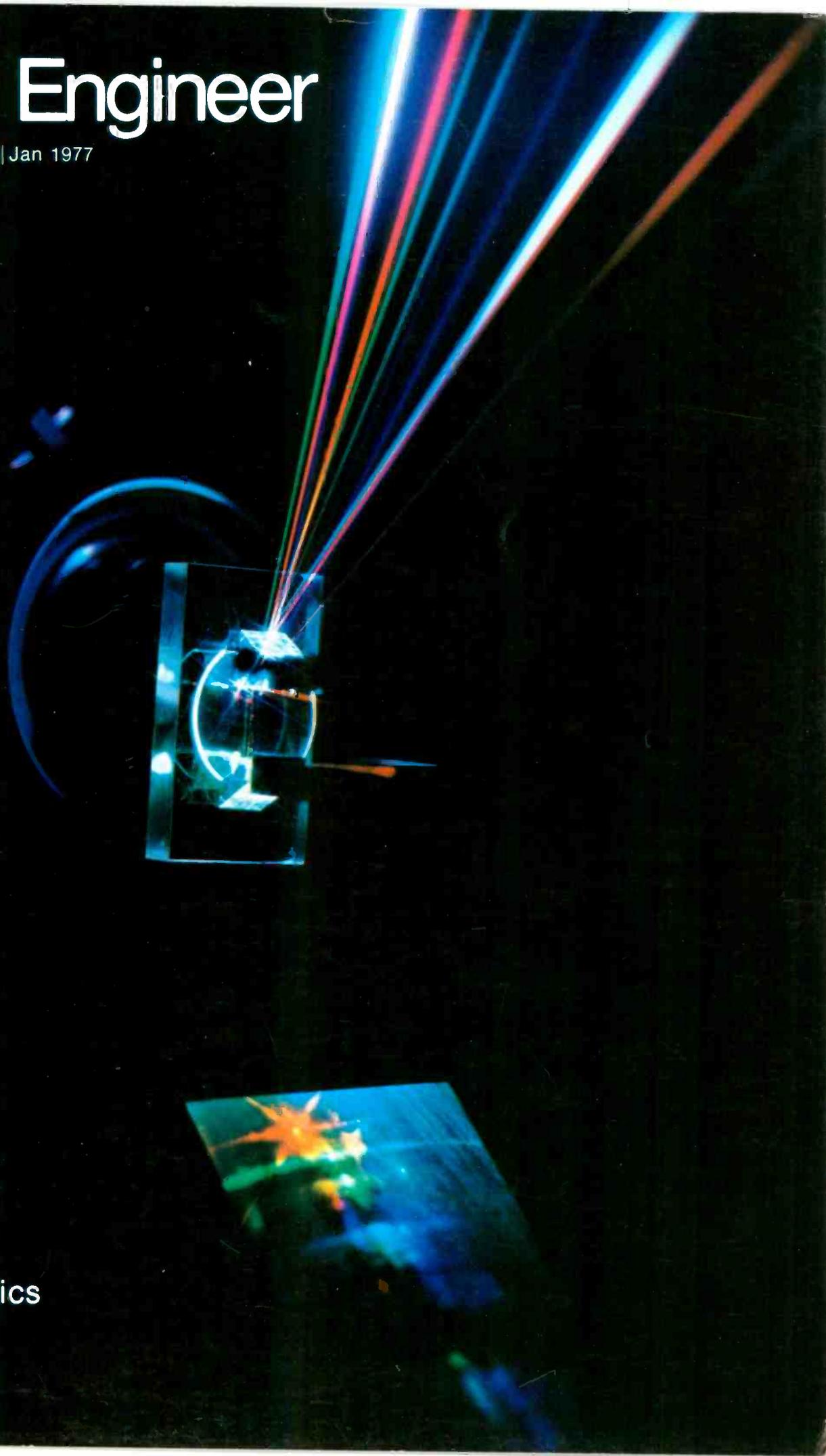


RCA Engineer

Vol 22 | No. 4 | Dec 1976 | Jan 1977



electro-optics

RCA Engineer

A technical journal published by
RCA Research and Engineering
Bldg. 204-2
Cherry Hill, N.J. 08101
Tel. PY-4254 (609-779-4254)
Indexed annually in the Apr/May issue.

RCA Engineer Staff

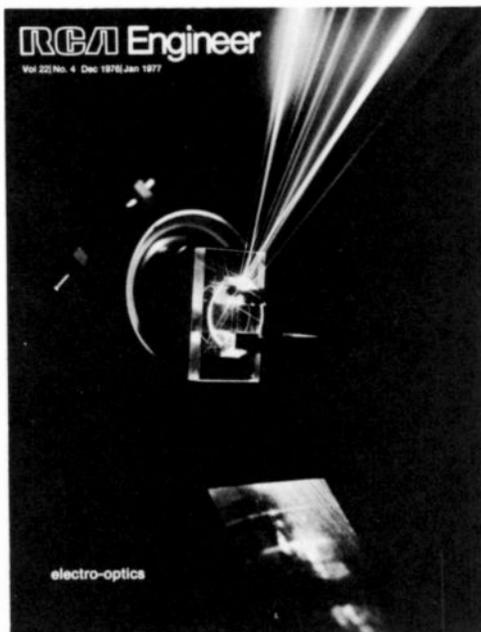
John Phillips	Editor
Bill Lauffer	Assistant Editor
Joan Toothill	Art Editor
Frank Strobl	Contributing Editor
Pat Gibson	Composition
Joyce Davis	Editorial Secretary

Editorial Advisory Board

Jay Brandinger	Div. VP, Engineering, Consumer Electronics
Bill Brook	VP, Engineering, RCA Americom
Joe Donahue	Div. VP, Engineering, Picture Tube Division
Hans Jenny	Manager, Technical Information Programs
Arch Luther	Chief Engineer, Engineering, Broadcast Systems
Howie Rosenthal	Staff VP, Engineering
Carl Turner	Div. VP, Solid State Power Devices
Joe Volpe	Chief Engineer, Engineering, Missile and Surface Radar
Bill Underwood	Director, Engineering Professional Programs
Bill Webster	VP, Laboratories

Consulting Editors

Ed Burke	Ldr., Presentation Services, Missile and Surface Radar Division
Walt Dannen	Mgr., News and Information, Solid State Division
Charlie Foster	Mgr., Scientific Publications, Laboratories



Our cover shows an electro-optic thin-film waveguide transmitting three different color laser beams. Waveguides of this type have been used to make switches and modulators capable of placing information on light at microwave frequencies, while requiring drive powers at least an order of magnitude less than those required by previous methods. (See J. Hammer, p. 71.)

Photo credit: Tom Cook, RCA Laboratories, Princeton, N.J. Experimental setup: Clyde Neil

- To disseminate to RCA engineers technical information of professional value
- To publish in an appropriate manner important technical developments at RCA, and the role of the engineer
- To serve as a medium of interchange of technical information between various groups at RCA
- To create a community of engineering interest within the company by stressing the interrelated nature of all technical contributions
- To help publicize engineering achievements in a manner that will promote the interests and reputation of RCA in the engineering field
- To provide a convenient means by which the RCA engineer may review his professional work before associates and engineering management
- To announce outstanding and unusual achievements of RCA engineers in a manner most likely to enhance their prestige and professional status.

Electro-optics

Electro-optics is one of the twentieth century's most important and pervasive technologies.

The most obvious and well publicized electro-optic system is television, and the enhancement of this system has been a major RCA thrust for at least three decades. Creative engineering efforts have extended our electro-optics capabilities far beyond the entertainment sphere to allow man to "see" inside the human body, at the bottom of oil wells, far outside the solar system; and in almost total darkness.

But well beyond extending man's vision, electro-optics has opened important new vistas in the transmission, processing, and storage of information by using the laser's ability to concentrate beams of coherent optical energy. For example, optical fibers promise to replace both wires and radio waves used for communications; optical memories are being pushed toward packing densities approaching the wavelength of light; and integrated optical devices have been developed to access these memories.

Finally, the exploitation of solar power in answer to our energy needs will define an even further expansion of electro-optics.

RCA has been, and continues to be, a leader in electro-optics technology and products, and the growth in applications seems certain. Thus, our broad technological base, coupled with sound business judgement, offers excellent opportunities for us in the future.

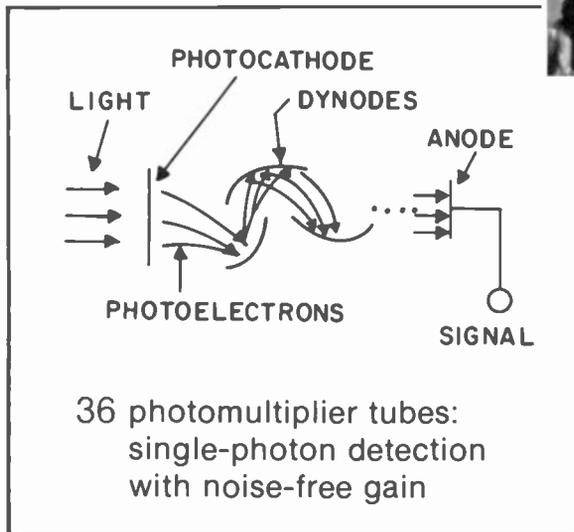
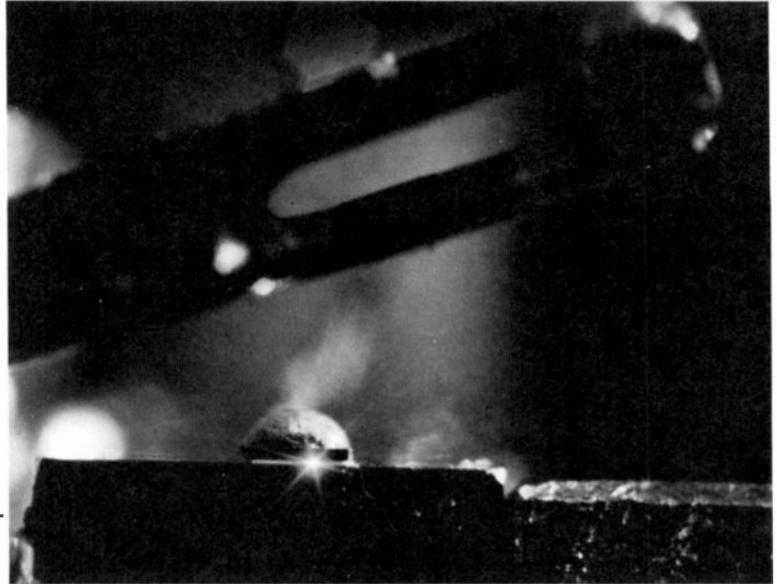
R. E. Simon

Ralph E. Simon
Division Vice-President
Electro-Optics and Devices
Solid State Div.
Lancaster, Pa.



electro-optics—

it all starts with sources, detectors, and media



16 semiconductor lasers find
communications applications

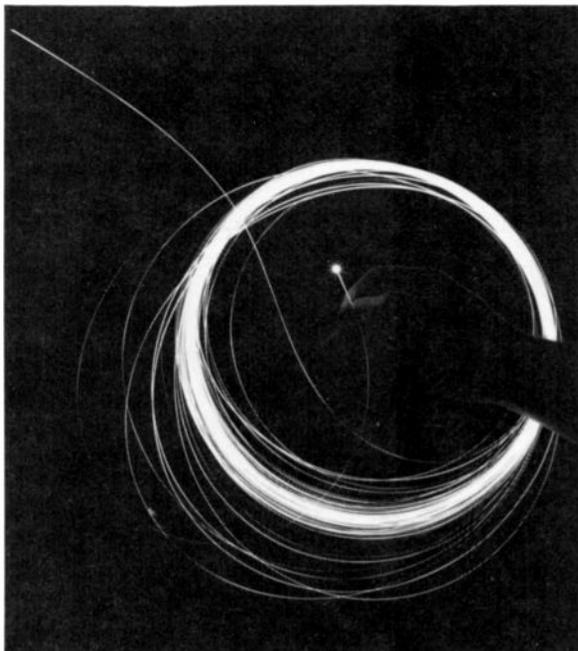
Learn

78 design digital filters

Upcoming issues

Electro-optics is too big a subject for one issue, so our Apr/May issue will be about E-O systems—image tracking, optical recording, the hand-held laser rangefinder, and CCTV surveillance, among others.

Microprocessors are the biggest thing to hit electronics since the transistor. Our next issue (Feb/Mar) tells how they'll affect your particular line of work and how to get involved, besides giving a host of applications and information on RCA's COSMAC microprocessor.



56 fiber-optics passes its first
large-scale test

RCA Engineer

R.E. Flory	4	on the job/off the job Electronic astronomy
		engineer and the corporation
O.F. Whitehead D.E. Hutchison	8	Technical information: where to get it
H.E. Haynes	14	Where the electro-optics action is at RCA
	16	light sources
I. Gorog	18	Light-generating devices: an introduction
K.G. Hernqvist R.W. Longsderff	19	Inexpensive He-Ne laser tube construction
H. Kressel H. Lockwood M. Ettenberg I. Ladany	22	Light sources for fiber-optic communications
I. Gorog P.V. Goedertier J.D. Knox I. Ladany J.P. Wittke A.H. Firester	29	Applying injection lasers to information scanning
		light detectors
E.D. Savoye	36	Light-detecting devices: an introduction
R.G. Neuhauser	40	The silicon-target vidicon
G.A. Robinson	47	The silicon intensifier target tube: seeing in the dark
R.J. McIntyre P.P. Webb	52	Avalanche photodiodes: no longer a laboratory curiosity
	56	light transmission media and devices
D.G. Herzog	58	Optical transmission media and devices: an introduction
J.P. Wittke	60	Optical-fiber communications links
D.A. de Wolf	64	Optical propagation through turbulent air
J.M. Hammer	71	Integrated optics and optical waveguide modulators
		special—short course
L. Shapiro	78	Digital electronics, Part III—Digital filters
		general interest papers
R.E. Simonds N.B. Mills J.F. Eagan	91	Operation of the RCA Frequency Bureau
R.E. Simonds	94	How the Communications Act affects you
J. Lewin	98	Ground-control system for Satcom satellites
		departments
	105	Pen and Podium
	107	Patents
	109	Dates and Deadlines
	110	News and Highlights

on the job/off the job

Electronic astronomy

R.E. Flory

Star gazing and planet watching can be made more pleasurable by using photography and electronics to add new dimensions to amateur astronomy.

An engineer working in television has a natural conditioning in the grand-sounding concept of the "extension of man's vision." A life-long interest in photography and a familiarity with television equipment combined to lead me to an interest in astro-photography, often with an electronic twist.

What are the characteristics of astro-photography that make it different from Sunday afternoon snapshot shooting or the slide photography of your last vacation trip? Generally, astro-photographic subjects are very dim and/or are apparently very small. The apparent small size is, of course, more a manifestation of the great distances involved. In my pursuit of these elusive images, I have used three basic techniques: television, image intensification, and conventional photography. Each of these techniques seems to have a place in assisting me to get records of different classes of astral objects.

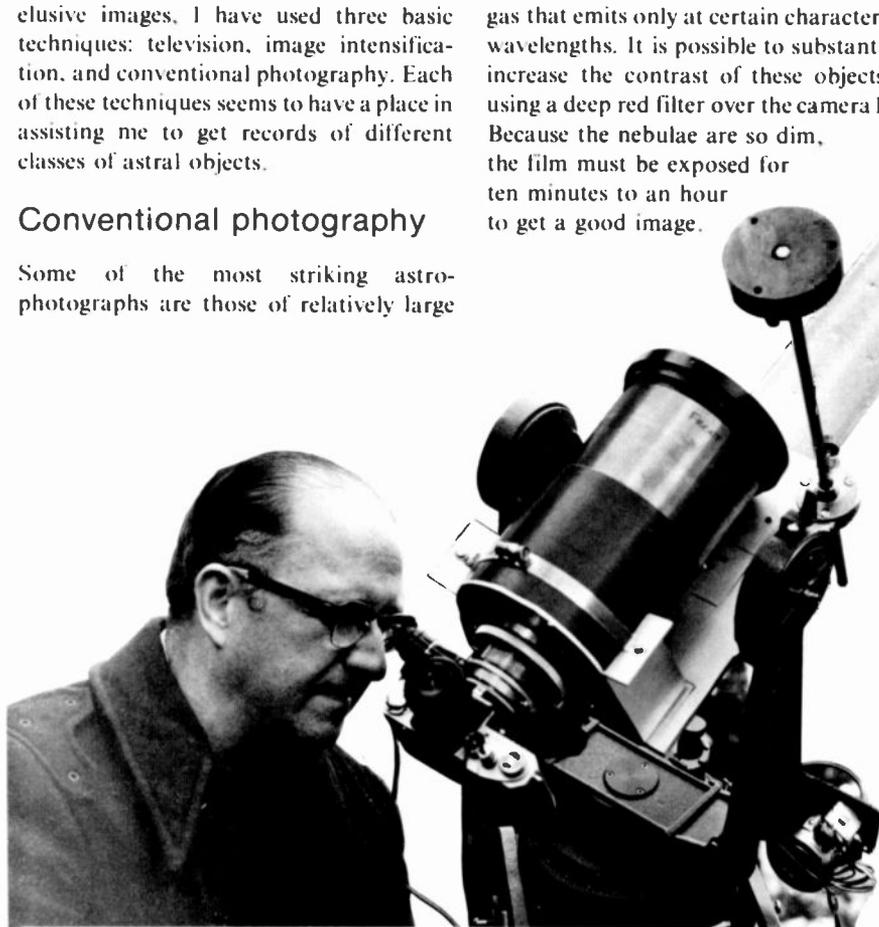
Conventional photography

Some of the most striking astro-photographs are those of relatively large

astral objects of a size that can be discerned with the naked eye or with binoculars. Taking such photos involves the use of conventional photographic optics and long exposure times to make a record of very dim objects. One such class of objects are diffuse galactic nebulae. They are often several degrees in extent but are so dim that they are seldom seen in populated areas, if at all. In central New Jersey, it is rare to experience a night so dark that you cannot see your hand before your face. Yet, extreme darkness is a requirement for good photographs of these nebulae.

Fortunately, there is a large class of these nebulae that emit only red light because they are composed of glowing hydrogen gas that emits only at certain characteristic wavelengths. It is possible to substantially increase the contrast of these objects by using a deep red filter over the camera lens. Because the nebulae are so dim, the film must be exposed for ten minutes to an hour to get a good image.

The choice of film for such photographs becomes very important. Conventional photographic film has a significant defect that prevents satisfactory use for many astro-photos. This is what is called reciprocity failure. Simply stated, it is that the product of exposure time and intensity is not a constant over all light intensities. The result is that such films as TRI-X cannot record objects so dim that they require more than about five minutes exposure. Special films have been developed that overcome this problem but at the expense of more picture grain and poor keeping properties. These films have also been made with more red sensitivity and higher contrast than usual; both characteristics aid in recording these dim objects.



Bob Flory joined RCA Laboratories in 1959 and has worked mainly on television and imaging and display systems. He participated in the early VideoDisc and Holotape projects. He was involved in the design of the RCA Space Mountain exhibit in Walt Disney World and in the installation of an RCA television system on Kish Island, Iran. He is presently working on advanced video recording techniques, one of which is the high-density video recording system for tv broadcast use. Mr. Flory received an RCA Laboratories Achievement Award, and he shared the 1972 David Sarnoff Team Award in Science. In 1976, he was appointed Fellow of The Technical Staff.

Contact him at:
**Communications Research Laboratory
RCA Laboratories
Princeton, N.J.
Ext. 3320**

Reprint RE-22-4-13

Final manuscript received October 14, 1976.



Fig. 1
"North American" nebula photographed with Kodak 130aE low-reciprocity-failure film using a Wratten 25 filter, a 300-mm f/2.5 lens, and with an exposure time of 20 minutes. The picture is approximately four degrees wide.

Fig. 1 shows one such object, photographed in September 1976, when it was straight overhead in the constellation of Cygnus. This is the so-called North American nebula and its companion, the Pelican. This picture was made with twenty-minute red-light exposure, on 35-mm film, using a 300-mm focal length lens having an f number of 2.5. This is a rather impressive lens taken from a military surplus reconnaissance camera. However, 135-mm or 200-mm camera lenses mounted on a 35-mm camera can produce good results.

To achieve clear photographs with such long exposures, it is necessary to cancel out the apparent movement of sky objects due to the earth's rotation. This is done by mounting the camera on a platform that rotates about an axis parallel to the earth's axis once in twenty-four hours. The photo

on the first page of this article shows the equipment I use for nebula photography. Because it is impossible to achieve sufficient mechanical perfection in gearing and to eliminate errors due to atmospheric refraction and mislocation of the camera's axis, it is necessary to monitor the exposure through an auxiliary telescope. This telescope must be of high magnification, but quality is relatively unimportant, as it is used only to keep a "guide" star centered on crosshairs in the eyepiece.

One recent technological improvement here is the use of an LED to illuminate the crosshairs. The star selected for guiding is often rather dim, and it is a real observational advantage to have the cross-hair image appear in red and the star image in white. The correcting movements are made by changing the rate at which the

clockwork rotates the platform axis. This is usually done by means of a variable frequency oscillator driving a synchronous motor.

Image intensification

Many sky objects are somewhat brighter than the diffuse nebulae but are very much smaller. These include the globular clusters and planetary nebulae. The former objects are clusters of stars so close together, and so far from us, that they appear to be packed into a ball. The planetary nebulae often have the form of a disc or a ring and have an appearance in the telescope somewhat similar to the planets of our solar system. All of these objects are many thousands of light years away from earth. A light year is defined as that distance that light travels in one year's time and is almost 6×10^{12} miles. At such great distances, these objects become very small.

10 arc-minutes

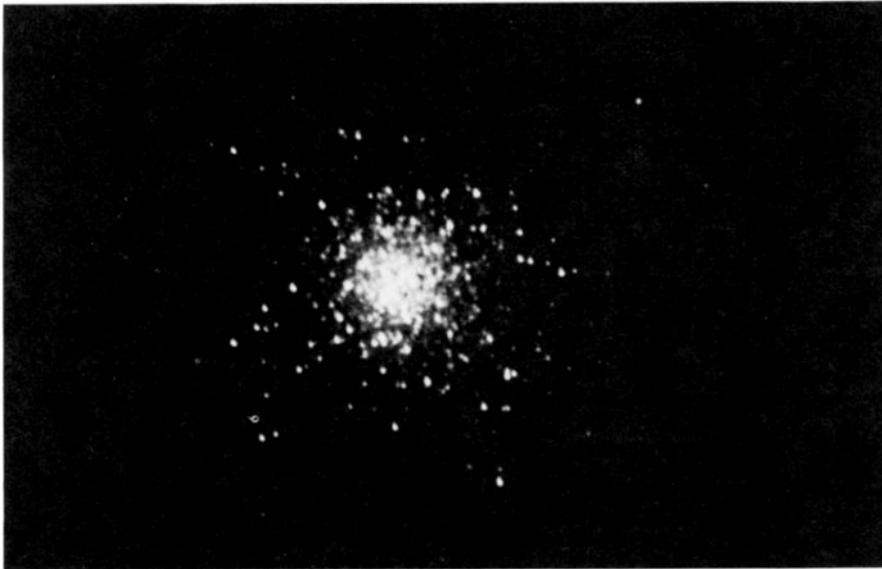


Fig. 2
Globular cluster in Hercules taken in a one-second exposure on Kodak TRI-X film, using an image intensifier. The telescope is a Celestron with a 2000-mm focal length at $f/10$. Most stars fall inside a ten arc-minute diameter sphere.

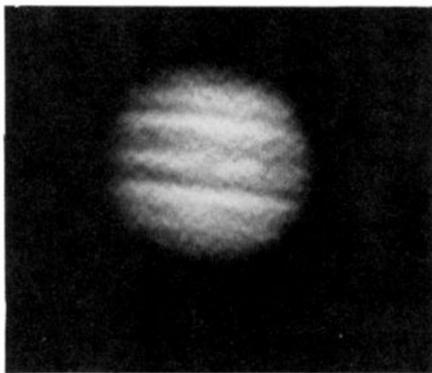


Fig. 3
Planet Jupiter photographed from a tv monitor in 1/30 second on Kodak TRI-X film. The camera used contained a 2/3-inch silicon vidicon. The telescope is a Celestron with a 4000-mm focal length at $f/20$. The planet has a diameter of about 45 arc-seconds.

A well known example of a globular cluster is the one in the constellation of Hercules. Most of the stars in this cluster fall within a sphere having an angular diameter of ten arc-minutes. Such an object is small enough to require a true telescope to resolve. I use a 2000-mm-focal-length Cassegrain mirror system of 200-mm aperture made by Celestron, Inc. Even though the object is a cluster of stars, the great distance and the f number of 10 for this telescope combine to make it a dim object. The manufacturer of my telescope has published a picture of this object which required a ten-minute exposure on the super-fast, low-reciprocity-failure film. I have used a three-stage RCA image in-



Fig. 4
Planet Saturn taken from tv monitor with same equipment and conditions as in Fig. 3. The planet diameter is 20 arc-seconds.

tensifier and a $1:1/f/2$ relay system to image the intensifier output onto amateur TRI-X 35-mm film. With this equipment, I have obtained an identical result with only a two-second exposure.

Such short exposures are a great advantage with small objects because it is extremely difficult to make the precise position corrections needed for long exposures with long focal lengths. Fig. 2 shows the Hercules cluster as photographed with the image intensifier. It "boggles the mind" to think that after the light traveled 22,000 years (1.32×10^{17} miles), I only had to admit light to my telescope for a few seconds to see the object in some detail.

Television

For small objects of reasonable brightness, television has been a convenient viewing and recording aid. At the very high magnifications needed to view the planets or the moon, vibration of the telescope is a serious problem, either in the case of simple observation or photography. In fact, the operation of the camera shutter is often sufficient excitation to cause vibrations which may last an appreciable part of the many-second exposure required to photograph the planets with any detail.

Objects as small as a few arc-seconds are hard to observe because of air turbulence—observers often sit for hours to get lucid glimpses when the air becomes calm for a few seconds. Television has been a useful tool for planetary observation because of the ability to operate the telescope completely "hands-off." When used in conjunction with a video tape recorder, it is possible to accumulate an evening's observation, and then, in the comfort of an armchair, playback the tape to watch the clear views of the planet.

The surface features of the planets tend to be of rather low contrast, and the resolution roll-off of a telescope is a gradual function. It is possible to apply simple video processing, such as contrast enhancement or video peaking, to improve the images. Figs. 3 and 4 show planetary images photographed from a nine-inch monitor. Both have had contrast enhanced considerably by the simple expedient of operating the contrast control higher than appropriate for a linear display. Scan lines are oriented vertically in these images to allow simple video peaking to enhance the largely horizontal detail in both planets.

The telescope was operated at $f/20$ (4000-mm focal length) by the use of a 35-mm camera "tel-extender" adapter. The vidicon is a 2/3-inch silicon tube mounted in a home-made camera. The camera was made in two pieces to allow a minimum weight to be imposed on the telescope. A further sensitivity gain was achieved by turning camera beam current on only on every sixth tv field, allowing light integration for the rest of the time. The raster vertical scan was collapsed by a factor of two to get more scan lines in a given image size on the vidicon. The monitor scan was similarly reduced to achieve the proper aspect ratio.

At the same scale of magnification, the

moon is an especially interesting object. I have often found willing audiences for television views of the moon and brighter planets. With the moon, it is interesting to "travel" around by allowing the earth's rotation to scan the camera view over the moon's surface. Figs. 5, 6, and 7 are monitor photographs made from a video tape recording of a quarter moon. The camera was again the 2/3-inch silicon vidicon, but with normal 60-field scan.

My most ambitious recent project was to televise the eclipsing, or occulting, of a star by the planet Mars. On April 7, 1976, Mars passed in front of a star in the constellation of Gemini. We are accustomed to think of planets as fixed in the field of stars on a given night. We are aware, however, that they move over a period of days. An object as small as six arc-seconds, as Mars was at this time, does not often pass very close to a star, so that we cannot see its motions minute-by-minute.

On this occasion, however, the star approached and passed behind Mars, all within half an hour. I planned to video tape the entire sequence, and then make a time-lapse motion picture sequence that would compress the event from about one half hour to about one half minute. Mars' small angular diameter of six arc-seconds dictated a focal length of about 10,000 mm and an *f* number of 50. For such a large *f* number and such a dim object, a super-sensitive camera was necessary. I had in my possession an RCA TC1030 silicon intensifier target camera which I was using on a research project. Rather than let it sit idle on the evening of April 7, I installed it on my telescope, with the impressive results shown in Fig. 8. In this figure I have superimposed six negatives, taken from the monitor during video tape playback. The somewhat erratic shape of both objects is due to air turbulence. Mars is not a round object in this view, being in a "gibbous" phase, not fully illuminated in earth view. The ability to see this event is impressive if one considers that the objects are the equivalent of a basketball and a baseball viewed from about ten miles away. (Of course, we are accustomed to viewing the progress of baseballs and basketballs at much greater distances with the aid of television.) The gain over photographic exposures is about 60:1, since one second would be a reasonable photographic exposure, as compared to 60 per second tv images.



Fig. 5
Short mountain range on the moon photographed with equipment and conditions described in Fig. 6. Picture width is 5 arc-minutes.

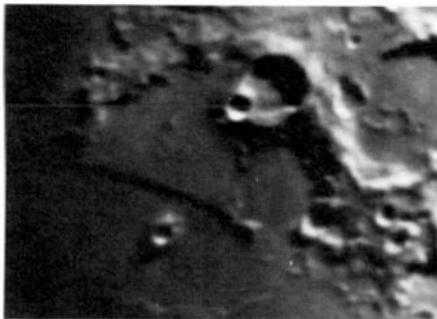


Fig. 7
Moon's straight wall, which is about 60 miles long and 800 feet high. It is probably due to a collapse of the floor of a very old crater. Photographed with same equipment and conditions as in Fig. 6. Picture width 5 arc-minutes.

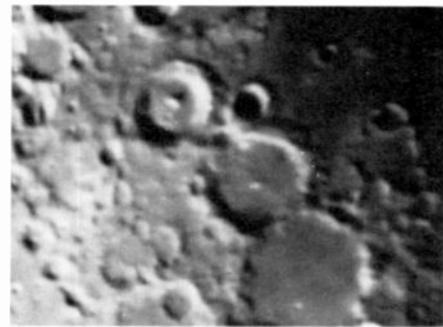


Fig. 6
Moon photograph made from video tape shows three large craters near the center of the moon's face. The largest of the three attached craters, Ptolemaeus, is 115 miles across. The picture was made with TRI-X film in 1/30 second. Equipment used include a 2/3-inch silicon vidicon camera and a Celestron telescope with a 4000-mm focal length, at *f*/20. Picture width is about 5 arc-minutes.

Summary

Astronomy is a character-building hobby. It teaches patience and equanimity in the face of disappointment. Some unique events occur once in a lifetime. Other views are available only when the earth, sun, and moon cooperate. All of them are available only when the weather and family plans allow. I have found that I have been able to use electronic imaging aids to enhance the limited time and non-optimum location available to me in pursuit of this rewarding hobby.

Bibliography

Astronomy magazine, published by Astro Media Corp., 411 E. Mason St., Milwaukee, Wisc., has been running a series, "Photography in Astronomy," since April 1976. The following, applying to this article, is one in the series:

Davis, J., Tobin, W., and Eaton, J. "Red light sky photography." *Astronomy*, Vol. 4, No. 8 (Aug 1976)

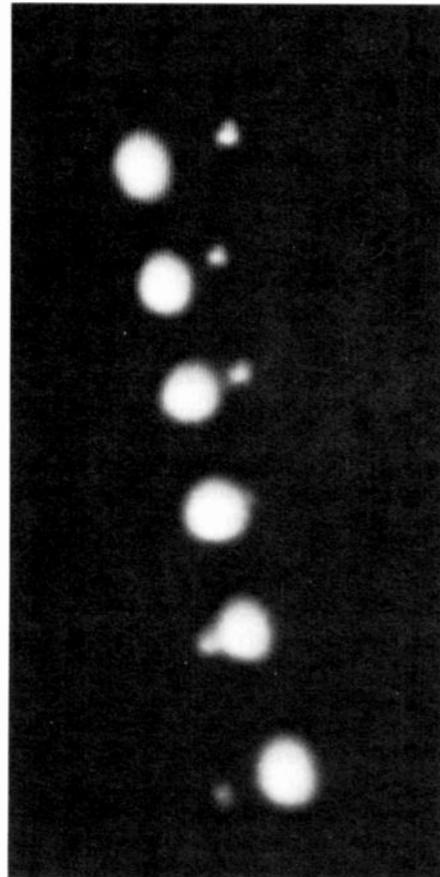


Fig. 8
Time lapse sequence shows occultation of a star by Mars. Six monitor photos, taken from a video tape playback, are superimposed to illustrate eclipse sequence. Exposures are about two minutes apart. Equipment used included a silicon intensifier target camera and a Celestron telescope with a focal length of 10,000 mm and an *f* number of 50. Photos were taken on TRI-X film at exposure times of 1/60 second.



Technical information: where to get it

The information explosion has produced such a vast number of sources that RCA's libraries have had to add to their traditional services to include on-line computer searches and resource sharing.

O.F. Whitehead
D.E. Hutchison

Technological events that will shape the world tomorrow are being announced in the literature today. Because of the tremendous amount of work that gets reported, engineers and scientists are probably finding it difficult to keep up with the literature. In RCA's libraries alone, there are over 1300 different periodicals, and this is just the tip of the iceberg. These same libraries have thousands of books and reports as well as access to the documentation of thousands of engineers and scientists in the U.S. and abroad.

Information is the basic commodity of the RCA Libraries. Such a collection, however, is of little value unless it is organized to enable you to find information when you want it quickly. Read on, while we tell you how you can find the information you need—when you need it.

Answers to questions

The key to the contents of your library is, of course, the card catalog, but if you can't find what you want, or you are not familiar with the card catalog, don't hesitate to ask your librarian for help. Librarians are

accustomed to answering questions. Here are a few of the more unusual ones we have answered at the Camden library during the past year:

- What is the length of the boundary between East and West Germany?
- When were the first radio sets marketed?
- How do you spell the word from a popular movie—super something or other?
- How do you apply for a Rhodes scholarship?
- What is the name of the plane geometric figure that looks like a four-leaf clover?
- What was the price of gold on February 1, 1975?
- Can you identify the source of "Selling is the art of overcoming objections"?
- Where can I get an illustration of a spacecraft passing through the Van Allen radiation belt?

Yes, your librarian can probably come up with answers to all kinds of questions you might need answered, but basically the purpose of a company library is to provide the right information, in the right place, at the right time. This requires a knowledge of

Reprint RE-22-4-19
Final manuscript received October 1, 1976.

RCA's programs and products, careful selection of resource materials, and speedy well-organized services to place these documents in the hands of readers.

Information services

The activities of a library are so varied in the fields of science and technology that the librarian must be capable of carrying on and supervising diversified operations. Where the size of the staff is minimal, which is the case in many of RCA's libraries, the librarian executes many operations. Some of the directly user-oriented ones are:

Resources: Selection and purchase of books, periodicals, conference proceedings, government reports, and other publications.

Services: Supervising readers' services, reference requests, and interlibrary loan. Executing literature searches. Organizing non-book files, e.g., company reports, contract reports, NASA literature, federal and state documents, and engineering notebooks.

Publications: Compiling Tables of Contents of current periodicals and Library Bulletins of recent acquisitions, including book reviews.

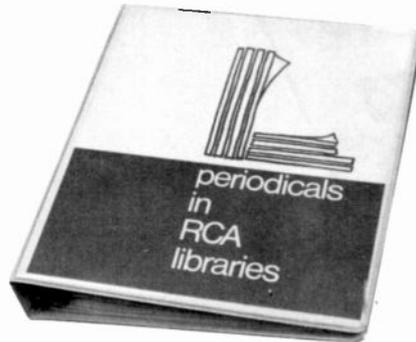
These main responsibilities are required in almost every library. Where staff assistance is available, providing these services is not a problem and additional services may be rendered.

New services

While the services listed above have been available for many years, the nature of the corporate library is changing, and RCA libraries will need to change, too. Automation is making information more accessible, and with the increased volume of information available, literature searching will, of necessity, need to be done with computers.

The library profession has recognized the importance of library cooperation and resource sharing, and RCA's libraries are also seeing this need. With the increased cost of journals and conference proceedings, it is no longer possible for every library to subscribe to every information source of possible interest to their readers. But, through cooperation with other RCA libraries, expensive acquisitions can be shared. In spite of budgetary restraints, our libraries will remain committed to providing the same service, but by doing it with more borrowing and less purchasing. It was this need that prompted the compilation of *Periodicals in RCA Libraries* (described later), and RCA Technical Information Systems is planning to publish similar compilations as needed.

The need for literature review before beginning research is vital and well recognized within RCA, but also evident within the company is the notion that engineers and scientists who spend very much time in the library are probably wasting time or sleeping! This probably springs from the work ethic of our Puritan heritage and the fact that reading is usually done for pleasure. Management apathy has sometimes placed a professional librarian in the role of "keeper of the books" rather than "information specialist." Some re-evaluation of the library's place in the corporate structure is beginning to take place in industry, and corporate libraries are being called upon to meet the demands made by their users. High-technology companies must keep abreast of scientific and engineering developments but, oftentimes, the corporate library is relegated to an insignificant place in the corporate structure and its potential services are overlooked. You can assist your librarian and increase the value of library services by expressing your needs to your librarian and to your management.



If it's not in your library, this compilation will tell you which of the other RCA libraries may have that periodical.

On-line search systems

The escalation of publications in the fields of physics, chemistry, and electronics has made machine-readable data bases a necessary ingredient in handling this volume of contemporary literature effectively. Approximately 200 such data bases are presently available, a number of which would adequately meet the needs of RCA's engineers and scientists. Access to these programs at RCA has just begun with our entry into two such search systems; librarians are training in their use now and the systems should be on-line shortly.

The two systems presently available are System Development Corporation's *ORBIT III* and Lockheed Missiles & Space Information Systems' *DIALOG*. Through these systems, you can search files representing millions of articles, reports, books, current and completed research projects, and other kinds of information. You may review the results of the search by having part or all of the citations printed on-line, or in the case of long bibliographies, you may have the results printed on a high-speed printer and airmailed to you on the same day. In most cases the output includes an abstract as well as a complete citation. Should you want a copy of the complete document, this would be obtained through normal sources.

Some of the data bases that can be accessed through these systems include:

CHEMCON

Chemical Abstracts Condensates, prepared by Chemical Abstracts Service.

Olive Whitehead has administered the RCA Camden library since 1969. Her professional duties there include selecting new books, reports, and journals, in addition to reference work and literature searching. She is also highly active in the Philadelphia Chapter of the Special Libraries Association.

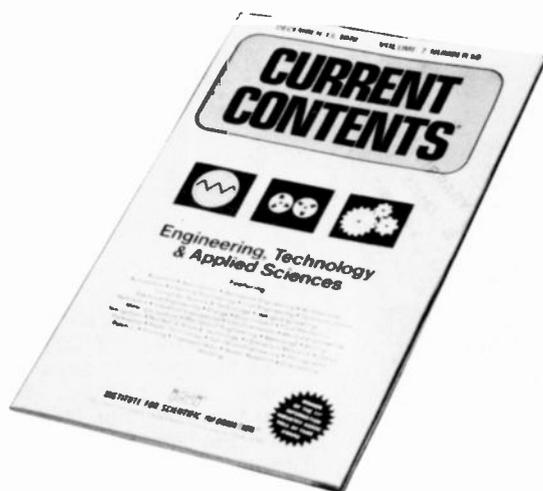
Contact her at:
Engineering Library
Government Communications Systems
Camden, N.J.
Ext. PC-3488

Doris Hutchison joined Corporate Engineering in 1970 and is currently Administrator of Technical Information Systems. This group is concerned with publishing *RCA Technical Abstracts* and is also responsible for providing services to the RCA Library Network, including publication of a *Union Catalog of Periodicals in RCA Libraries*.

Contact her at:
Technical Information Systems
Corporate Engineering
Cherry Hill, N.J.
Ext. PY-5412

Doris Hutchison (left) and Olive Whitehead at a computer terminal used for on-line literature searching.





A quick scan of *Current Contents* will tell what articles may interest you in this week's publications.

Covers biochemistry, organic chemistry, macromolecular chemistry, applied chemistry and chemical engineering.

COMPENDEX

Prepared by Engineering Index. Corresponds to Engineering Index Monthly. Covers civil-environmental-geological engineering; mining-metals-petroleum-fuel engineering; mechanical-automotive-nuclear-aerospace engineering; electrical-electronics-control engineering; chemical-agricultural-food engineering; and industrial engineering, management, mathematics, physics, and instruments.

SSIE

Prepared by the Smithsonian Science Information Exchange. Covers on-going and recently completed research in the life, physical and social sciences—both basic and applied research projects.

INSPEC

Prepared by Institution of Electrical Engineers. INSPEC data bases include: *Physics Abstracts*, *Electrical and Electronic Abstracts*, and *Computers and Control Abstracts*.

SCISEARCH

Produced by the Institute for Scientific Information. Over 850,000 citations from 2,500 periodicals in the physical and life sciences.

No library can subscribe to all the journals.

One of the most valuable resources in the RCA libraries are the periodicals received by paid subscription or by controlled circulation.

At some locations, *Current Contents*, a weekly compilation of articles in various journals, provides an easy scanning

medium for recent issues. The Institute for Scientific Information compiles the weekly issues of *Current Contents* from advance copies of title pages of periodicals supplied by the publishers, or from the title pages of journals obtained by subscription. *Current Contents* appears in several editions of related subjects. The two editions closely allied with RCA's interests are *Engineering, Technology and Applied Sciences* and *Physical and Chemical Sciences*, each covering approximately 700 journals.

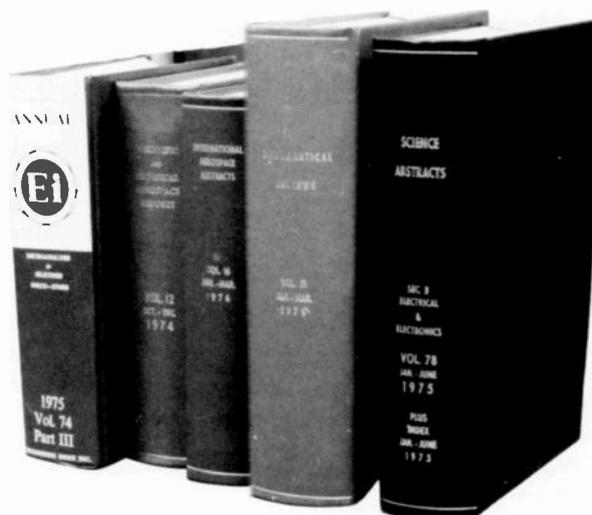
Several RCA libraries duplicate tables of contents from current issues of magazines as they are received and compile a "Tables of Contents" which is circulated to interested persons.

It is impossible for the library at each RCA location to subscribe to all the journals that might carry articles of interest. To assist in locating periodicals quickly, RCA Technical Information Systems has prepared *Periodicals in RCA Libraries*, a union catalog of subscriptions and holdings of journals and selected serials in each of the RCA libraries. If there is no source within RCA for a journal, there are several tools that provide listings of periodical holdings in other organizations. (Consult your local library for titles.)

Interlibrary loan

"Neither a borrower, nor a lender be."

