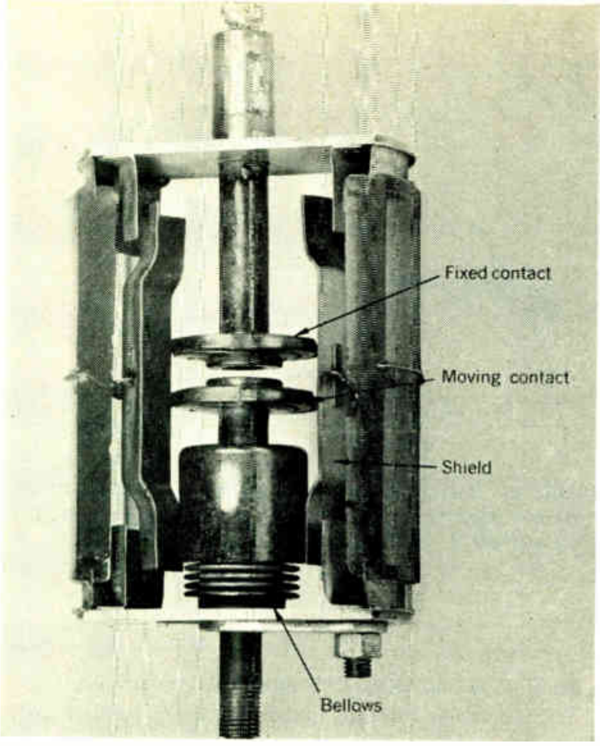


FIGURE 28. Section of a vacuum interrupter.

price (which has to be compared with that of the spontaneously appearing arc), the outlay for cooling, the reliability (especially after long use and when overloaded), and the size.

However, if there is a metallic contact in parallel with the semiconducting unit to carry the continuous current, an arc will appear when it opens and the arc voltage must be increased to a value high enough to ensure current commutation to the semiconducting device. When this arc has been extinguished, the semiconducting unit may start to interrupt the residual current and to disconnect the circuit reliably.

Obviously it would be much simpler, cheaper, and more reliable if the current were interrupted immediately by the first arc, which is better qualified to do so, need not be manufactured, and cannot be damaged by overload.

Despite this negative result, however, a semiconductor can be used to aid the arc during the current-zero interval in a way that does not need high energy dissipation in the semiconductor but, in fact, prevents it. A system that has been in service for some time involves the use of nonlinear resistors in air-blast breakers.

Conclusions and outlook

At last the question of future development arises. It is not possible to prophesy seriously, but only to disclose and extrapolate present trends and to weigh certain possibilities and difficulties.

From the point of view of the user, the present trend is toward faster, more reliable, and smaller breakers, which need less maintenance and are able to carry and interrupt higher currents at higher voltages under more severe conditions. The manufacturers know that a new type of breaker must also be cheaper (per MVA) if it is to be successful on the market.

It is difficult to predict whether one of the present interrupting principles (magnetic-blast, oil, air-blast, SF_6, vacuum) will displace the others. Probably an equilibrium with regard to applications will be found, but with wide overlapping areas. Perhaps magnetic-blast breakers will be displaced by vacuum breakers in certain applications and indoor SF_6 cubicles will be successful in the cities. But most probably the principles used today will remain and coexist for the next decade. There is no question that each scheme is able to interrupt any desired power and, although SF_6 seems to be a nearly ideal arc medium and isolant, surely the physical features around current zero will not be the only decisive factors. Each of the mediums discussed has its merits and disadvantages with respect to current interruption and behavior in service. Therefore, not the medium itself, but how the designer exploits its advantages and compensates its weak points, will decide the success of models developed during the coming years.

Whether techniques based on new principles will succeed is more difficult to estimate. It is not probable that a better multiatomic gas than SF_6 will be found, but perhaps a better liquid than oil is feasible. At present the difficulties of vacuum interrupters at high currents and voltages seem to be inherent in their physics, and their expensive manufacture and vacuum supervision are drawbacks, but predicting a future invention means inventing it—at least in this case. Air-blast breakers tend toward higher pressures; nobody knows the optimum value.

The arc plasma has and will maintain the widest range of fast-changing resistivity. Therefore, the next generation will build arc breakers too, and it does not seem likely that semiconductors will be more successful in high-voltage and high-power applications, but they may find use in certain special cases, such as transformer tap changers in locomotives and a few applications of contactors. Probably they will be successful as an aid around current zero of an arc or a synchronous interrupter.

Synchronous breakers can hardly avoid the switching arc, but they may reduce the time and current of arcing or burning of a semiconductor. There are some prototypes of this kind, but they are not yet on the market. Price and reliability will dictate whether and how far this road will be taken, and whether the role of the arc will be limited to a short time interval around current zero, where it also could be replaced by a semiconductor, since a disconnector in series with a high-voltage synchronous switch seems to be necessary in any case.

The author acknowledges information and the permission to show photographs from the following companies: AEG, Germany; AEI-GEC, England; ASEA, Sweden; Brown Boveri, Germany and Switzerland; Calor-Emag, Germany; General Electric, U.S.A.; Magrini, Italy; Merlin & Gerin, France; Siemens, Germany; and Sprecher & Schuh, Switzerland.

REFERENCES

43. Rieder, W., *Plasma und Lichtbogen.* Braunschweig: Vieweg, 1967.

44. Leber, R., and Treptow, A., "Leistungsschalter für Hoch- und Niederspannung," *ETZ*, vol. A 89, pp. 489–492, 1968.

45. Barkan, P., Lafferty, J. M., Lee, T. H., and Talento, J. L., "Development of contact materials for vacuum interrupters," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-89, to be published.