Despite Shakespeare's admonition, nearly every RCA library must participate in interlibrary loans to supplement its resources. Cooperation between RCA locations is excellent, borrowing and lending books, conference proceedings, reports, and copies of periodical articles as often as possible. According to circumstances, university libraries, state libraries, area reference libraries, and special libraries may be contacted for literature to supplement RCA's resources, and they may in turn call upon any RCA library for assistance. There is a great deal of variation in regulations within libraries controlling interlibrary loans, which will govern how and where the interlibrary loan request is submitted. In placing a request for a loan, it is necessary to have accurate and complete references; otherwise the request may not be honored.



Indexing sources provide a quick route to hundreds of thousands of articles. Specifics are on next page.

Book selection

The right book, in the right place, at the right time.

The librarian selects the "right book" by scanning publishers' announcements, book reviews in periodicals, and the *Weekly Record*, which lists recently-published American books. But most important are the recommendations from RCA personnel on which titles should be added. While building their resources to fill present and anticipated future needs, all RCA libraries acquire numerous conference proceedings, reports from various government agencies and scientific organizations, and the periodicals most closely allied to their interest. Announcements of new acquisitions are made through *Library Bulletins*, displays of book jackets, and "new book" shelves.

Indexing sources

Have you ever tried to locate an article that you read 2 or 3 years ago in an IEEE publication?

Or was it in *Bell Systems Technical Journal*? But now you need that article quickly and don't have time to go back and look at journals until you find the right one. Here are some of the indexing sources that you might consult:

IEEE Author and Subject Indexes to Publications.

Recent annual issues cover all IEEE technical periodicals, technical items from IEEE magazines, conference papers, and IEEE standards.

Engineering Index Monthly and Annual
About 85,000 items, selected from 2000 journals, are abstracted annually and classified under broad subjects with extensive cross-references.

Applied Science and Technology Index
A subject index to about 225 English-language periodicals in the fields of aeronautics and space science, automation, chemistry, earth sciences, electricity and electronics, engineering, materials, mathematics, physics, telecommunications, and related subjects.

Mathematical Reviews.

This monthly journal, which abstracts the world literature in mathematical research, is prepared by the American Mathematical Society.



Where to find what's new
in government research reports.
Biweekly publication has abstracts of all unclassified government reports having unlimited distribution.

Physics Abstracts.

An abstracting vehicle which covers the whole field of physics, drawing from works in all countries and in all languages. Publishes 85,000 items per year.

Electrical and Electronics Abstracts.

A monthly abstracting publication, prepared by the Institution of Electrical Engineers (British). Publishes 40,000 items per year.

Computer and Control Abstracts.

Includes literature from worldwide sources on computer science and control engineering.

Science Citation Index

This index, published by the Institute for Scientific Information (ISI), lists all documents which are being referred to in current literature over a specific period of time. It is arranged alphabetically by the names of the first authors of the documents being cited. Below each of the cited items is a list of the newer articles that cite them, also arranged alphabetically by the names of their authors.

Companion to the *Citation Index* is the *Source Index*, which is an alphabetically arranged author index of more than 420,000 articles and other editorial material indexed each year by ISI.

Chemical Abstracts

Under the auspices of the American Chemical Society, this publication provides excellent abstracts in English of the world's literature on chemistry and chemical engineering. In 1975, *CA* contained over 300,000 abstracts from 1000 primary journals and numerous secondary ones.

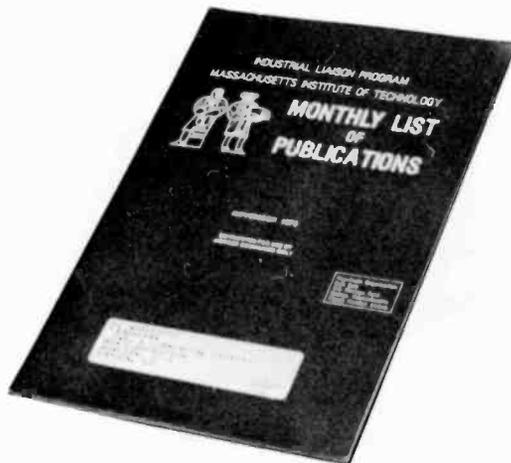
External resources

Government agencies, professional associations, and research organizations are important sources.

Several important resources and services outside the Company can be approached through RCA libraries. Among those frequently called on by RCA librarians are various government agencies, professional associations, and research organizations.

The National Technical Information Service (NTIS), Springfield, Virginia, has over 900,000 reports for sale, representing Government-sponsored research, development, and engineering reports and other analyses by Federal agencies, their contractors, or grantees. Its customers are supplied with about 4 million documents and microforms annually. Many RCA locations receive the NTIS biweekly publication, *Government Reports Announcements and Index*, which abstracts new research reports that are unclassified and have unlimited distribution. For rapid searching, this publication carries indexes in the following categories: subject, corporate author, personal author, title, contract number, and accession report number. To secure prompt service from NTIS, most RCA locations maintain a deposit account there.

The services of the Defense Documentation Center (DDC), Alexandria, Virginia, are available to RCA as contractor to various military agencies. However, access to this data bank requires registration of an active contract on the pertinent DoD form (DD 1540), commonly called "Field of Interest Register." The technical reports currently in the DDC collection total more than a million titles, of which 670,000 are under computer control for quick retrieval.



MIT research is available through the Institute's *Monthly List of Publications*.

These include more than 65,000 titles in the field of electronics and electrical engineering and 90,000 in the subject of physics. Current classified reports and unclassified reports with distribution limitations are announced in the Confidential DDC *Technical Abstract Bulletin*.

At most RCA locations, the Library is the designated point-of-contact between the Defense Documentation Center and Company personnel. Bibliographies supplied by DDC take three forms: *Report Bibliography*, prepared in response to a specific request for references to technical reports having a particular kind of relationship in subject matter; *Scheduled Bibliography*, prepared for subject areas for which requests are anticipated, and *Rapid Response Bibliographies*, prepared for user organizations having access to TELEX.

Registration with DDC also provides access to the services offered by the Defense-sponsored Information Analysis Centers and the major DoD technical libraries. There are eight DoD/DSA Information Analysis Centers that collect, store, review, evaluate, synthesize, and disseminate authoritative scientific and technical information in a format useful to scientists, engineers, and technicians. In their literature surveys, these centers review technical reports from DoD, other Government agencies, industry, and academic institutions; open literature, including foreign sources; and unpublished papers and similar sources.

Another source of Federal publications is the Superintendent of Documents, Government Printing Office, Washington,

D.C., or the regional GPO bookstores located in about 20 cities. Announcements of a portion of the output of the Government Printing Office is made in the *Monthly Catalog, United States Government Publications*. These publications may range from Abacus to Zirconium in subject, from elementary to postgraduate in scope, from pamphlets to multivolume sets in format. All purchases from the Government Printing Office require payment in advance through Deposit Accounts, usually maintained by the RCA libraries.

When the National Aeronautics and Space Administration was established in 1958, its charter stipulated that one of its functions should be "to provide for the widest practicable and appropriate dissemination of information concerning NASA's activities and their results." Out of this requirement came an extensive publication program of technical research reports, technical translations, contractor reports, bibliographies, and nontechnical descriptions of all aspects of space flights from Mercury through Skylab. Several RCA libraries receive NASA publications regularly, and unclassified documents can be purchased from the National Technical Information Service.

To provide bibliographical control for this extensive literature, NASA established an indexing and abstracting service to publish *Scientific and Technical Aerospace Reports*, frequently called "STAR." This abstracting journal, issued semimonthly, gives worldwide coverage to aerospace-report literature. By arrangement between NASA and the American Institute of Aeronautics and Astronautics, the publica-

tion *International Aerospace Abstracts* (IAA) abstracts books and scientific and trade journals, to complement STAR's report coverage.

One of the most valuable non-governmental organizations whose resources are available through RCA libraries is the Industrial Liaison Office of the Massachusetts Institute of Technology. Through this office, reports covering research at MIT may be obtained. Frequent symposia on important current studies are open to RCA personnel, and consulting services with scientists on the staff at MIT can be arranged. The Industrial Liaison Office issues a *Monthly List of Publications*, which announces and abstracts current reports on research and educational efforts in progress at MIT.

RCA's libraries

Meet your librarian—see what your library can do for you.

A recent survey of RCA's larger libraries revealed that many are plagued by similar problems—mainly, not enough staff or money to serve the engineers at their location to the extent that they would like. Depending on the requirements of the RCA division, the library staff size, and budgetary constraints, library service varies from location to location. Some libraries issue a Library Bulletin listing new acquisitions, while others display new items in the library, requiring readers to go to the library to see what is new. Some libraries circulate periodicals to those interested, while others require users to peruse them in the library, or will copy papers of interest. New acquisitions are based on recommendations of a Library Committee at some locations, and at others they are by request of the users, or at the discretion of the librarian, or a combination of all three.

If you have not been a library user in the past, start now by visiting your local library and meeting your librarian. She will be happy to show you the library, tell you what services are available to you, and help with your information needs. If your library circulates a list of new acquisitions, get on the distribution list. If not, drop in often enough to be an early borrower of new titles. Since RCA libraries do vary in scope, you will have to get acquainted with yours to find out just what services are available. Table I should give you a start in this direction.

Table I
The RCA library system has direct access to over 80,000 books and more than 300 periodicals.

<i>Location</i>	<i>Librarian</i>	<i>Principal subjects</i>	<i>Books & bound periodicals</i>	<i>Periodicals received</i>	<i>Microfilm reader</i>	<i>Microfiche reader</i>
AS, Burlington	Veronica Hsu, Adm., Library Resources	Electrical & Electronic Engineering Computer Technology Mathematics Physics Management Business	5,200	150	No	Yes
GCS, Camden	Olive Whitehead, Librarian Virginia Mattice, Asst. Librarian	Electronics Physics & Optics Telecommunications Computer Technology	13,500	250	Yes	Yes
AE, Hightstown	Mary Pfann, Librarian	Electronics and Communications Computer Science Materials Engineering Space Science Systems Engineering Mechanical Engineering Optics Mathematics Semiconductor Physics	3,660	143	No	Yes
CE, Indianapolis	Estella B. Perkins, Librarian	Electronics TV & Radio Engineering Physics Mathematics Electrical Engineering Mechanical Engineering	2,370	143	Yes	No
SSD/PTD, Lancaster	Mary Kathryn Noll, Adm., Library Services	Electronics Television Physics & Optics Chemistry Materials	6,650	140	Yes	Yes
MSR, Moorestown	Natalie J. Mamchur, Librarian	Radar & Associated Electronics Mathematics Military Strategy Business	13,500	250	Yes	Yes
Labs, Princeton	Wendy Chu, Librarian Halina Kan, Associate Librarian	Radio & Television Physics Chemistry Mathematics Electronics Solid State Physics Computer Science	30,000	275	Yes	Yes
RCA Ltd., Ste. Anne de Bellevue	Gladys Donaldson, Adm., Company Library	Solid State Physics Plasma Physics Communications Aeronautics & Space Laser Physics Business	3,500	250	Yes	Yes
SSD, Somerville	Esther Jankovics, Librarian	Electronics Physics Chemistry Mathematics	5,740	150	Yes	Yes

Where the electro-optics action is at RCA

H.E. Haynes *Starting with "talkies" in the twenties, RCA's involvement with electro-optics broadened to the point where it covers almost the entire corporation.*

"Electro-optics" is a term that can mean rather different things to different people; however, for the purposes of this issue of the *Engineer*, its scope is limited somewhat arbitrarily to the categories of light sources, light detectors, and devices and media for transmitting and controlling light. (IR and UV are included under the heading of "light.") Such areas as systems applications, electron optics, image processing, and the psychophysics of vision are, for practical reasons and with some regret, not included.

RCA came into existence nearly 60 years ago as a communications company, and it continues to be a world leader in this field. Having such a genesis, it might be surmised that its first venture into electro-optics would have been brought about by the advent of television, but interestingly this is not the case. What was probably RCA's first significant plunge into electro-optics was its role in the development of optical sound film recording, the technique that changed the movies into "talkies" almost overnight in the late twenties. An engineering group called Photophone Advanced Development, headed by the late E.W. Kellogg, was a pioneer in this technology. (Incidentally, this group has evolved, through many stages, into the Advanced Technology Laboratories in Camden.) The group represented, in a sense, a microcosm of RCA's present electro-optics activities, since these engineers brought forth innovations in all of the electro-optics categories mentioned above. Light sources, light modulators, beam deflectors, unique optical-system designs, photodetectors, photographic technology—all were essential to produce a successful motion-picture sound-recording and reproduction system.

Television, which obviously relies very heavily upon electro-optics technology, launched a revolution of much greater magnitude in the forties; here RCA's role is far better known. And then in the fifties, television experienced the full impact of color, calling for electro-optics technology of even greater complexity and sophistication. As Ralph Simon points out in the inside cover message for this issue, developments in television continue unabated.

In the more recent past, an explosive broadening has taken place in the E-O field. Completely new phenomena and devices, many of them resulting either directly or indirectly from military and space programs, have brought on this new expansion. Lasers, semiconductor light sources and detectors, charge-coupled imaging arrays, new types of camera tubes, fiber optics, holography, liquid crystals, optically active materials—the list is long and it continues to grow.

The lion's share of E-O technology applications are concerned with some aspect of information handling—sensing, displaying, transmitting, recording, processing—



Harold E. Haynes has been Staff Technical Advisor in Government Engineering since 1972, with his major work emphasis on E-O and related activities throughout the Government Systems Division. He has been with RCA since 1940, and has been engaged in numerous advanced development programs in electronics, recording systems, electro-optics, color reproduction systems and the electro-optical aspects of a precision CRT-based computerized photocomposition system.

Contact him at: **Government Engineering, Government Systems Div., Moorestown, N.J., Ext. 3843**

and so it is hardly surprising to see that electro-optics applications have found their way into a great many parts of an information-oriented company such as ours. A glance at the accompanying diagram tells just how widely the technology is being applied at RCA. The family tree at the top shows the wide range of electro-optics technology and devices that are available to draw on. Each of the organizational units listed is in some way involved with some facet of electro-optics. Upward arrows indicate those parts of RCA that are contributing to specific areas of E-O technology, while downward arrows indicate those RCA organization that are using this E-O technology to their benefit. Complicated as it may appear, the chart is still only an incomplete snapshot at one point in time, and cannot be completely comprehensive. The overall picture it imparts—one of widespread and varied activities and their interrelations—is the real message.

This issue contains articles relating to several—but far from all—of the topics shown in the diagram. It is divided into three sections on each of the E-O technology areas, each section beginning with a brief introductory summary. Whether your attention is drawn to the "big picture" or to some of its details, or perhaps both, you should enjoy this guided tour through some of the complex and changing world of electro-optics at RCA.

Light Sources

An Electro-optic light source can be as basic as sunlight, but many systems require the laser's coherent light, narrow spectral linewidth, or high source radiance.

Although there are a number of types of lasers, the operating mechanism is essentially the same for all. First, electrons are excited to higher energy levels by an external means, such as an electric field or flash of lights. Since these higher levels are unstable, the electrons spontaneously drop back to lower energy levels. In doing so, they emit photons of a wavelength proportional to their energy drop. This ordinary fluorescence provides incoherent light.

But, if these electrons can stimulate other electrons with light at a wavelength corresponding to one of their preferred energy drops, we have coherent radiation. This **Light Amplification by Stimulated Emission of Radiation** takes place in a resonant cavity called the Fabry-Perot resonator. There, the initially radiated lightwaves reflect back and forth, stimulating more and more radiation. A small fraction of this light beam can be used as the laser's output by making one end of the resonant cavity semireflective.

Laser types

Laser action can take place in many different media (it has even been reported in Jello*), but solid, gas, and semiconductor lasers are among the most common.

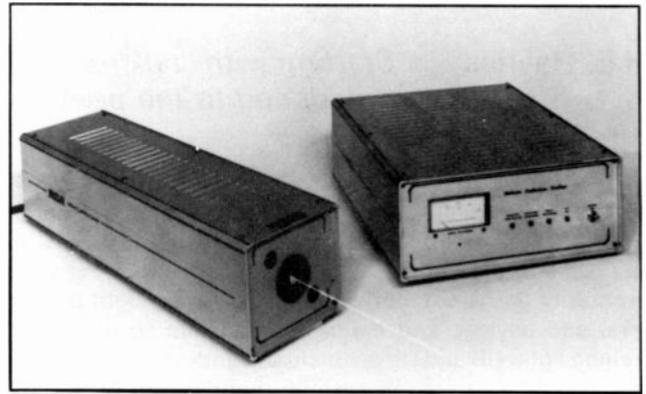
Solid lasers are more precisely "lasers based on ions in solid materials."

Solids, such as the ruby, were the first materials to show laser activity. In a ruby laser a flashlamp excites the chromium impurity ions in the ruby's aluminum-oxide host. Light is produced as the chromium ions return to their base-level energy. The ruby crystal acts as the resonating cavity; its ends are coated with reflective surfaces, one of which allows a small fraction of the light to pass through as the output signal.

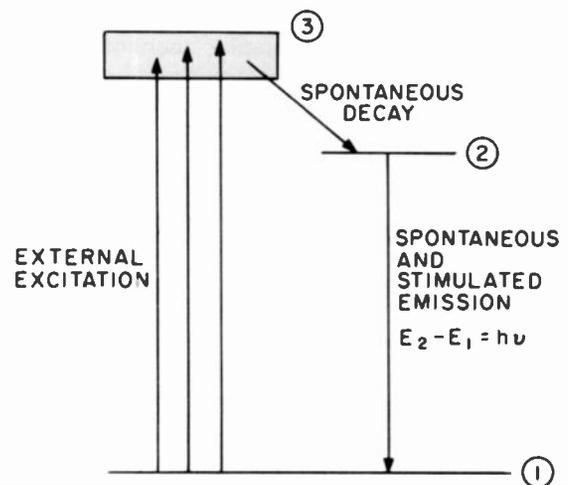
Gas lasers are low-cost, but require a high-voltage dc supply.

There are literally thousands of types of gas lasers, with the gas parameters determining the wavelength of the light produced. Not all are practical, though, and He-Cd, He-Ne and CO₂ are the dominant types. In gas lasers, the electrons are accelerated by a high voltage, usually on the order of a few kilovolts, and transfer energy to the gas molecules by collision. Mirrors at the ends of the gas tube form the resonating cavity for light amplification.

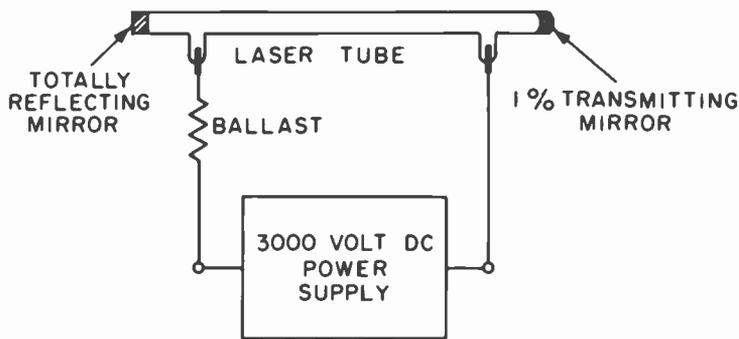
*Hansch, T.W.; Pernier, M.; and Schawlow, A.L.; "Laser action of dyes in gelatin," *J. Quant. Elec.*, Vol. QE-7 No. 1 (Jan 1971).



Size contrast: He-Cd gas laser and power supply (above); and GaAs solid-state injection laser, shown both unmounted (through the eye of a needle) and in final-product form.



Simplified energy-level diagram for a solid (ruby) laser. A flashlamp excites the ruby's impurity chromium ions to the broad energy band at 3, from which they decay to their preferred level at 2, giving off heat as they do. The ions then drop down to their base level at 1, emitting light at a frequency proportional to the energy change from 2 to 1. If the ion population is distributed among the levels properly, the spontaneously emitted photons will then stimulate additional radiation.

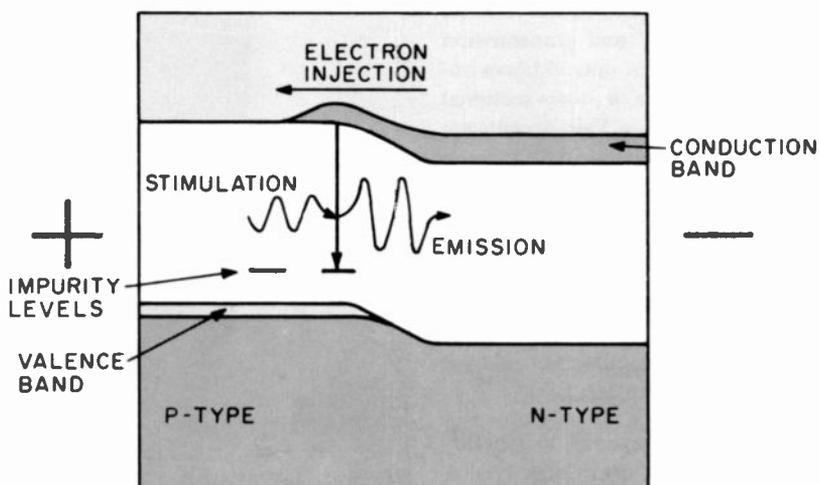


Gas laser system. High-voltage dc supply excites the gas ions in the tube; light is produced as the ions drop to lower energy levels. Hemispherical mirrors provide the resonant cavity for the stimulation of additional lightwaves.

Semiconductor lasers operate at low supply voltages and are highly efficient.

In certain semiconductors, such as GaAs, light emission takes place at the p-n junction when an external potential drives both electrons and holes into the junction. The holes and electrons recombine very rapidly by transitions from the conduction level to impurities near the valence band; each transition produces light at a frequency proportional to the energy drop.

So far, we have an LED emitting incoherent light. But, by optically polishing the material in the direction perpendicular to the plane of the p-n junction, we create a resonant cavity. Light traveling back and forth between the polished sides stimulates other transitions until all the light is oriented in the same direction. This coherent light beam then can be coupled out by making one of the polished sides an imperfect reflector.



Solid-state laser energy-level diagram. In the forward-biased p-n junction, electrons injected into the p-type region recombine with holes. Each recombination produces lightwaves, which then reflect back and forth between the optically polished sides of the device, stimulating more emissions as they do.

Laser fusion—maybe

One area of laser research outside of RCA's communication-information activities, but one that could greatly affect us all, is laser fusion. Such a system may produce "micro-explosions" of clean thermonuclear power when laser energy is focused on a target pellet. This laser-produced thermonuclear burn, which was demonstrated last summer, was the second of seven "milestones" that must be passed on the road to a commercial laser-fusion power reactor. The breakeven point, where thermonuclear output equals laser-energy input, is scheduled for 1982. Despite the tremendous progress made toward laser fusion in a relatively short time, its future remains cloudy—ERDA has said that its research funds are highly contingent upon having the breakeven point in sight by 1982.

Projected specifications for a fusion power plant of the late 1990s include a peak power of up to 500 TW (500×10^{12} W) and energy of 500-1000 kJ. For comparison, the most powerful system presently available, the Livermore Laboratories' Argus laser, is capable of focusing over 2 TW. Their new 20-beam Shiva system, scheduled for completion late this year, should be capable of delivering 10 kJ in subnanosecond pulses. However, automatic pointing and focusing systems may help multiple-beam systems become capable of the energy and power levels needed.

an introduction

Light-generating devices

I. Gorog

Electro-optics activities at RCA emphasize information-handling applications, with optical recording, reading, and communications all being actively pursued. Even though optical systems in all three of these areas had been successfully operated before the discovery of lasers, cost and performance considerations suggest that in the future virtually all electro-optical information-handling systems will use laser sources.

The key laser characteristic used in most applications is their very high source radiance. (Another important laser characteristic is the narrow spectral linewidth of the emitted radiation, which is used in many interferometric applications. But, although laser interferometric systems are important measurement tools, they represent a very small fraction of the total potential laser systems market.) A state-of-the-art 1-mW gas laser has an output-beam cross-sectional area of about 10^{-2} cm², a beam divergence of 10^{-1} radians, and thus a corresponding source radiance of about 10^5 W/cm²/steradian. The source radiance is important primarily because it determines the ultimate achievable spot irradiance (incident power per unit area).

For tv-rate information-recording, for example, the sensitivity of available recording materials implies recording spot-irradiance requirements in the range of 1 W/cm² to 10^3 W/cm². The high power density is required with the grainless recording media used for very-high-density recording applications; the low-power limit corresponds to low-resolution silver halide films. The high-power end of this range of requirements cannot be achieved even with the most powerful incoherent lamps, but low-power (≤ 1 -W output) lasers can easily be focused to provide the needed power density. Even though the power required to expose low-resolution

Optical information-handling systems need the laser's high source radiance. Both gas and injection lasers are currently being considered for these applications.

silver halide can be achieved with an incoherent source, overall system considerations make lasers the preferred alternative even in this case.

The power required for information readouts is determined by signal-to-noise considerations and is relatively insensitive to the density of the recorded information. The spot power required for reading a typical fm-coded tv signal recording is approximately 10^{-4} W. Currently operating recording systems employ spot sizes in the range of 10^{-8} cm². In order to deliver 10^{-4} W into a 10^{-8} cm² spot, a typical state-of-the-art optical system requires a source with a radiance of approximately 10^4 W/cm²/steradian. A low-power laser can easily supply this radiance. For comparison, state-of-the-art light-emitting diodes (LEDs) operate in the range of 10^2 - 10^1 W/cm²/steradian, and tungsten bulbs provide about 10 W/cm²/steradian.

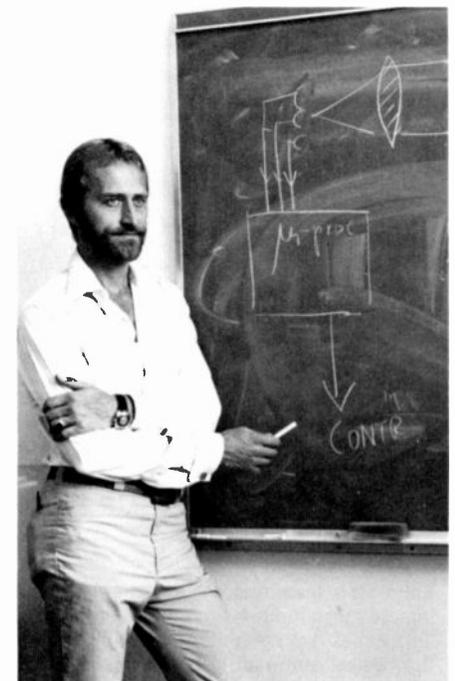
The light source requirements of optical fiber communications systems are determined by the required signal-to-noise ratio, the bandwidth, and transmission medium chosen. Typical optical fibers under development have a cross-sectional area of about 10^{-4} cm², a light-acceptance cone angle of 10^{-1} radian, and transmission losses in the vicinity of 10 dB/km. For short-distance (< 1 km), narrowband (< 10 MHz) communication applications, standard LEDs are suitable. Special-purpose LEDs have been developed for higher-capacity intermediate-range applications, but lasers must be used for long-haul, wide-band applications.

RCA's laser development and production activities include work on He-Cd, He-Ne, and CO₂ gas lasers, and on AlGaAs/GaAs injection lasers and LEDs. The application areas for the He-Cd product line are in information recording; low-cost He-Ne lasers are the primary sources in high-density information-retrieval systems, and the CO₂ lasers developed by RCA Ltd. in

Canada are being considered for line-of-sight terrestrial and outer-space communications. The GaAs lasers and LEDs are being developed primarily for fiber-optic communications; however, they are also expected to find extensive applications in information-retrieval systems.

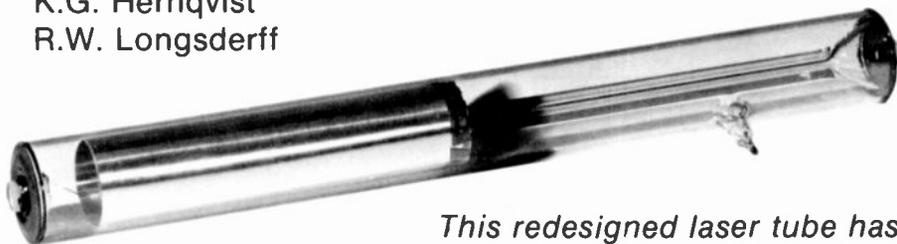
Istvan Gorog is Head of the Optical Electronics Research Group at RCA Laboratories. His main areas of interest have been quantum electronics and electro-optical systems. Dr. Gorog's research activities have included lasers and laser systems, holography, pre-recorded-video recording and playback techniques, displays, and investigation of the psychophysical aspects of electronic imaging. His current activities include product and process development related to the RCA VideoDisc, manufacturing instrumentation, and electrochromics.

Contact him at
Optical Electronics Research Group
RCA Laboratories
Princeton, N.J.
Ext. 3202



Inexpensive He-Ne laser tube construction

K.G. Hernqvist
R.W. Longsdorff



This redesigned laser tube has a potential price well below the current \$100 level. Low-cost materials and automated assembly should make this cost breakthrough possible.

The emerging large-volume laser markets will affect laser production techniques and manufacturing costs greatly. Currently, there is a great potential for using lasers in information-handling systems, such as scanners for point-of-sale checkout systems, readers and memories, character and pattern recognition systems, and copiers. The He-Ne laser, today's most widely used gas laser, sells for approximately \$100 per unit. This price must be lowered significantly if the lasers are to be produced and sold at high-volume levels.

As a source of lower-power, high-quality visible laser radiation for these applications, the He-Ne laser is unsurpassed because of its simple construction and method of operation. Using gas lasers also simplifies the end product, since they require no add-on optics or heat sinks.

Making cost predictions for the He-Ne laser is helped by the fact that it is basically similar to a gas discharge diode, such as the neon glow tube that has been manufactured for some time. It differs mainly in its optics, which require high alignment accuracy and thus new production techniques. The materials cost of a He-Ne laser can be predicted with reasonable accuracy, but the labor cost, which depends strongly on production rate and labor-saving machinery, is less certain. In this paper we describe a He-Ne laser tube constructed with the least-expensive parts and labor-saving production techniques and then review the estimated materials and labor costs for production quantities.

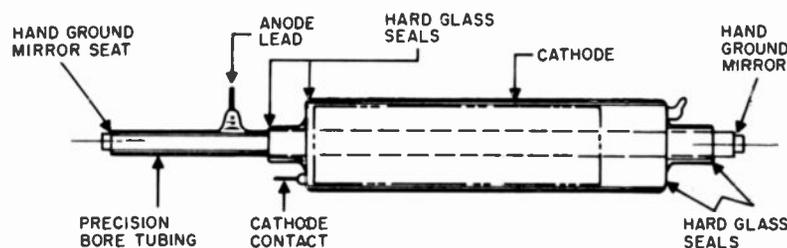
The most critical He-Ne laser production costs are in the high-precision areas associated with the optical cavity. These areas include:

- 1) mirror grinding and coating;
- 2) precision aperture for mode control;
- 3) seats for aligning the cavity accurately.

The art of mirror-making is beyond the scope of this paper; suffice it to say that

techniques for grinding and coating for mass production are well established in the optical industry. Traditionally, functions 2) and 3) have been performed by glass parts that require more expensive production methods than metal parts. In the He-Ne laser tube construction described in this paper,¹ these functions are taken over by metal parts that can be inexpensively fabricated by drilling and pressing.

Standard—



Low-cost—

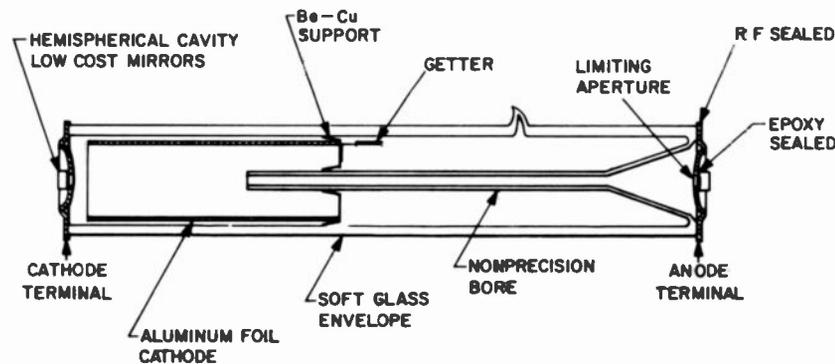


Fig. 1
Low-cost laser uses non-precision bore, aluminum-foil cathode, and rf-sealed metal endcaps on soft glass envelope. Metal endcaps and BeCu bore support represent substantial savings over the glass parts used in the standard laser.

Reprint RE-22-4-2
Final manuscript received April 13, 1976.

He-Ne laser tube construction

At the very minimum a He-Ne laser tube consists of the following parts:

- 1) a tube envelope that also serves as a platform for the cavity mirrors;
- 2) a discharge bore and supports;
- 3) cathode and anode electrodes;
- 4) an aperture for optical mode selection.

In the low-cost laser (Fig. 1), the envelope consists of a soft glass tube with metal endcaps, both inexpensive materials. Simple glass-to-metal sealing using either radio-frequency (rf) or fritting (low-temperature-melting glass) methods are applicable. The metal endcaps are supplied with spherical depressions as conventional seats for mirror-tuning and epoxy-sealing. One endcap has an oversized aperture, the other a precision aperture for mode selection. These metal parts can be manufactured at relatively low cost using standard high-volume metal-forming techniques.

Using the aperture in the endplate for mode selection makes the discharge-bore diameter uncritical, as long as it is somewhat oversized. This produces a savings over an expensive precision diameter. The bore needs to be supported at two ends; one of the supports also isolates the two electrodes electrically. This is most easily accomplished by flaring out the bore at one end and letting it become part of the end seal.

A thin-tongued beryllium-copper disc supports the other end. Another method of supporting the bore uses a number of metal discs alternately connected to prevent electrical breakdown outside the bore. This method does not require any sealing to the bore, and has already been used in He-Cd lasers.²

Since one of the metal endcaps can serve as the anode, the one tube part left to discuss is the cathode. Since the cathode power dissipation in a low-power He-Ne laser is small (less than 1 watt), a piece of thin aluminum tubing or foil may serve as a cathode if the surface area is large enough to support the emission current as normal¹ cathode drop. An alternative method uses an aluminum-film coating on the inside of the glass envelope as a cathode; depositing this film may become a part of the tube processing.

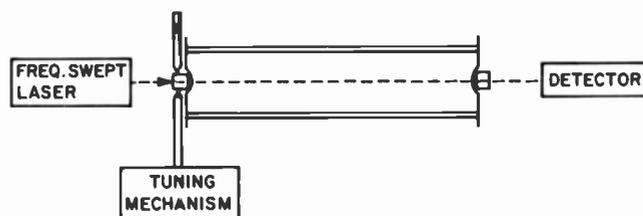


Fig. 2
Production alignment method transmits an external frequency-swept laser beam through the laser cavity, optimizing it for maximum transmission.

Tube assembly and processing

The new laser tube shown in Fig. 1 was assembled by stacking the parts on a mandrel that aligns the endplates and the bore. The endplates were then rf-sealed, a process well-suited to mass production.

For large-scale production it is more convenient to align the laser cavity and epoxy the mirrors to the endplates if the tube is not operating; techniques for doing this have been described elsewhere.⁴ This method treats the laser cavity as a Fabry-Perot resonator and optimizes the transmission of an external frequency-swept laser beam through it (Fig. 2). It is easy to visualize adapting this method for large-scale production.

From this point on, processing in mass-production machinery may proceed as previously demonstrated⁵ for small argon lasers.

Performance and life test

The major emphasis of this work was producing lasers at low cost, so the effects of low cost on performance were examined in detail.

Devices employing major components of the low-cost systems were constructed and life-tested. Both evaporated and foil aluminum were found to be as good or better than conventionally processed aluminum cathodes. Lasers using these cathodes were life-tested to more than 6500 hours with no performance degradation.

Devices employing all the aspects of the proposed low-cost laser have been con-

structed, and their performance was even better than conventional helium-neon lasers in many cases. Several of these devices were life-tested to more than 2500 hours with no failures recorded. Using low-cost lasers also lowers overall system cost because their lower operating voltage (roughly 60% of a conventional He-Ne laser's) and starting voltage in turn lower the power-supply cost.

See Table I for a summary of operating characteristics.

Estimated materials and manufacturing costs

The volume of laser production to date has permitted the use of hard-glass construction techniques. However, for low-cost construction, expensive hard-glass operations cannot be tolerated. Without high-temperature, hard-glass seals, the tube envelope becomes a simple piece of nonprecision glass tubing cut to length. Such glass tubing can be purchased in large quantities for approximately 20% of the cost of hard glass. This glass also lends itself to automated parts-making.

Using a limiting aperture eliminates the need for precision, hand-shrunk plasma

Table I
Typical operating characteristics of low-cost He-Ne laser.

Wavelength	632.8 nm
Power output	1.5 mW
Beam diameter	0.6 mm
DC operating voltage (anode-to-cathode)	1200 V
DC operating current	4 mA
Cylinder outline	1" diameter by 11.4" long

bore tubing. Again, standard low-cost soft-glass tubing cut to length provides bore materials at a cost substantially less than hard glass. Automatic rotary-sealing techniques can add the "funnel" at one end of the bore tubing for plasma isolation. These concepts allow a complete bore assembly to be produced for less than 10% of the cost of a hard-glass precision bore.

The cathode can be constructed by using aluminum foil or evaporating aluminum directly onto the inner surface of the tube envelope. Both these methods eliminate the expensive machining or forming currently used.

The mirror seats and end-closure caps are formed-metal parts fabricated with automated stamping equipment, so the cost of these parts is almost that of the material. With proper design one part can perform the functions of many others—end-closure, mirror alignment seat, electrode contact, and limiting aperture. The beryllium-copper spacer that supports the plasma tabulation at its non-flared end is also fabricated by common small-parts forming techniques.

Mirrors are by far the most expensive material/part used in the laser system. However, unique mirror construction techniques that are conducive to low-cost, high-volume production have been reported.⁶ It is estimated that using these techniques could reduce mirror costs to less than 10% of their current value.

These parts represent the major components of a low-cost laser. Miscellaneous contacts, getters, epoxy, etc. complete the parts list. In summary, vendor price-quotations based on purchasing materials for 100,000 units show that the materials/parts cost for the laser described here amounts to about \$1.15 per laser, exclusive of mirrors.

Assembly and processing costs

Using the techniques already described, the labor content needed to produce the low-cost laser was estimated in order to compare unit-cost estimates with current laser costs and the demands of high-volume markets.

The major components of the laser (plasma bore tubing, bore spacer, end-caps, cathode and tube envelope) all

come together in one basic operation. Using rf techniques along with automated annealing furnace lines permits this operation to be completed with less than 10 minutes of unskilled labor. Proper fixturing and assembly layout should hold down scrap levels to less than 5%.

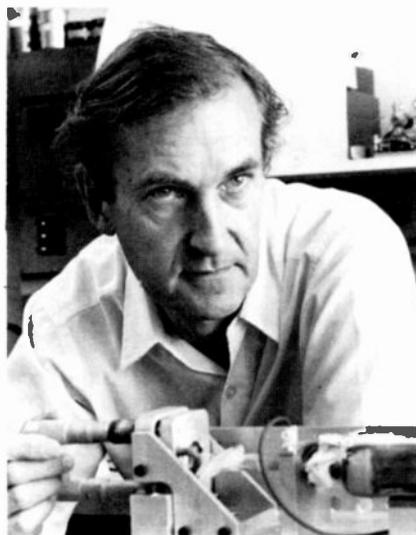
The mirror-alignment techniques previously described consume an estimated 12 minutes of labor time, although automated alignment equipment could reduce this amount further.

Exhausting and backfilling the laser assembly usually involves time-consuming bake-out procedures and gas-filling and pumping cycles. Proper equipment design, including computer-programmed processing control, should hold labor for this operation to approximately 5 minutes per laser.

The labor content of all the remaining miscellaneous operations such as test aging, inspection, parts preparation, etc. should not exceed 15 minutes of direct

Karl Hernqvist joined RCA Laboratories in 1952 and became a Fellow of the Laboratories in 1969. In 1956 he independently conceived the thermionic converter and reduced it to practice. Dr. Hernqvist is presently doing research on gas lasers and high-pressure mercury lamps. He has received five RCA Laboratories Achievement Awards and the 1974 David Sarnoff Award.

Contact him at:
**Systems Research Laboratory
RCA Laboratories
Princeton, N.J.
Ext. 2932**



labor, resulting in a total of 50 minutes of labor per laser.

Using these estimates for material and labor, and applying nominal mark-ups and profit considerations, one can see that it is possible to set an actual selling price well below current laser selling prices.

Acknowledgment

The authors wish to express their thanks to W. Lynch for his encouragement and support during the progress of this work.

References

1. Hernqvist, K.G., "Gas laser tube," U.S. Pat. 3,904,986.
2. Hernqvist, K.G.: "Vented-bore He-Cd lasers," *RCA Review* Vol. 34 (1974) p. 401.
3. Cobine, J.D., "Gaseous Conductors," Dover Publications, New York, 1957, p. 217.
4. Hernqvist, K.G. and Firester, A.H.: "Pre-alignment of gas laser cavities," *Rev. Sci. Instr.* Vol. 46, (1975), p. 1040.
5. Hernqvist, K.G.: "Advances in gas laser technology," *RCA Engineer*, Vol. 15, No. 5 (Feb-Mar 1970) p. 14.
6. Firester, A.H., Heller, M.F., and Wittke, J.P.: "Inexpensive laser mirrors," *Amer. J. of Physics*, Vol. 41, (1973) p. 1202.

Dick Longsdorff came to RCA Lancaster in 1959 and joined the Gas Laser Group in 1971 as a development engineer, aiding in the development of He-Ne laser packaging for construction-alignment purposes. He was also responsible for developing the low-cost laser. This line of responsibility continues with his present assignment—cost reduction on high-production photomultipliers.

Contact him at:
**Electro-optics Group
Solid State Division
Lancaster, Pa.
Ext. 2028**



Light sources for fiber-optic communications

H. Kressel|H. Lockwood
M. Ettenberg|I. Ladany

Research has shown that heterojunction lasers are ready for practical systems applications. Heterojunction lasers and LEDs have now been operating continuously for over two years.

System requirements for optical communications are largely determined by the available components. Although through-the-atmosphere systems have some applicability, they are generally restricted to small installations where their inherent limitations are tolerable. The more important optical-transmission media will be glass or plastic fibers, and thus, the properties of the fiber (absorption, dispersion, mechanical strength, etc.) are all-important. The other determining factors in the system are the availability of compatible sources, fibers, and detectors. Fortunately, this compatibility exists, and it is the reason why optical communications is such a rapidly expanding field. It is generally true that commercially available optical fibers have low loss and small dispersion (both modal and material) in the wavelength range of 0.8 μm to 1.1 μm . Therefore, in this article we discuss the current status of semiconductor light sources spanning this wavelength region, and since there are applications for both coherent and incoherent sources, we will treat both lasers and LEDs (light-emitting diodes).

Heterojunction structure

It is only since the advent of the first practical AlGaAs/GaAs heterojunction laser structures¹⁻⁵ and the subsequent technological progress that the potential of laser diodes for optical communications has been realized. Prior to that, GaAs homojunction lasers had such high threshold and operating currents that they were impractical for room-temperature operation. In some heterojunction devices that we describe, threshold current densities have been reduced two orders of magnitude, down to 500 A/cm² from the early homojunction values of about 50,000 A/cm². This progress has made practical a compact high-radiance continuous-wave (cw) laser that can be directly modulated to hundreds of MHz and has a power output

in the range of 10 mW. Since optical fibers with losses under 20 dB/km are now commercially available, multi-kilometer, high-data-rate transmission systems (even without repeaters) are expected to become competitive with coaxial systems.

There are also many applications for fiber-optic links over short distances and/or at low data rates (e.g., local distribution systems). The LED satisfies many of the requirements of such systems, and since it is not a threshold device, its output power is less sensitive than the laser's to small changes in operating current or ambient temperature. Much LED technology is comparable to that of lasers, so the two devices have tended to develop together. State-of-the-art LEDs now deliver several milliwatts of output power at modulation rates of 100 to 200 megahertz.

Choosing the alloy system

The most highly engineered and tested lasers and LEDs for optical communications are those derived from the ternary alloy system Al_xGa_{1-x}As. Thin layers (0.05 to 1 μm) are sequentially grown with a high degree of perfection by liquid phase epitaxy⁶ on substrates of GaAs. By varying the alloy composition of the recombination region, the emission wavelength can be varied over a useful range of 0.8 to 0.9 μm . Emission, and indeed lasing, can be achieved at still shorter wavelengths, but attenuation and dispersion in the fiber begin to become appreciable at shorter wavelengths. The purest (OH-free) fibers available have attenuation and dispersion values that decrease to negligible values in the 1.0 to 1.1 μm wavelength region, so there is a real

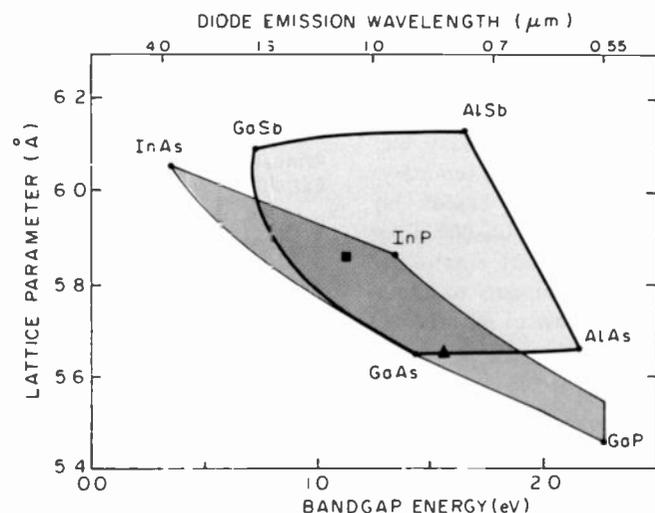


Fig. 1
Extent of band gaps and lattice parameters covered by the quaternary alloys Al_xGa_{1-x}As_ySb_{1-y} (light area) and In_xGa_{1-x}As_yP_{1-y} (dark area). The boundaries of the areas represent ternary, and the vertices binary, alloys. Two particularly useful alloys are indicated: the square locates In_{0.80}Ga_{0.20}As_{0.35}P_{0.65} the lattice parameter of which matches InP substrates and would make diodes that emit at 1.1 micrometers; the triangle shows the position of Al_{0.1}Ga_{0.9}As, which matches the GaAs lattice and emits at about 0.82 micrometer.

interest in developing new alloy systems and a heterojunction technology for emission in this range also.

In the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ system the lattice parameter varies by only 0.14% as x goes from zero to unity. Heterojunctions in this system consequently have negligible strain-induced defects and demonstrate long-term reliability. More typical of most ternary alloys, the bandgap and lattice parameter vary significantly between the extremes of the binary alloys from which they are derived. $\text{In}_x\text{Ga}_{1-x}\text{As}$, $\text{In}_x\text{Ga}_{1-x}\text{P}$ and $\text{GaAs}_x\text{P}_{1-x}$, are examples of such alloys. As a result, heterojunction structures in these materials are inevitably strained to a high degree unless compositional grading (most easily controlled in vapor-phase epitaxy) is used.

An alternative approach to obtaining efficient device performance at 1 to $1.1\ \mu\text{m}$ is to fabricate heterojunction structures in quaternary alloys, such as $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ and $\text{Ga}_x\text{Al}_{1-x}\text{As}_y\text{Sb}_{1-y}$. In these alloys, which cover the desired emission range, the bandgap and lattice parameter can be independently adjusted (Fig. 1). At some cost in simplicity, this extra degree of freedom permits the fabrication of strain-free heterojunction devices. As evidenced by a growing literature on the subject, progress toward useful $\sim 1.1\ \mu\text{m}$ lasers and LEDs is being made, although much work remains in ensuring their reliability, which is handicapped by lattice defects. Whether the complexity of fabrication will be justified by the moderate improvement in attenuation and dispersion will ultimately be determined by cost and long-term reliability.

The problem of lattice-parameter match at heterojunctions can be discussed in terms of mismatch-dislocation networks. For example, for a 1% lattice mismatch at an interface, a dislocation will be generated over approximately 100 atom planes. Since the dislocation core consists of non-radiative recombination centers, such a high dislocation density will depress the device's internal quantum efficiency. Furthermore, dislocations are not always confined to the lattice-mismatched interface, but can propagate through multilayer structures. The effect of dislocations on radiative efficiency is dramatically illustrated in Fig. 2. There, we compare a transmission electron photomicrograph of a dislocation array in a mismatched heterojunction structure and a

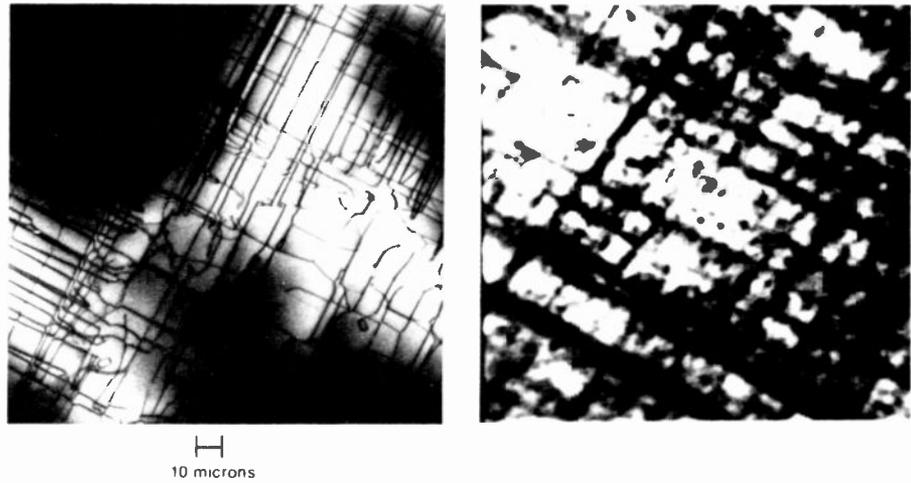


Fig. 2 **Misfit-dislocation arrays** in a compositionally graded $\text{In}_x\text{Ga}_{1-x}\text{P}$ vapor-grown epitaxial layer on a GaP substrate. The left-hand photo is a transmission electron micrograph (Ref. 7). At the right is a cathodoluminescence scan, in which dislocations near the surface appear as non-radiative regions. Photos are not to scale.

Michael Ettenberg joined RCA Laboratories in 1969 and has been involved in studies concerning factors that affect the reliability of electroluminescent diodes, improved liquid phase techniques for the synthesis of III-V compounds, and theoretical studies of the thermodynamics of crystal synthesis.

Ivan Ladany has worked in GaP, GaAlP and GaAs luminescent-diode research since coming to RCA Laboratories in 1966. Recently he has worked on improved GaAs infrared diodes, green-emitting GaP LEDs and III-V compound growth on insulating substrates using liquid-phase epitaxy. At present, he is devoting most of his time to injection laser development, having made significant contributions to the development of long-lived room-temperature cw lasers.

Reprint RE-22-4-3
This article originally appeared in similar form in *Physics Today*, May 1976

Harry Lockwood was one of the early investigators of injection luminescence in III-V compounds, in particular GaAs and GaAsP. This work preceded the injection laser development. Dr. Lockwood joined RCA Laboratories in 1969 and is currently working on material-synthesis problems and luminescent devices, including injection lasers.

Henry Kressel is presently Head of Semiconductor Device Research at RCA Laboratories. Dr. Kressel pioneered in the field of $(\text{AlGa})\text{As-GaAs}$ heterojunction devices, in particular laser diodes, and has been actively engaged in the study of devices and luminescent processes in various III-V compound materials.

Contact him at:
Semiconductor Device Research
RCA Laboratories
Princeton, N.J.
Ext. 2427

Authors Ettenberg, Ladany, Kressel, and Lockwood.



cathodoluminescence scan of a similar structure. The dark lines and spots in the cathodoluminescence micrograph are areas of low radiative efficiency, corresponding to dislocations that lie parallel and perpendicular to the plane of the surface viewed. Therefore, in designing heterojunction structures for LEDs and lasers, it is extremely important to choose either a totally lattice-matched system or remove dislocations from the active region through compositional grading.

Continuous-operation lasers

Although there are numerous potential configurations for the cw laser diodes, the symmetric double-heterojunction with a stripe contact has been widely adopted as the laser geometry most suitable for optical communications. The schematic and actual cross section of a typical laser, Fig. 3, shows the optical cavity (recombination region) defined by its higher bandgap dielectric walls as well as the outer n- and p-type GaAs regions to which ohmic contact is made.

The efficient operation of a laser diode requires effective confinement of minority carriers and radiation to the optical cavity, which is also the recombination region of the device. Both functions are provided by heterojunctions. Radiation confinement is provided by the dielectric discontinuity, while carrier confinement results from a potential barrier created by the bandgap difference between the materials that form the heterojunction. The average carrier-pair density injected into the recombination region of a double heterojunction device for a current density J is

$$\Delta N \cong J\tau_e / e d$$

where e is the electron charge, τ is the carrier lifetime and d is the width of the recombination region (i.e., the heterojunction spacing). In a typical GaAs laser diode, ΔN at lasing threshold is $\sim 1.5 \times 10^{18} \text{ cm}^{-3}$ at room temperature.⁸ To minimize the threshold current density, we restrict the recombination region's width by placing the heterojunction forming the potential barrier for minority carriers at the diffusion length from the injecting heterojunction. It is, however, essential that the heterojunction interfaces be relatively defect-free in order to prevent excessive nonradiative recombination of the injected carriers.

The nonradiative loss of carriers at an

interface is characterized by the recombination velocity S of that interface. Under the usual laser operating conditions we can express the effective recombination rate due to the presence of the interface as

$$\frac{1}{\tau_{\text{eff}}} \cong \frac{1}{\tau_0} + \frac{2S}{d}$$

where $1/\tau_0$ is the recombination rate in the absence of an interface. The internal quantum efficiency is given by

$$\eta_i \cong \tau_{\text{eff}} / \tau_0$$

so that

$$\eta \cong (1 + 2S\tau_0 / d)^{-1}$$

In typical cw laser diodes, $d = 0.3 \mu\text{m}$ and $\tau_0 = 10^{-9}$. Therefore, for an internal quantum efficiency of 50% (a reasonable lower limit), we would require $S \leq 2 \times 10^4 \text{ cm/s}$. The single most important contribution to S is from nonradiative recombination states introduced at the heterojunction due to the lattice-parameter mismatch between the two materials. If this mismatch is kept below 0.1%, S will be under $2 \times 10^4 \text{ cm/s}$ and cw operation can be anticipated. Experimental data concerning S in GaAs-Al_{0.3}Ga_{0.7}As heterojunctions indicate that $S \approx 5 \times 10^3 \text{ cm/s}$ in practical laser structures (where $\Delta a_0/a_0 \leq 0.07\%$), a value that is fully satisfactory for narrow recombination region devices.⁹

To ensure wave propagation along the plane of the junction, as well as low threshold current density and high efficiency, there must be a means of confining

stimulated radiation to the region of inverted population (or its close proximity). Two heterojunctions, as indicated in Fig. 3, provide a controlled degree of radiation confinement because of the higher refractive index in the lower-bandgap-recombination region. The fraction of the radiation confined depends on the heterojunction spacing d and the refractive index steps Δn at the lasing wavelength. In Al_{0.3}Ga_{0.7}As/GaAs structures, $\Delta n \cong 0.612x$ at $\lambda \cong 0.9 \mu\text{m}$. In general, it is desirable to equalize the refractive index steps Δn at each heterojunction to prevent the loss of waveguiding that can occur in thin asymmetrical waveguides. However, even with the symmetric double-heterojunction laser, wave confinement within the heterojunction is gradually reduced as its spacing becomes small.

Low-threshold operation

The fraction of the radiation confined to the recombination region of the double-heterojunction laser affects the radiation pattern and threshold current density. The radiation pattern (far-field intensity distribution) is affected because it is determined by the effective source size (near-field intensity distribution); the threshold current density J_{th} is determined by the gain at threshold—only that portion of the optical flux within the recombination region is amplified. Fig. 4 shows calculated and experimental values of J_{th} as a function of d for various Δn values corresponding to varying heterojunction barrier heights. The lowest J_{th} value¹¹ of

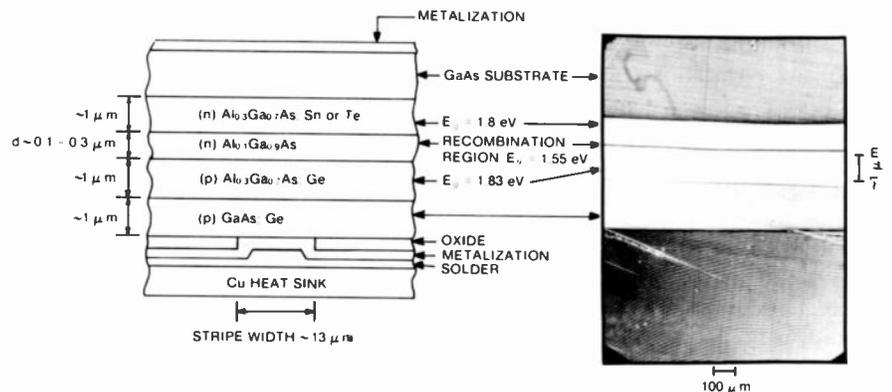


Fig. 3 The cross section of a typical laser in a schematic illustration (left, not to scale) and a Nomarski photomicrograph of a sample that has been polished at a shallow angle to produce high magnification in the transverse direction (right). The "terracing" effect evident on the lower portion of the micrograph, a growth artifact, causes some interface roughness.

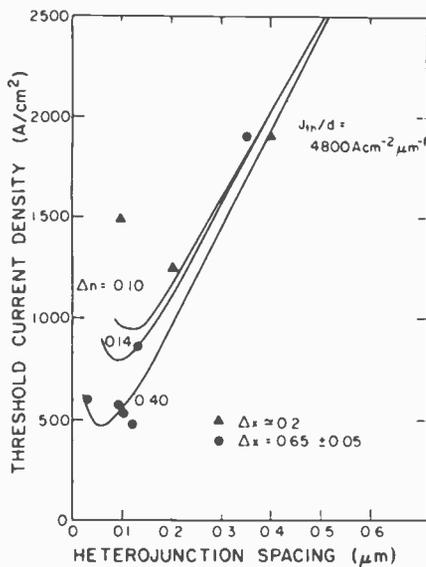


Fig. 4
Threshold current density as a function of the heterojunction spacing for $\text{Al}_x\text{Ga}_{1-x}\text{As}$ double-heterojunction lasers. The experimental data points are for aluminum-concentration steps Δx of 0.20 (triangles) and 0.65 (circles). The theoretical curves are for the discontinuities in the refractive index Δn shown, where the relation $\Delta n = 0.62 \Delta x$ has been assumed to apply.

475 A cm^{-2} is obtained with $d = 0.1 \mu\text{m}$ and $\Delta n \cong 0.4$ (corresponding to an $\text{Al}_{0.6}\text{Ga}_{0.4}\text{As}/\text{GaAs}/\text{Al}_{0.6}\text{Ga}_{0.4}\text{As}$ structure). It should be noted that the barrier height also affects the high-temperature performance of lasers.¹²

The maximum desirable threshold density for cw operation is below 2000 A/cm^2 , and double-heterojunction laser diodes have so far routinely operated cw at room temperature with 0 to 12% AlAs concentration in the active recombination region. Because of the corresponding bandgap variation, this produces an emission wavelength range of 0.9 to $0.8 \mu\text{m}$. Room-temperature cw operation of lasers at a wavelength of 1.0 to $1.1 \mu\text{m}$ has also been obtained with $\text{Ga}_{1-x}\text{Al}_x\text{As}/\text{Sb}_{1-x}\text{As}$ diodes,¹³ where $J_{th} \cong 2000 \text{ A/cm}^2$ was achieved, and for $\text{InGaP}/\text{InGaAs}$ diodes.¹⁴

Constructing the stripe-geometry laser

Laser diodes are prepared by cleaving two parallel facets to form the Fabry-Perot cavity. The cw laser diode uses a stripe geometry to define the lateral dimension (Fig. 5) for several reasons:

- The radiation is emitted from a small region, which simplifies coupling of the radiation into fibers with a low numerical aperture.
- The operating current can be minimized because it is relatively simple to form a small active area with photolithographic contacting procedures.
- The thermal dissipation of the diode is improved because the heat-generating active region is imbedded in an inactive semiconductor medium.
- The small active diode area makes it simpler to obtain a reasonably defect-free area.
- The active region is isolated from an open surface along its two major dimensions, a factor believed to be important for reliable long-term operation.

In the simplest stripe-contact structure, the active area is defined by opening a stripe in a deposited SiO_2 film. The surface of the diode is then metalized, with the ohmic contact formed only in the open area of the surface. Other methods that have been used to define the stripe contact include various implantation methods that increase the lateral resistance, and narrow mesas.¹⁵

Diodes for cw operation are generally designed with the thin p-side mounted on copper heat sinks to minimize the thermal resistance of the structure, using a soft solder such as indium to minimize strains in the devices.

Laser-diode operation

The lateral width of the emitting region can be adjusted for a desired operating level by adjusting the stripe width. For example, 100 mW of cw power is obtainable (from one facet) for a strip $100 \mu\text{m}$ wide. However, for the typical power levels needed in optical communications (5-10 mW), stripe widths of $\sim 13 \mu\text{m}$ are used. This dimension represents a suitable compromise between low operating currents and an appropriate power-emission level. A typical curve of power output as a function of diode current is shown in Fig. 6. The junction temperature for such devices is only a few degrees above the heat-sink temperature. For example, a temperature differential of 7 K is calculated with a typical power input of 0.5 W at a diode current of 0.3 A and a thermal resistance of 14 K/W.

The electromagnetic modes of the laser diode cavity are separable into two independent sets with transverse electric (TE) and transverse magnetic (TM) polarization. The mode numbers m , s , and q define the number of sinusoidal half-wave variations along the three axes of the cavity transverse, lateral and longitudinal, respectively.

The allowed *longitudinal* modes are determined from the average index of refraction and the dispersion seen by the propagating wave. The Fabry-Perot mode

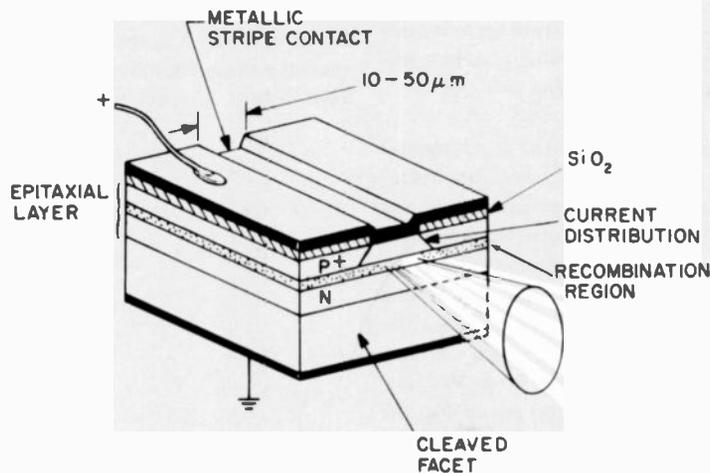


Fig. 5
Typical cw heterojunction laser; devices are now commercially available.

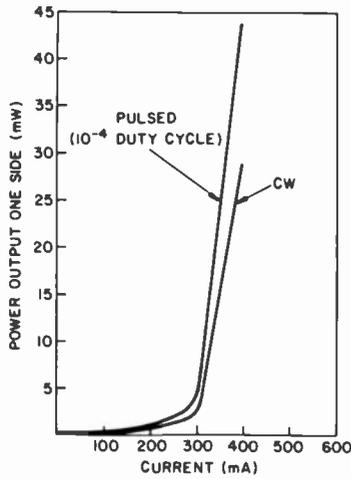


Fig. 6
Optical power output from one facet of a typical AlGaAs cw laser as a function of the drive current. Measured at room temperature in pulse and cw operation.

spacing is several angstrom units in typical laser diodes. The *lateral* modes are dependent on the method used to define the two edges of the diode. Generally, in stripe-geometry lasers only low-order modes are excited; their mode spacings are 0.1 to 0.2 Å and they appear as satellites to each longitudinal mode. The *transverse* modes depend on the dielectric variation perpendicular to the junction plane. In the devices discussed here, only the fundamental mode is excited, a condition achieved by restricting the width of the waveguiding region (i.e., heterojunction spacing) to values well under 1 μm. Therefore, the far-field radiation pattern consists of a single lobe in the direction perpendicular to the junction. (Higher-order transverse modes would give rise to "rabbit-ear" lobes, undesirable for fiber coupling.)

For a laser operating in the fundamental transverse mode, the full angular beamwidth at half power *perpendicular* to the junction plane is a function of the near-field radiation distribution. The narrower the emitting region in the direction perpendicular to the junction plane, the larger the beamwidth. In practical cw laser diodes the beamwidth is ~ 30 to 50°. The beamwidth in the direction *parallel* to the junction plane (lateral direction) is typically 5 to 10° and varies only slightly with the diode topology and internal geometry. At least one-half of the power emitted from one facet can be coupled into a multi-mode step-index fiber with $NA = 0.14$ and a core diameter of 80 μm.

While fundamental transverse mode operation is easily achieved, most narrow stripe laser diodes operate with several longitudinal modes, and therefore emit over a 10- to 30-Å spectral width, but some units can emit several milliwatts in the fundamental lateral and transverse mode and a single longitudinal mode. The linewidth is 0.15 Å and the power emitted is 3 mW (Fig. 7).

Methods of modulating the laser output vary widely, depending upon the application. For fast-pulse modulation, the diode is biased with a current near the threshold current and then pulsed to an appropriate level above threshold. If no bias is applied, there is a lasing delay (related to the spontaneous carrier lifetime) before the carrier population becomes fully inverted and the device turns on.¹⁶ This delay is several nanoseconds long in a typical situation, but vanishes if the laser is biased to threshold.

LED structure

The spectral bandwidth of the LED is typically 1 to 2 kT (300 to 600 Å) at room temperature, hence 1 to 2 orders of magnitude broader than that of the typical laser diode. Because of the spectral dispersion in fibers, this may limit the bandwidth for long-distance fiber communications using LEDs. Furthermore, the coupling efficiency of LEDs into low-numerical-aperture fibers is much lower than for laser diodes. However, the LED has the advantage of simpler construction and less temperature dependence of its emitted power. For example, the spontaneous output from an LED may decrease by only a factor of 1.5 to 2 as the diode temperature increases from room temperature to 100°C (at constant current), while the output of a

laser diode would typically decrease by more than a factor of 3.

LED topology is designed to minimize internal reabsorption of the radiation, allow high-current-density operation and maximize the coupling efficiency into fibers. While the structures used are applicable to all materials, most of the work on communications-type devices reported so far has been on Al_xGa_{1-x}As.

Two basic diode configurations for optical communications have been reported: see emitters¹⁷ and edge emitters.^{18,19} In the *surface emitter*, the recombination region is placed close to a heat sink and a well is etched through the GaAs substrate to accommodate a fiber. The emission from such a diode is essentially isotropic. The *edge-emitting* heterojunction structures, similar to the geometry of Fig. 5, use the partial internal waveguiding of the spontaneous radiation due to the heterojunctions to obtain improved directionality of the emitted power in the direction perpendicular to the junction plane. The lateral width of the emitting region is adjusted for the fiber dimension, but is typically 50-100 μm. Fig. 8 shows an edge-emitting LED.

LED performance

Surface-emitting and edge-emitting structures provide several milliwatts of power output in the 0.8- to 0.9-μm spectral range, operated at drive currents of 100 to 200 μA (2000 to 4000 A/cm²). The coupling loss into step-index fibers with a numerical aperture of 0.14 is about 17 to 20 dB for surface emitters and 12 to 16 dB for edge-emitting diodes, compared to about 3 dB for an injection laser. Since the coupling loss decreases as NA^{-2} , much more power

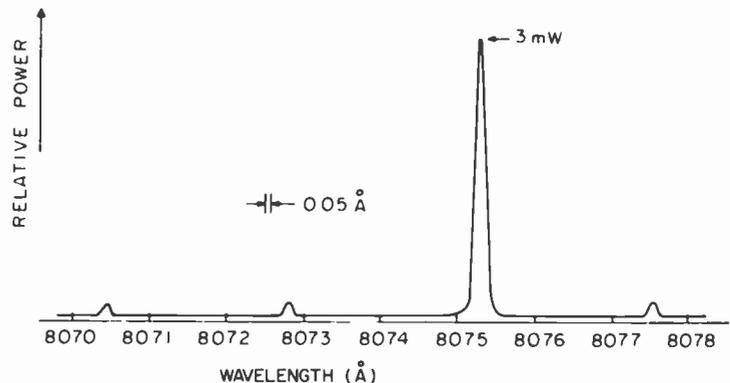


Fig. 7
Spectral output from one facet of an AlGaAs cw laser. It emits 3 mW when operating in a single longitudinal mode and in the fundamental transverse and lateral mode.

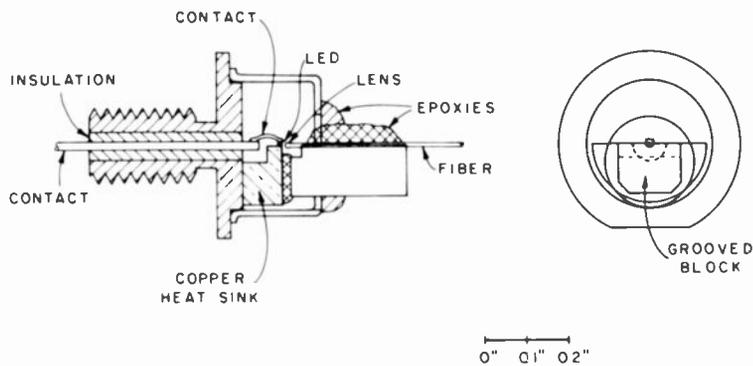


Fig. 8 Edge-emitting LED designed for fiber-optic communications. Short fiber extending to right couples to the long fiber.

can, of course, be coupled into larger-numerical-aperture fibers. But with these coupling losses, $\text{Al}_{0.1}\text{Ga}_{0.9}\text{As}$ double-heterojunction LEDs can provide about $100 \mu\text{W}$ into an 0.14-NA , $80\text{-}\mu\text{m}$ -diameter fiber at drive currents of about 200 mA with an applied voltage of 1.7 V .

Turning to the modulation problem, the relation between the optical power output of an LED (with constant peak current) and the modulation frequency, ω , is given by²⁰

$$\frac{P(\omega)}{P_0} = \frac{1}{[1 + (\omega\tau)^2]^{1/2}}$$

where τ is the injected carrier lifetime in the recombination region, and P_0 is the dc power emission value. (However, parasitic circuit elements can reduce the modulated power range below this value.)

It is evident that a high-speed diode requires the lowest value of τ possible without sacrificing internal quantum efficiency. Low values of τ are obtained at high doping levels; but in GaAs and related compounds, a high density of nonradiative centers is formed when the dopant concentration approaches the solubility limit at the growth temperature. Of the devices reported so far, germanium-doped double-heterojunction LEDs have exhibited modulation capability (at the 3-dB point) to about 200 MHz .¹⁸ The use of Ge is advantageous because it can be incorporated into GaAs to concentrations of about $2 \times 10^{19} \text{ cm}^{-3}$, thus providing lifetimes on the order of one nanosecond without unduly reducing the internal quantum efficiency.

With regard to LEDs for 1.0- to 1.1- μm wavelengths, surface emission outputs of

about 1 mW have been achieved using (InGa)As structures.²¹ Further progress is expected, particularly with the use of the lattice-matched $\text{In}(\text{GaAs})_x(\text{InGa})_y\text{P}$ heterojunction structures.²²

Failure causes

In any practical optical communications system, component reliability is of great concern. It has been a major research goal to identify and correct the myriad of failure mechanisms that seemed to plague early electroluminescent devices.²³ The failure modes have since been identified as either facet- or bulk-related; they can be of gradual or catastrophic nature. Facet degradation is specifically a laser problem because the facets are the mirrors that define the Fabry-Perot resonator. Bulk degradation, on the other hand, can occur in both LEDs and lasers. Most failure mechanisms have been eliminated as the overall technology (crystal growth, device fabrication) has matured, but active concern still exists for those remaining few which may limit ultimate operating life. Insufficient data exist for mean time to failure which, for telephone applications, is in the 100,000-hr range. A comprehensive model for laser facet failure does not exist, but considerable phenomenological data have been accumulated.

Intense optical fields

Facet failure due to intense optical fields is a well-known phenomenon in solid-state lasers; it occurs in all types of semiconductor lasers under varying conditions when the optical power density in the recombination region reaches the order of 10^6 W/cm^2 . The appearance of the damaged laser facets suggests local dissociation of the material, as well as "cracking" in some cases.

The critical damage level is also a function of the pulse length, t , decreasing as $t^{-1/2}$, over the range of 20 to 2000 ns. It is not surprising, therefore, that facet damage can occur in room-temperature laser diodes operating cw at their maximum emission levels (even with relatively low total power). Because of the nonuniform radiation distribution in the plane of the junction in stripe-contact lasers, it is difficult to establish precise linear power density criteria. However, the damage threshold for 100-ns pulses is about 10 times higher than it is in cw operation of diodes selected from the same group.²⁴

In addition to the dependence on optical flux density and pulse length, it has been found that ambient conditions, specifically moisture, and surface flaws (scratches, dirt particles) can lead to premature facet failure. In order to remove these limitations, lasers must be operated at a specific maximum power level and the facets must be passivated with protective dielectric coatings for isolation from their surroundings. For cw lasers the operating range of linear power density is about $1 \text{ mW}/\mu\text{m}$ of stripe width. Facet failure, at least in its early stages, commonly manifests itself as a decrease in differential quantum efficiency without an increase in threshold current density.

Bulk degradation

Bulk degradation, the other failure mode, is accompanied by an increase in threshold current density and a decrease in differential quantum efficiency caused by a reduction in *internal* quantum efficiency and an increase in absorption coefficient. The reduction in output power may be small, but the device may turn off completely if it is a laser operated near threshold. In either case, a small adjustment in current will restore the output to its initial value. With such a feedback system, a definition of operating life becomes somewhat arbitrary and system-dependent.

The available evidence suggests that this gradual degradation process results from an increase in the concentration of non-radiative centers in the recombination region.²⁵ These defects are initiated by the growth of flaws that are initially present in the recombination region of the diode. Point defects may also diffuse from external flawed regions adjacent to the active

region. Detrimental flaws include dislocations and impurity precipitates. One prominent effect in some degraded lasers is the formation of "dark lines" in regions where the luminescence is gradually extinguished.^{25,26} These regions have been identified as large dislocation networks that, having started as smaller pre-existing dislocation networks, grow by the migration of vacancies or interstitials.²⁷ In addition, more dispersed nonradiative centers, such as native point defects, apparently contribute to the degradation process.

It has been suggested that a multi-phonon emission process resulting from non-radiative electron-hole recombination gives rise to an intense vibration of the center, which reduces its displacement energy.²⁸ (Whether any point defects are actually formed within the recombination region remains unclear.) Hence, if non-radiative electron-hole recombination occurs, say at the damaged surface of a diode (as in the case of the sawed-edge diode experiment described in Ref. 29), it accelerates the motion of point defects into the device's active region.

The stoichiometry of the material in or close to the active region also can affect the local density of point defects;³⁰ for example, the vacancy concentration under certain conditions may be substantially above the equilibrium value. A modification of the initial stoichiometry may account for the improved degradation resistance of diodes fabricated with $\text{Al}_{0.1}\text{Ga}_{0.9}\text{As}$, rather than GaAs, in the recombination region. Finally, regions where nonradiative recombination occurs will tend to grow in size, leading to the strongly nonuniform degradation process that is commonly observed.

Reliable devices available

Enormous progress has been made since 1970 in eliminating many degradation mechanisms in lasers and LEDs by careful attention to the liquid-phase epitaxial growth and device processing. Facet passivation with dielectric coatings has virtually eliminated facet damage as an important failure mode in cw lasers. The ultimate operating lifetime of state-of-the-art devices remains undetermined, but accumulated data exists on devices in operation at constant current for well in excess of 20,000 hr without a serious drop in output power. Two examples of life data

taken over many hours of operation on more recently fabricated diodes show that

- for a lot of (AlGa)As LEDs emitting about 1 mW at 0.8 μm , the emitted power remained constant within 5% over 18,000 hours of operation; and
- for a lot of (AlGa)As cw lasers emitting between 5 and 10 mW from one facet at a wavelength of 0.82 μm , the maximum deviation was less than 25% in 14,000 hours.

These curves contrast sharply to those of a few years back, when the best operating life under similar conditions was a few hundred hours.

Conclusions

Heterojunction structures of AlGaAs emitting in the 0.8- to 0.85- μm spectral range have progressed to the point where practical systems applications of lasers and LEDs are becoming possible. Extensive research concerning properties of materials and methods of synthesis, as well as identifying the major factors affecting the operating lifetime of these devices, has made commercial products possible. Research is now under way on devices emitting in the 1.0- to 1.1- μm range, which offers some potential advantages in reduced fiber absorption and dispersion. These devices involve more complex material problems than the (AlGa)As devices because of the need to use dissimilar alloys to match the lattice.

References

1. Kressel, H.; and Nelson, H.; "Close-confinement gallium arsenide pn-junction lasers with reduced optical loss at room temperature," *RCA Review*, Vol. 30, (1969) p. 106.
2. Hayashi, I.; Panish, M.B.; and Foy, P.W.; "A low-threshold room-temperature injection laser," *IEEE J. Quantum Electron.*, Vol. 5, (1969) p. 211.
3. Alferov, Zh. I.; Andreev, V.M.; Portnoi, E.I.; and Trukan, M.K.; "AlAs-GaAs heterojunction injection lasers with a low room-temperature threshold," *Sov. Phys. Semiconductors*, Vol. 3 (1969) p. 1328. (Trans., *Sov. Phys. Semiconductors*, Vol. 3 (1970) p. 1107.)
4. Kressel, H.; and Hawrylo, F.Z.; "Fabry-Perot structure AlGaAs injection lasers with room-temperature threshold current densities of 2530A cm^2 ," *Appl. Phys. Lett.*, Vol. 17 (1970) p. 169.
5. For recent papers see *J. Crystal Growth*, Vol. 24 (1974), a special issue on liquid-phase epitaxy. The apparatus used in our laboratory is described by H.F. Lockwood and H. Kressel on pp. 97-105 of that issue and in Lockwood, H.F.; and Ettenberg, M.; "Thin-solution multiple-layer epitaxy," *J. Crystal Growth*, Vol. 15 (1972) p. 82.
6. Olsen, G.H.; "Interfacial lattice mismatch efforts in III-V compounds," *J. Crystal Growth*, Vol. 31 (1975) p. 223.
7. Stern, F.A.; "Gain-current relation for GaAs lasers with n-type and undoped active layers," *IEEE J. Quantum Electron.*, Vol. 9 (1973) p. 290. For an introductory treatment to stimulated emission in semiconductors, see Yariv, A.; *Quantum Electronics* 2nd ed., John Wiley 1975, or Pankove, J.I.; *Optical Processes of Semiconductors*, Prentice-Hall, Englewood Cliffs, NJ 1971. For detailed discussion of device requirements for optical communications see Miller, S.F.; Lingye, I.; and Marcanti, E.A.J.; "Research toward optical-

- fiber transmission systems. Part II: Devices and systems considerations," *Proc. IEEE*, Vol. 61 (1973) pp. 1726-1751.
9. Ettenberg, M.; and Kressel, H.; "Interfacial recombination at (AlGa)As GaAs heterojunction structures," *J. Appl. Phys.*, Apr 1976.
10. Analyses of mode guiding in thin structures are discussed in Butler, J.K.; Kressel, H.; and Ladany, I.; "Internal optical losses in very thin cw heterojunction laser diodes," *IEEE J. Quantum Electron.*, Vol. 11 (1975) p. 402; Dumke, W.P.; "The angular beam divergence in double-heterojunction lasers, with very thin active regions," *ibid.*, p. 400; Selway, P.R.; and Goodwin, A.R.; "The properties of double heterostructure lasers with very narrow active regions," *J. Phys. D.*, Vol. 5 (1972) p. 904; and Casey, H.C. Jr.; Panish, M.B.; and Merr, J.L.; "Beam divergence of the emission from double-heterostructure injection lasers," *J. Appl. Phys.*, Vol. 44 (1973) p. 5470.
11. Ettenberg, M.; "Very-low-threshold double-heterojunction AlGaAs injection lasers," *Appl. Phys. Lett.*, Vol. 27 (1975) p. 652.
12. Goodwin, A.R.; Peters, J.R.; Pion, M.; Thompson, G.H.B.; and Witeway, J.F.A.; "Threshold temperature characteristics of double-heterostructure GaAlAs lasers," *J. Appl. Phys.*, Vol. 46 (1975) p. 3126.
13. Nahory, R.F.; Pollak, M.A.; Beeke, F.D.; DeWinter, J.C.; and Dixon, R.W.; "Continuous operation of 1.0- μm -wavelength GaAs_{1-x}Sb_xAlGa_{1-x}As_{1-x}Sb_x double-heterostructure injection lasers at room temperature," *Appl. Phys. Lett.*, Vol. 28 (1976) p. 19.
14. Nuese, C.J., et al., to be published.
15. Various aspects of the technology are discussed in papers in the special issue on semiconductor lasers, *IEEE J. Quantum Electron.*, Vol. 11 (Jul 1975) pp. 381-562.
16. Konnerth, K.; and Lanza, C.; "Delay between current pulse and light emission of a gallium arsenide injection laser," *Appl. Phys. Lett.*, Vol. 4 (1964) p. 120; and Ripper, J.F.; "Measurement of spontaneous carrier lifetime from stimulated emission delays in semiconductor lasers," *J. Appl. Phys.*, Vol. 43 (1972) p. 1762.
17. Burrus, C.A.; and Miller, B.I.; "Small-area, double-heterostructure aluminum-gallium arsenide electro-luminescent diode sources for optical-fiber transmission lines," *Opt. Commun.*, Vol. 41 (1971) p. 307.
18. Ettenberg, M.; Lockwood, H.F.; Witke, J.; and Kressel, H.; "High radiance, high speed AlGaAs heterojunction diodes for optical communications," *Technical Digest, 1973 International Electron Devices Meeting*, Washington, DC, p. 317.
19. Witke, J.P.; Ettenberg, M.; and Kressel, M.; "High radiance LED for single-fiber optical links," *RCA Review*, Vol. 37, No. 2 (Jun 1976).
20. Liu, Y.S.; and Smith, D.S.; "The frequency response of an amplitude modulated GaAs luminescence diode," *Proc. IEEE*, Vol. 63 (1975) p. 542. J. Witke independently derived this expression for heterojunction structures.
21. Nuese, C.J.; and Engstrom, R.E.; "Efficient 1.06- μm emission from InGaAs electro-luminescent diodes," *IEEE Trans. Electron. Devices*, Vol. 129 (1972) p. 1067.
22. Nuese, C.J.; and Olsen, G.H.; "Room temperature heterojunction laser diodes of InGaAs InGaP with emission wavelength between 0.9 and 1.15 μm ," *Appl. Phys. Lett.*, Vol. 26 (1975) p. 528.
23. A comprehensive review of the literature until 1973 was presented in Kressel, H.; and Lockwood, H.F.; "A review of gradual degradation phenomena in electro-luminescent diodes," *J. de Physique*, (3, Suppl.), Vol. 35 (1974) p. - 23.
24. Kressel, H.; and Ladany, I.; "Reliability aspects and facet damage in high-power emission from (AlGa)As cw laser diodes at room temperature," *RCA Review*, Vol. 36 (1975) p. 230.
25. DeLoach, B.C.; Hakk, B.W.; Hartman, R.L.; and D'Asara, I.A.; "Degradation of cw GaAs double-heterostructure lasers," *Proc. IEEE*, Vol. 61 (1973) p. 1042.
26. Itoh, R.; Nakashima, H.; Kishino, S.; and Nakada, O.; "Degradation sources in GaAs-AlGaAs double-heterostructure lasers," *IEEE J. Quantum Electron.*, Vol. 11 (1975) p. 551.
27. Petroll, P.; and Hartman, R.L.; "Defect structure introduced during operation of heterojunction GaAs lasers," *Appl. Phys. Lett.*, Vol. 23 (1973) p. 469.
28. Gold, R.D.; and Weisberg, L.R.; "Permanent degradation of GaAs tunnel diodes," *Solid-State Electron.*, Vol. 7 (1964) p. 811.
29. Ladany, I.; and Kressel, H.; "Influence of device fabrication parameters on gradual degradation of (AlGa)As cw laser diodes," *Appl. Phys. Lett.*, Vol. 25 (1974) p. 708.
30. Ettenberg, M.; and Kressel, H.; "Heterojunction diodes of (AlGa)As-GaAs with improved degradation resistance," *Appl. Phys. Lett.*, Vol. 26 (1975) p. 478. The effect of dopants on reliability were studied by McMullin, P.G.; Blum, J.; Shih, K.K.; Smith, A.W.; and Woolhouse, G.R.; "Effect of doping on degradation of GaAs-AlGaAs injection lasers," *Appl. Phys. Lett.*, Vol. 24 (1974) p. 595.

Applying injection lasers to information scanning

I. Gorog|P.V. Goedertier
J.D. Knox|I. Ladany
J.P. Wittke|A.H. Firester

Experimental investigations have shown that cw room-temperature GaAlAs injection lasers compare favorably with He-Ne gas lasers for use in information-retrieval systems.

One of the most important future areas of laser application lies in high-density information storage and retrieval. High-density optical storage systems are apt to be both read-only and read/write types, and lasers are generally assumed to provide the optimal light source for both applications. However, apart from holographic memories, these systems do not use the temporal coherence properties commonly associated with lasers, but use only the laser's extremely high radiance. Gas lasers, such as the low-power He-Ne type, are generally used for optical-information storage systems, but long-lived, room-temperature cw injection lasers have recently become serious contenders for these applications.

Indeed, our experience with these small, efficient, readily modulated devices indicates that injection lasers may well

replace gas lasers as the preferred source in high-density information-retrieval systems and in flying-spot scanner applications. Possible systems that are currently receiving considerable attention include facsimile transmission terminals and computerized point-of-sale checkout systems.

Because the optical characteristics of injection lasers differ rather markedly from those of a typical gas laser, the optical systems that must be used with them also differ. In the next section we consider what these injection laser characteristics are and what sort of optics is required to produce useful optical beams from them. Then we will present experimental results showing how these (suitably modified) characteristics can be applied to high-density read-out systems and to a facsimile scanner.

Characteristics of injection lasers

For the applications of present interest, an injection laser that can be run (quasi-)continuously without cryogenic cooling is required. Fig. 1 is a schematic sketch of a state-of-the-art injection laser structure for such cw operation. Those used in this study are planar-stripe, double-heterojunction (AlGa)As lasers that have demonstrated good reliability.¹ In these devices, the effective laser width is defined by the narrow stripe contact that confines the current to a small part of the total junction area. In the lasers used for the present experiments, the stripe width is $\sim 13 \mu\text{m}$, and the laser length is $\sim 500 \mu\text{m}$, as defined by cleaved end (reflecting) surfaces. Current-spreading between the stripe and opposing contact leads to an emission width several μm wider than the metallic stripe. Although the electron-hole recombination region is only $0.2\text{--}0.5 \mu\text{m}$ thick, the light is emitted from a region $0.7\text{--}1.0 \mu\text{m}$ high surrounding the junction plane at the laser facet. Thus, the emitting region is slit-shaped, some 15-20 times as wide as it is high.

The narrow height of the emitting region, on the order of a wavelength, forces the laser to oscillate in a single transverse mode (in the direction perpendicular to the junction plane). The beam thus emerges, in this direction, as a diffraction-limited beam whose full width at half-power is about 40° . Laterally, the active region is many wavelengths wide, and several lateral modes can be sustained. Therefore, the far-field beam pattern in the plane of the junction often shows some structure and

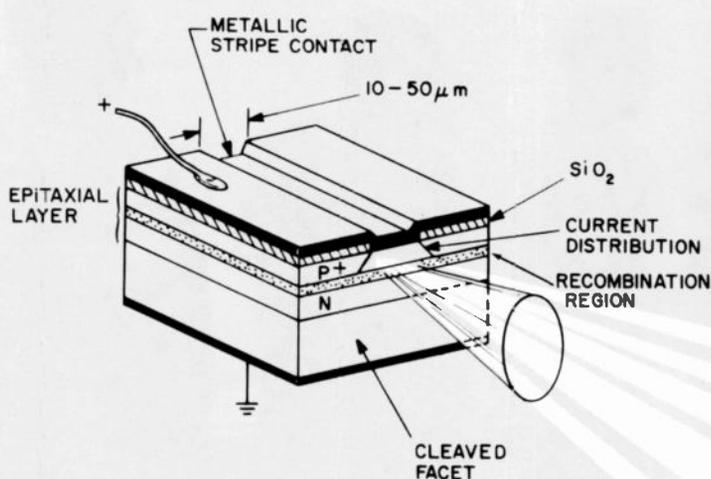


Fig. 1
Stripe-geometry injection laser has slit-shaped emitting region, 15 to 20 times as wide as it is high.

Reprint RE-22-4-4
This article appeared in *Applied Optics*, June 1976, in similar form.

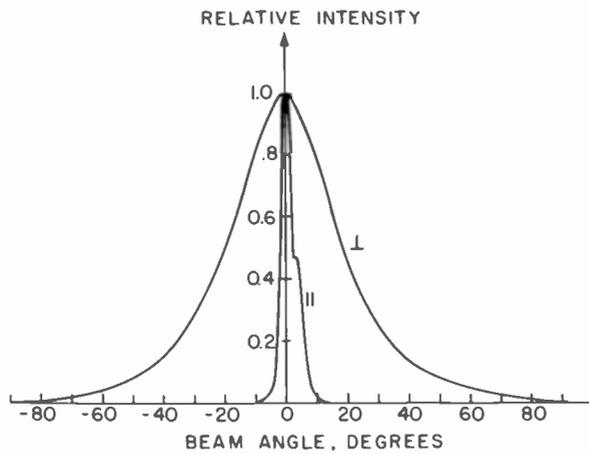


Fig. 2
Far-field pattern is not circular, and so requires additional optics to produce a circular spot.

represents the emission in several lateral modes, in a beam pattern perhaps 10° in width. This behavior is illustrated in Fig. 2.

Because the laser source region is thus elongated, rather than circular in cross-section, simple imaging of the source onto a target plane produces a similarly elongated image spot. While one can envision systems in which such a spot shape is desirable, most applications require a round spot. Focusing the laser beam into such a spot requires anamorphic optics. One simple system, to be described in more detail in connection with the flying-spot document scanner, consists of a microscope objective lens, placed near the emitting laser facet,

Arthur H. Firester has been involved with electro-optics for some time—his doctoral dissertation was on the modulation of light by optically pumped alkali-metal vapors. His recent research has been in nonlinear optical phenomena and their possible application to image processing; he is also working with coherent-light optical problems and holography.

Joseph D. Knox has worked extensively in the design and development of optical scanning systems. In addition to the one described here, these systems have been used for laser deflection displays, light valve displays and IC mask and wafer inspection. Dr. Knox's other activities include the design and fabrication of acousto-optic deflectors, modulators, and cavity dumpers for visible and infrared lasers.

Peter V. Goedertier was the co-discover of the He-Ne "cascade" laser and a pioneer in the development of the cross-pumped YAG:Nd:Cr laser. His other research projects have ranged from ion physics and early gaseous and solid-state optical masers to electro-optics and the development of various optical systems.

The biographies of **Istvan Gorog**, **James Wittke**, and **Ivan Ladany** appear with their other articles in this issue.

Authors **Firester**, **Gorog**, **Goedertier**, **Wittke**, **Knox**, and **Ladany**
 Contact them at: **Systems Research Laboratory, RCA Laboratories, Princeton, N.J. Ext. 3202.**



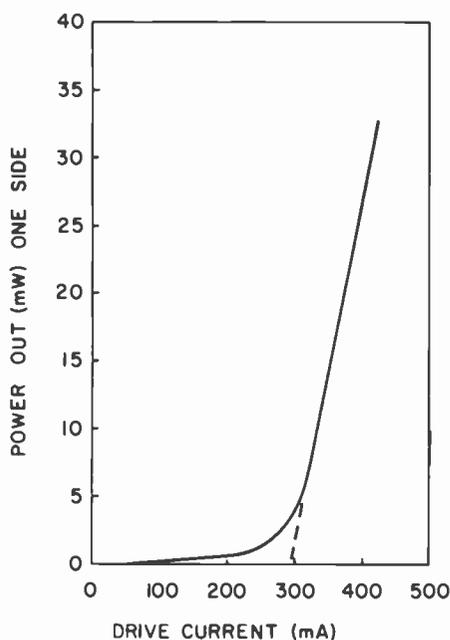


Fig. 3
Past threshold, output power of injection laser is a linear function of drive current, so modulating drive current will modulate output power.

that images the facet onto the target. This is followed by a "telescope," made from two cylindrical lenses, that provides different magnifications in the plane of the cylinder lens axes and the perpendicular plane. In this way, the elongated source can be imaged into the symmetric, round spot usually required.

Another important laser characteristic is the available optical power. The lasers operate in an incoherent mode below the threshold drive current of about 300 mA. Above threshold, the output power is a linear function of the drive current, as shown in Fig. 3. This permits the output power to be simply modulated by modulating the drive current. Lasers operate at 1.5-2.0V across the junction at currents between threshold and perhaps 100 mA above threshold, giving output powers of 10 mW or more. Thus, the overall laser efficiency is in the 1-5% range. Typical gas lasers have efficiencies less than 0.1% and require a high voltage to maintain the discharge, so the advantage of using an injection laser when feasible is obvious.

The noise characteristics of the laser output are of great significance for any information-handling application. These were evaluated for both unmodulated and pure-tone-modulated lasers by detecting the laser output with a fast photomultiplier

and scanning the photomultiplier output with an rf spectrum analyzer. In the 0-100 MHz spectral region tested, the measured noise was within a few dB of the value predicted from the shot noise formula:

$$\langle i_n^2 \rangle = 2eIGB$$

Here $\langle i_n^2 \rangle$ is the average mean-square current fluctuation at the photodetector output, e is the electronic charge, I is the photocurrent at the detector output, G is the detector gain, and B is the analyzer bandwidth. For example, when the laser

output was 5 mW and the photodetector was arranged to produce 5- μ A output current with a gain of 2.5×10^5 , the measured noise level into 50 Ω in a 3 kHz bandwidth was -102 dBm.

Finally, since a laser is a threshold device, to maintain constant output power, the drive current above threshold must be held stable. Since the threshold current of an injection laser is temperature-dependent,² either the laser temperature must be stabilized or an optical feedback loop on the drive current must be used to prevent

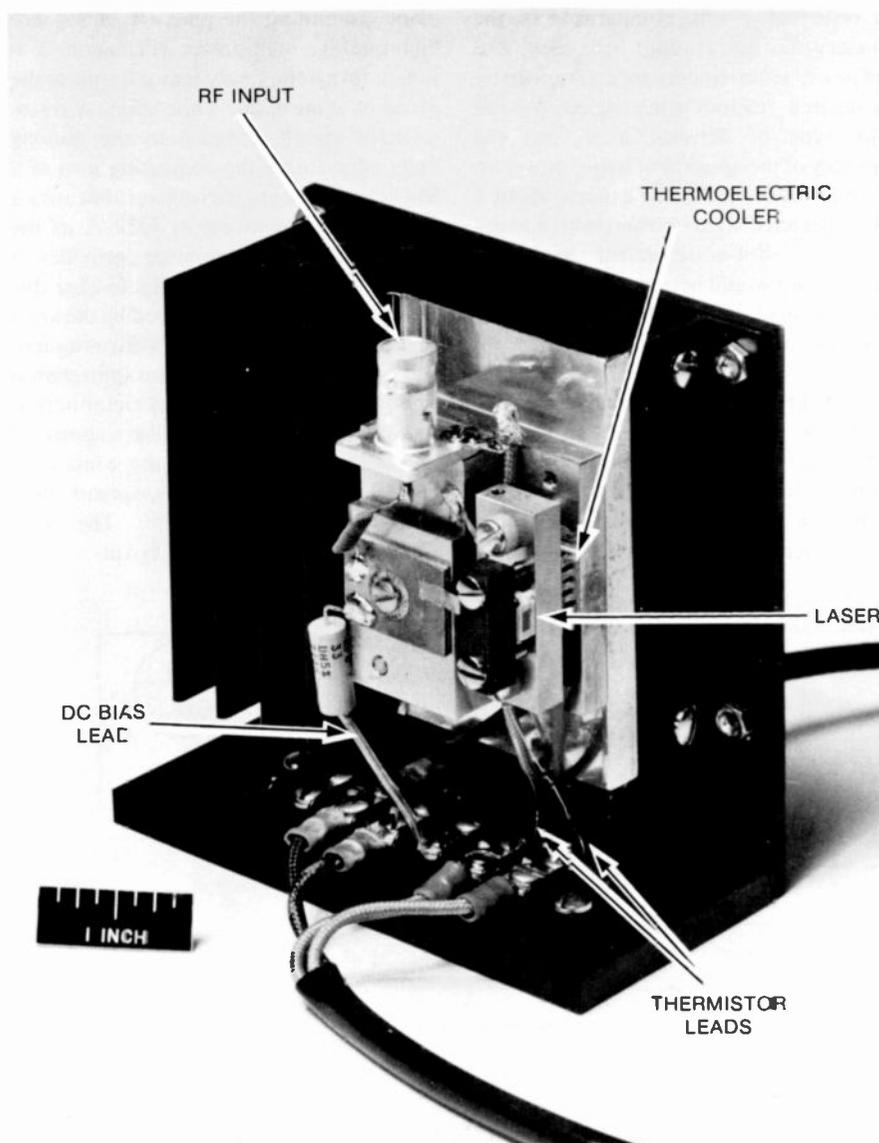


Fig. 4
Developmental cw injection laser package uses thermistor/thermoelectric cooler system to stabilize laser temperature.

output drifts with changing ambient conditions. We have stabilized the laser temperature using a thermistor sensor in a control loop that includes a small thermoelectric cooler. This system can provide a constant laser power output with a constant injection current. Fig. 4 shows a complete injection-laser package.

High-density information retrieval

High-density optical storage systems can be designed in a number of ways. One method records the information on a planar surface in the form of small pits or depressions. The pit dimensions can be on the order of $1 \mu\text{m}$, comparable to the wavelength of the readout light used. The spot power requirements are determined by the desired readout rate, signal-to-noise ratio, type of detector used, and the efficiency of the optics. For a spot power on the order of 0.1 mW , one expects about $1 \mu\text{W}$ of detected signal power; with a state-of-the-art, shot-noise-limited avalanche detector, this would permit digital information to be read out at 250 Mb/sec , with bit error rates $\leq 10^{-10}$.

If the transverse mode structure of the (AlGa)As laser resembled an axially symmetric Gaussian spot, most of the output could be focused into a spot whose diameter would be approximately one-half of a wavelength, or approximately 4200 \AA .

As discussed above, the injection laser's thin-junction geometry makes the output highly asymmetric; it also can exhibit some structure (multi-lateral mode). Since it was therefore difficult to predict the laser's actual focusing properties analytically, a series of knife-edge scan experiments were performed to obtain quantitative data.

Fig. 5 shows the experimental arrangement. The injection laser output is collected and shaped by an anamorphic lens system that consists of a low-power microscope objective and a two-cylindrical-lens telescope. The low-power objective acts as a collimator and the cylindrical lenses expand the beam in the plane containing the junction. A second, high-quality, high-power microscope objective forms the finely focused spot in the plane of a moveable knife edge. A retro-reflector rigidly coupled to the moving knife edge forms the measuring arm of a Michelson-type interferometer that uses a He-Ne laser operating at 6328 \AA as the source. The interferometer provides a precise calibration for the knife-edge displacements. The light reflected by the knife edge is collected by the spot-forming lens and projected through a beam splitter onto a photodetector. The photodetector output is thus the spatial step-function response of the beam; differentiating the knife-edge response then gives the line-spread function (spot profile) easily. The spot diameters deduced from these

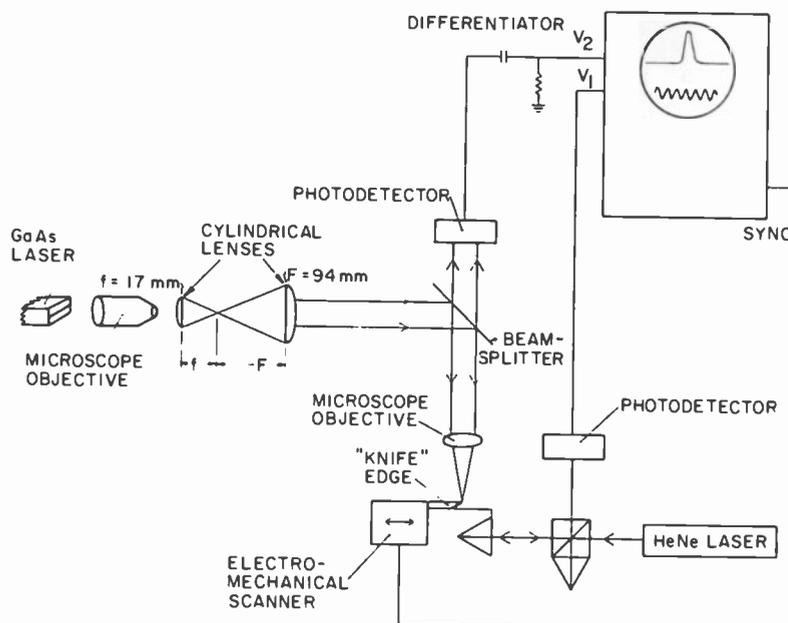


Fig. 5
Experimental system determined laser's focusing properties by measuring micrometer-sized spots.

Table 1
Spot size data for GaAlAs cw injection laser. Output power was 6.6 mW and focused spot power 0.6 mW , giving overall efficiency about equal to the He-Ne system.

Lens magnification/ numerical aperture	Spot size (full width at half intensity)
$\times 50/0.85$	$.64 \mu\text{m}$
$\times 40/0.65$	$1.03 \mu\text{m}$
$\times 20/0.40$	$1.7 \mu\text{m}$

measurements are listed in Table 1. In the range studied, the spot power was a constant 0.6 mW ; the optical insertion loss between the laser and the focused spot was approximately 10 dB . Because of the high injection laser efficiency, the overall efficiency, defined as the power in the focused spot divided by the laser output power, is about the same for an injection laser as it is for a He-Ne system.

Flying-spot scanner system

A second category of systems that can profitably use injection-laser sources includes flying-spot document scanners. In such systems, the object to be scanned is held stationary, (at least in one dimension,) and the light beam is moved across the object. The resolution requirement for document scanning is in the vicinity of $250 \mu\text{m}$. This corresponds to a scanner resolution of approximately 100 lines per inch, which is the highest resolution demanded by a human viewer from an 18-inch viewing distance.¹

Fig. 6 shows an experimental flying-spot scanner. The injection laser output first passes through a low-power microscope objective (10X) and then through a cylindrical beam expander. Ideally, the microscope objective is so chosen that it both produces the required magnification and has its entrance aperture filled by the diffraction-limited wide beam emitted in the plane perpendicular to the junction. This plane is arbitrarily chosen to coincide with the horizontal scan plane; its only active imaging element is the low-power objective that magnifies the thin dimension a few hundred times. In order to minimize aberrations and facilitate focus adjustment, another spherical lens could be used in conjunction with the microscope objective. In the plane of the junction, which

was chosen to coincide with the plane of the vertical scan, the laser beam is imaged by the combination of the microscope objective and a cylindrical beam expander. One may understand the operation of this arrangement by recalling that in the junction plane the effective source size is approximately 20 times larger than in the plane perpendicular to it and, therefore, the required magnification is approximately 20 times smaller. Our design philosophy is based on the idea that we wish to operate the scanner in both the vertical and horizontal directions in a diffraction-limited mode because this results in both minimum lens and mirror aperture sizes and a minimum deflector tilt angle.

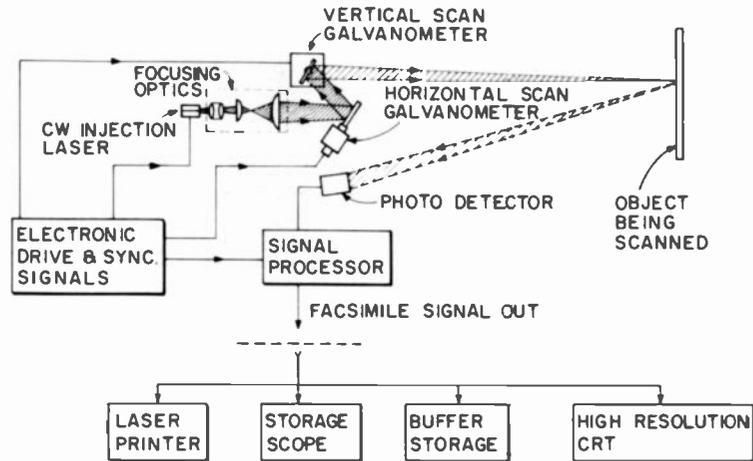


Fig. 6
Scanner systems produced resolutions over 100 lines per inch, and implied that very-high-quality scans of a few hundred lines per inch could be produced.

The line-spread function obtained with the above arrangement is shown in Fig. 7. This figure was obtained by scanning the spot across a narrow slit mounted in the scan plane. The spot profile shows reasonable cylindrical symmetry and has a full width at half maximum of approximately $100 \mu\text{m}$. The optical efficiency of the system, defined as the spot power in the scan plane divided by the laser output, was again approximately 10%. A number of experimental scanners were constructed, using nominally similar lasers. The efficiency for all of these units was about the same; the spot size varied from 100 to $200 \mu\text{m}$.

Our flying-spot scanner⁴ typically scanned a standard document ($8\frac{1}{2} \times 11$ in.) from a distance of 1 meter. Its vertical resolution was limited by the number of scan lines to 1000 lines per picture height. The horizontal response can be computed from the spot profile data. With reference to an $8\frac{1}{2}$ -inch-wide image, the horizontal modulation transfer function, as calculated from the data of Fig. 7, is shown in Fig. 8. Our results suggest that very high quality (few hundred lines per inch) image transmitters can be constructed using injection-laser scanners.

A recently developed application area for flying-spot laser scanners is reading coded product information in point-of-sale (POS) systems. The resolution requirement in POS applications is about the same as for document readers (about $200 \mu\text{m}$), but the POS system has the unique requirements that the depth of focus should be several inches and that adequate means of ambient-light discrimination must be provided. The depth of focus for a diffraction-limited system is completely determined by the spot size and the

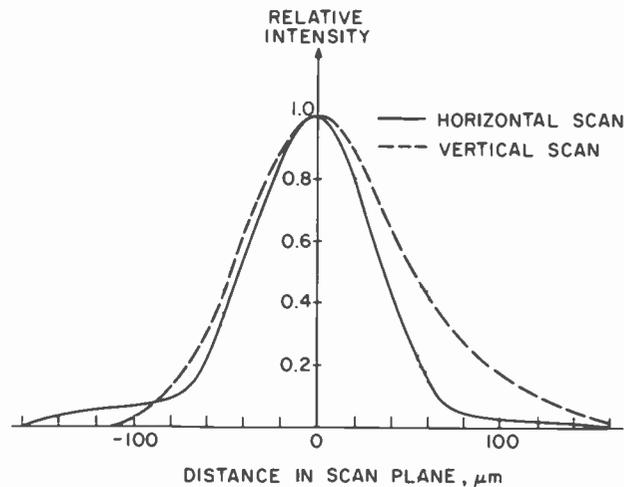


Fig. 7
Line-spread function shows reasonable cylindrical symmetry.

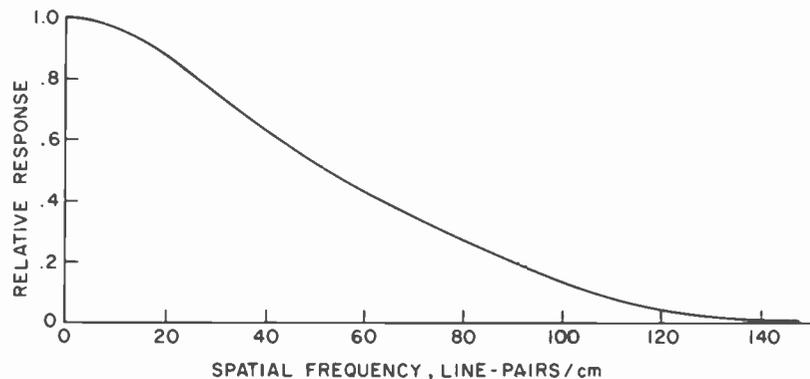


Fig. 8
Horizontal modulation transfer function.

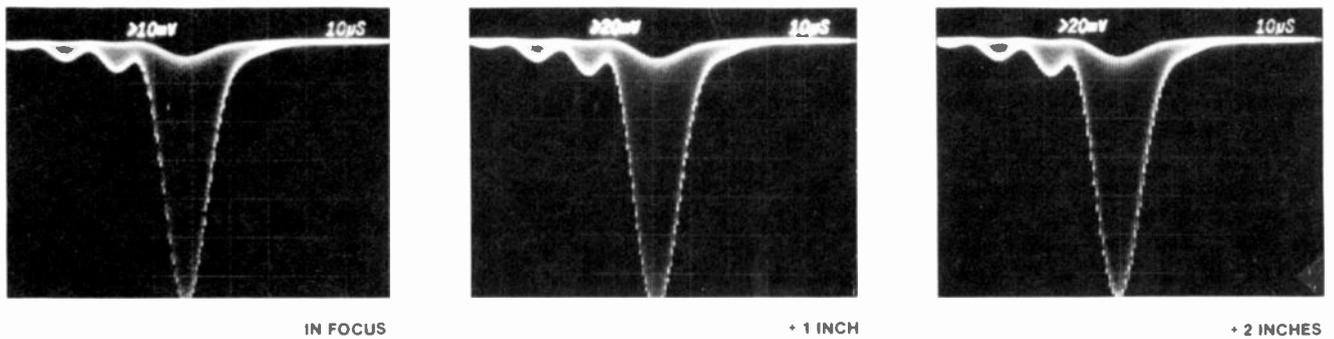


Fig. 9 Spot profiles obtained by moving the spot out of the nominal scan plane showed that the scanner had a satisfactory depth of focus.

wavelength. Let us assume that the field distribution in the focal plane is approximately Gaussian and denote the spot radius where intensity drops to $1/e^2$ times its peak by w_0 . Then one can show⁵ that the radius of the $1/e^2$ intensity point will increase to $2^{1/2} w_0$ for an axial focus error of

$$z = \pm \pi w_0^2 / \lambda$$

where λ is the light wavelength. For $\lambda = 8200 \text{ \AA}$ and $w_0 = 100 \mu\text{m}$, $z = 15.3 \text{ cm}$. Fig. 9 shows a set of spot profiles obtained by focusing an approximately 200- μm -radius spot in various planes in the vicinity of the nominal scan plane. We found that an operating range in excess of 4 inches is available without significantly increasing the beam diameter.

Ambient-light discrimination

In principle, narrowband interference filters can effectively discriminate between the spectrally-narrow laser output and the broadband ambient background. However, the actual acceptable bandwidth of the required interference filter is often determined, not by the laser linewidth, but by the scan angle. As the angle of incidence varies, the peak transmission of an interference filter also varies. For example, if a photodetector is located 1 meter from a standard-size document, the angle-dependent detuning is approximately 100 \AA from the center of the document to its top or bottom. This implies that, for this geometry, in order to achieve reasonable shading (i.e., an incidence-angle-independent detected signal) interference filters with a few hundred angstroms half-

power bandwidth would be required. Such broadband filters provide very poor discrimination against typical ambient light levels.

We have achieved very effective discrimination against unwanted ambient light by using a modulated laser in conjunction with synchronous detection of the return signal. Fig. 10 is a block diagram of an experimental synchronous flying-spot scanner system designed to read standard-size documents at a nominal rate of 12 s/page, with 1000 scan lines resolution. The nominal bandwidth was 50 kHz (12 ms/line, 500 cycles/line, 2 ms flyback time). The injection laser was modulated at 1.2 MHz, and the modulation signal was also used as the reference signal in the synchronous detector. The output current from the optical detector is fed to the

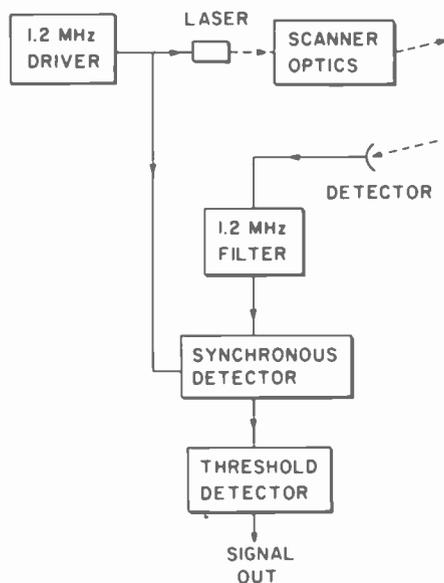


Fig. 10 Synchronous detector arrangement discriminates against ambient light.

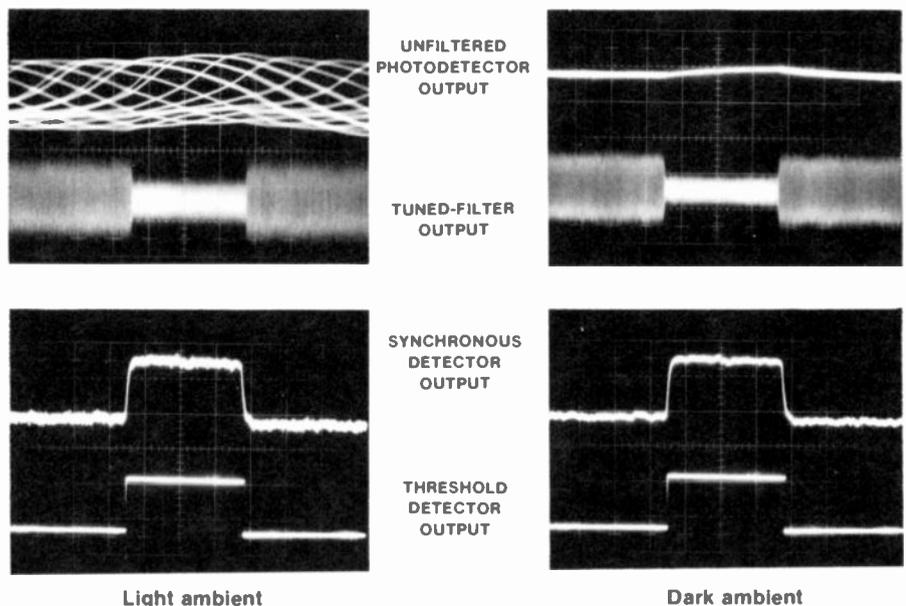
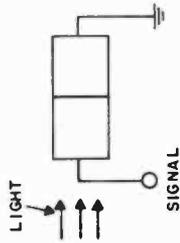


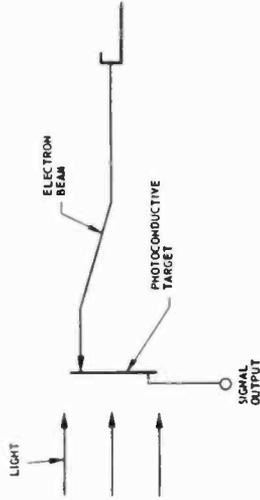
Fig. 11 Oscilloscope traces for scan of black bar on white background.

Nonimaging

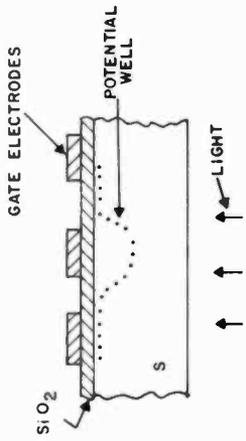


Photoconductor, in which light detection produces a transfer of minority carriers *within* solid material, e.g., this p-n junction.

Imaging



Vidicon uses an array of p-n junctions as the photoconductive element; scanning beam reads out the image sequentially.



Charge-coupled device uses Si detectors, storing the charge in potential wells under MOS gates. Readout is via a charge-transfer process—clock voltages applied to the gates move the charge from well to well.

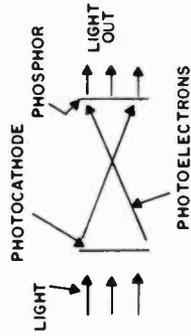
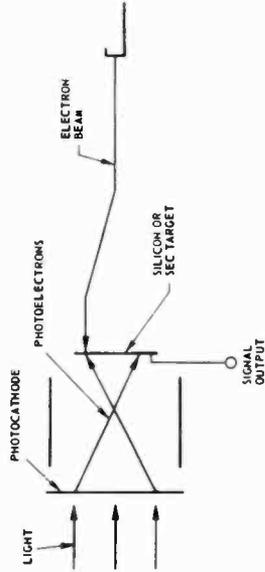
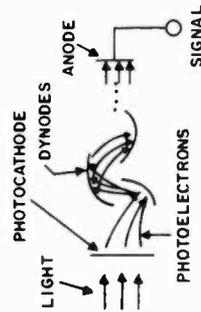


Image tubes convert the optical image to an electron-emission pattern by means of their photocathodes.



SIT tube combines a photocathode detector and Si target to produce a nearly ideal device.



Photomultiplier, in which light detection produces the emission of electrons into vacuum, as in this photomultiplier tube. Successive secondary emission along the series of dynodes produces amplification.

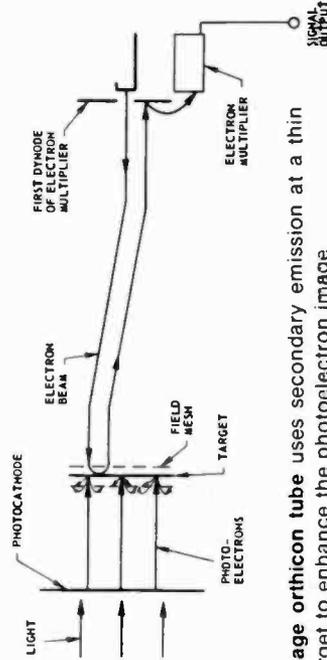


Image orthicon tube uses secondary emission at a thin target to enhance the photoelectron image.

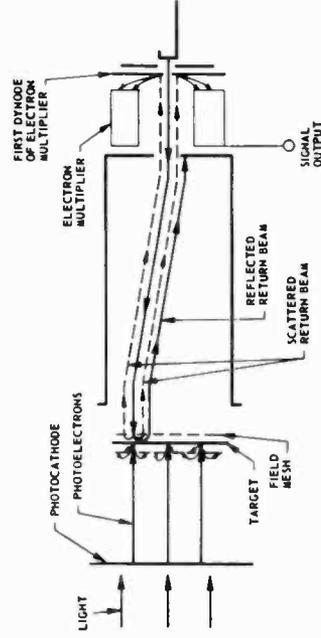


Image isocon is a modified image orthocon that eliminates much of the electron-scan-beam noise by using a return-beam separator. With the separator, only those electrons that have actually been scattered after hitting the target get amplified.

Fig. 1
Light detectors can be divided into two classes—photoconductors and photoemitters—with both imaging and nonimaging types within each class.

an introduction Light-detecting devices:

E.D. Savoye

Two different classes of light-detecting devices, photoconductors and photoemitters, are approaching the ideal detector from different directions.

The ideal light-detecting device must satisfy two basic criteria:

- 1) Each incoming photon must contribute to the detected signal, i.e., the device should have unity quantum efficiency (Q.E.).
- 2) The detector or detector/amplifier combination must have sufficient sensitivity to produce a useful output signal for a single-photon input.

The first of these criteria is important in order to preserve the signal-to-noise ratio (S/N) of the photon signal; the second is required (in addition) for detecting and imaging low-light-level signals, which may have only a single photon event per resolution element per integration time.

Practical light-detecting devices have approached the ideal in two different configurations, which define the two broad classes of photodetectors:

- 1) *photoconductors*, in which photo-excitation causes a transfer of minority carriers within a solid material, e.g. across a p-n junction; and
- 2) *photoemitters*, in which photo-excitation produces the emission of electrons into vacuum.

In order to produce a broad range of practical devices in both classes, trade-offs have been made in a number of areas, including choice of materials, physical configuration, and operating conditions.

The principal advantage of using photoconductors lies in their high quantum efficiency, provided by the fact that internal collection of photoexcited carriers can be highly efficient when there is no

energy barrier against their collection. Their principal disadvantage is at low light levels, where the signals are too small to be detected above the system noise.

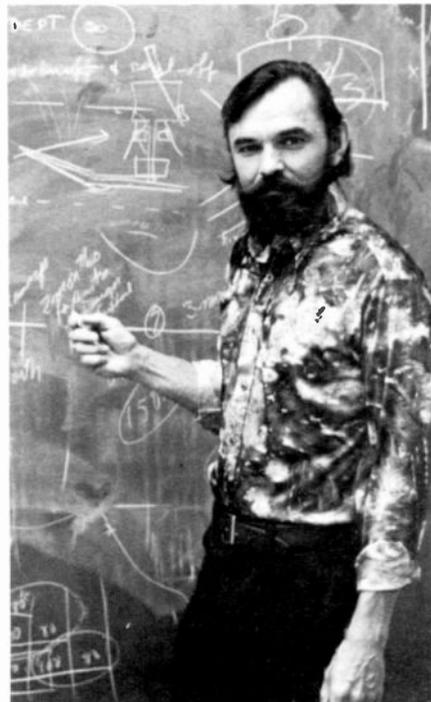
In contrast, photoemissive detectors are relatively limited in quantum efficiency because of the energy barrier against the emission of electrons at the surface of the detector material. However, they have the advantage that photoelectrons, once emitted, can readily be accelerated in vacuum to produce essentially noise-free gain by secondary emission or by impact-ionization in a solid material. Devices based on photoemitters are thereby capable of detecting single-photoelectron events.

The principal practical devices that have evolved based on these two classes of photodetectors are summarized in Fig. 1, and discussed in the following pages.

Dick Savoye is presently Manager of Engineering for Electro-optics Products at RCA Lancaster. Dr. Savoye began working on photoemissive devices for RCA in 1966, and received the RCA Laboratories Outstanding Achievement Award in 1968 for his work on tunneling photoemitters. He has supervised advanced development work on Si vidicons and CCDs, negative electron affinity photoemitters and secondary emitters, and the SIT tube.

Contact him at:
**Electro-optics Products Engineering
Solid State Division
Lancaster, Pa.
Ext. 2020**

Reprint RE-22-4-5
Final manuscript received October 15, 1976.





"Tea-cup" photomultiplier is shorter and more efficient than the "venetian-blind" PMT it is expected to replace. The three-inch tube shown is the forerunner of a family of tubes designed for scintillation counting in medical electronics equipment.

Nonimaging detectors

Nonimaging photoconductive detectors are based upon the p-n junction.

The photoconductive detector, represented here by a single p-n junction, also exists in many other configurations, such as p-i-n junctions, avalanche detectors, simple photoconductor materials with appropriate electrical contacts, Schottky barrier detectors, etc. In general, however, incident light photoexcites carriers within the device to produce the output signal. In the case of the avalanche detector, internal gain greatly increases the low-light-level capability of the device.

Photoconductive devices of various types are being developed and manufactured at RCA's Montreal plant and, with the proliferation of various electro-optical devices, have quadrupled their sales in the past two years. Two typical end uses are military optical-tracking equipment, using a quadrant p-i-n detector, and a range-finder that uses both a YAG laser and an avalanche detector with a risetime of less than 1 ns. Other military applications include proximity fuses and weapon-fire-simulators. This last equipment uses a laser beam as the "ammunition" and detectors located at critical positions on military targets as hit indicators.

A very promising consumer application lies in automatic ranging for photographic cameras. In these systems an ir-emitting diode is the source and a p-i-n detector senses the time it takes for the reflected radiation to return from the subject to the camera. This information is used to control the camera focus. One may also anticipate a large demand for nonimaging detectors with the expected explosive growth in the whole field of optical communications.

Nonimaging photoemissive detectors are represented by the photomultiplier tube, which has very high noise-free gain.

The photomultiplier tube (PMT) represents this category. In this type of tube, electrons emitted from the photocathode are accelerated to the first dynode, where they land with several hundred eV energy, which is sufficient to give rise to several (4-7) secondary electrons emitted per incident primary. The secondaries in turn are accelerated to the second dynode, where further multiplication occurs. Using about 10 such stages can give essentially noise-free gain on the order of 10^6 , sufficient for detecting a single photoelectron from the cathode.

PMTs are important in a broad range of applications, including astronomy, laser detection, spectroscopy, and medical electronics. RCA has been particularly successful in dominating the latter field; our annual sales of 3-inch PMTs in gamma cameras exceed \$3M. These gamma-ray imaging cameras are now to be found in almost all major hospitals. A typical camera uses 37 PMTs, set up in a hexagonal cluster arrangement, to sense the gamma-ray scintillations coming from radioactive isotopes introduced into the patient.

Imaging detectors

Imaging photoconductive detectors store the optical image as a charge pattern.

This class is represented by the vidicon and the charge-coupled device (CCD). In both devices, an optical image is detected and stored as a charge pattern within the photoconductive element, which is an array of p-n junctions for the vidicon. The CCD uses Si as the detector, storing charges under MOS gates. In each case the



Vidicon tube finds use in broadcast, industrial, and surveillance applications, depending upon the quality level of the tube.

image is read out sequentially by a scanning process—an electron beam for the vidicon and the charge-transfer process for the CCD.

Today, the largest application of vidicons is in surveillance, both industrial and military. Particular vidicon types are also used by the military in tv-guided missiles. Broadcast tv uses the PbO vidicon for "live" tv; film-pickup for tv reproduction uses the Sb_2S_3 vidicon.

An important application for the CCD is anticipated in tv journalism, where portability is very important. The world-wide market in camera tubes is approximately \$100M annually, of which RCA's share is presently about 10%.

Imaging photoemissive detectors are excellent for low-light-level operation.

In these devices the optical image is converted to an electron-emission pattern by the photocathode. In the image tube, the emitted electrons are accelerated to strike a phosphor screen, giving rise to a light-emission pattern that corresponds to the original optical image. Image tubes can be made with fiber-optic input and output faceplates. This construction permits series



CCD camera's small size and light weight (1 kg without lens) and ruggedness make it valuable in tv journalism.

coupling of the image tubes with minimum resolution loss and electron gain figures of 30-50, measured from photocathode to photocathode. Three such stages can provide enough gain so that a single photoelectron emitted from the input photocathode produces a visible output at the final phosphor screen. This makes operation at very low light levels possible.

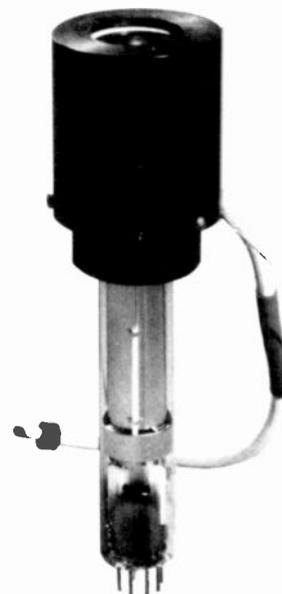
Image tubes were developed primarily with government sponsorship for use in military night operations. Initially, they used an auxiliary infrared illuminator—the original “snooperscope” and “sniperscope.” The ready detectability of the infrared source, however, led to the development of “passive” viewers, which produce images with the bare minimum of natural light available even on moonless cloudy nights. Many police departments have taken advantage of the capability of these devices for nighttime surveillance. Today's uses, however, have expanded to include many scientific and commercial applications. For example, magnetically focused image tubes, because of their minimum of image distortion, have proved to be very important in augmenting the photographic plate in astronomy. Since the image tube has an effective quantum efficiency about 10 times that of the photographic plate, its use is equivalent to increasing the telescope's diameter by a factor of three. Also, combined with a tv-camera tube, the image tube can aid in reproducing the very dim fluorescent images found in x-ray inspection systems.

The image orthicon was the quality pick-up device during the early years of television.

In these tubes, the photoelectron image is enhanced by secondary emission at a thin target and the charge pattern is scanned off by an electron beam. The return-beam signal is amplified by further secondary emission, but noise in the electron beam prevents the image orthicon from providing useful signals at very low light levels. A modified image orthicon, called the image isocon, eliminates much of the electron-scan-beam noise by a novel return-beam separator that permits the amplification of only those return-beam electrons that have actually been scattered after hitting the target. The image isocon thus approaches an ideal device, but it is still limited by target secondary-emission noise and electron-beam separation noise.

The recently developed SIT (silicon-intensifier target) tube is very close to an ideal device. The SIT tube combines the photocathode detector with a Si-vidicon type of target. Photoelectrons are accelerated to strike the Si target at several keV, giving rise to an impact-ionization gain of several thousand within the target. The device thus functions at low light levels, and, when coupled to an image tube to produce the I-SIT configuration, provides a distinguishable output signal for a single-photoelectron input.

In either the SIT or I-SIT configuration, the overall gain of the system can be varied electronically by means of the accelerating potential, making both devices useful over a broad range of light levels. SIT tubes, because of this adaptability to different light levels, are becoming more and more useful in industrial surveillance. Such a



SIT tubes operate at light levels nearly down to photoelectron noise limit.

typical industrial application is in parking lots, where it is then not necessary to provide a high level of illumination.

Approaching the ideal

The potential of realizing the ideal photodetector, combining high quantum efficiency with detecting single-photon events, is highly promising. This ideal is being approached from two major directions:

- 1) CCD detectors can, in principle, operate at very low light levels. With further development it is possible that we may produce a cooled, buried-channel, distributed-gate on-chip amplifier device that would be capable of providing a distinguishable output for a single-photon input.
- 2) Photoemissive detectors can already detect single-photoelectron events, and the so-called “negative electron affinity” devices can provide revolutionary advances in photoemissive quantum efficiency. In fact, laboratory experiments have demonstrated quantum efficiencies near those of photoconductors.

With further development it is possible that one or both of these methods may closely approach the performance of an ideal photodetector.

The silicon-target vidicon

R.G. Neuhauser

With 1800 diodes per inch on a silicon wafer, this vidicon is replacing the standard vidicon in many applications and also performing in completely new situations where previous camera tubes were useless.

The concept of using a large-scale array of diodes as a mosaic photosensitive target for a camera tube was first proposed at Bell Laboratories in the late 1960's. Rapid expansions in silicon technology and large-scale integrated-circuit technology eventually produced a vidicon tube with an array of more than 600,000 individual diodes on a silicon wafer. In this design, the silicon generates charge carriers from the light of an optical image focused on the wafer; the diodes store the carriers until they are scanned out as video information.

Robert Neuhauser has been employed in television camera tube development, manufacturing, and application work since 1959, when he joined RCA in Lancaster. He is a fellow of the Society of Motion Picture and Television Engineers and received the Society's 1964 David Sarnoff Award. He is also the author of 19 technical articles on television camera tubes.

Contact him at:
**Application Engineering
Electro-Optics and Devices
Solid State Division
Lancaster, Pa.
Ext. 2223**



The tube's advantages

The silicon-diode vidicon is rugged and its sensitivity is excellent. The silicon detector has a quantum efficiency (charge carriers detected per incident photon) of nearly 100 percent in the visible portion of the spectrum. This sensitivity extends in the infrared to nearly 1200 nanometers, and also well into the ultraviolet. In addition, the tube is immune to any image burn-in from intense light, even an unattenuated image of the sun.

The most recent development with this type of tube is a reduced-blooming target that greatly reduces the highlight-blooming characteristic of the silicon-target tubes. (Blooming is a lateral spreading of the highlight image that occurs when the signal amplitude is limited by either complete discharge of the target or by the use of insufficient beam current to handle the signal that is being developed.) Continuing improvements in manufacturing efficiency and techniques have reduced the number and prominence of blemishes resulting from faulty diodes and have also produced a lower-cost tube. The mechanical strength of the tube has been increased by mounting the target as a part of the faceplate assembly. All of these features expand the usefulness of the silicon-target tube in existing applications and extend its use to new areas.

The silicon target is now used in both the original 1-inch-diameter tubes and in smaller 2/3-inch-diameter tubes. With some modification it can also be used in silicon-intensifier-type (SIT) tubes.¹

The electronic mechanism

At first glance, the silicon-target vidicon looks disarmingly like the more familiar

camera tube, the antimony trisulfide vidicon. The same bulb, basing, and electron gun is used for both types; only the target is different. The target, of course, is where most of the action is. In any vidicon the target has two distinct functions: 1) to convert a pattern of light and dark (the image) into an electrical charge pattern, and 2) to accumulate and store this pattern until the electron beam and scanning mechanism get around to reading out the signal.

The conventional vidicon uses a continuous photoconductive coating as its target.

Fig. 1 illustrates the conventional photoconductive vidicon target, a soot-like film of antimony trisulfide evaporated over a transparent conductive coating on the inner surface of the faceplate glass. The target is a capacitor that has been charged by connecting the scanned surface to ground potential by means of a moving electron beam. At each point of illumination, the light produces electrical conduction, which partially discharges the capacitance in that area. When the scanning electron beam reaches this discharged area, it deposits electrons. This recharging current flows through the target circuit and is used as the output signal. The thinness of the photoconductive layer and its high resistance prevents any significant lateral leakage of the individual image charges.

The silicon target is an array of discrete diodes.

The silicon target, Fig. 2, is a separate structure, independent of the faceplate. The beam side of the silicon wafer contains an array of diodes that performs the storage function; light detection takes place throughout the n region. Silicon is highly absorbent in the visible spectrum, so most of the incident light is absorbed and does generate charge carriers.

However, as illustrated in Fig. 3, the charge-isolation mechanism in the silicon-

Reprint RE-22-4-7
Final manuscript received June 17, 1976.

target is different from that in the conventional target. Although silicon is highly conductive, the diodes, as will be seen, are reverse-biased. They are thus like little islands that store electrons left by the beam as it charges them negatively to the potential of the cathode of the electron gun. These charges on the diodes are prevented from wandering about by the insulating effect of the reverse-biased junction.

The remainder of the target action of the silicon-target vidicon is better understood by looking at Fig. 4, which shows a small section of the target with its diode cell charged; an energy-level plot taken through the section is at the top of the figure. Somewhere in the bulk of the silicon, an incoming photon of light yields its energy to the release of a pair of charge carriers; the negative electron moves to the left. The left side of the target has been doped to facilitate the capture of electrons. The positive hole moves toward the right, "falls" up into the electron-rich p-region, and reduces the stored free-electron inventory there by one. Only when the scanning electron beam recharges the diodes does current flow through the E_T -circuit (Fig. 3). This recharging current constitutes the video signal.

Silicon-target construction

Starting with a wafer of n-type silicon, and then using techniques familiar to the transistor and integrated-circuit manufacturer, the wafer is oxidized, treated with photoresist, exposed to a pattern of dots, and then etched, leaving openings in the oxide corresponding to the dots. A p-type dopant is then introduced through these openings into the underlying silicon; the result is the diode array. By leaving the oxide on to cover the area between diode centers, the scanning beam is prevented from landing outside of the diode and thus finding a short-circuit path through the silicon.

There are two methods of shielding the oxide from the scanning beam.

It is also important, however, to keep the oxide from being exposed to the beam, for if it is, it will accept and store electron charge negatively to the point where it will repel any further incoming beam; i.e., even the openings into the diode centers will be pinched-off from receiving any more electrons. To prevent this, a pattern of raised conducting "beam-landing pads" have

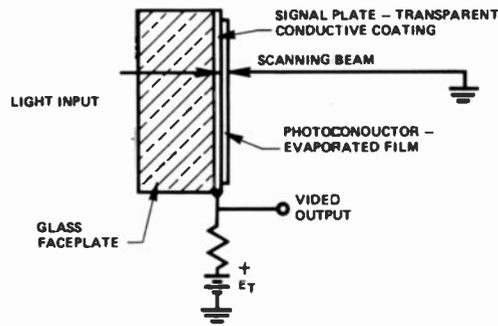


Fig. 1 **Conventional vidicon target** consists of a photoconductive film evaporated onto the glass faceplate. Target acts as a capacitor that is charged by connecting it to ground through the scanning beam. Lighted areas conduct, and so partially discharge the capacitance in that area, so the scanning beam will deposit electrons there. This recharging current is used as the tube's output signal.

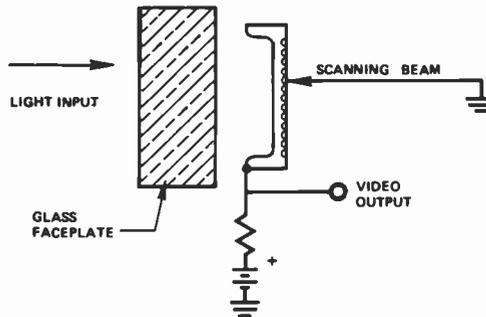


Fig. 2 **Silicon target** is independent of faceplate. Beam side of the wafer consists of an array of diodes that are discharged by the presence of light. Although the structure (see Figs. 3 and 4) is different from the vidicon, the beam's recharging current is similarly used as the video output.

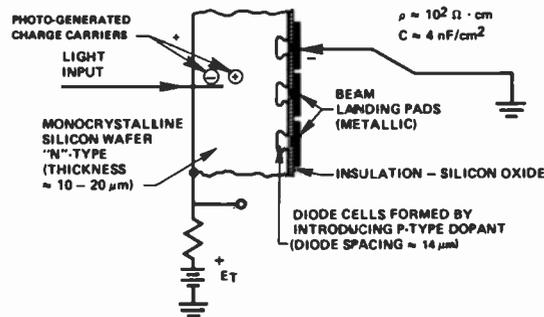


Fig. 3 **Reverse-biased silicon diodes** act as islands that store electrons left by the scanning beam as it charges them to the negative potential of the electron-gun cathode. Highly conductive beam-landing pads shield oxide layer from electron beam.

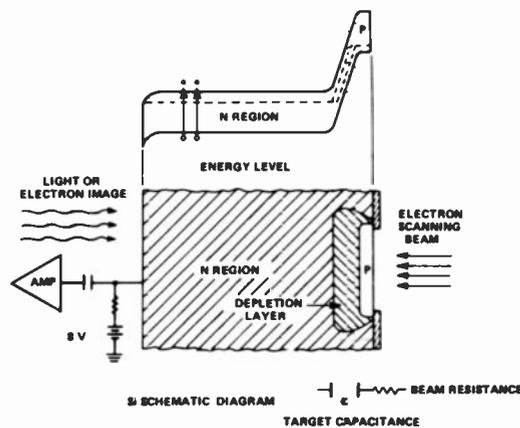


Fig. 4 **Each photon striking the target** produces a pair of charge carriers in the bulk of the silicon layer. The electron is attracted by the positively doped left-hand side of the target, while the positive hole moves into the p region and reduces the stored free-electron inventory there by one. The target capacitance stores the electron charge there until the scanning beam replaces this electron and is read out as the video signal when it does so.

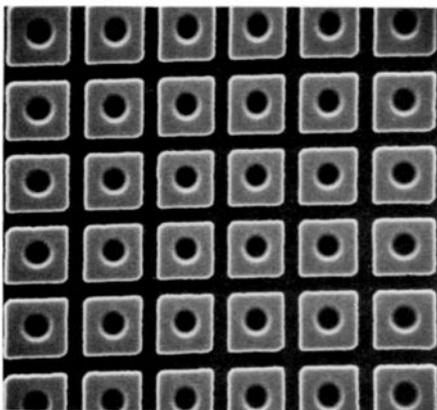


Fig. 5
Scanned side of the target in an SEM photograph. Diodes are circles, beam-landing pads are squares, and the oxide layer behind them is background. Diodes are on 0.014-mm centers; patterned structure must isolate diodes from each other, yet shield the oxide layer from the scanning beam as completely as possible.

been added to cover all but a small portion of the exposed oxide (Fig. 3). All of this is in a pattern that repeats at approximately 1800 to the inch!

Fig. 5 is a scanning-electron-microscope photograph of the scanned side of a portion of the completed RCA silicon-diode target. The precision fabrication of these pads gives nearly complete shielding of the silicon dioxide from the scanning beam, yet still isolates each diode pad. An alternative means of protecting the oxide insulator, preventing it from charging, is shown in Fig. 6. Some manufacturers use this method, in which the oxide and diode structure is overlaid with a "somewhat conducting" layer called a "resistive sea," instead of individual highly conductive beam-landing pads. It is apparent that this overlayer interconnects the diodes, which,

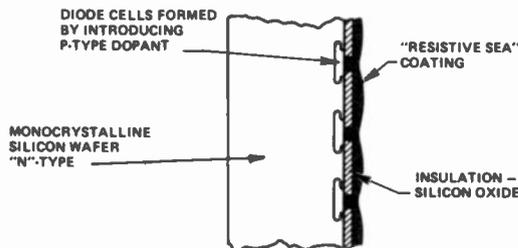


Fig. 6
Resistive-sea construction is another method of isolating diodes and shielding the oxide layer. A relatively easy-to-manufacture "somewhat conducting" coating performs the task, but lowers resolution because of diode interconnection.

up to now, had been isolated, and that the layer must have better conductivity than the oxide or it will be useless.

The resistive-sea technique is advantageous from a manufacturing standpoint in that it is not a pattern structure, as are the landing pads, so it does not need to be registered with the diode pattern. Being unpatterned, the resistive sea does not contribute discrete defective areas (spot blemishes) of its own and even tends to blend or smooth out some of the less-profound spot blemishes in the underlying pattern. However, using a resistive sea results in a loss of resolution, and with high-contrast images, the lateral leakage path appears to promote premature blooming.

Silicon-target characteristics

Capacitance stores the image charge.

The preceding sections have described the target's first function—converting a light pattern into an electrical pattern. The second function, pattern storage, is accomplished in the silicon target by storing the image charge with the capacitance of the diodes. The diode storage capacitance is the combination of the capacitance to the silicon through the depletion zones and the capacitance of the diode pads to the silicon across the silicon oxide insulator (Fig. 4). The storage capacitance is approximately 4500 pF in the 1-inch tube and 2200 pF in the 2/3-inch tubes. These values are somewhat dependent on the target voltage; the depletion regions contract at lower voltages, with a resultant increase in capacitance.

Dark current is typically about 10 nA.

Dark current, the signal developed in the absence of light, has several components:

the reverse-bias diode leakage, thermal-carrier generation in the n material, surface injection from the diode perimeter, and conductivity through the oxide insulator. Dark current is temperature-dependent, and doubles for approximately each 9°C increase in temperature. At 30°C, a typical 1-inch tube operates with a dark current of between 7 and 12 nA.

Cosmetic characteristics vary from tube to tube.

The problem of constructing a silicon wafer incorporating 600,000 similarly functioning diodes of relatively consistent and stable operating characteristics is no easy task. The level of blemishes has been and is being improved, and tubes are sold with varying grades of freedom from blemishes. Blemishes remain very stable with tube operating life as long as the target voltage remains unchanged.

There are several typical varieties of blemishes; they are caused by different factors. "White spots" appear under non-illuminated conditions and are also visible when the target is illuminated. Two major causes of white spots are shorted diodes and "unscheduled" holes in the silicon oxide. "Black spots" appear only when the target is illuminated, and may be black or black and white. Larger black spots often have a white halo around a black core. The major causes of black spots are inoperative diodes, missing diodes, connected pads, and particles on the target. Larger, low-contrast defects, such as streaks and smudges, can also be present. These may be caused by silicon-crystal defects, scratches, or contamination.

A low-level crosshatch pattern may also be discernible under some conditions; this effect is only noticeable when the tube is operated at very low signal currents and high amplifier gain. Shading or variations in sensitivity may be observed under certain conditions; this effect is only noticeable when longer wavelength light is used. At these wavelengths, the silicon absorbs only part of the light, and the sensitivity variations represent thickness variations in the silicon wafer.

The tube has an optimum target voltage.

The silicon-target vidicon tube uses the same gun design as a standard vidicon. Voltages are similar, except that the silicon-target tube has limited wall-mesh and focus-electrode voltage ratings and a low target voltage. The target voltage is

fixed and must not be changed. Even momentary increases in target voltage during tube operation will increase the apparent size and contrast of spots and will affect beam landing by charging the SiO₂ insulator. The optimum target voltage for the RCA silicon-target tube is 8V, for the following reasons.

The upper limit of target voltage is determined by excessive dark current and reverse-bias diode breakdown. The lower limit is determined by ineffective collection of the generated charge carriers, high lag, the inability of the tube to handle reasonable levels of signal current, and lower resolution. Fig. 7 illustrates the dark current and the charge-carrier collection efficiency in terms of signal output, both as a function of the target voltage.

High dark current is generally undesirable, and should be held to a minimum. Excessive diode leakage can appear as white spots when the target voltage reaches 10 to 12 V; hence, the voltage should not exceed 8 to 10 V. The volt-ampere signal output curve of Fig. 7 shows that the carrier collection efficiency increases rapidly to a maximum at about 4 V. At this value, the voltage generated at the depletion region is substantially greater than the energy of the generated carriers, and most of them are collected. No more "sensitivity" will be achieved at higher target voltages. Lag also decreases slightly at higher target voltages as the larger depletion depth reduces the effective capacitance.

If the target voltage is set too low, the charge that can be stored in the target capacitance will be insufficient to produce a high, noise-free signal current. At 8 V, the signal current reaches a limit at about 1100 nA; this voltage has been chosen to produce an optimum tradeoff among the various factors to be considered.

Performance characteristics show advantages and disadvantages.

The most obvious advantages of a silicon-target vidicon are its sensitivity and its resistance to any image burn-in or damage resulting from overexposure. Directly focusing the tube on the sun for several minutes, for example, will not permanently damage the target, and it is virtually impossible to produce any type of retained image burn-in on this type of tube if it is operated at the correct fixed target voltage.

The sensitivity approaches unity quantum

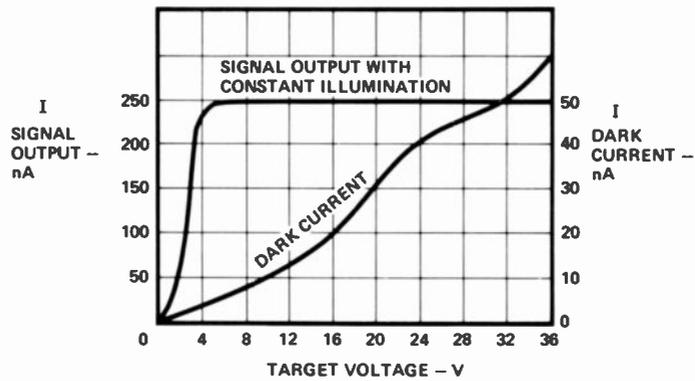


Fig. 7 Target voltage affects dark current and signal output. Optimum operating point is 8 V.

yield (one charge carrier detected per incident photon) throughout the entire visible spectrum and into the infrared, and exceeds that of any other vidicon-type photoconductor in total sensitivity. However, this wide spectral range of sensitivity can be a liability as well as an asset. It is sometimes necessary to restrict the light entering the tube to either the visible or the infrared. This is because lenses are not corrected for both the visible and infrared spectrum, and the lens coatings that are commonly used to reduce the internal lens reflections in the visible portion of the spectrum are not effective in the infrared. In fact, they begin to act as mirrors for infrared light. Internal reflections in the optical system then cause lack of contrast and flare around highlighted areas of the picture. To obtain maximum resolution, it is desirable to exclude light of those wavelengths for which the lens is not corrected. However, the tube's infrared sensitivity can be used to detect hot objects or to view infrared-illuminated scenes where it is impossible or undesirable to use visible illumination.

Table I Sensitivities of photoconductive camera tubes. Values are microamperes per lumen of 2856 K tungsten light.

	Total sensitivity	Infrared excluded
Silicon target vidicon	4350	1100
PbO	450	450
PbO (extended red)	700	550
Vidicon	250*	225

*Medium sensitivity operation

The high sensitivity might also appear to be desirable in a color camera, where lack of sensitivity is usually a major limitation. Table I compares the sensitivities of the various vidicon-type camera tubes in microamperes per lumen of 2856 K tungsten light. An examination of the spectral sensitivity curves of Fig. 8 shows that the silicon-target tube's sensitivity in the blue portion of the spectrum is nearly identical to the sensitivity of the lead-oxide photoconductor of the Vistacon and Plumbicon² tubes used in most color-tv cameras.

Since the blue channel is usually the limiting factor in either sensitivity (noise) or lag in color cameras, the silicon target does not promise to increase the overall sensitivity of color television cameras by any significant amount. A major increase in sensitivity can be achieved, though, in the red channel of a color camera, where the signal can be increased by a factor of 4 above the lead oxide tube's or a factor of 2 above the extended-red lead-oxide tube's. One version of the silicon-target tube with

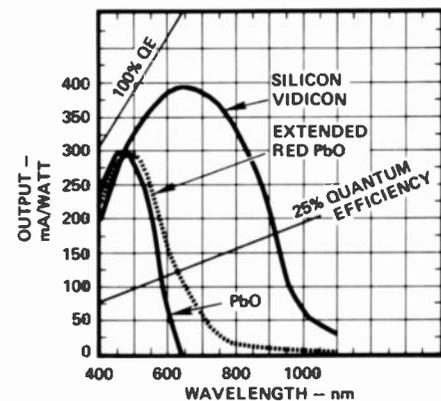


Fig. 8 Spectral response of silicon-target tube is quite close to standard tubes' in blue region, but is much higher in red region.

very tight blemish specifications is being used in the red channel of a well-known three-tube color camera, where its excellent red spectral response and high signal level produce noticeable improvements in noise level and red colorimetry.

"Blooming" is a problem being conquered.

The signal amplitude from a 1-inch silicon-target tube is limited at about 1100 nA. This limiting occurs when the capacitance is fully discharged, and is an asset, since it prevents the intense highlights from developing a very high signal spike that could overload the camera's video-amplifier system. This limiting action also produces a second effect called "blooming," which is a lateral spreading of the highlight image. It occurs when the signal amplitude is limited by either complete discharge of the target or by the use of insufficient beam current to handle the signal being developed on the target.

Blooming is caused by the relatively long lifetime and mobility of the hole carriers and the lack of trapping states in the silicon. Since carriers are free to move laterally to adjacent negatively charged diodes when the field of the diodes under the illumination is reduced to zero, excess hole carriers do so when the highlight illumination completely discharges the target at such a highlighted point. Because of their excellent mobility and relatively long lifetime, these carriers will not "disappear" or recombine with free electrons, but will be accounted for elsewhere on the target. Therefore, instead of producing a high-amplitude signal at the point of highlight illumination, the carriers spread laterally and produce a large white blob that is completely devoid of picture detail.

A new line of reduced-blooming targets and tubes has been developed that promises to eventually replace the standard tubes. The comparison photos of Fig. 9 show how the new tube reduces blooming in a practical situation.

Light-sensitivity control is via a lens-iris/filter system.

The silicon-target tube, along with the lead-oxide tube, has a fixed sensitivity. Unlike the Sb_2S_3 vidicon, its sensitivity cannot be reduced or varied by changing the target voltage or the voltage on any other electrode on the tube. Changes in illumination level must be accommodated by changing



Fig. 9 Improved picture of a reduced-blooming silicon-target tube (left) compared with picture from a conventional silicon-target tube (right).

the quantity of light impinging on the tube by using a lens iris or appropriate light-attenuating filters. Either method presents some operational problems in general-purpose surveillance-camera use. The tube is sensitive enough to produce an acceptable picture with as little as 0.1 footcandle of illumination with an $f/1.4$ lens (full daylight illumination is 5×10^3 footcandles).

If the silicon-target tubes are to be useful over this entire 50,000:1 illumination range, the iris of a conventional lens system that can only handle an illumination range of 250:1 must be supplemented with filters capable of handling an additional illumination range of 200:1. To accommodate this wide range of illumination, a series of lenses that employ a graded neutral-density dot in the center of the lens has been developed by various manufacturers. This further attenuates the light level at high f numbers and provides automatic operation with scene illumination levels varying over a 10,000:1 range.

Lag is lower than for many other camera tubes.

Lag characteristics determine the ability of a tube to capture motion. Carriers have a very long lifetime in silicon because there are very few areas where they can be trapped, i.e., there are very few sites where recombination of carriers can take place. However, the lifetime is much shorter than the television frame time. Since there is very little charge-trapping taking place in the target, the conduction mechanism starts and stops very quickly as the light changes at any point on the target. Therefore any lag effects are caused primarily by the storage capacitance of the target and the beam characteristics.

Referring to the equivalent circuit of Fig. 4, the beam resistance approaches infinity as the charge voltage on the diode pad reaches 0 V during the beam read-out period. This high-RC series circuit produces a noticeable delay in the signal read-out at low signal levels. (The effect is negligible

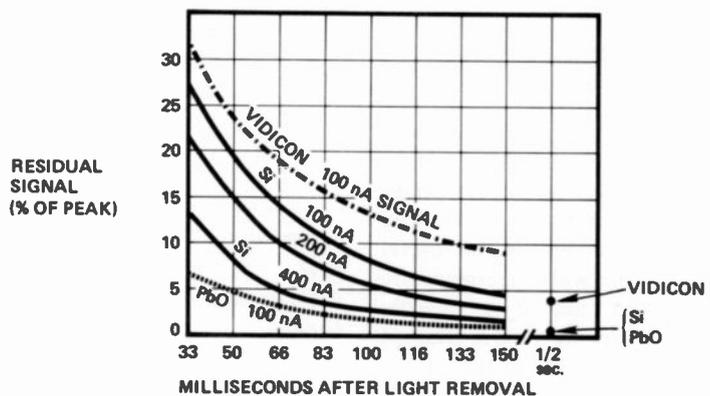


Fig. 10 Lag is nearly undetectable in an Si tube operating with a 400-nA signal current. The tube's high sensitivity and finite dark current make this possible.

when large signals are being generated.) Fig. 10 shows the comparative lag characteristics of the various types of tubes. The "early" part of the curve results in loss of "dynamic" resolution; the latter part of the curves illustrate the relative amount of "smear" of a moving object. Operating at the higher signal currents obviously decreases the lag.

Since the silicon-target tube has much greater sensitivity than many other camera tubes, the lag for a given scene illumination can be much lower. In fact, it is nearly undetectable when signal currents of 500 nA are used. The finite dark current of the silicon target also contributes to lower lag because it provides a bias or fixed dark-signal level that is well above zero-signal level. Thus, it is easier for the beam to recharge the storage capacitance in dark areas of the picture.

The capacitance of the silicon target in a 1-inch tube is approximately 4500 pF. (By comparison, the 30-millimeter lead-oxide tube has a capacitance of 1500 pF, and the 1-inch antimony trisulphide vidicon, 1200 pF.) A small silicon-target tube will have a reduced capacitance in direct proportion to the reduction in scan area.

Resolution is presently lower than for standard vidicons

The resolution in the 1-inch silicon tube is limited partially by the number of diodes and partially by the electron-optics of the beam-forming and deflecting systems. The 1829 diodes per inch on the RCA target corresponds to 700 diodes per picture height in the 1-inch tube, which limits the resolution of the tube to approximately 700

tv lines per picture height. Resolution, however, is less in the small-sized tube.

Resolution is best defined by a modulation-transfer-function curve or, as it is commonly called, an amplitude-response curve. A typical response curve for a 1-inch tube is shown in Fig. 11, along with the responses for an antimony trisulphide vidicon and a 2/3-inch silicon tube. The silicon-tube amplitude response could be increased somewhat with higher focusing fields and electron-gun voltages. (Note the improvement generally achieved in Sb_2S_3 vidicons at higher voltages.) However, at the present state of the art, either X-rays or gas ions generated in the tube at these higher voltages strike the silicon target and cause the dark current to rise rapidly and permanently. At a wall mesh voltage of 700 volts or more, the dark current will rise to unacceptable levels in a few hundred hours. At a maximum mesh electrode rating of 500 volts, the predicted life of silicon targets is on the order of 50,000 hours.

Camera design considerations

The 8-V operating target voltage places a special requirement on the deflecting and focusing coil system. These coils should be designed so that the beam-landing error will be less than 0.5 V. This limit is necessary to assure that the full target voltage will be applied uniformly over the entire target area. The dark current of a silicon target is reasonably low, in the neighborhood of 10 nA at 30°C; this current is only 3% of a typical 300-nA signal and does change with temperature. Because, as described above, target-voltage adjustment cannot be used as a means to reduce dark current, the camera should be designed to keep the temperature of the tube from ranging too widely, a condition that could lead to unacceptably high dark current.

The silicon-target tube is a linear device, and signal output up to the signal-limitation point is directly proportional to the amount of light input. This condition produces an accurate signal that is an exact analog of the light intensity at any point on the image. Lead-oxide tubes have a similar linear characteristic. However, the signal from a linear tube must be processed by black stretching, or "gamma correction," if proper tonal values are desired on the reproducing picture tube. This action, of course, stretches the blacks and enhances the noticeability of any lag.

Quality levels and tube life

It is a continual challenge to manufacture a silicon tube that is completely free of defects. Since defects are caused primarily by shorted, inoperative, or bridged diodes, available silicon tubes are classified according to the blemish defect level, and price varies inversely as the number of defects. At present, few silicon-target tubes can be made that approach the quality required for tv broadcast use in the color channel producing the bulk of the picture detail and brightness signal.

The life of a silicon-target tube operating within the maximum voltage ratings is primarily determined by the life of the electron gun, which is in the neighborhood of 15,000 hours. Sensitivity, lag, and all other important performance characteristics remain constant through life, and no residual-image signals accumulate with life. Blemishes do not grow during life and only occasionally will a diode short produce a white spot.

Typical uses

Silicon-target tubes are replacing standard vidicons in many areas.

The list of uses for the silicon-target vidicon is growing constantly. One group of applications involves upgrading the performance, especially sensitivity, of existing general-purpose tv systems. Property protection is a good example of an application in which existing camera installations are being retrofitted with the silicon tube. The results of the retrofit are quite dramatic, especially where the lighting is produced by a red-rich tungsten source.

The fact that the target is not destroyed when it is focused on the sun (or other intense source) has been the basis for yet other retrofits. Of course, if the tube receives too much light, the picture washes out completely. But, the tube has the important feature that, as soon as the overlighting transient is removed, the silicon vidicon is back in normal operation.

The similarity of the silicon vidicon's gamma characteristic to that of the Plumbicon² and Vistacon lead-oxide tubes has led to another color-camera design improvement. The problem has been with the red deficiency of lead oxide in color cameras. Silicon tubes, with red sensitivity to spare, are used in the red channels of these cameras with the lead-oxide tubes remaining in the green and blue sockets.

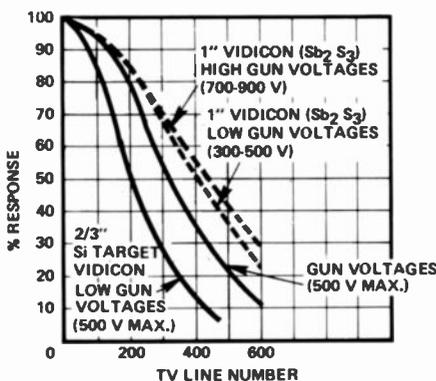


Fig. 11
Amplitude-response curve shows resolution that is presently lower than for standard vidicons.

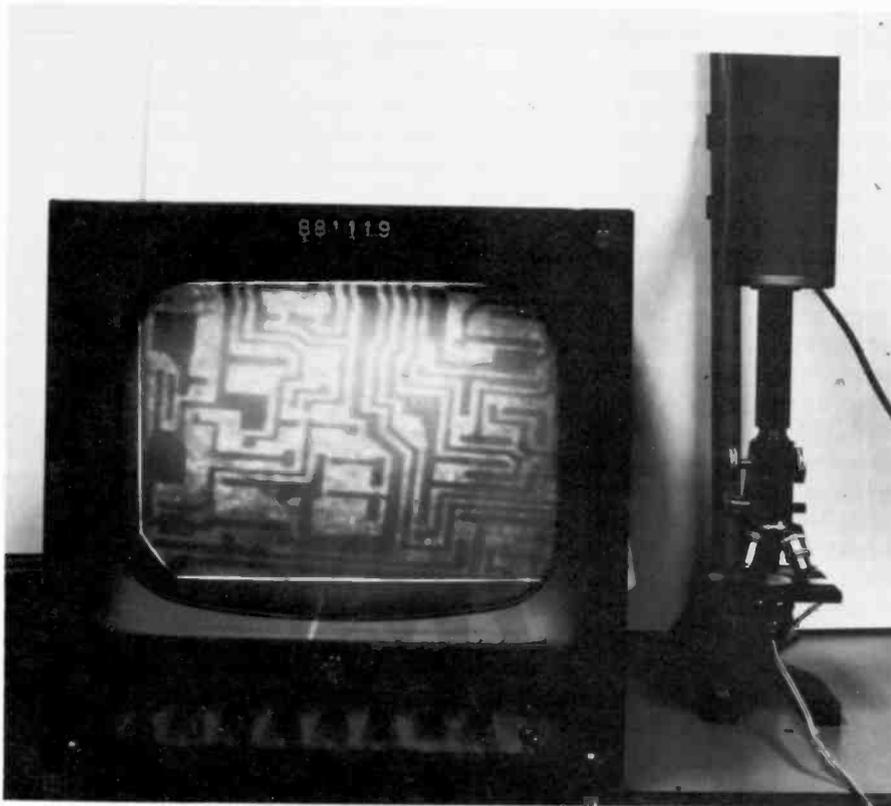


Fig. 12
Inspection application uses silicon vidicon's ir-sensitivity to "see" through silicon IC wafer.

The tubes are also finding uses in entirely new areas.

The secondary category of uses includes new, novel and/or previously impractical forms of television cameras. For example:

A silicon-target vidicon camera coupled to a microscope can "see" through materials that are opaque to the eye—if the materials transmit infrared. One such material is the silicon wafer itself used in making the camera tube or in manufacturing transistor and integrated-circuit devices (Fig. 12).

The behavior of gallium-arsenide light-emitting diodes operating at their typical 9000 Å can be monitored in real time by the use of the silicon-target vidicon.

Several companies are working with tv-style spectrometers that aid in pollution control by scanning the spectra produced in chemical phosphorescence or during rapid oxidation (flame spectra), Fig. 13. The scanning rates are extremely fast when compared with mechanical-scanning and data-recording schemes, and the spectrometer outputs are well suited to oscilloscope display and computer processing.

Extensions to performance and future prospects

Silicon-target tubes with very good ultraviolet response can be made. Their response to ultraviolet light seems to be limited primarily by the optical transmission characteristics of the tube's faceplate and secondarily by the surface

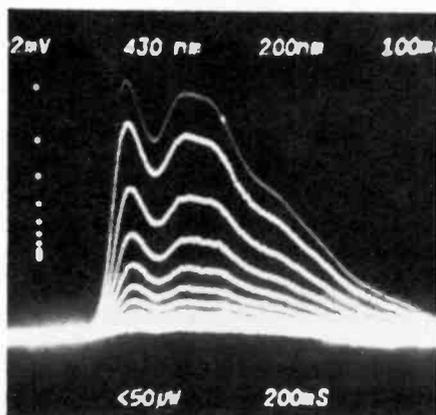


Fig. 13
Rapid-scanning tv-style spectrometer uses silicon target tube as its detector. These curves measure phosphorescence delay; other applications are with flame spectra. (Photo courtesy Tektronix.)

characteristics of the silicon target. (Strongly absorbed short-wavelength light generates carriers in the silicon surface layers that are unavoidably n⁺ type, and carrier recombination takes place rather easily in this region. This effect begins to limit the effective sensitivity of the tube at the shorter wavelengths.) Reasonably good sensitivity out to wavelengths of 200 nm has been measured on developmental models.

The infrared sensitivity of the target can be increased by increasing its thickness and hence the absorption of the more weakly absorbed longer wavelengths. This increase in thickness is accompanied by some loss in resolution, as the carriers diffuse laterally before falling under the influence of the field generated by the depletion regions surrounding the diode sites. Special tubes with enhanced ultraviolet and infrared sensitivity are currently available on a custom basis, awaiting the development of a viable market and a definition of general performance specifications that the tubes would have to meet in new and unique situations.

The silicon-target vidicon is replacing the conventional vidicon in many applications, but more importantly, by virtue of its very wide spectral response and other unique characteristics, it is performing in completely new situations where previous camera tubes were useless. Because the target relies on silicon technology and the fabrication techniques of large-scale integrated-circuit technology, its future development will be advanced and aided by the continuing progress in these allied fields.

Acknowledgments

Much credit must be given to P.R. Rule, G.A. Robinson, and the late P.D. Huston, who assisted in preparing the lecture on the silicon-target tube that is the basis of this paper. The assistance of F. Wallace and R. Phillips in interpreting the performance characteristics of the target and the reduced blooming tube is gratefully acknowledged.

References

1. Robinson, G.A.; "The silicon intensifier target (SIT) tube," *this issue*.
2. Plumbicon is a registered trademark of N.V. Philips.
3. Neuhauser, R.G.; and Miller, L.D.; "Beam landing error and signal output uniformity of vidicons," *J. SMPTE*, (Mar 1958) p. 149.
4. Rodgers, R.L., III; "Beam-scanned silicon targets for camera tubes," ST-4936, RCA Solid State Division, Somerville, N.J.

The silicon intensifier target tube: seeing in the dark

G.A. Robinson

The SIT tube's excellent low-light-level characteristics have led to applications ranging from police surveillance to internal inspection of jet engine parts.

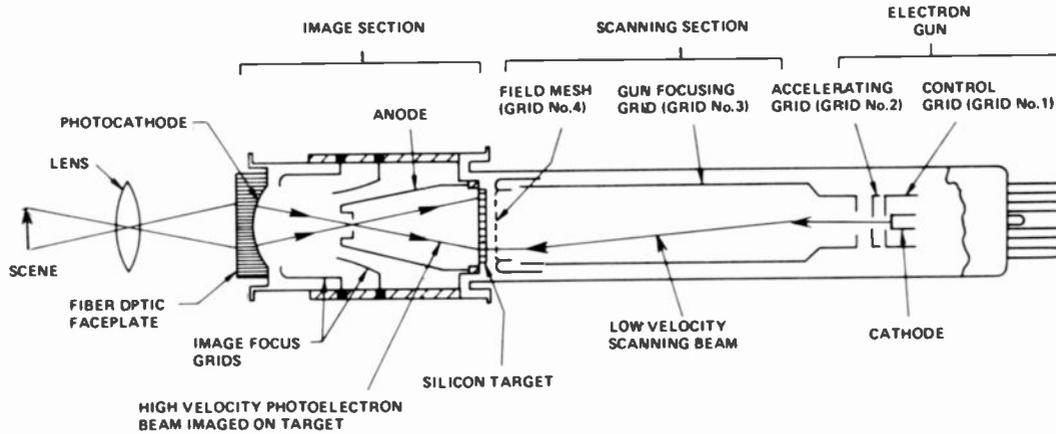


Fig 1
Light enters the SIT tube through a fiber-optic faceplate, which transfers the flat-scene image onto the curved photocathode. The light then travels through the focusing grids and strikes the target, which is a matrix of over 1800 silicon diodes per inch. The image is typically stored there and read out by the scanning beam every 1/30 second.



George Robinson has contributed to the design of flying-spot kinescopes, 1.5-inch magnetic, 1-inch hybrid and all-electrostatic vidicons since joining RCA in 1955. His specialties include electron-gun work, high-resolution vidicons, and camera tubes for military systems and unusual environments. In his present position he is involved with applications at low light levels.

Contact him at:
Applications Engineering
Electro-Optics Products
Solid State Division
Lancaster, Pa.
Ext. 2073

The idea of using a silicon-diode-array target in an intensifier tube (SIT) has received wide acceptance in the last few years in cameras operating at low light levels. This paper discusses some of the characteristics of the SIT tube that make it the leading low-light-level camera tube in use today. Elsewhere in this issue, R.G. Neuhauser¹ discusses how the silicon target is used in vidicon tubes.

Photocathode + silicon-diode array

The SIT tube (Fig. 1) is a photodetector using a photocathode as the light-sensitive surface and a silicon-diode-array target as the surface upon which an image charge pattern can be stored. The SIT tube's gain comes from the high number of hole-electron pairs generated in the target when it is bombarded by high-energy electrons from the photocathode. In the silicon-vidicon target some fraction of the incoming photons creates a hole-electron pair, but in the silicon target of the SIT tube each photoelectron can create many hole-electron pairs. A tube operating with 9000 volts across the intensifier section will have an electron gain of about 1600.

The relationship of gain to intensifier voltage is shown in Fig. 2. By introducing an energy-absorbing "buffer layer" in front

of the target, the useful gain characteristic is kept above 3000 volts, where the intensifier section performs best. The buffer layer selectively absorbs photoelectrons. If all photoelectrons had equal energy, the

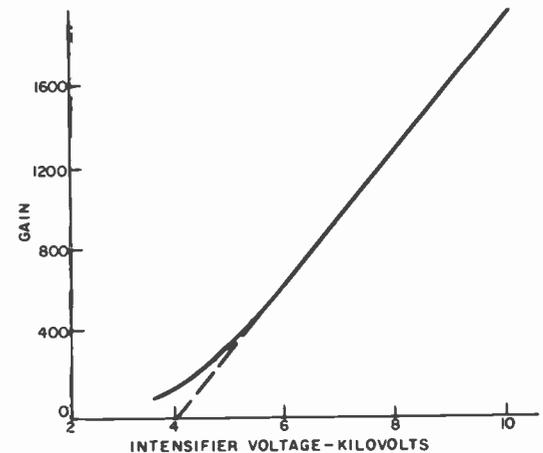


Fig. 2
Electron gain (solid line) is an essentially linear function of the intensifier voltage. An energy-absorbing "buffer layer" keeps the gain above 3 kV, where the intensifier section performs best. Deviation from linearity (the dashed line) is caused by high-energy photons penetrating the buffer layer.

Reprint RE-22-4-6
Final manuscript received October 22, 1976.

gain would be expected to continue its linear decrease as voltage is reduced until a cutoff voltage is reached. But as Fig. 2 shows, some gain exists below the projected cutoff voltage, suggesting that some higher-energy electrons are getting through the buffer layer.

An SIT tube with a gain of 1600 and a photocathode responsivity of $140 \mu\text{A}/\text{lm}$ will be approximately 50 times more sensitive to tungsten illumination than a silicon-target vidicon with a published sensitivity of $4350 \mu\text{A}/\text{lm}$.

Photocathode

The photocathode surface is the basic S-20 multi-alkali (Na-K-Cs-Sb), and is placed on the inside of a curved fiber-optic faceplate. Since fiber-optic plates have poor ultraviolet transmission, the tube response is low to the shorter wavelengths. The 340-nanometer cutoff can be seen in

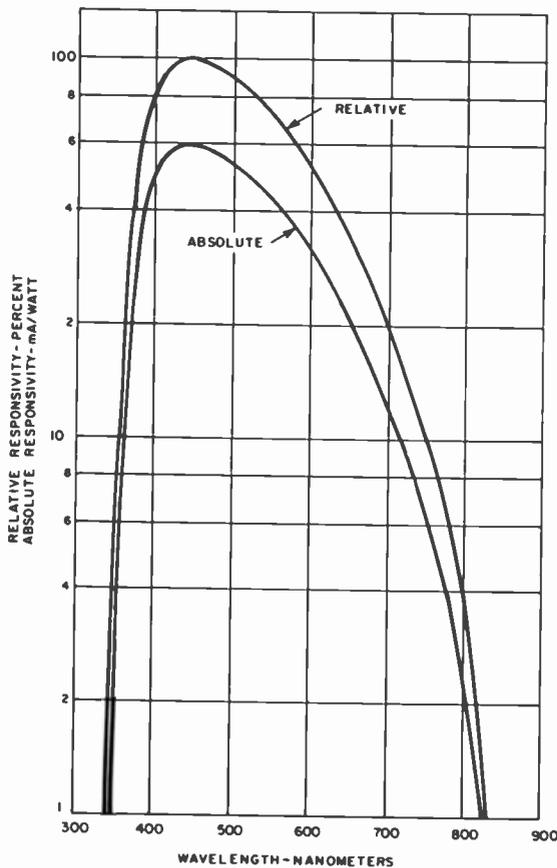


Fig. 3
Photocathode response is limited by the low uv transmission of the fiber-optic faceplate. (This can be corrected by using a uv-sensitive scintillator, however.) Tube also does not have the near-ir response of the vidicon.

the response curve of Fig. 3. Some users overcome this deficiency by applying an ultraviolet-sensitive scintillator to the outside of the faceplate.

The response to the longer wavelengths is limited by tube processing of the S-20 photocathode; the extended-red (ERMA) photocathode is not presently obtainable in the SIT tube. Because the silicon target is bombarded by electrons rather than photons, the basic near-ir response of the target, which is shown in the vidicon, is lost.

Image section

The image section of the SIT tube inverts the photoelectron image and focuses it onto the silicon target. The electron optics of this process requires a spherical photocathode surface, but a conventional lens system focuses the image of a scene onto a flat plane. Therefore, it is convenient to have a fiber-optic faceplate transmit the flat-scene image onto the curved photocathode. The faceplate, which is thicker at its edge than at its center, introduces a fixed "shading" signal into a flat-light field because of the transmission difference across the faceplate. This shading accounts for a drop in signal output of approximately 17% at the corners. Also, geometric distortion increases by about 2% as a result of the transition from a plane image to a spherical surface. This distortion is displayed as "pin-cushion" effect on the camera monitor.

Resolution

The matrix structure of the silicon target limits the resolution performance of the SIT tube. The existing target has a density of over 1800 diodes per inch, or about 35 line pairs per millimeter of resolution. In a nominal 16-mm optical image ($1/2 \times 3/8$ in.) this condition will result in a limiting resolution of about 700 tv lines per picture height. Higher resolution is obtained by using a larger-sized target rather than a target of higher diode density. A 27-mm target with a limiting-resolution capability in excess of 1000 tv lines per picture height is currently being used in 1.5-in.-bulb-dia. tubes.

The contrast transfer function (CTF), or square-wave amplitude response characteristic, can be useful in determining resolution performance. The complete curve shown in Fig. 4 can be obtained in the

laboratory by viewing high-contrast patterns of parallel white and black bars. This curve shows the relationship of output signal to bar width. The signal reduction associated with small images is significant in that the signal-to-noise ratio is directly affected and is important in determining the low-light limitation of operation. This limitation may be reached at light levels higher than expected because actual scenes are not made up of parallel black and white bars. The effect upon amplitude response when looking at points rather than bars can be approximated by squaring the CTF characteristic curve, Fig. 4.

Resolution also degrades as light level is lowered because lower output signals affect the signal-to-noise ratio. Fig. 5 shows the relationship of light level and limiting resolution; the curve was made using a static scene consisting of black and white bars. Two different contrast levels are shown: the 100% level is typical of a laboratory-type evaluation, while the 30% level corresponds more closely to typical outdoor scenes.

Lag

Resolution is only one of the important characteristics in low-light-level operation. A second very important characteristic is lag, which becomes worse as the signal level decreases. Lag is the residual signal measured in the dark and is expressed as the percentage of the original signal present after three fields of scanning in the dark.

SIT tubes exhibit no photoconductive lag, but there is some capacitive lag resulting from the finite time it takes for the electron beam to remove accumulated charge from the target. A target with high capacitance will store relatively more charge with less voltage change than will a low-capacitance target ($C = dQ/dV$). However, it will take a longer time for the beam to discharge the signal because of the electron velocity distribution within the beam. It is for this reason that lag increases as light level (and charge) decreases. Fig. 6 shows typical "third-field" lag for a 4804 tube as a function of light level.

It is possible to improve lag by artificially raising the "zero-signal" voltage so that the beam electrons can discharge the target more effectively. This can be done either by simulating an increased dark current (using bias lighting) or by actually increasing the dark current (increasing target voltage or raising target temperature).

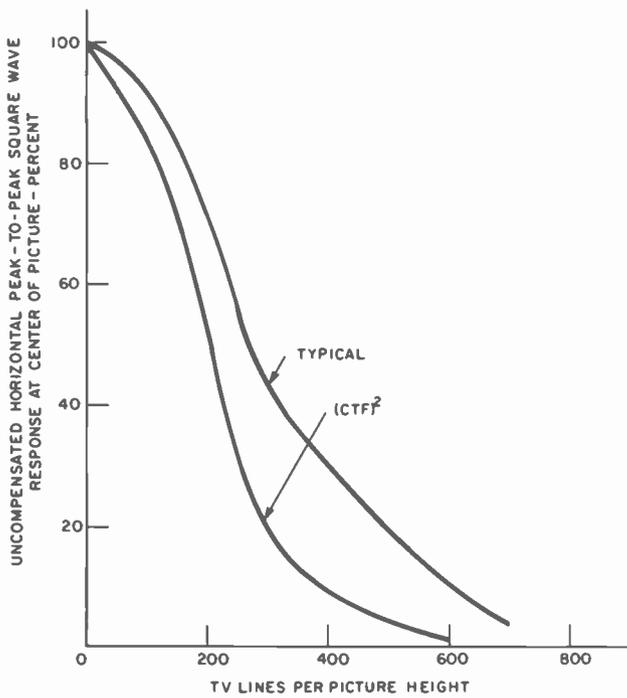


Fig. 4 Response to black-and-white parallel-bar pattern depends on bar width. Curve is called the *contrast transfer function (CTF)*; squaring it gives the effect upon amplitude response when the tube sees points instead of bars.

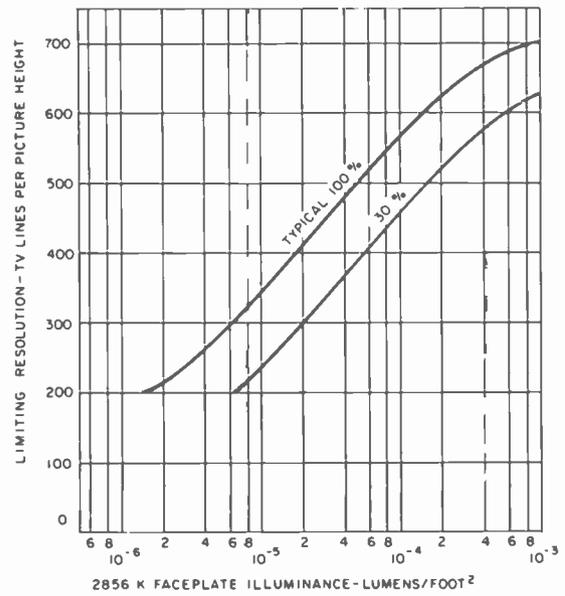


Fig. 5 Low light levels lower resolution—100% and 30% levels correspond to laboratory and outdoor environments.

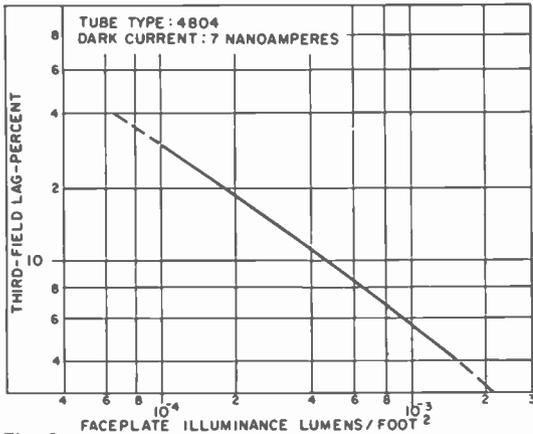


Fig. 6 Lag is a measure of the residual signal present after scanning the field in the dark. Curve shows lag for a standard 4804 tube after scanning three fields in the dark.

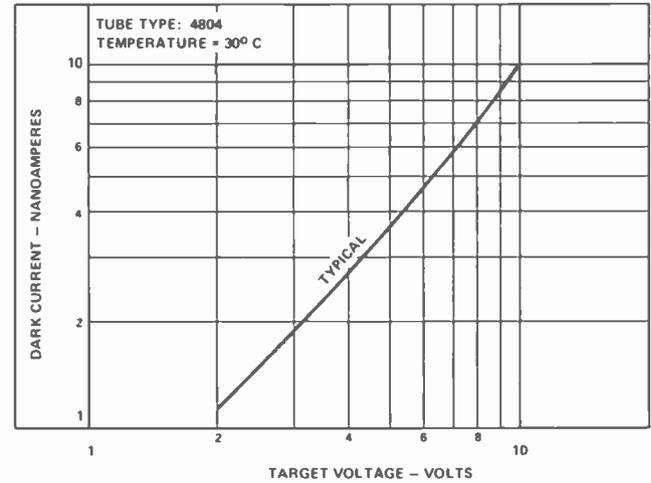


Fig. 7 Dark current is caused by thermally-developed carriers in the target, and is seen as a background to the image. Dark current is a function of both target voltage and temperature.

Dark current

Dark current manifests itself as a background to the image. This background is a result of thermally developed carriers in the target and is usually not a problem except for a low-level grain or mottling that is visible only at extremely low signal levels. It is normally desirable to keep the dark current low so that the dark portion of the scene will have a voltage reference near zero. There may be circumstances when the dark current will be increased to compromise performance, such as to improve lag, as discussed previously. If this is done by increasing target voltage above the nominal 8 V, then the maximum signal

capability of the tube will be increased as well. Normally this will be of no concern to the user, because the tube generally has sufficient signal-handling capability. However, whenever the target voltage is increased, the maximum value should not exceed 15 V because of the increased prominence of target defects. Fig. 7 shows the typical relationship of dark current to target voltage. This curve is for a 30°C operating temperature; dark current increases with increased operating temperature.

Operation at low light levels

The SIT tube is basically a low-light-level device. Even though it has a variable-gain

feature, as shown in Fig. 2, around-the-clock operation should not be attempted without auxiliary light-control capability. Rather than operate the tube at high illumination for long periods, it is better to incorporate the variable-gain feature of the tube into the camera design as a fast-acting automatic light control (ALC). As an example, the nominal operating voltage might be chosen as about 4500 volts, where the gain is about 200 (3 lens stops away from full gain). At this operating point the tube will have a good signal-to-noise ratio, the picture will be pleasing, and the tube will have long life. The voltage (gain) can be varied from this operating point to handle light levels changing up or down by as much as a factor of 8 in less than a second.

Then, at a slower rate, lens-iris adjustment and/or filter insertion can bring the light level back to the value that will allow 4500-volt operation.

Tube life

When caution is exercised in using the SIT tube, reduced thermionic cathode emission will end its life, as in other camera tubes. A calculated mean-time-to-failure (MTTF) in excess of 5500 hours (to 90-percent confidence) has been maintained throughout the last three years of production. Faceplate exposure must be controlled for long life. By following recommended camera design considerations, the operating time before the onset of damage will be 2000 hr. Protection must be provided against two types of damage mechanisms: target damage from high-energy electron bombardment and photocathode damage from ion bombardment.

Target damage

In addition to time and illumination, the energy of photoelectrons that bombard the target determines the extent of target damage. Excessive exposure will cause a permanent increase of dark current at the point of impact on the target. If the damage becomes severe, it will be evident in the dark portions of a scene or even in total darkness with the high voltage completely removed, Fig. 8. Damage can be related to signal level on the target and can be controlled by keeping the signal within bounds.

Damage is most apt to occur where small, intense sources are present. However, in



Fig. 8
Severe target burn caused by excessive exposure. Damage is evident here even with the high voltage removed.

such an application it is usually necessary to obtain information from the dark background surrounding the small source, so it is not practical to reduce gain and signal level just to protect the target from the exposure of the small area. If prolonged operation is necessary under these conditions, it is best to move the camera so that the small area does not remain in one spot.

Target damage is most likely to occur in an unattended camera. In an attended and correctly operating camera, the overexposure of a small spot will usually bloom to an unusable degree before a damage level is reached. If the over-illuminated area becomes large and takes up a significant portion of the picture area, the chance for photocathode damage increases.

Photocathode damage

Ion damage to the photocathode results in a poorly defined dark spot in the center of the picture, Fig. 9. This dark spot is actually an area of reduced photocathode sensitivity and cannot be seen in the dark or with the high voltage removed. The damage results from the bombardment of positive ions, originating from collisions between photoelectrons and residual gas molecules, that are accelerated toward the negative potential of the photocathode. In normal operation the number of photoelectrons is never high enough to generate a damaging number of ions; it is only when the photocathode current becomes excessive that the number of ions reaches a damaging level.

Ion generation is a function of numbers of photoelectrons rather than energy. It is possible to have such a small voltage (as low as 100 volts) on the image section that,

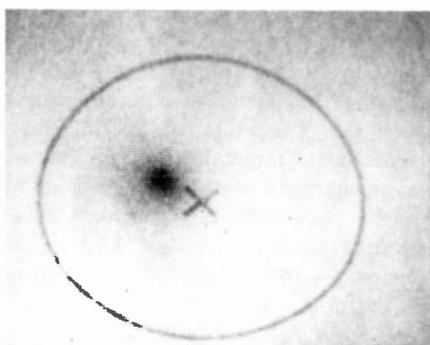


Fig. 9
Photocathode ion damage produces a poorly defined dark spot in the center of the image.

although no picture is present, photoelectrons are flowing as a result of light exposure. Therefore, the only sure methods of photocathode protection are to completely remove all photocathode voltage or to limit the light level.

S/N at low light levels

The degradation of resolution at low light levels is closely related to signal-to-noise ratio. It is desirable to have a camera system that is so quiet and a tube so sensitive that performance will be limited only by the number of photons available from the scene. The SIT tube comes close to reaching this goal. Several factors are involved in evaluating the low-light-level limit of operation: the detector quantum efficiency and its integration with the spectral distribution of available photons; the reflectivity and contrast of the scene; the lens aperture; the solid angle corresponding to the picture element; and the integration time. All these factors are important in establishing an S/N at the target.

When the camera system processes this information, it contributes its own additional noise, the significance of which depends upon the actual signal level coming out of the target. Fig. 10 shows a typical S/N characteristic for a tube and camera. Note that at the higher light levels the S/N does not continue to increase along the "photocathode-limited" line. When full signal output is obtained from the tube, any further light-level increase is accompanied by a gain reduction brought about by decreasing the intensifier voltage. As this lower voltage cuts into the photoelectron energy distribution, the number of primary electrons entering the target through the buffer layer will only be sufficient to maintain signal level, thus flattening the S/N characteristic.

Signal integration

The silicon target stores the charge image until it is scanned off. For broadcast systems and most closed-circuit systems in the United States this integrating time is nominally $1/30$ s. If the photon flux input can be integrated for a longer time, more information will be stored; the results will have an improved signal-to-noise ratio but a loss of motion perception. Operation in an extended integrating mode will be limited by dark-current build-up, which is usually proportional to integrating time.

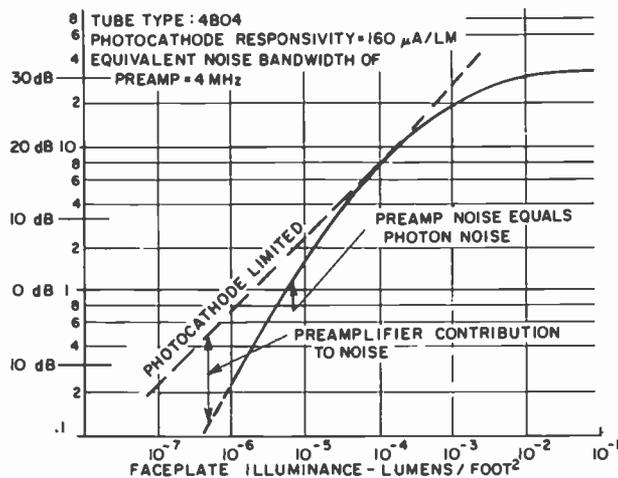


Fig. 10
Typical S/N characteristic for tube and camera. At higher illumination levels, intensifier voltage decreases to limit gain. This lowers the number of primary photoelectrons striking the target, so the S/N flattens out from the photocathode-limited line.

When the tube is operated in this mode, increased dark current resulting from build-up can usually be tolerated up to about a one-second integrating time. Longer times will require target cooling of about 20°C for each order increase in integrating time. In addition, for integrating times in excess of ten minutes, it has been found useful to turn off the gun heater until just before ready to read out the information. Signals have been successfully stored in this manner for up to eight hours before being read out.

Blooming

Blooming is the spread of a highlight image and is associated with most camera tubes. It occurs when portions of the pictures are overloaded with excessive light; the overloaded area then appears larger than it should because excess charge on the target spills over into adjacent areas. Low-light-level scenes are particularly susceptible to blooming because the general content of a scene may have detail with a rather narrow contrast range except for an occasional bright light or flash.

Recently developed targets with a new reduced-blooming feature are now being used with significantly improved results. The 4804H reduced-blooming tube, at extreme overload, can handle intensities 20 times greater than those that can be handled by the conventional 4804 tube.

However, spot intensities greater than 1000 times full signal introduce additional concerns for the user of the reduced-blooming

tube: either further distortion of the spot image as a result of lens flare (internal reflections) or potential damage to the target from overexposure.

Alternate configurations

Although most camera applications involving simple surveillance can use a 16-mm SIT tube with no modification, there are circumstances requiring additional features or characteristics. At extremely low light levels, such as those found outdoors with no artificial illumination, it is necessary to have the pickup device introduce as little noise as possible to the incoming image signal and to have enough gain available so that the output signal-to-noise ratio is not degraded by amplifier noise.

An image-intensifier tube can provide additional gain to an SIT tube so that operation can be realized up to the photoelectron noise limit. The fiber-optic output element of the image tube couples directly to the SIT photocathode through the SIT tube's fiber-optic faceplate. The additional white-light gain of over 20:1 is the ratio of the SIT tube photocathode current to the photocathode current of the image-intensifier.

For additional light collection, it is possible to increase the size of the input aperture from 16 mm to 40 mm. In applications where space is at a premium, an SIT with a smaller, 2/3-inch gun section may be useful. Tubes are also available for unique applications that require either gate or zoom in the intensifier section. Where

improved resolution is required, a 27-millimeter target is available that provides resolution in excess of 1000 tv lines per picture height.

Applications

Many military, medical, and scientific applications, in addition to surveillance applications, have developed for low-light-level television cameras employing the SIT tube.

Airborne cameras used in gunfire control take advantage of the tube's ability to provide useful information under adverse conditions of low contrast and low light level. Cameras on aircraft carriers operating under similar adverse conditions are used to help guide aircraft landings. Other shipboard cameras are used for night maneuvers and harbor piloting. These applications encounter less rapid motion, but present a challenge to blooming control because of bright lights in the harbor or running lights on other ships. Submarine periscopes are also being outfitted with low-light-level cameras.

Parking-lot surveillance applications in both the private and public security domains are well known, as are the search-and-observe functions of the cameras in law-enforcement applications.

Various businesses have found uses for low-light-level cameras, including observing activity in film-processing plants and inspecting internal parts of jet engines during maintenance. Cameras go underwater to aid in oil drilling. Airlines use low-light-level cameras with low-level X-ray systems for baggage inspection. Fishermen use airborne cameras to locate fish schools in the ocean by observing plankton fluorescence.

Scientists can use low-light-level cameras to observe the nocturnal habits of birds and animals or to advance their knowledge in astronomy. Medical applications find low-light-level cameras used in low-light X-ray systems and eye fundus investigations.

References

1. Neuhauser, R.G.; "The silicon target vidicons: new features and expanded uses," this issue.
2. Engstrom, R.W., and Robinson, G.A.; "Choose the tube for I. V.," *Electro-optical Systems Design*, Jun 1971.
3. Rodgers, R.L., III; "Beam scanned silicon targets for camera tubes," IEF, Intercon, Mar 1973.
4. Mesner, M., and Sensenig, W.; "Final report SIT sensor measurement," RCA Astro-Electronics, unpublished.

Avalanche photodiodes: no longer a laboratory curiosity

R.J. McIntyre | P.P. Webb

New APD configurations, including large-area arrays and detector/preamplifier modules, are adding numerous applications to a fast-growing list that includes laser rangefinders and detectors for optical communications.

Avalanche photodiodes (APDs), which were little more than laboratory curiosities a few years ago, have begun to find their way into many electro-optical systems. In this paper we wish to summarize some of these applications, and to point out some other applications in which APDs have not yet been exploited to their full capabilities.

Reprint RE-22-4-15 | Final manuscript received December 13, 1976.

Robert McIntyre has contributed to the understanding of avalanche multiplication in semiconductor diodes, generating theories to explain the noise spectral density and gain distribution in avalanche photodiodes, among others. The photosensor R&D program under his direction has led to the establishment of a rapidly expanding business in this area for RCA Ltd.

Contact him at:
Electro-Optics Department
Solid State Div.
RCA Ltd.
Ste. Anne de Bellevue, Que.
Ext. 340



Basic properties of avalanche photodiodes

The properties of both p-i-n¹ and avalanche photodiodes^{2,3} have been described fully elsewhere. The more important properties of RCA's avalanche photodiodes are as follows:

Paul Webb has been involved in the development of Ge(Li) diodes for nuclear radiation detection in addition to various types of silicon photodiodes. Most recently, he participated in developing a successful process for fabricating large-area, fully-depleted "reach-through" avalanche diodes, both single-element and quadrant.

Contact him at:
Electro-Optics Department
Solid State Div.
RCA Ltd.
Ste. Anne de Bellevue, Que.
Ext. 340



Quantum efficiency is high.

With proper design, avalanche photodiodes can be made with quantum efficiencies greater than 70% from the ultraviolet right through the visible to about 1 μm . With selective coatings, quantum efficiencies of 80 to 95% can probably be achieved at any desired wavelength in the 0.5- to 0.9- μm range. At 1.06 μm the absorption coefficient of silicon is not high enough to allow really high quantum efficiencies, but *QEs* in excess of 30% have been achieved with a wide (200 μm) depletion-layer device.

APDs are fast.

The response time of an avalanche photodiode is limited either by the carrier transit time or, for very fast diodes, by the time for multiplication. For devices designed for use below 0.9 μm , such as the C30884, the depletion layer is 55 μm or less in width, and the total carrier transit time is less than 1 ns. Devices designed for use at 1.06 μm are a little slower, with response times of 3 ns for a 100- μm depletion-layer width (C30817, C30872, C30895), ranging up to 10 ns for devices with 200- μm depletion width. For very fast diodes, (narrow depletion layers), the ultimate speed limitation is the avalanche buildup time. This gives a frequency response of the form $M(\omega) = M_0 / (1 + \omega^2 M_0 \tau_1^2)^{1/2}$, where τ_1 is 4 to 5 $\times 10^{-13}$ s. Thus, the device has a gain-bandwidth product limitation of 300 to 400 GHz.

Their size is small.

Because of the extreme doping uniformity required ($\sim 0.1\%$) to give a respectably

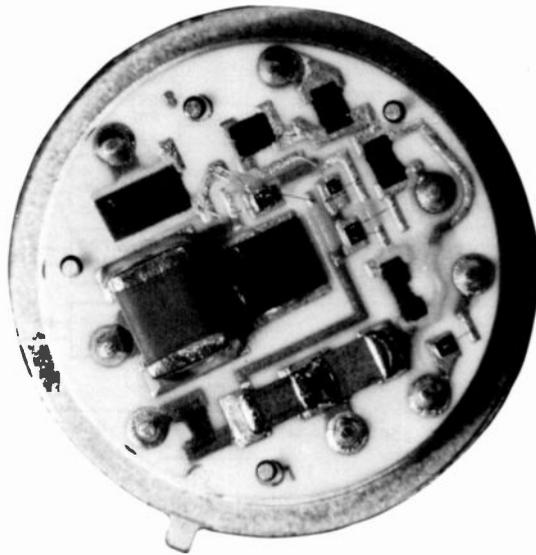


Fig. 1
Avalanche detector-preamplifier module designed for laser rangefinder. Detector chip in center has a sensitive area of 0.5 sq mm.

uniform detector, high-gain avalanche detectors cannot be made very large with any acceptable yield. Most devices of 1 mm diameter or less can be operated at gains up to about 200, with some capable of going much higher. Larger devices are normally limited to somewhat lower gains. The largest device made to date in any quantity, with an area of 5×5 mm, is designed to operate at a gain of 40.

Capacitance depends on the junction area and depletion layer width.

Small devices are in the 0.5- to 2-pF range. Larger devices are about 1 pF/mm².

APDs are quiet or noisy, depending on how you look at it.

APDs are rather noisy amplifiers, having excess noise factors (the ratio of the output-noise power to the multiplied input-noise power) on the order of 3 to 5 (depending on gain), as compared to 1.1 to 1.4 for a photomultiplier (PMT). However, under background-limited conditions, which includes most applications, the signal-to-noise ratio is proportional to η/F , where η is the quantum efficiency and F is the excess noise factor. Because of the higher quantum efficiency of the APD, it is competitive with the PMT throughout the visible range and is better than the PMT in the near infrared (0.8-1.1 μm). Dark currents are low enough ($\sim 10^{-11}$ A/mm² before multiplication at room temperature) that the devices are background-limited at very low light levels. Like photomultipliers, the dark current can be reduced by cooling if necessary.

Applications of avalanche photodiodes

The following partial list of applications shows where APDs have been used to good advantage. New ones are arising every day.

One of the biggest current markets for APDs is for laser rangefinders using either pulsed or swept-frequency modulated cw lasers.

RCA's APDs have been particularly popular for use at 1.06 μm because of their relatively high quantum efficiency ($\approx 20\%$). Fig. 1 shows an APD detector module, with an integrated thick-film preamplifier, designed for this purpose. The module has a responsivity of greater than 4×10^5 V/W, a bandwidth of over 15 MHz, and an NEP

(Noise Equivalent Power) in the dark of less than 10^{-13} W/Hz^{1/2}. This particular module is used in the AN/CVS-5 hand-held laser rangefinder developed by Automated Systems in Burlington for U.S. Army ECOM.⁴

APDs are particularly appropriate detectors for optical communications, both for line-of-sight systems and with fiber optics.

Fig. 2 shows both the measured and calculated pulse responses for a detector designed specifically for use with a 400 Mb/s PCM line-of-sight system at 1.06 μm , which was being evaluated by NASA for possible space use. Using this detector at a gain of 300 to 400 straight into 50 Ω (i.e., no low-noise preamplifier), bit-error-rates (BERs) of less than 10^{-6} were measured with average optical signals of about 90 nW (i.e., ~ 500 photoelectrons, or 2400 photons/pulse). This was considerably better than could be achieved with specially developed cooled photomultipliers that, when new, had quantum efficiencies in the range of 1 to 2% at 1.06 μm . The use of a good low-noise preamplifier should reduce the required signal level by about a factor of two.

Detectors for use with fiber optics are now being developed. Devices with sensitive diameters of 0.020 in. have been fabricated with NEPs of about 10^{-15} W/Hz^{1/2} at 0.8 to 0.9 μm at a gain of about 200 (i.e., 120 A/W).⁵ These devices have response times under 1 ns and show effective dark currents equivalent to less than 1 thermally generated electron per μs , or one every 100 bits in a 100 Mb/s system. Thus, these

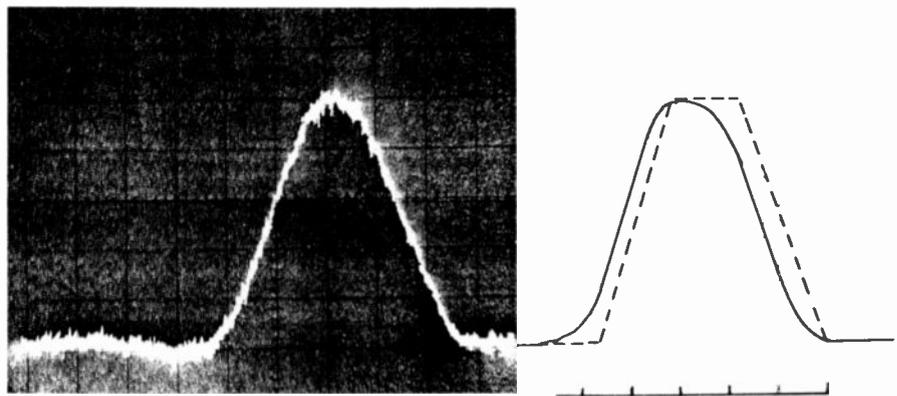


Fig. 2
Response to 100 ps pulse at 1.06 micrometers for detector with 80-micrometer depletion layer width. Time scale = 0.5ns/div; gain = 470. Dashed curve shows theoretical pulse shape on the basis of transit-time effects alone. Solid curve includes effects of avalanche build-up time.

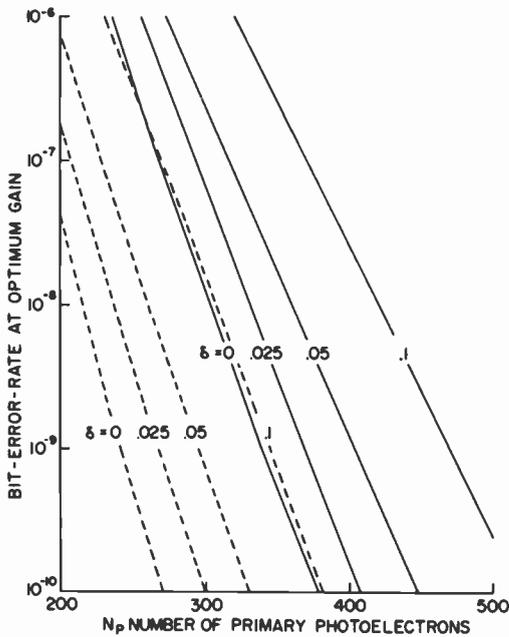


Fig. 3
Very low bit-error rates have been calculated for an optical communications system at optimum gain. Here, δ is the ratio of the 'off' to 'on' pulse. Broken lines: using a preamplifier with a noise equivalent charge of 800 electrons; Solid lines: using a preamplifier with a noise equivalent charge of 2000 electrons.

devices will be limited by the noise in the signal itself, unless the on/off ratio of the pulse signal is greater than about $4 \times 10^4:1$, which is unlikely.

Fig. 3 shows how the BER varies with on/off ratio and amplifier noise as a function of signal level. The amplifier noise-equivalent-charge values of 800 and 2000 would correspond roughly to what is achievable today with ≈ 20 Mb/s and ≈ 100 Mb/s systems respectively. Thus, a BER of 10^{-10} should be achievable with about 425 photons per pulse (385 photoelectrons per pulse), or an average optical power of about 5 nW in a 100 Mb/s system. Since laser diodes are capable of coupling greater than 1 mW into a single fiber, total fiber attenuations of ≈ 55 dB can be tolerated without affecting the error rate appreciably. With the most recent fibers (<5 dB/km) this can mean a repeater spacing of over 10 km in a fiber-optics communications link.

Fig. 4 shows the design of a prototype detector module for use with fiber bundles. This unit, which will accept a fiber bundle up to 0.045 inch in diameter (1 mm^2), has an NEP of about $1.5 \times 10^{-13} \text{ W/Hz}^{1/2}$ at $1.06 \mu\text{m}$, or about 5×10^{-14} at $0.9 \mu\text{m}$. Such units

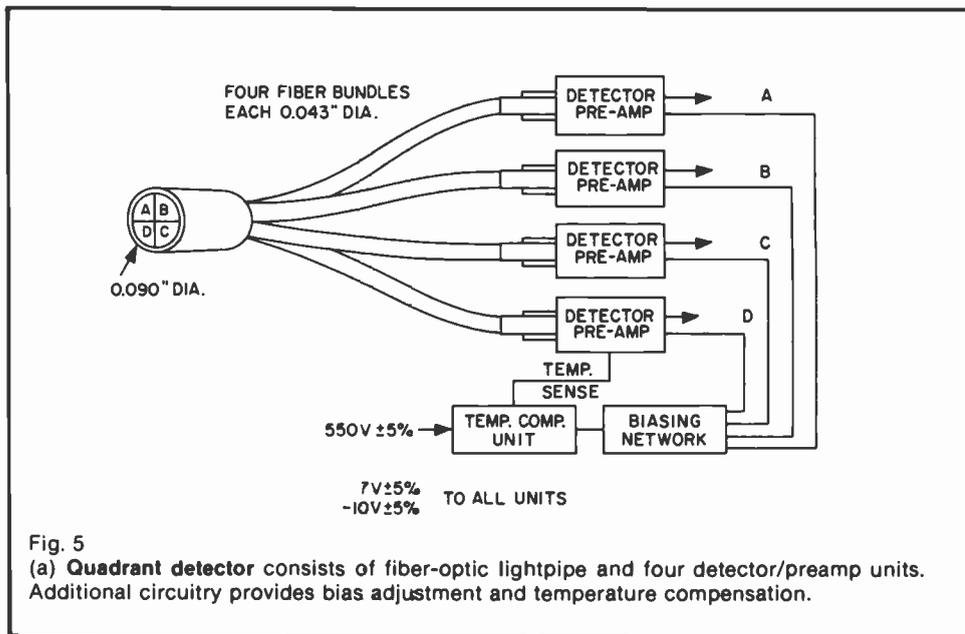


Fig. 5
(a) Quadrant detector consists of fiber-optic lightpipe and four detector/preamp units. Additional circuitry provides bias adjustment and temperature compensation.

are not just for optical communications—one customer used 10 of them to make a 10-element linear array approximately $1 \text{ mm} \times 10 \text{ mm}$ in size. Fig. 5 shows how four units are used with a quadrant light-pipe to give an avalanche quadrant detector with a 0.090-in.-diameter sensitive area.⁶ Obviously, with a sufficient number of detector modules, any desired detector geometry is possible.

Large-area detectors can be built up from smaller modules.

Avalanche diodes up to $5 \text{ mm} \times 5 \text{ mm}$ in area have been fabricated. Fig. 6 shows the design of a detector module in which the sensitive area has increased from $5 \text{ mm} \times 5$

mm to $1 \text{ cm} \times 1 \text{ cm}$ through the use of a tapered light-pipe. This module is designed so that the module dimensions are only a few thousandths of an inch greater than those of the light-pipe. Thus, any desired number of these modules can be mounted in an $n \times m$ array, with a dead space between detectors of less than 0.010 inch.

As soft X-ray detectors, APDs operate at very high count rates.

Avalanche photodiodes are excellent detectors of soft X-rays in the 1 keV to 10-keV range.⁷ Although the achievable resolutions are not particularly good because of the difficulty of making uniform detectors, resolutions of about 10 to 15%

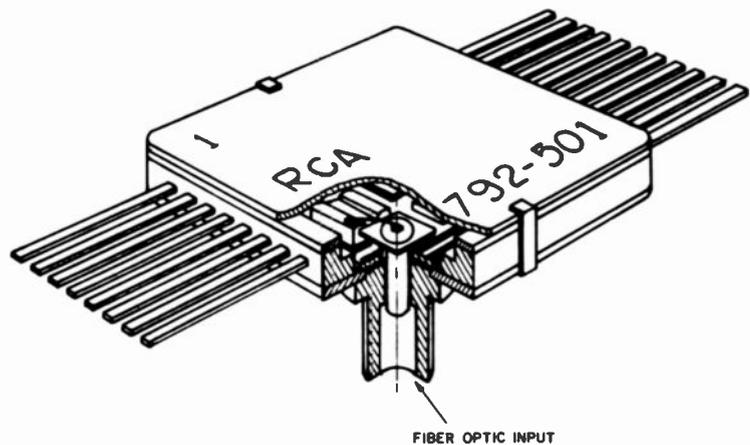
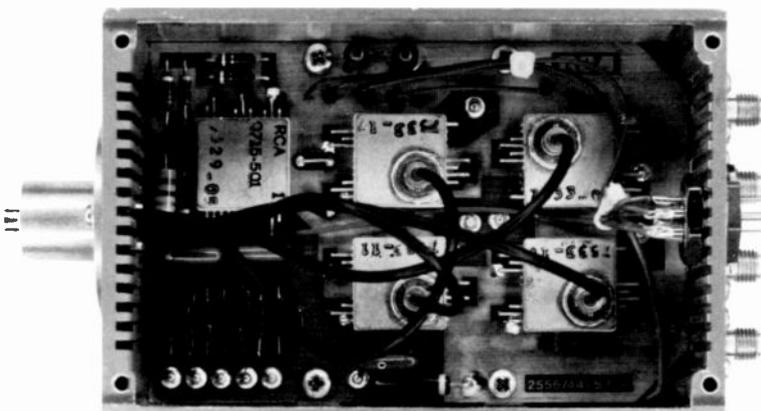
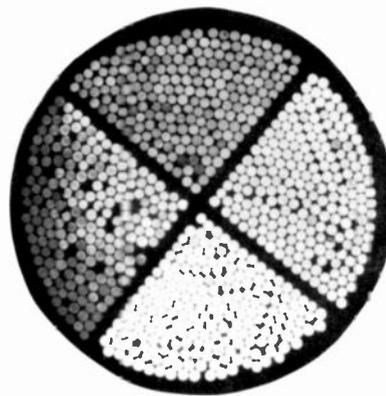


Fig. 4
Developmental detector-preamplifier designed for use with fiber-optic bundles up to 0.045 inch in diameter. The bundle, with its protective coating, is cemented into the collar at the bottom, and interfaces with a clad light-pipe that carries the signal to the APD chip and preamplifier. Package is standard $5/8" \times 5/8"$ flatpack.



(b) Assembled test unit with cover removed.



(c) End view of quadrant fiber bundle.

are possible in the 1- to 10-keV range. These devices have the advantage over non-multiplying solid-state detectors of being usable at much higher count rates (up to a few million/sec), and thus could have applications in high-speed monitoring of X-rays.

Cooling reduces dark current for noise-free low-light-level operation.

For narrow bandwidths, the noise-equivalent power of an APD, at room temperature, is limited by its dark current. This can be reduced by cooling. When sufficiently cooled, the dark current becomes insignificant and the NEP can get down to the 10^{-16} to 10^{-17} range, so the noise

in the signal usually determines the noise limit. A gain of 20 to 25, for which the excess noise factor is about 2.5, is sufficient to reach this limit.

APDs can count single photons at high rates in a manner similar to a Geiger tube.

If cooled sufficiently, say to 77K, the thermal-generation rate of electrons in the device is reduced to a few per second. Under these conditions the APD can be used above the breakdown voltage like a Geiger tube, with single-photon detection efficiencies in excess of 50%. A quench circuit is required to "reset" the device after breakdown. If this is sufficiently fast, an APD can be used for single-photon

counting at rates up to a few million per second.⁸

Conclusions

APDs are gradually being exploited in many electro-optic systems in which they are the most sensitive detectors available. The next few years should see a rapid expansion in both the number and varieties of their applications.

Acknowledgments

The authors would like to acknowledge the contributions of their co-workers—J. Conradi, R. Cardinal, J. Bignet and M. Teare—who assisted in the design, fabrication, and measurement of the devices described here.

References

1. Doyle, T.; and Sprigings, H.C.; "Silicon photosensors and their applications," *RCA Engineer*, Vol. 19, No. 6 (Apr/May 1974) p. 60.
2. McIntyre, R.J.; Sprigings, H.C.; and Webb, P.P.; "Solid state detectors for laser applications," *RCA Engineer*, Vol. 15, No. 5 (Feb/Mar 1970) p. 32.
3. Webb, P.P.; McIntyre, R.J.; and Conradi, J.; "Properties of avalanche photodiodes," *RCA Review*, Vol. 35, No. 2 (Jun 1974) pp. 234-278.
4. Woodward, J. "Hand-held laser rangefinder," to be published in *RCA Engineer* Vol. 22-6 (Apr/May 1977).
5. Conradi, J.; Webb, P.P.; and McIntyre, R.J.; "Silicon reach-through avalanche photodiodes for fiber-optic applications," *Proc First European Conference on Optical Fibre Communications*, London (Sep 16, 1975) p. 128.
6. McIntyre, R.J.; Webb, P.P.; Teare, M.; and Bignet, J.; "A quadrant avalanche detector for laser applications," *Proc. Electro-Optics 1976 Conf.*, New York, (Sep 1976).
7. Webb, P.P.; and McIntyre, R.J.; "Large area reach-through avalanche diodes for X-ray spectroscopy," *IEEE Trans. Nucl. Sc. Vol. NS-23*, No. 1 pp. 138-144.
8. McIntyre, R.J.; "On the avalanche initiation probability of avalanche diodes above the breakdown voltages *IEEE Trans. Elec. Dev.*, Vol. ED-20, No. 7, (Jul 1973) pp. 637-641.

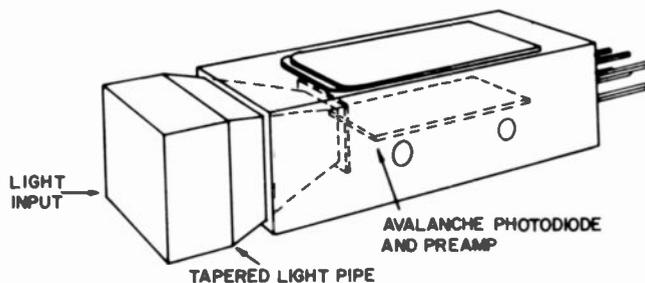
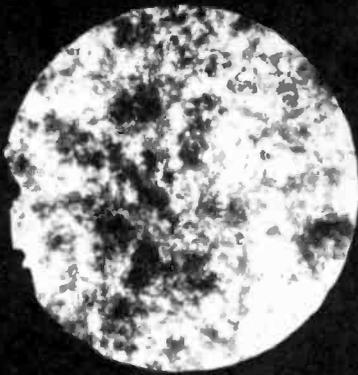


Fig. 6
Large-area (1cm X 1cm) stackable APD module. Tapered light-pipe funnels light down to 5 mm X 5 mm APD and preamplifier, indicated by broken lines.

The atmosphere

Distorted laser beam, once coherent, after passing through 1500 meters of the atmosphere. A still-coherent beam would appear as a pure white circle.

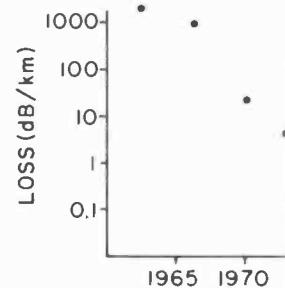
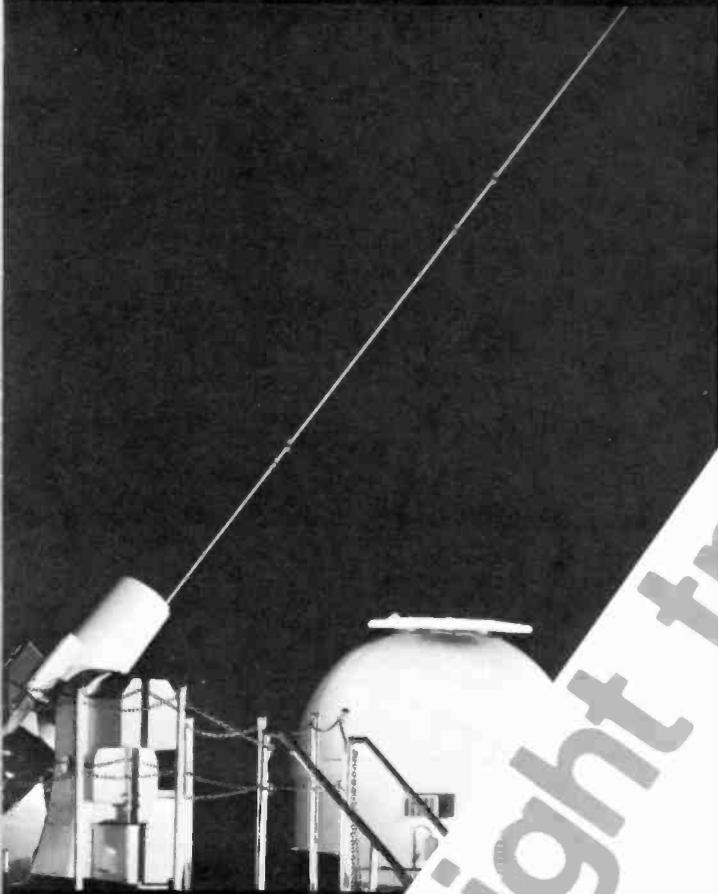


Laser satellite tracker uses both the atmosphere and free space as its transmission media. The atmosphere is a low-loss medium, but can cause signal distortions. This makes it good for tracking and ranging, but poor for lightwave communication.

Correcting the atmosphere's distortion

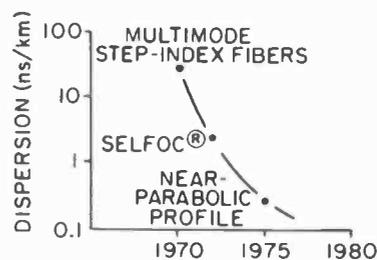
The possibility of correcting the type of lightwave distortions shown at the left with real-time adaptive optics seems to be distinct. Recent developments with coherent laser beams correct the phase error of time-varying distortions by using reflected returns from "glints" at the focus of the phase front to compensate for atmospheric effects and then reshape the wavefront.

The Defense Advanced Research Projects Agency currently has several programs underway in this area, including work on a high-frequency (10-kHz) deformable mirror and large optics with as many as 10^4 to 10^6 control elements. The most far-reaching implications of adaptive optics are not necessarily defense related, however. Other possibilities besides long-range laser communication are beam-focusing in laser-fusion systems and tight-tolerance integrated-circuit manufacturing via aberration compensation.

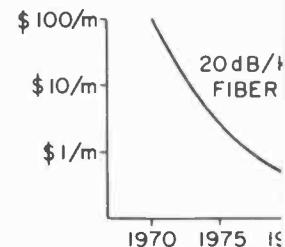


Signal losses have dropped magnitude in about 15 years. For comparison, RG-58 coaxial cable has 230 dB/km at 10 MHz, 60 dB/km at 1 GHz, and 15 dB/km at 1 GHz.

fiber optics progress



Dispersion, or pulse-spreading, has dropped by a factor greater than 100. Caused by light of different wavelengths travelling at different velocities in the fiber, it is not a problem in single-mode fibers.



Selling price of cable has dropped significantly. Corning Glass pilot-plant experience to give \$0.10 per meter when the 100,000 km of communication per year.

(Source: Lockie, Electro-optical Systems Design, Oct 1976)

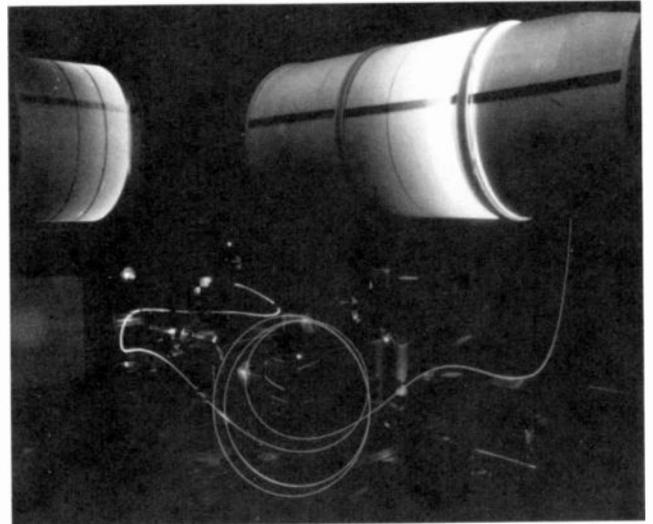
Fiber optics

How it has progressed...
...and what this means to communications

The dollar figures behind fiber-optic communications now look very promising for metropolitan areas, where the current underground distribution systems are very crowded. Extending the present level of service by digging for additional conventional lines is expensive, but fiber optics' great bandwidth can transmit many more communication channels than coaxial cable of the same diameter. The losses in coaxial systems also require repeaters, which in turn require manholes for servicing. Fiber-optic systems would cut down on both of these expensive items considerably.

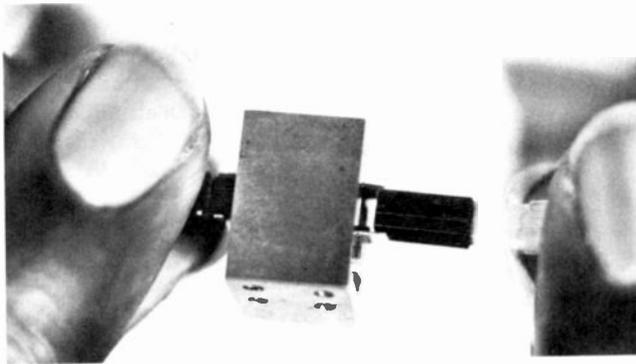
The largest-scale test so far of an urban lightwave communications system has been a Bell Labs/Western Electric setup in Atlanta that proved the technical feasibility of such a system. The experimental system there transmitted and received digitized voice, television, and data signals at 44.7 Mb/s, equivalent to 672 telephone channels per fiber. Their results indicated that "highly reliable transmission may be obtained over distances of about 7 km before regeneration is necessary. Since telephone central offices are often spaced closer than this, lightwave communications systems could connect many of these offices without the need for intermediate repeaters in manholes."

Fiber optics has finally reached the point where it is doing an awful lot more than pose for interesting pictures like the ones below. Businesses now think fiber optics is looking good, and they're talking dollars, not aesthetics.



Courtesy Corning Glass

Transparent fibers will be jacketed in actual use, and the optimum transmission wave-lengths are above the visible region. These facts may prove disappointing to those who felt fiber-optic phone conversations would be much more spectacular than ones moving through copper wires.

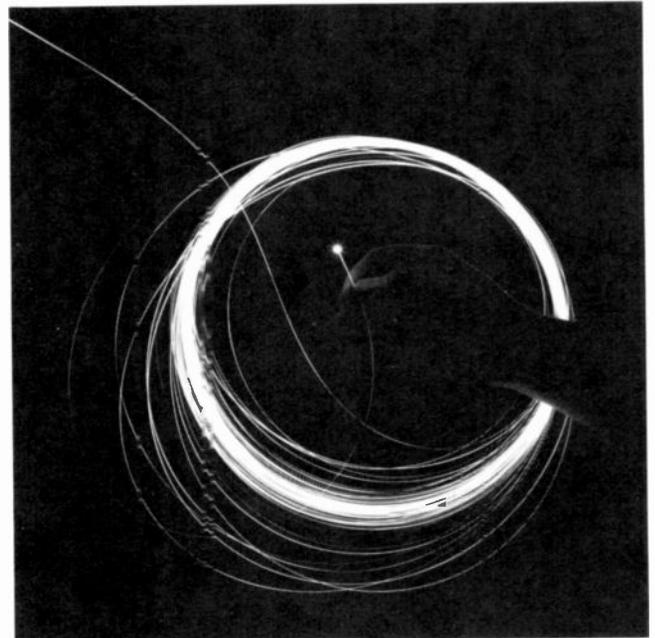


Courtesy Bell Laboratories

Practical lightguide cable connectors will be a necessity for field splicing. This prototype aligns 144 fibers within 0.0001 inch.

What's a fiber optic?

A fairly confusing word, we think. Have you even seen one? Is it adjective or noun, fiber-optics or fiber optic, or what? Optical fibers, lightguides, and lightwave communications are good replacement terms for fiber optics; they all get the point across with less confusion.



Courtesy Bell Laboratories

Over a kilometer long, the fiber on these reels glows with laser light during a test. Several companies are now producing communications-grade (20 dB/km loss) fibers at these lengths.

1975 1980

three orders of
years. For com-
le has a loss of
km at 100 MHz,

km SINGLE
CABLE

1980

has dropped
s Works has used
ve an estimate of
market calls for
ions-grade fiber

an introduction

Optical transmission devices and media

D.G. Herzog

Optical information must travel through a medium between source and detector. Our understanding of media and propagation remained constant for years after Einstein's unifying theory until the invention of precision laser and radar systems forced a number of technological developments to occur.

For any optical transmission and detection to take place, three elements—a light source, a propagation medium, and a detector—are needed, as shown in Fig. 1. An understanding and a definition of the light source is necessary to predict and control any optical transmission through a medium. Conversely, a detection criterion is also necessary because of the interplay between the transmission medium and the detector. Light sources are treated in another section of this issue (see Gorog, p. 18); detectors are also treated separately (see Savoye, p. 36). In this discussion, we are concerned with optical transmission devices and media.

Early concepts

Early man accepted that he "saw," but he considered vision to be perhaps a "mechanical transfer" from the object to the viewer. The concept of light as radiating waves was far too difficult to understand because it deviated from man's sensory experience. The early Greek philosophers considered vision as something projected from the eye that fell on the object, resulting in particles or small "repliculars" returned to the eye for analyzing, with light somehow playing a part. However, the Greeks also explained observations of the heavens by way of superficial analogies with human and animal behavior—with very little experiment. These philosophers

observed the stars with intellectual gratification, but felt that the laboratory smacked of a workshop for gain.

From the theory to experiment

Yet it was from the workshops—with carefully planned experiments and observations—that the understanding and untangling of the puzzling nature of light finally came. Galileo Galilei (1564-1642) learned how to manage the optical-transmission character of light by producing images that could be both magnified and demagnified. This provided the necessary control so that other experimenters could go further. Isaac Newton (1642-1727) initiated experiments into the nature of light and its effect as a wave phenomenon, leading to the concept that light had a sound-like method of propagation. A.J. Fresnel (1788-1827) established the initial mathematical basis from which James Clerk Maxwell (1831-1879) was able to provide his propagation equations.* In

1902, Max Planck demonstrated that light energy was inversely proportional to wavelength.

During this same time frame, other experimenters established that light had a particle nature by measuring the minute amounts of momentum associated with light. Earlier observers had already noted that pressure from the sun's light apparently kept the tails of comets always pointing away from the sun.

Thus, two camps were established—one theorized that light was a particle-transmission phenomenon; the other thought that light was a wave-propagation phenomenon. Finally, in 1905, Albert Einstein completed the basis of our modern understanding of light transmission. His unifying theory held that the energy of light is not distributed evenly over the whole wave front, as the classical picture assumed, but is rather concentrated or localized in small discrete regions. Each wavefront of light, therefore, could be imagined to be studded with photons.

In Einstein's $E = mc^2$ equation, light travels at the speed c (2.99×10^5 km/s) according to Maxwell's equation. Light of energy E (per unit wavelength) thereby appears as a particle of mass E/c^2 .

Thus, Einstein provided the concept of light as we know it today—a small region of the electromagnetic spectrum, produced by the transitions of electrons from states of higher energy to states of lower energy, and travelling in basically a straight line from source to detector.

Until the last 25 years, little was accomplished past this ideal propagation

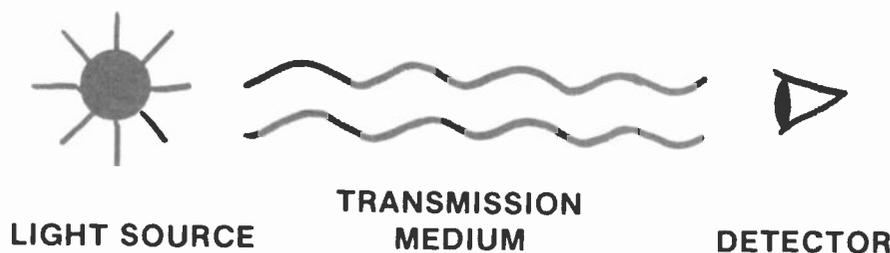


Fig. 1
Necessary ingredients for an optical propagation system.

*Sometimes labor pays off in more than one area. Quoting from a letter by Maxwell, "I made out the equations in the country before I had any suspicions of the nearness between the... velocity of propagation of magnetic effects and that of light so that I think I have reason to believe that the magnetic and luminiferous media are identical." (At that time, luminiferous media was also known as ether, some yet-unknown media through which light waves were propagated like ripples caused by a pebble tossed in water.)

model. Astronomers suffered with the atmospheric distortion and did the best they could. Technologists rolled up their sleeves, though, and made a start during World War II when the need to understand radar forced some practical understanding of electromagnetic-wave propagation through the atmosphere.

With the discovery of the laser in 1961, scientists finally had a powerful, coherent light source—the measurement tool and driving force needed to extend propagation studies. However, the laser met with so many initial limitations that practical applications were slow in coming.

It has happened many times in the past that when the force of producing a product meets a technical limitation, many technological developments occur. So it was with many laser and laser-associated electro-optic systems over the past fifteen years.

Recent developments

Fig. 2 summarizes the areas when understanding, technology, components and products have expanded significantly over this fifteen-year period.

We understand turbulent-media propagation better today.

Our understanding of propagation in free space has remained constant since Einstein established his criteria. However, our understanding of transmission and propagation through a turbulent medium, such as the atmosphere or the sea, has advanced significantly, leading to practical techniques to counter and correct for the disruption and noise generated by the medium. These approaches promise to improve astronomical resolution through the atmosphere by an order of magnitude. DeWolf (p. 64) discusses the problem areas involved in establishing, and working with, a propagation model for random media.

Lens management has advanced significantly.

Optical lensing management has also advanced significantly because of the need to optically manage laser light, as well as the need for night-vision, sensing, and military-reconnaissance applications. The need was coupled through computer technology to develop the significant refinements necessary for the versatile electro-optic applications of today and tomorrow.

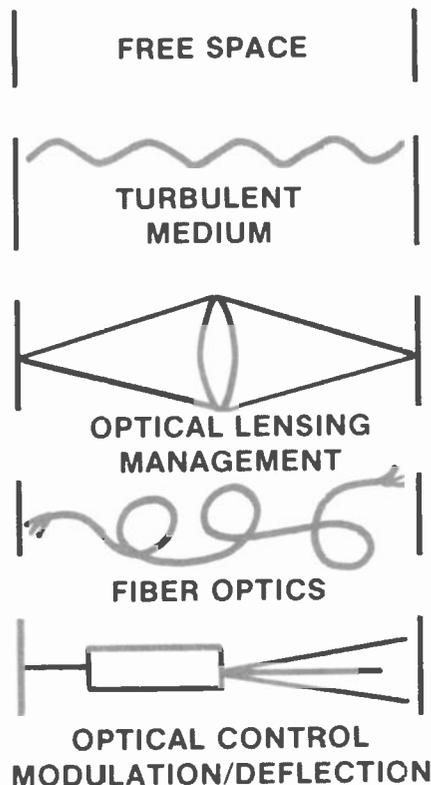


Fig. 2
Types of transmission media.

Fiber optics shows great promise.

Fiber-optic technology shows great promise of providing low-cost, wider-bandwidth, more-reliable communications, and so many corporations and governments are investing heavily in its development. For example, in the U.S. alone, more than \$10 million will be invested into developing fiber optics this year. *Optical Spectra* (Oct 1976) projects that, by the turn of the century, fiber optics will represent more than \$10 billion a year in U.S. business alone.

The fiber optic rod—an optical analog to the electrical wire—guides an optical beam around twists, corners, and over extremely long distances with low loss. Recent advances have improved transmissivity by many orders of magnitude. An optical fiber 0.01 inch in diameter (no thicker than a hair) has a bandwidth of 500 MHz and is therefore capable of carrying more communication information than the six-inch-diameter communication trunk lines that join cities and continents. Optical fibers are discussed in more depth by Wittke (p. 00).

We now have practical optical control of light.

Optical control of light through modulation and deflection has developed, in some

cases, far beyond the capability envisioned even by many far-sighted people a few years ago. Optical modulators that are almost too small to be seen and consume little power are being introduced. Integrated optical devices—the optical equivalent of integrated circuits—have been developed to modulate, steer, switch, trunk, and add and subtract optical beams. Hammer (p. 71) discusses some of the integrated-optics work carried on at RCA Laboratories.

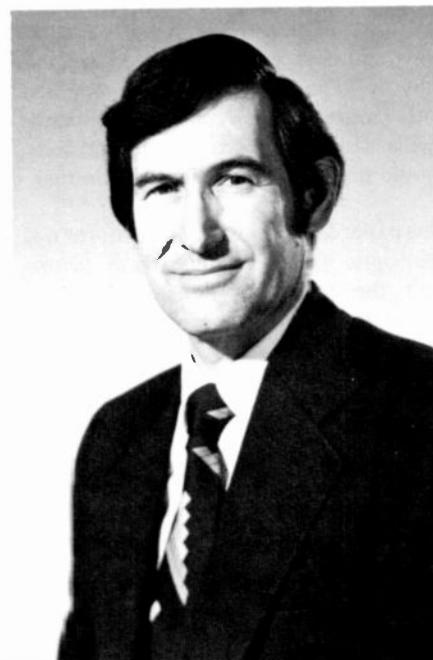
Acousto-optics can provide low-cost beam steering.

Solid-state deflection techniques that combine acoustical and optical interaction show promise of providing the beam-steering management at low cost and small size, both of which have previously limited successful product development in many electro-optics applications.

The acousto-optics phenomenon modifies the optical transmission through a medium via an acoustic interaction with the medium. This phenomenon imparts its frequency and amplitude characteristics into the optical beam, so it may truly be called an "optical transmission modifier."

Don Herzog, as Manager of the ATL's Systems and Applications Group, has experience in a number of electro-optic and laser-associated areas. He has worked on an injection-laser intrusion alarm system, laser communication systems, optical tracking and ranging systems, and now recording systems.

Contact him at:
Systems and Applications Group
Advanced Technology Laboratory
Camden, N.J.
Ext. PC 3171



Optical-fiber communications links

J.P. Wittke

Before optimizing the design of a practical fiber link, it is necessary to understand its signal-to-noise considerations, frequency response, and dispersion effects.

Optical-fiber communications systems carry information in the form of intensity-modulated light along transparent optical fiber waveguides. They are of growing interest world-wide because of their very low loss and potentially very high bandwidth. Fibers are commercially available now with losses under 20 dB/km, and within a year this figure will probably drop to under 10 dB/km, or even lower. (For comparison, RG 59/U coax has a loss of 66 dB/km at 30 MHz.) Dispersion effects in present fibers strongly limit their useful bandwidths, but with continued development of graded-index and single-mode fibers, bandwidths approaching a gigahertz can be expected. These features, combined with small size, low weight, and freedom from most pick-up and interference problems, make fiber-optics look very attractive for future communications systems.

Cost, of course, will be a deciding factor in adopting such systems. A commercial cabled bundle of 6 independent fibers is presently available for about \$4 per ft, or 75 cents per fiber-ft, which is about 5 times as expensive as conventional coaxial cable of equivalent bandwidth, such as the type used in cable-tv installations. This is an introductory price, however, that clearly will drop significantly with volume production (and competition). Fiber optics can thus be expected to be cost-competitive with more conventional transmission media within a few years, and, ultimately, should prove to be significantly cheaper.

This paper discusses noise and distortion in fiber-optic systems. Since these factors limit the performance achievable, understanding them makes performance estimates possible for systems being designed.

The basic function of a communications link is to transmit information from one point to another without introducing unacceptable degradation in the transmitted signal. The specification of what constitutes "unacceptable degradation" depends upon the way in which the information is sent. In an analog link, the information is contained in the detailed shape of the transmitted electromagnetic wave. In a digital system, the signal consists of a series of discrete symbols (waveforms) which must be distinguished from each other on reception. Noise and distortion therefore affect analog and digital systems somewhat differently.

Noise and distortion

To see where the major sources of noise and distortion in an optical link are, consider the system represented in Fig. 1. The (electrical) input signal is first applied to a signal shaper/encoder. For analog transmission, this element pre-distorts (pre-equalizes) the signal to compensate for unavoidable distortions introduced later in the system; for analog signals being transmitted digitally, the shaper/encoder

performs the analog-to-digital conversion and generates the appropriate digital symbols; and for pure digital systems, it may detect the incoming data stream and regenerate and retiming the symbols appropriately for the optical driver. After the electrical signal is thus suitably "shaped," it is applied to a source driver, which modulates the current flowing through the optical source to produce the desired optical-signal output. This assumes a current-modulated optical source, such as a semiconductor incoherent light-emitting diode (LED) or a (properly biased) injection laser.

The light from the source is coupled into the fiber, which transmits it to a (square-law) optical detector at the receiving end of the system. The electrical signal generated in the detector is amplified and fed to a signal shaper/decoder, which converts the raw electrical signal into the proper form for use. In long links, repeaters may be required. These would essentially act as the receiver in the above system, except the output would be applied to another source driver and optical source, whose output light would be coupled into a following fiber section, etc.

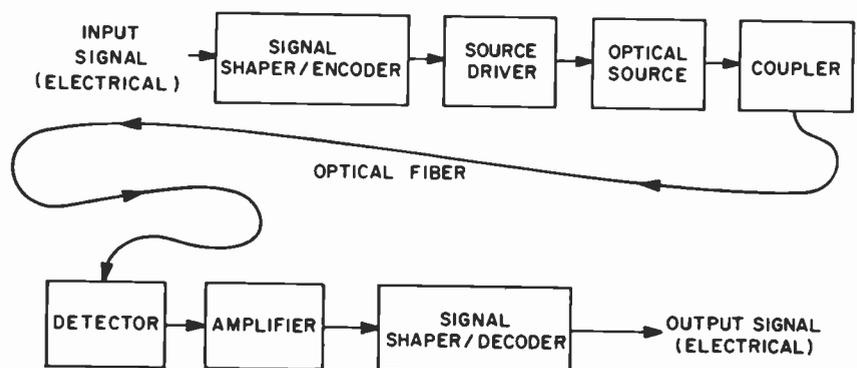


Fig. 1 Elements of a point-to-point fiber-optic communications link. Most of the noise in the system is associated with the optical receiver.

Reprint RE-22-4-10

A lengthier version of this paper appeared in the Proceedings of the 1975 S.P.I.E. Annual Meeting.

Detector noise

We shall assume that the input signal and the electronics in front of the optical source are noiseless. This is generally, (but not invariably) a very good approximation, since the electrical signal at the source is usually at a high level compared to the associated electrical-noise sources. Similarly, noise introduced by the source and spurious light coupled into the fiber are negligible at the detector in a well-designed system.

The basic remaining noise is associated with the optical receiver, which has two main noise sources to be considered. One is associated with the discreteness of the electronic charge and the "quantization" of the light upon photodetection. This is generally called "shot," or "quantum," noise and arises from the essentially statistical nature of the process by which the incoming light generates electrical carriers in the detector. The noise current of this shot noise has a zero average value, but a mean square value of

$$\langle i_Q^2 \rangle = 2eIBG^2F_0 \quad (1)$$

Here e is the elemental electronic charge (1.60×10^{-19} coulombs), I is the average photocurrent (in amperes) generated by the incident light signal, and B is the (information) bandwidth (in hertz) in which the noise current is observed. The last two factors depend on the nature of the photodetector: G is the current gain within the detector, and F_0 is the equivalent "noise factor" associated with the gain mechanism. In a photoconductive detector, such as a silicon PIN diode, there is no gain mechanism, and $G = F_0 = 1$. In an avalanche photodiode, on the other hand, G , adjustable via the bias voltage, can exceed 100, while the noise factor F_0 is generally gain-dependent. For our purposes, the commonly used approximation of $F_0 = G^{1/2}$ for avalanche photodiodes¹ will be assumed valid.

Even if an avalanche detector (with gain) is used, the signal level leaving the photodetector proper will be at a low level, and more amplification is required. Thermal (Johnson) noise fluctuations in the equivalent input resistor of the required amplifier, and amplifier noise, form the second significant noise source. This noise can also be represented by a mean square "thermal" noise current, $\langle i_T^2 \rangle$:

$$\langle i_T^2 \rangle = (4kT/R_{eq})BF_1 \quad (2)$$

Here k is Boltzmann's constant (1.36×10^{-23} J/°K), T is the (absolute) temperature (in °K) of the equivalent detector load resistance R_{eq} (in ohms), and F_1 is the noise factor of the amplifier.

The noise component of "dark," or leakage, currents that occur in the detector in the absence of illumination must also be considered. However, with proper detector choice and operation, "dark" noise currents are smaller than those discussed above, and may be neglected.

With these simplifying approximations, the total mean square noise current $\langle i_N^2 \rangle$ can be written

$$\begin{aligned} \langle i_N^2 \rangle &= \langle i_Q^2 \rangle + \langle i_T^2 \rangle \\ &= 2eIBG^2F_0 + 4kTBF_1/R_{eq} \quad (3) \end{aligned}$$

If an optical power P (transmitted down the fiber) is incident on the photodetector, the generated signal photocurrent I is given by

$$I = (e/hc) \lambda \eta P, \quad (4)$$

where h is Planck's constant (6.624×10^{-34} J-sec), c is the velocity of light (3×10^8 m/s), λ is the wavelength of the light (in meters), and η is the quantum efficiency of the detector (for converting optical photons into mobile electrons). For a wavelength of 830 nm. (where fiber losses are low and efficient (GaAl)As light sources are available), $\eta \approx 0.85$ for silicon and, from Eq. 4,

$$I(\text{amps}) \approx 0.57 P(\text{watts}). \quad (5)$$

For a sinusoidally modulated carrier with a modulation index m , the mean-square signal current $\langle i_S^2 \rangle$ is given by

$$\langle i_S^2 \rangle = \frac{1}{2} I^2 G^2 m^2 \quad (6)$$

(Note that throughout the above development, both noise and signal currents have been referred to the input load resistor at the amplifier following the photodetector proper.) The system signal-to-noise ratio can then be written

$$(S/N) = \frac{\frac{1}{2} I^2 G^2 m^2}{2eIBG^2F_0 + (4kTBF_1/R_{eq})} \quad (7)$$

Since both noise terms are proportional to the system bandwidth B , this can be rewritten in the form

$$(S/N) \times B = \frac{m^2 I^2}{4eIF_0 + (8kTK_1/R_{eq}G^2)} \quad (8)$$

Eq. 8 is shown plotted in Fig. 2, for the (reasonable) parameter values $m = 1$ and $F_1 = 4$ (6 dB), again with the previously assumed detector sensitivity of $(I/P) = 0.57$ A/W. Two gain values are shown: $G = 1$ corresponds to a PIN diode, and $G = 100$ to a "typical" avalanche detector. The curves are drawn for three values of equivalent load resistor: 50, 4K, and 1M ohms. These cover the gamut of interest, from low load resistance, where high (hundreds of megahertz) analog bandwidth are available without the need for equalization, to a very high load resistance, where digital signals can be detected with a good signal-to-noise ratio, but the pulse shapes are badly distorted and require equalization at high data rates.*

Fig. 2 shows that shot noise dominates at higher signal levels, while thermal noise dominates at low levels. Moreover, at low

*The advantages of using a high equivalent load resistor to improve the signal-to-noise ratio are discussed in Ref. 2.

James P. Wittke has studied optical fiber communications systems extensively, making measurements of the transmission properties of various fibers and on the frequency response of light emitting diodes, and developing techniques for coupling the light from semiconductor light sources into fibers. Currently, he is investigating applications of cw injection lasers for optical information handling and working on problems connected with laser processing of materials.

Contact him at:
Systems Research Laboratory
RCA Laboratories
Princeton, N.J.
Ext. 3261



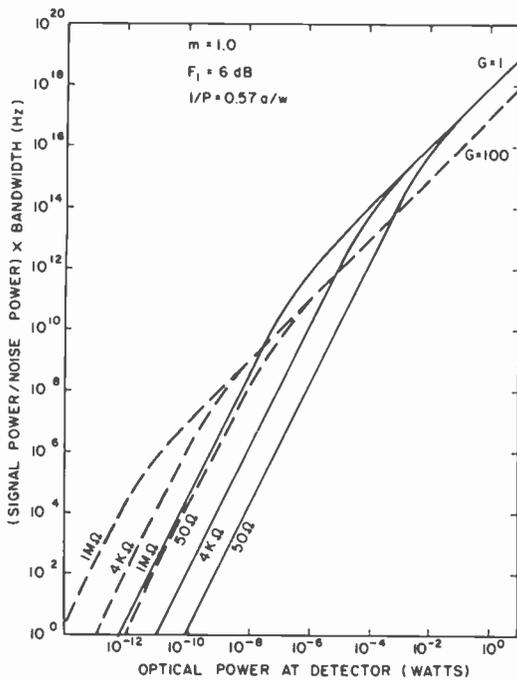


Fig. 2 Typical S/N-bandwidth product available for different detector gains and load resistors. Thermal noise dominates at low levels; shot noise dominates at higher levels.

received-light levels, the gain of an avalanche detector makes it superior to the PIN diode, while at high levels, the noise associated with the avalanche amplification process makes the PIN better.

Source power and fiber losses

A long-lived LED source can have an optical output power of the order of a few milliwatts. How much of this can be coupled into a fiber transmission line depends on LED geometry, fiber acceptance angle (numerical aperture) and effective fiber diameter. Typical values might range from about 100 microwatts into a small-aperture ($NA = 0.1$) fiber up to several milliwatts into a large-aperture ($NA = 0.6$) fiber. An injection laser can readily couple ten milliwatts into an $NA = 0.15$ fiber.

How much of the power coupled into the fiber actually reaches the detector depends primarily upon fiber losses. With present fibers, losses in the 800-900 nm low-loss wavelength region range from 500 dB/km for low-cost, high-aperture glass fiber bundles to under 20 dB/km for high-cost, low-aperture quartz fibers. Research fibers have shown losses of about 2 dB/km.

Required S/N ratio

The required S/N ratio for the communications link is one of its most important specifications. For example, a S/N ratio of 43 dB for analog television signals gives good cable-tv quality, while a 25 dB ratio would give unacceptable performance in most tv applications. On the other hand, when binary digital pulses are being transmitted, the bit error rate (BER, the probability of error per transmitted bit) determines acceptable performance. An S/N of 20 dB corresponds to a BER of better than 10^{-6} and, at $S/N = 22$ dB, the BER drops to 10^{-10} . Roughly, then, good analog transmission corresponds to $S/N = 40$ dB (noise current 1% of signal current), while $S/N = 25$ dB will provide essentially error-free digital transmission.

Source and detector nonlinearities

Consider now a second basic difficulty that must be overcome in a communications link—signal distortion. Distortion can be introduced in any of a number of places in the system of Fig. 1. There are two types of distortion associated with the light sources: amplitude- and frequency-dependent nonlinearities. In LEDs at small drive currents, nonradiative process that saturate as the drive current is increased can introduce distortion unless the diode is biased into, and operated in, a linear portion of its characteristic. At high drive levels, diode heating can occur, with an attendant reduction in radiative efficiency, again leading to distortion.

The frequency response of an LED is determined by the recombination time of the injected carriers and by the diode capacitance. LEDs with response times as short as 1.1 ns have been made, making them useful at frequencies of hundreds of megahertz. (See Fig. 3.) Their high-frequency fall-off in output, however, will distort signals with such high-frequency components.

If an injection-laser light source is used, the associated nonlinearities are somewhat different. Since a laser is a threshold device, its output is (approximately) a linear function of the drive current only above its threshold, and if the current is allowed to drop below that threshold, signal "clipping" can occur. Since stimulated, rather than spontaneous, emission occurs in a laser, the recombination time does not control the response speed in the same simple way as it does in an LED; injection lasers have been modulated at rates of over 1 GHz.⁶ However, an internal response resonance⁷ can lead to signal distortion.

The limited frequency response in detectors also leads to distortion. This frequency limitation is generally associated with the time, or the spread of times, required to collect the photogenerated carriers. PIN detectors are available with response times well under 0.1 ns, and hence present no real limitation to presently considered systems. Avalanche diodes, on the other hand, generally have somewhat slower response times, on the order of 2-3 ns, and hence can distort signals with very-high-frequency components.

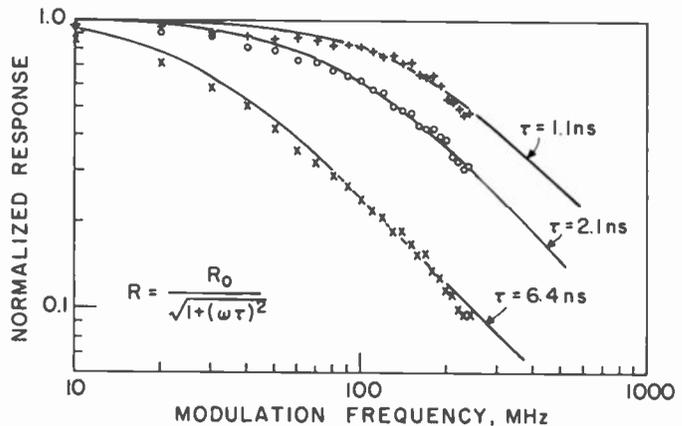


Fig. 3 Frequency response for three developmental LEDs. Their high-frequency fall-off presently limits distortion-free use to hundreds of MHz; injection lasers, however, have been modulated at over 1 GHz.

Fiber-induced distortion

In many systems, the distortion introduced by dispersive effects in the fiber can be the most important. There are two major types of distortion to be considered—material dispersion and multimode distortion.

Material dispersion depends upon both the fiber material and the light source.

When the light from the source is distributed over a band of wavelengths, material dispersion in the fiber will cause a "chromatic" spread of propagation times that will distort the shape of the received signal. Low-loss fibers are often made of (nearly pure) quartz, while some of the higher-loss, larger-aperture fibers are made of borosilicate and other lower-melting glasses. Gloge⁸ has shown that, expressed as a spread of propagation times, the material dispersion is about 4 ns/km for LEDs, with their spectral widths of 30-50 nm, and only about 0.1 ns/km for the spectrally purer injection lasers. This material dispersion is lower for silica fibers than for borosilicate and other higher-index glass fibers by about 40%. In general, with laser sources, material dispersion is negligible; however, with LEDs, it cannot be ignored at high modulation frequencies.

Multimode distortion is caused by changes in propagation velocities with changes in mode.

The ultimate in fiber bandwidth is achieved by making the fiber core so small that only one (low-loss) mode can propagate. This results in core diameters of only a few micrometers, however, and it is extremely difficult to couple significant amounts of light into such a single-mode fiber in a stable, reliable way. For this reason, most present effort is being spent on multimode fibers. The propagation velocities of optical signals are different for the different modes of such fibers, leading to distortion of the optical signal. It is more serious for step-index than graded-index fibers. Both types of fiber can be modeled by the radial index of refraction distribution

$$n(r) = n_0 [1 - 2\Delta (r/a)^\alpha]^{1/2} \quad (9)$$

taking $\alpha \rightarrow \infty$ for the step-index case.

The numerical aperture of a fiber is defined as the sine of the half-angle of its acceptance cone. For a step-index fiber, this can be expressed as

$$NA = (2 n_0 \Delta)^{1/2} \quad (10)$$

The solid angle of the light acceptance cone, and hence, to a first approximation, the amount of light that can be coupled into a fiber, varies as $(NA)^2$, and thus, for a step-index fiber at least, as Δ . If a preliminary link design indicates that the system is power limited (inadequate S/N), one should consider fibers with a larger Δ , (larger NA). However, if dispersion limits link performance, smaller-aperture fibers are advantageous, as we now show. Gloge and Marcattili⁹ have shown that the spread of propagation times for a step-index fiber is

$$T \approx (n_0 L/c) \Delta \quad (11)$$

while for a parabolically-graded-index fiber, it is only

$$T \approx (n_0 L/c)(\Delta^2/2). \quad (12)$$

Since Δ , the (maximum) fractional index change across the fiber, is generally only about 0.01, the advantages of graded-index fibers are obvious.

A note of caution is due here. The above discussion assumes that the distortion varies linearly with transmission-path length. For fibers greater than about 1/2 km in length, it has been observed that dispersive effects scale in a non-linear way with fiber length, due to mode-mixing and other effects.¹⁰

Eq. 11 indicates that one has less dispersion in fibers with lower Δ , i.e., less index difference between core and cladding. However, going to such a fiber has two serious disadvantages: 1) the guiding becomes weaker, and losses at bends and imperfections can become serious; and 2) it greatly reduces the acceptance cone for light that can propagate in the fiber, as mentioned above. In a digital system, where distortion is tolerable until it begins to cause decision errors at the receiver/regenerator, other criteria are generally employed. Intersymbol interference between succeeding symbols is of vital importance, and such concepts as "eye-diagrams"¹¹ become useful. The pulse distortion will depend critically upon the detailed frequency spectrum of the digital symbols transmitted (and on the phase relations between them). This means that the designer must choose that set of symbols which, when convoluted with the *full* (amplitude and phase) modulation transfer

function, will result in the least intersymbol interference.

Conclusions

In an optical fiber communications link, noise and distortion play vital roles. Most noise is associated with the photodetector. If a PIN detector is used, the most serious noise source is the thermal noise in the detector load resistor, indicating that as high a load resistor as feasible should be used. With an avalanche detector, its gain generally makes the shot, or quantum, noise in the detection process dominate. Avalanche detectors are more sensitive, but their high bias-voltage requirements, higher cost and need for temperature stabilization may make them less desirable for links where the received light is strong enough to permit use of PIN detectors.

Using injection lasers makes greatly increased signal strengths possible. Higher modulation capabilities and lower dispersions also result from using lasers. However, their threshold nature makes them require careful temperature stabilization, with added circuit complexity, high power requirements, and higher cost. The choice of fiber will depend upon the required attenuation characteristics, numerical aperture, and dispersion properties. In general, one expects properly fabricated graded-index fibers to be superior to comparable step-index ones for most purposes.

References

- Anderson, L.K. and McMurty, B.J.; "High-speed photodetectors." *Proc. IEEE*, Vol. 54 (1966) p. 1335.
- Goell, J.E.; "An optical repeater with high-impedance input amplifier." *Bell Syst. Tech. J.* Vol. 53, (1974) p. 629.
- Ettenberg, M., Lockwood, H.F., Wittke, J.P. and Kressel, H.; "High radiance, high speed $Al_2Ga_{1-x}As$ heterojunction diodes for optical communications." *Tech. Digest, Int. Elec. Devices Meeting*, (1974) p. 317.
- Schwartz, M.; *Information Transmission, Modulation, and Noise*, McGraw-Hill Book Co., N.Y., 2nd Edition, 1959; Ch. 5.
- Lockwood, H.F., Wittke, J.P., and Ettenberg, M.; "LED for high data rate optical communications." *Opt. Comm.* Vol. 16, (1976) p. 193.
- Schicketanz, D.; "Modulation von galliumarsenid-laserdioden." *Siemens Forsch.-u. Entwickl.-Ber.* Vol. 2 (1973) p. 218.
- McCumber, D.E.; "Intensity fluctuations in the output of cw laser oscillators—1." *Phys. Rev.*, 141, 306 (1966).
- Gloge, D.; "Dispersion in weakly guiding fibers." *Appl. Opt.* Vol. 10 (1971) p. 2442.
- Gloge, D. and Marcattili, E.A.J.; "Multimode theory of graded-core fibers." *Bell Syst. Tech. J.*, Vol. 52 (1973) p. 1563.
- Chinnock, E.L., Cohen, L.G., Holden, W.S., Standley, R.D., and Keck, D.B.; "The length dependence of pulse spreading in CGW-Bell-10 optical fiber." *Proc. IEEE* Vol. 61 (1973) p. 1499.
- Lucky, R.W., Salz, J. and Weldon, E.J., Jr.; *Principles of Data Communication*, McGraw-Hill Book Co., N.Y., 1968, Ch. 4.
- Oliver, B.M., Pierce, J.R. and Shannon, C.E.; "Philosophy of PCM." *Proc. IRE* Vol. 36 (1948) p. 1324.

Optical propagation through turbulent air

D.A. de Wolf

Turbulent air causes optical rays to deviate from otherwise straight paths by irregular slight undulations, and it also changes their phase velocities irregularly. These fluctuations cause a host of unpleasant degradations to communications signals.

If you have any prior notion about "atmospheric turbulence," you probably have in mind some blurry stars and indistinct features of the moon viewed through telescopes. That is not a bad starting point. Before 1945, no one appears to have given much thought about the subject beyond this, and astronomers usually shrugged their shoulders and toted their equipment up to high, poorly accessible mountain peaks where the effects of the atmosphere

David A. de Wolf has been at RCA Laboratories since 1962, working on wave propagation in the atmosphere, radiation from antennas, communications in planetary atmospheres, and diffraction optics, among other areas. Dr. de Wolf is one of the early contributors to the understanding of multiple-scattering effects on high-frequency waves propagating through random media such as turbulent air.

Contact him at:
Materials Research Laboratory
RCA Laboratories
Princeton, N.J.
Ext. 3023



are much reduced.¹ This was a brute-force approach, highly effective at the time, and this article would end here were it not for the development of high-powered radars after World War II, and—around 1960—the advent of the laser, enabling strong light beams to propagate through the atmosphere. It is not always practical to place huge radars on mountaintops, and laser beams are often required elsewhere.

This created an interest in understanding the influence of turbulent air on electromagnetic waves propagating through the atmosphere, and lately this interest has increased because the prospect of doing something about it appears imminent. The capabilities of present-day computers and processors are such that novel techniques are being contemplated: techniques that allow one to gather, process, and make real-time corrections to maneuverable multifaceted lenses to compensate for the distortions caused by atmospheric turbulence. But I am getting ahead of myself,

for I have yet to tell you what these distortions are.

The simplest description is often the best, and always the most attractive to the non-initiated! It seems helpful to use a radar-like picture because it covers the effects on both radar and optical waves. One can consider a point transmitter spewing out a beam in a certain sector of the sky at some angle as in Fig. 1. You will find that the power received at some location at distance R , elevation angle θ , and azimuth ϕ from this transmitter is the absolute square of a complex electromagnetic field that consists of three obvious factors in free space:

- 1) a $1/R$ dependence on the distance;
- 2) an angular gain factor telling you how much this field at θ, ϕ is less than it would be on-axis at the same distance;

Reprint RE-22-4-12
Final manuscript received May 27, 1976.

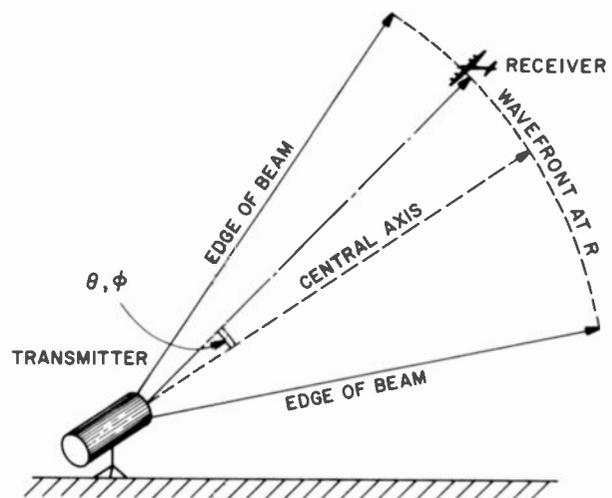


Fig. 1
In free-space propagation, power received is a function of radius, angle and phase.

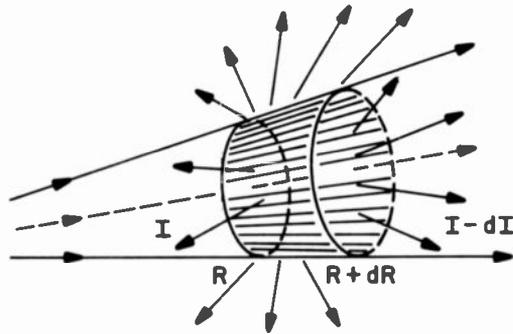


Fig. 2
Intensity losses are caused by absorption and turbidity.

3) a phase factor essentially equal to the number of wavelengths in distance R times 2π .

Laser-beam propagation in free space can be more complicated in the presence of focusing optics, but for this survey it is sufficient to group lasers under point sources and to ignore the extra complexity of focusing. The above power is really an energy flux and is measured² in W/m^2 or some other outmoded non-MKS system of units. At large distances, objects filling only a tiny arc of angle "see" a practically-plane constant-phase surface; for such objects one may consider the source to be infinitely far away. On the other hand, the gain factor does vary with θ and ϕ at constant distance R but—again—for distant objects the change in this factor over the small arc will be quite negligible. The amplitude will appear, merely, to change as $1/R$ with increasing distance—as it should, for energy must be conserved from one sphere to the next. By "amplitude," we really mean the factors 1 and 2 that multiply the phase factor 3 above.

How the atmosphere differs from free space

In free space no more factors are needed. In a medium such as air, water, or glass, through which light can propagate, a fourth factor is needed to help express the changes caused by the medium. This fourth factor can affect both amplitude and phase, and consequently it is complex. We name it $u(R, \theta, \phi)$ and note that it is unity, of course, in free space. This survey focuses entirely upon this new factor $u(R, \theta, \phi)$.

Atmospheric gases absorb radiation.

Now for atmospheric effects: first of all, certain atmospheric gases—particularly

water vapor and carbon dioxide—are able to absorb electromagnetic energy at optical frequencies from the beam and convert this energy to internal processes ultimately resulting in heat, i.e., convert light into radiation at entirely different portions of the electromagnetic spectrum. Not only are detectors at the receiver insensitive to such radiation; more significantly, such radiation is reradiated more or less isotropically from each air molecule and thus essentially lost. Because such absorption losses are usually proportional to the amount of optical radiation incident upon a spherical slab and to the thickness of the slab, these losses give rise to an exponential decrease in intensity $I(R)$ from that at the transmitter (in any direction over and above the $1/R^2$ loss due to distance). If we write $I(R) = I(0) \exp(-2\alpha R)$, then α represents the loss per unit distance. For simplicity, I have assumed we're looking along the axis, and the medium is uniform. (See Fig. 2). This simple formula is easily extended for other directions and non-uniform media (but editorial restraint prevents me from doing it here!) The major point here is that all that $u(R, \theta, \phi)$ consists of is a simple attenuating factor, $\exp(-\alpha R)$, for $\theta = 0$, $\phi = 0$ in this case of uniform absorption.

Turbidity can scatter radiation.

Atmospheric turbidity—to be carefully distinguished from the concept of turbulence—is so similar in concept to absorption that confusion is easy. Turbidity implies the presence of air particles in the beam that do not absorb, but scatter, incident optical radiation in all directions so that—again—scattered radiation for all practical purposes does not reach the receiver.³ Like absorption, its effect is expressed by a loss per unit length. The

Definitions

C_n^2	refractive-index structure constant
C_T^2	temperature-index structure constant
e	vapor pressure
f	frequency
h	altitude
I	intensity of radiation
k	wave number
L	pathlength
L_m	largest possible scalelength
l_0	smallest possible scalelength
l_F	Fresnel radius
n	refractive index
δn	deviation of n from free-space value
p	pressure
r	spatial coordinate
R	distance from source
T	absolute temperature
U	float velocity
u	field factor used to characterize an electromagnetic field <i>not</i> travelling in free space
α	absorption-loss coefficient
θ	angular coordinate (elevation)
θ_d	angle of ray diffraction
θ_r	angle of ray deflection
λ	wavelength
ρ	radius vector
ρ_c	coherence length
α_T^2	scattering constant
$\Phi(k)$	turbulence spectrum
ϕ	angular coordinate (azimuth)
ψ	phase angle
<i>auto-correlation</i>	correlating an ordered series of observations with the same series in altered order

study of these loss quantities is obviously of paramount importance to those interested in solar energy.

The atmosphere's refractive index is not constant.

The next item on the list is refraction. In perfectly uniform air (i.e., in the absence of density gradients, temperature, pressure differences, etc.), the wavelength is not just λ . According to a famous optical theorem, the original incident free-space radiation is absorbed by the medium and reradiated "instantaneously" in the same direction at a new wavelength $\lambda/[n(r)]$ where $n(r)$, the refractive index, varies very slowly over many wavelengths. (Otherwise, the concept

of "medium" makes little sense.) Of course $n(r)$ is very close to unity for such sparse media as atmospheric gases.

Herein lies a new problem: the atmosphere is not uniform, but its density [and therefore the deviation of $n(r)$ from unity] varies with altitude because of gravitational forces. The optical pathlength [proportional to the refractive index $n(r)$] decreases somewhat with altitude, and the locus of all locations with the same optical pathlength as the distance R along the central axis of radiation is no longer a sphere. The situation of Fig. 3 holds. Because rays are defined as light trajectories perpendicular to the those loci, the wavefronts, it follows that rays are now deflected towards regions of higher $n(r)$, i.e., greater atmospheric density, i.e., downwards. Therefore, the apparent location of stars close to the horizon may be several microradians higher than their real directional location. Fortunately, this effect is not dependent on frequency (to a good approximation), and the density gradient is slow enough that distortion of observed objects is negligible in clear-atmosphere conditions.

This finally leads us to the following: suppose now that we allow for local fluctuations in the refractive index $n(r)$ around the mean profile, and suppose also that these fluctuations occur on some scale much less than the e-folding length of 8 km (or so) of the altitude profile of air density. If, on the other hand, these fluctuations are not microscopic (e.g., at the molecular level), then a situation as sketched in Fig. 4 may occur. Rays are deflected irregularly from their otherwise straight paths and their optical pathlengths also deviate from what they would have been in free space.

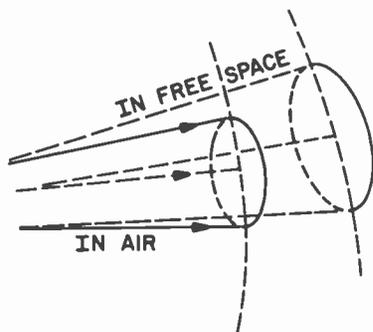


Fig. 3 Rays are deflected toward denser air because the atmosphere's refractive index changes with altitude.

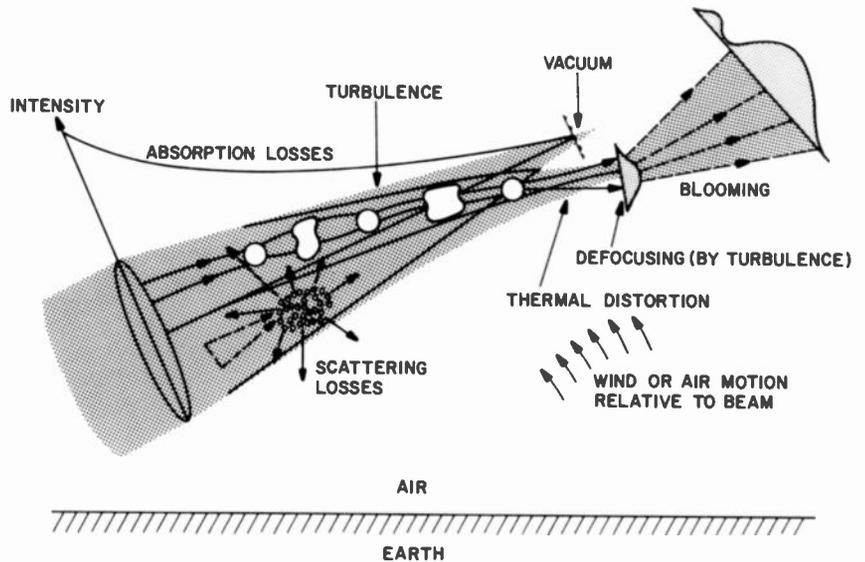


Fig. 5 A laser beam traveling through air faces a number of distorting effects.

This situation is a typical effect of *clear-air turbulence*, which produces just this type of refractive-index fluctuation.

Changes to the signal

All these effects, and some others, are summarized for laser beams in Fig. 5. Let us see what these irregular deflections and changes in optical pathlength do to our signal parameter $u(r)$:

Phase changes are caused by refractive-index variations.

Consider a narrow ray bundle passing through a bunch of fluctuations in $n(r)$. By "narrow," I mean that each ray in the bundle goes through almost the same string of fluctuations, e.g., as in Fig. 6. The optical pathlength, or *phase*, is 2π times the

number of wavelengths in the actual pathlength R . Because $\lambda \propto n^{-1}$, it follows that phase is proportional to refractive index. Indeed, phase in the turbulent medium differs from that in free space, namely $2\pi R/\lambda$, because R is replaced by nR if n is a constant, and by a sum of $n_i \Delta R_i$ for each path segment of length ΔR_i with its own refractive index $n_i \approx 1 + \delta n_i$. From this sum you can see that the phase increment over and beyond the free-space phase involves a sum of $\delta n_i \Delta R_i$. This sum behaves like a Gaussian random variable with zero mean (δn_i can be positive or negative), and therefore the total phase fluctuates about the quiescent free-space mean.

Because phase is a means of expressing the concept of a wavefront, the coherence of a wavefront is strongly related to coherence

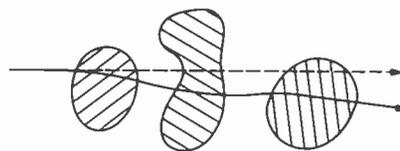


Fig. 4 Clear-air turbulence deflects rays from their normal straight path.

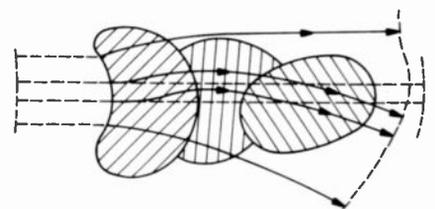


Fig. 6 Phase fluctuations are caused by variations in phase velocity when light travels through irregular refractive-index changes.

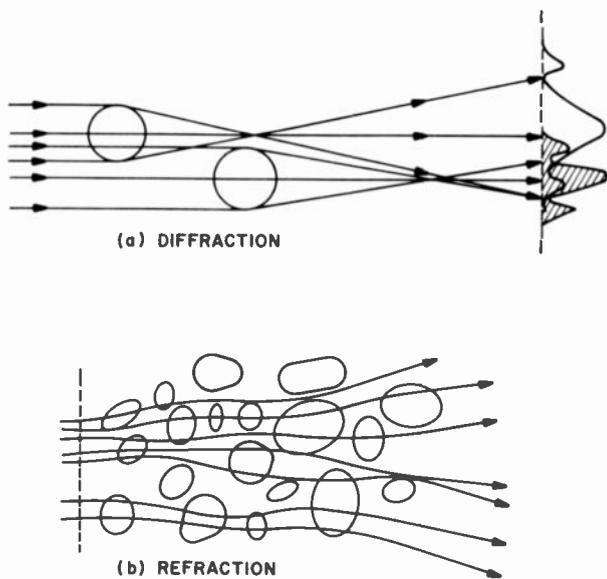


Fig. 7
Small-angle scattering redistributes energy unequally and so produces amplitude changes.

of phase. ("Coherence" is just a measure of how closely one point of a wavefront follows motions of a nearby point.) And, because phase is linearly related to refractive index, it follows that phase coherence is determined by refractive-index coherence, i.e., by the spatial or temporal correlation of the refractive indices at two locations. Indeed, this autocorrelation of $n(r)$ is a basic property we must know in order to predict wave behavior.

Amplitude changes can be caused by small-angle scattering.

Although none of the energy-loss mechanisms depicted in Fig. 2 hold here, it is possible that energy is redistributed unequally at distance R by two small-angle scattering mechanisms. One is by diffraction (edge scattering) from more-or-less coherent fluctuations; the other is cumulative refractive deflection in a random-walk fashion from a string of such regions. These are illustrated in Fig. 7; discussion will resume later.

Imaging and communications parameters

In a typical imaging application, light is gathered by a lens and focused to a small spot at a distance f beyond the lens. The lens "sees" the electromagnetic field incident upon it, and in particular it senses the extra factor $u(r)$ over and above the normal free-space field (see discussion

above). The mathematical property of focusing entails basically that the electromagnetic field in the focal plane at f beyond the spot is a Fourier transform of that part of the factor $u(r)$ corresponding to points on the lens [$u(r)$ is set zero elsewhere]. The focal-plane field is therefore linearly related to the factor $u(r)$, which contains the turbulent phase correction (and also amplitude corrections).

Workers in optics often prefer to examine the *mutual transfer function* (MTF) of the lens: this quantity is the spatial Fourier transform of the intensity in the focal plane. To say it another way: analyze the intensity in the focal plane into sine-wave components; the coefficient of each component forms a function of the wavelength of that component that is just the above-mentioned MTF. As the MTF is a Fourier transform of the square of a quantity that is linear in $u(r)$, it follows mathematically that *the MTF is linearly dependent upon the autocorrelation of $u(r)$* . For long exposure times you can show that the lens MTF is simply a product of this autocorrelation with the free-space lens MTF. The only point I wish to stress in this context is that the deterioration of image quality (blurring) is determined by parameters such as the *wave-field spatial autocorrelation*.

Another parameter is illustrated in Fig. 8: a very narrow beam twists and bends as it passes through a turbulent medium. By "very narrow," I mean that all three rays

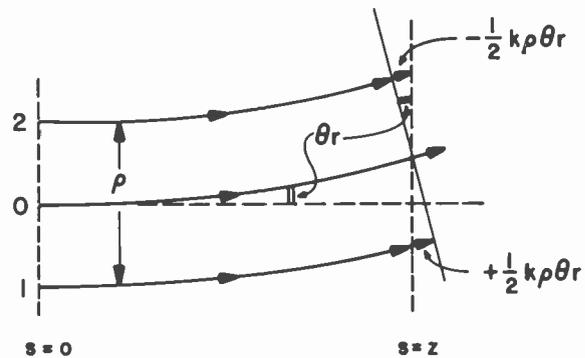


Fig. 8
Turbulent medium deflects a very narrow light beam.

sketched in Fig. 8 "see" essentially the same refractive index at each path point s . The phase front is approximately plane and perpendicular to the middle ray everywhere. By remembering that the phase of rays 1 and 2 at s can be related to that of ray 0 at $s = z$ by simple geometry, one obtains

$$\psi_2(z) \approx \psi_0(z) + \frac{1}{2} k \rho \theta_r$$

$$\psi_1(z) \approx \psi_0(z) - \frac{1}{2} k \rho \theta_r \quad (1)$$

It follows that the ray-angle of deflection is

$$\theta_r \approx [\psi_2(z) - \psi_1(z)] / k \rho \quad (2)$$

The above should also make clear that θ_r , the cumulative angle change, is a fluctuating quantity. More importantly, this angular deflection is given by the phase difference $\psi_2(z) - \psi_1(z)$. Thus, certain statistics, such as the variance of this difference, known as the *phase-structure function*, are important for determining directional variations. Note that the mean is zero, so it is not very interesting.

In another application, narrow laser beams are directed at objects to be illuminated. The important intensity factor at the illuminated object is $I(r)$, where r is at $z = L$, and $r = (\rho, L)$. The laser forms a spot in the $z = L$ plane, and the intensity $I(\rho, L)$ drops off sharply as ρ varies from the spot center. Problem: how do we describe the spot center? One method⁵ is operational—just

weight the intensity with the radius vector ρ to obtain a formal spot radius ρ_0 that fluctuates in turbulent air because the intensity fluctuates both as a function of ρ and of time t . Thus, statistics of ρ_0 must be calculated, and in particular $\langle \rho_0 \rangle$ and $\langle \rho_0^2 \rangle$ are of interest. These two require knowledge of the intensity and its spatial autocorrelation, two other important parameters.

Ideally, we would wish to know all the statistics of these quantities, and that would entail knowledge not only of intensity and phase autocorrelations (which also imply mean square, of course), but also of crosscorrelations and a veritable Pandora's box of higher statistical moments. I wished only to point out here that a good deal of basic information is contained in these few lowest-order statistics of phase and intensity.

Amplitude and phase fluctuations

We have now reduced the problem of dealing with imaging and communication in a random medium to one of understanding the fluctuations of phase ψ and amplitude A (or intensity $I \propto A^2$) of the standardized electromagnetic-field factor $u(r)$. In particular, I have argued, with some intentional oversimplification, that much of what we need to know is given by phase and amplitude means and variances, and several spatial autocorrelations, in planes normal to the line of sight. Having hammered home this point, let me proceed to discuss the basics of phase and amplitude in random media:

Phase fluctuations can be very large over long paths.

I equated "phase" with "optical pathlength," and indicated that the latter differs from a free-space optical pathlength by weighting each path segment by the refractive index. The phase increment (difference of optical pathlengths in random medium and in free space) has path segments weighted by δn , the deviation of n from its free-space value of unity. It can be argued that this increment is a normal random variable with zero mean. It can also be shown that its variance is given by

$$\begin{aligned} \text{Variance of } \psi &= k^2 L L_0 \langle (\delta n)^2 \rangle \\ &= k^2 L L_0^5 C_n^2 \end{aligned} \quad (3)$$

where k is the wavenumber, L the pathlength, L_0 a representative scalelength of the fluctuations (theory shows L_0 is the

largest possible scalelength), $\langle (\delta n)^2 \rangle$ is a measure of the strength of the fluctuating mechanism, and C_n^2 will be discussed shortly. Observations indicate that $\delta n \approx \leq 10^{-6}$ and $L_0 \approx 1$ m, so that it follows that the variance of ψ can be very large at optical frequencies for kilometer-long paths or so.

Angular deflection depends upon the "refractive-index structure constant."

Given the above knowledge of ψ , and the fact that the angular deflection θ_r is given by Eq. 2, it can be inferred that θ_r is also a normal random variable with zero mean. Only, in this case, the pathlength is weighted with $\delta n(1) - \delta n(2)$, the difference of the refractive index at points $\rho/2$ above and below the path segment. In order to calculate further, a model for the statistical behavior of this difference is needed. Turbulence theory gives the fundamental "two-thirds" law:⁶ when ρ lies between the microscale l_0 (smallest possible scalelength) and macroscale L_0 , then

$$\text{variance of } [\delta n(1) - \delta n(2)] = C_n^2 \rho^{2/3} \quad (4)$$

where C_n^2 is known as the *refractive-index structure constant*.⁷ It can then be shown, by delving into more detail (we shall not do that here), that

$$\text{variance of } \theta_r = C_n^2 L l_0^{-1/3} \quad (5)$$

Here, it is interesting to point out that, as ρ cannot exceed L_0 in Eq. 4, it follows that the variance of the left side of Eq. 4 tends to $\langle \delta n^2 \rangle$, and the right to $C_n^2 L_0^{2/3}$. Substitution for $\langle \delta n^2 \rangle$ in Eq. 3 leads to the alternative form for the variance of ψ . In the atmosphere, $l_0 \approx 1$ mm and $C_n^2 \approx \leq 10^{12} \text{m}^{-2/3}$ (queer unit, but that's what it is!); hence the rms angular deflection for $L \approx 1$ km can be as much as 0.1 mrad, according to Eq. 5. This may appear to be small, but it means that the rms transverse shift of a ray, $L\theta_r \approx 10$ cm, which is not necessarily negligible. There are other consequences to this Gaussian random ray dancing with rms shift ≈ 10 cm that will come into discussion; however, it is preferable to discuss intensity effects.

Eddies produce intensity scintillations.

Intensity scintillations, or amplitude fluctuations, are more difficult to clarify at an elementary level. In the scattering-of-light regime under which turbulent deflection falls, the wavelength λ is typically much shorter than any of the refractive-index fluctuation scalelengths l . In particular, $\lambda \ll l_0$, which of course implies $\lambda \ll l$. In Fig.

7a, the effect of one coherent fluctuation of diameter l —an *eddy* of size l , colloquially—is shown. It creates a *diffraction* pattern at $z = L$ because the "edges" bend light rays by an angle $\theta_d(l) \approx \lambda/l$ towards higher δn .

There aren't edges in reality, but a $1/e$ drop in strength does define the diameter l . The situation is a little different from classical slit diffraction. The edge ray, diffracted by angle $\theta_d(l) \sim \lambda/l$, meets an undeflected ray at $z = L$ that left $z = 0$ at a transverse separation $L\theta_d(l) \approx \lambda L/l$. If this separation is much less than the Fresnel radius $I_F = (\lambda L)^{1/2}$, it follows that there is no interference effect. But $L\theta_d(l) \ll I_F - \lambda L/l \ll (\lambda L)^{1/2} - l \gg I_F$. Here is the first complication. Only those eddies with $l \leq I_F$ can possibly give rise to interference, and hence diffraction effects, leading to scintillation. On the other hand, when $l \ll I_F$, the diffraction pattern will be as sketched for the top eddy in Fig. 7a: the diffracted energy is spread over many Fresnel zones, but is diffused to such a weak strength that it is of little importance.

Thus, it appears that out of the entire range of scale sizes l with $l_0 \leq l < L_0$ (with $l_0 \sim 1$ mm and $L_0 \sim 1$ m in the lower atmosphere), *only those eddies with $l \sim I_F$ contribute important amplitude fluctuations*. It is easy to convince oneself with these cited scalelength limits that $l_0 < I_F < L_0$, for km-length optical links; hence diffraction effects, superimposed on what otherwise appears to be a geometrical-optics medium, must be accounted for. However, as only a small portion of the spectrum of eddy sizes is affected, it follows that these diffraction effects remain very weak.

However, as C_n^2 increases, cumulative scattering and interference effects increase, and intensity scintillations increase. This can be clarified by returning to Eq. 5, which is a special case of

$$\text{variance of } \theta_r(l) \approx C_n^2 L l^{1/3} \quad (6)$$

for cumulative refraction by eddies with sizes l and larger. Remembering that $l \sim I_F$ [$\sim (\lambda L)^{1/2}$] for important diffraction effects by *one* eddy, we set $l \sim I_F$ in Eq. 6 and compare $L\theta_d(I_F)$ to $L\theta_d(l)$ to see if cumulative effects outweigh single-eddy diffraction. That is, squaring the above quantities, and rearranging the factors on each side somewhat, we compare

$$I_F^2 C_n^2 k^{-6} L^{11/6} \text{ to } I_F^2, \text{ or } C_n^2 k^7 L^{11/6} \text{ to } 1.$$

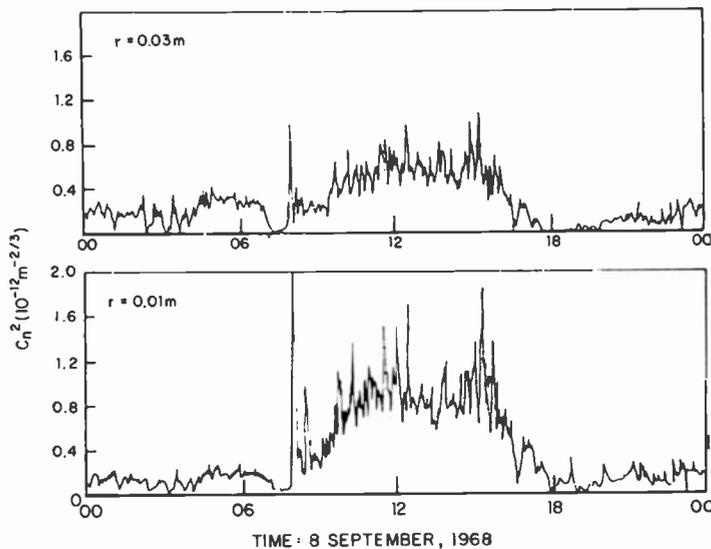


Fig. 9
Refractive-index structure constant is far from being a constant. "Turbulent-weather" day produces order-of-magnitude changes.

Thus, we note that the curious dimensionless combination $C_n^2 k^7 l^6 L^{11/6}$ —often referred to as σ_T^2 or something similar—determines whether or not the scattering is strong. When $\sigma_T^2 \gg 1$, it is strong and effects will be very noticeable. The physical meaning, I think, of these parameters, which you will encounter all through the literature, is thus clarified.

In summary, we note that several size and strength parameters govern phase and amplitude fluctuation. Three lengths are crucial:

- l_m : the microscale of turbulence,
- l_f : the (intermediate) Fresnel scale determining amplitude effects,
- L_m : the macroscale,

and three strength parameters,

- C_n^2 : the strength of refractive-index variations,
- σ_T^2 : the strength of amplitude scintillation,
- $C_n^2 k^2 L L_m^{-3}$: the strength of phase fluctuations.

Space precludes more exhaustive treatment but there are other important parameters. For example, there is a coherence length ρ_c such that when $|\rho_1 - \rho_2|$ exceeds ρ_c , the autocorrelation of the electromagnetic field will be very small. This wave-field coherence length is given by

$$\rho_c \sim (C_n^2 k^2 L)^{-1/4} \sim l_f \sigma_T^{-6/5} \quad (7)$$

It can be shown that l_f is the coherence length for weak amplitude (intensity scintillation), whereas when $\sigma_T^2 \gg 1$ (strong scintillation) the coherence length reduces to ρ_c , which is then less than l_f . For weak scintillation, the electromagnetic field is almost unchanged in amplitude; the phase $\delta\psi$ fluctuates over very many times 2π , and thus the field behaves like a log-normally distributed random variable. Refraction is weak, hence l_f is the important correlation length. For strong scintillation, ρ_c decreases way below l_f , and suddenly all kinds of rays that formerly were in phase with undetected ones interfere with these neighboring ones destructively. The field distribution then changes towards the expected Rayleigh distribution. The intermediate case, where $\sigma_T^2 \sim 1$, gives all the headaches to workers in this area; the statistics are then almost unpredictable.

What causes the fluctuations

We have seen that communications-system parameters are determined by the behavior of an appropriately normalized electromagnetic field, aside from normal free-space geometry and other parameters. Amplitude and phase of this u -function are determined in turbulent air by refractive-index fluctuations. It is time to explain what, in turn, causes these. For optical frequencies, a good approximation for n in terms of pressure p (millibars), temperature T (degrees Kelvin), and vapor pressure e (millibars), is

$$n = 1 + A(p/T) [1 + B(e/pT)] \quad (8)$$

where

$$A \approx 77.5 \times 10^6 [1 + 5.15 \times 10^{-1} \lambda^{-2} + 1.07 \times 10^{-4} \lambda^{-4}]$$

with wavelength λ in μm . B is a numerical constant unimportant here. Often—but not always!—humidity fluctuations or contributions can be ignored, and of course the higher-order λ contributions are unimportant at visible or infrared frequencies. Because pressure fluctuations are very short-lived it follows that

$$\delta n \approx -Ap\delta T/T^2 = (1-n)\delta T/T \quad (9)$$

and thus δn is proportional to δT at any altitude (remember that p is a strong function of altitude). In practice, one measures not refractive-index, but temperature fluctuations, and in particular for turbulence one finds the mean square of δT , or more precisely

$$\langle [\Pi(r_1) - \Pi(r_2)]^2 \rangle = C_T^2 \rho^2 \quad (10)$$

where $|r_1 - r_2| \equiv \rho$ in the so-called inertial subrange of turbulence ($l_0 < \rho < L_0$). Of course, this is just the fundamental "two-thirds" law mentioned in connection with Eq. 4. You see from Eq. 9 that C_n^2 is simply proportional to C_T^2 by the factor Ap/T^2 , and in practice that is how values of C_n^2 are obtained. However, a caveat is highly in order: there are often humidity fluctuations to be accounted for, and then the relationship between C_n^2 and C_T^2 is different.

Now back to C_n^2 , the refractive-index structure constant that is not a constant. The behavior of C_n^2 (altitude, time) is highly erratic. Fig. 9, taken from a representative publication,⁸ indicates that C_n^2 varies over several orders of magnitude during a typical "turbulent-weather" day. Fig. 10, taken from a now somewhat outdated work, shows the behavior of C_n with altitude (unfortunately the unit of the C_n axis is different). More recent work by its author⁹ indicates that C_n^2 itself is a very random variable. However, at ground level (and up to 150 m or so) a semi-empirical model has been worked out,¹⁰ predicting $C_n^2 \propto h^{-2.3}$ to $C_n^2 \propto h^{-3.3}$; the former extreme for quiescent (night) conditions and the latter for highly unstable (midday) conditions. These trends compare well with the initial portions of dawn/dusk and sunny-day curves, respectively, of Fig. 10.

The estimates of macro- and microscales of turbulence (L_o and l_o) are more complicated. Briefly, L_o is proportional to altitude h , but as the strength factor C_n^2 dies out rapidly with increasing h , $L_o \sim 1$ m appears to be a reasonable estimate for horizontal propagation just above the surface. But, for slant-path propagation, a variable L_o must be used. Fortunately, most system parameters can be expressed so that L_o does not enter into them: L_o is in Eq. 3, but not in most expressions in which the phase fluctuations usually enter into results! On the other hand, l_o is determined by a heat cascade process: the largest eddies break up into smaller ones, conserving their eddy-motion energy, until at some smallest size l_o they prefer to give up their energy as heat (and the eddy dies out). Thus l_o is a function of the Reynolds number. Fortunately, l_o enters into only some calculations and then usually as some fractional power such as $l_o^{-1/3}$. The reader is referred to the literature for judicious guesses.

Something has been said about the strength of δn^2 —see Eqs. 9 and 10—but the fundamental parameter is its *spectrum*. That is to say, for homogeneous turbulence one can write $\langle \delta n^2 \rangle$ as an integral of the spectrum $\Phi_n(k)$ for all possible wavenumbers k ($k = 2\pi/\lambda$). The fundamental theorem of turbulence similarity theory states (and it is usually quoted as the Kolmogorov or Obukhov-Kolmogorov spectrum) that the spectrum $\Phi_n(k)$ is proportional to $C_n^2 k^{-5/3}$. No form can be given for small wavenumbers $k < 2\pi/l_o$, and sometimes a Gaussian roll-off is given at the large k (small eddy) end, but is basically meaningless to do more than to absorb these unknown sections into the definitions of L_o and l_o . Filter factors usually enter into the calculations as factors of $\Phi_n(k)$ to prevent its behavior for both small or large k from seriously influencing the effect of the inertial subrange $2\pi/L_o < k < 2\pi/l_o$. Another caveat: one often finds an “ $-11/3$ ” law for $\Phi_n(k)$ cited in the literature. The reason is that—in contrast to the above—a three-dimensional $\Phi_n(k)$ can be defined, which differs of course by a factor $4\pi k^2$ for isotropic turbulence.

Bandwidth

Just a word about time effects. Most of the behavior with time is adequately described by the frozen-flow hypothesis: things vary at one location in time because a “frozen” irregular spatial structure floats by. Fre-

quently, scales are translated from l to frequency f by the simple relationship $f \sim U/l$, where U is the float velocity. Thus, Eq. 10 manifests itself at *one* detector but at *two* different times as a frequency effect $C_n^2 f^{2/3}$. In the lower atmosphere f is sandwiched between U/L_o and U/l_o , and so for $U \sim 10$ m/s, between 10 Hz and 10/kHz or so, but at lower wind velocities, this spectrum may shift.

Final comments

In summary, the subject has been developed actively in the last ten years. Pertinent literature is cited in Refs. 5 and 11, and even the casual peruser will notice that most of it is in only four or five journals: *J. Opt. Soc. Am.*, *Appl. Optics*, *Radio Science*, *IEEE Trans. Antennas and Propagation*, and the Russian journal *Radiofizika—IVUZ* (also available in translation, as *Radiophysics and Quantum Electrodynamics*, but with several years backlog). Tatarski's textbook (Ref. 6) is considered to be the standard primer, but is now somewhat dated as it was written in 1967.

The following recent development is interesting. With the advent of fast large-memory computers, it has become possible to process atmospheric-turbulence information and correct for it in real time. This is especially valuable for astronomical applications, where the major error-causing effects are the undulations (and

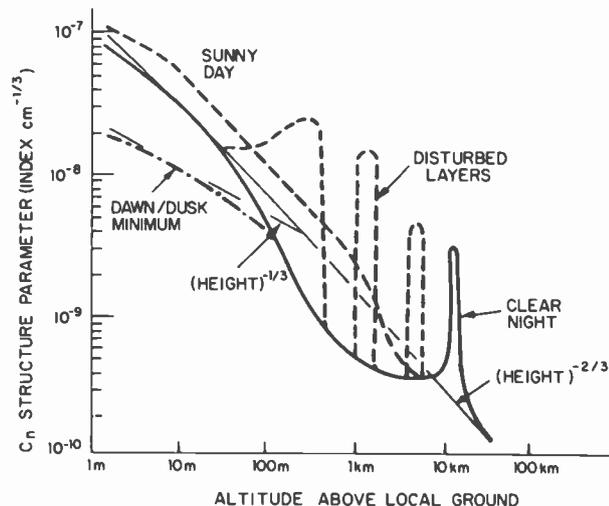


Fig. 10
“Constant” varies with altitude as well as time. Model assumes that local ground is level, not on a mountaintop. Direct comparison with Fig. 9 is impossible because of the different units.

overall retardations or advances) in the phase front. A telescope mirror is struck not by a plane wave but by an undulating wave. It is now possible to correct for this by using deformable mirror elements, process information about the wave-tilt striking each element, and then adjust each element in tilt by a servomechanism to obtain the sharpest image. The *raison d’être* of this development is—of course—the dominance of phase shifts (over other amplitude-modifying causes) caused by turbulent air layers.

References

1. At an altitude of 3 km, the density of the atmosphere has decayed to 60% of its sea-level value, and thus a good fraction of the atmosphere already lies under our astronomer.
2. There is a photometric coterie that insists upon foisting a confusing terminology and notation upon the scientific community—see the *RCA Electro-Optics Handbook*, Section 2, for a patient exegesis—but their recommendations are often (including here) ignored.
3. I am oversimplifying here. Actually, for very dense media such as clouds, rescattered radiation is quite important, to the extent that reradiation in other than the original direction can be comparable in magnitude to the remainder of the beam. On the other hand, this is not the case in the almost-clear atmosphere suitable for optical propagation.
4. Fried, D.L.; *J. Opt. Soc. Am.*, Vol. 56 (1966) p. 1372.
5. Fante, R.L.; *Proc. IEEE*, Vol. 63 (1975) p. 1669.
6. Tatarski, V.I.; “*Effects of the Turbulent Atmosphere on Wave Propagation*,” (transl.) U.S. Dept. of Commerce (NHS) Doc. 1168-50464.
7. Not unlike many names coined by present-day researchers, this one is a misnomer: C_n^2 is not constant at all, neither in time nor in space!
8. Lawrence, R.S.; Ochs, G.R.; and Clifford, S.F.; *J. Opt. Soc. Am.*, Vol. 60 (1970) p. 816.
9. Hulnagel, R.E.; *Digest of Proceedings, Topical Meeting on Optical Propagation Through Turbulence*, July 9-11, 1974, Boulder, Colo., Paper WA1.
10. Wyngaard, J.; Izumi, Y.; and Collins, S.A.; *J. Opt. Soc. Am.*, Vol. 61 (1971) p. 1646.
11. Prokhorov, A.; Bunkin, F.; Gochelashvily, K.; and Shishov, V.; *Proc. IEEE* Vol. 63, (1975) p. 790.

Integrated optics and optical waveguide modulators

J.M. Hammer

Optical communications requires all the switching and modulation functions of its electronic counterpart. Recently developed waveguide modulators are capable of operating below 10V and 0.5 mW/MHz.

The term *integrated optics* usually applies to a system performing a variety of operations on light guided in a single plane. Such a system should combine two or more active functions, such as modulation, spatial switching, frequency conversion, amplification, generation, and detection. An ultimate integrated-optics system, then, would be a monolithic optical waveguide containing a variety of active and passive subelements coupled together to achieve the desired functions. The driving signals are frequently thought of as being electrical, but may well be optical. Similarly, the output could be either an electrical signal or light. In the latter case, the output light could be sent on to a transmission system, such as a fiber-optic waveguide, or viewed directly.

Comparing electronics and integrated optics

Although the role of integrated optics is not yet fully established, the hope is that such circuits will not only replace electronics in certain applications, but will provide functions not previously available using electronic methods. An example of such a "new function" would be the direct man-machine interface made possible by reading out the light from an integrated optical processor. Some of the hoped-for advantages of integrated optics circuits are: ruggedness; insensitivity to temperature extremes, high humidity, and high ambient radiation; resistance to electrical interference and cross-talk; operation speeds limited by light transit times rather than

capacitive effects; bandwidth capabilities comparable to light frequencies; low drive power requirements; and finally, relative production simplicity. In addition, integrated-optics systems are compatible with fiber-optic communication and data-transmission systems, about which more is said below.

To date there have been few examples of experimental integrated optic circuits having more than one active operation on a single substrate. Since most work has been devoted to single discrete devices, this review is mainly concerned with thin-film dielectric optical waveguides, which can be used for active operations such as switching and modulation.

An encouragingly large number of discrete active operations on light traveling in planar optical guides has indeed been reported. These include switching¹ (deflection in the waveguide plane), modulation,² amplification,³ and generation.⁴ These active devices do in fact show many of the hoped-for advantages. In addition, a large number of important passive operations has been demonstrated. These include all the refractive effects analogous to those performed with conventional optics: diffraction, directional coupling between planar guides, and coupling light into and out of optical waveguides.

Advantages of optical communication

The related fields of optical communication and data transmission probably hold the most important applications for integrated optics in the near future. The idea of sending information using optical-frequency carriers even predates lasers, but the idea neared fruition when lasers

became available. This strong historical interest in light carrier systems is based on the high frequency of optical waves, which allows tremendous bandwidths to be contemplated. For example, a system using only 1% of the center frequency of 6000-Å light could transmit 5×10^{12} Hz (50,000 digital tv channels). The short wavelength allows the use of small antennas, a few millimeters as compared to a few meters for microwave systems. In addition, since light quanta are fairly energetic, efficient detectors such as PIN or avalanche detectors can be used. Because the atmospheric transmission of light is severely limited by absorptions and made hopeless by foul weather, it became necessary to consider closed transmission systems or "pipes" enclosing more-or-less conventional lenses. More recently, the advent of both multi- and single-mode fiber-optic transmission lines gave additional reason to study optical-frequency carrier systems.

Fiber-optic systems

The description of light propagating in a transparent fiber or dielectric film is similar to that for microwaves traveling in (or on) the dielectric waveguide already studied by microwave engineers.⁵ Thus, a good understanding of propagation of light in fibers and thin film guides was available at the outset of research in the area. Work on fiber-optics has progressed to the point where multi-mode fibers with losses less than 5 dB/km have been produced.⁶ While single-mode fibers with losses below 10 dB/km have also been demonstrated, the actual bandwidth of a fiber-optic system is restricted by the dispersion of the fiber. This amounts to 10^{-10} s/km for single-mode fibers and deteriorates to as much as 10^{-7} s/km for multi-mode fibers. The

Reprint RE-22-4-11
Final manuscript received June 18, 1976.

The research in this paper was jointly supported by the Office of Naval Research, Arlington, Va; Avionics Laboratory, Wright-Patterson AFB, O. and RCA Laboratories, Princeton, N.J.

single-mode dispersion, however, is low enough not to be an important limitation in a practical system. For an ultimate system, the large bandwidth can be efficiently used in an error- and noise-resistant coding scheme, such as pulse-code modulation. Such a system would consist of an integrated-optic transmitting terminal that encodes, multiplexes and amplifies the final light-pulse train and couples it to a fiber-optic transmission line. The fiber-optic line would be equipped with suitably spaced integrated-optic regenerative repeaters and terminate in an integrated-optics receiving terminal.

While such an elaborate system still requires further development of its components, such as efficient couplers between the fibers and the I/O devices, more modest coherent systems have been realized. For example, over short distances, as in an aircraft data-transmission link, bundles of single-mode fibers provide large information-handling capabilities and are still free of crosstalk, need little protection

Jacob Hammer has worked on low-noise microwave research, electron interactions with atoms, and lasers since joining RCA Laboratories in 1959. He has been studying problems in optical communication and waveguides since 1969. Dr. Hammer received the RCA Laboratories Outstanding Achievement Awards in 1962, 1964, and 1973.

Contact him at:
Communications Research Laboratory
RCA Laboratories
Princeton, N.J.
Ext. 3210

Author Hammer examining a waveguide modulator.



from the aircraft's environmental extremes, avoid ground-loop and other electromagnetic interference problems, and are light in weight. Systems such as these could be fed by simple waveguide modulators with effective bandwidths greater than 10 MHz per fiber. As another example, a simple cable-tv system can use similar modulators and accommodate one tv channel per fiber over distances on the order of 5 miles.²⁰

Components for optical communications—switches, modulators, and couplers

At this point we will describe the developments in switches, modulators, and fiber couplers that have been the main focus of the program at RCA Laboratories. The reason for this emphasis is that optical-waveguide modulators and switches are vital components of any sophisticated optical communication or data transmission system.⁷ Without electronically controlled optical switches, the switching in an optical system must be performed either mechanically or by detecting the light and then switching the detected signal electrically and regenerating a new optical signal. The disadvantages of both these approaches are obvious.

In the absence of good modulators, information can only be impressed on optical carriers by modulating the light source. This requirement places constraints on the possible sources that may be used, and, because of the very nonlinear pumping characteristic of lasers, presents serious problems in obtaining analog modulation at high frequencies. Using a modulator, therefore, widens the choice of light sources and allows the chosen light source to be optimized for such basic emission characteristics as, for example, efficiency and coherence.

The future of optical communications, therefore, is highly dependent on developing efficient, high-speed, light modulators and switches. Without these components, optical systems will never be as flexible as conventional electronic methods.

The RCA Laboratories program has produced devices that meet these requirements and also operate at voltages and powers consistent with those used in electronic integrated circuits. The frequency capabilities of these devices extend well into the microwave region. The waveguide

modulators are readily produced using mass-production techniques similar to those used to make semiconductor devices.

The Pockels effect

This paper covers only optical waveguide switches and modulators based on the use of linear electro-optic (Pockels) effect in essentially insulating crystalline materials. The Pockels effect is shown only by crystals that lack a center of symmetry. Briefly, the effect is a change in birefringence or a change in refractive index caused by, and linearly proportional to, applied electric field. Generally, refractive-index changes on the order of some parts in 10^4 can be obtained with high but finite electric fields. Clearly, to be useful, the index change must take place in the crystal volume through which the light flows.

Optical waveguides

Optical waveguides, which inherently restrict light flow to regions with at least one dimension roughly the size of an optical wavelength, are ideally suited to make optimal use of the Pockels effect. Their advantage is that a low voltage may be applied across the narrow gap to give a high field. At the same time, the volume occupied by the optical wave is small, so the total stored energy in the applied field is low. As a result, waveguide electro-optic modulators operate at far lower voltage and require far less power than bulk electro-optic modulators.

The linear electro-optic effect in crystals has an additional attractive feature in that the inherent frequency response is extremely high, extending well about the millimeter-wave region. It is thus clear why the interest in electro-optic waveguide modulators is high.

In discussing the application of the Pockels effect to optical waveguides, it is convenient to distinguish two broad classes of operation. The first class makes use of crystal and waveguide orientations so that the applied electric field produces birefringence. The guided wave thus undergoes a change in polarization that is proportional to the applied field. The second class uses orientations and polarizations in which the electric field produces a simple refractive-index change. Here the guided wave undergoes phase changes that are proportional to the applied field. Both of these classes can be

used in conjunction with a variety of electrode configurations to construct phase and amplitude modulators and switches.

Diffraction gratings + waveguides = optical switching

This section is devoted to modulators and switches based on electrode configurations that provide periodic index variations.^{8,9} This type of structure produces an electrically controlled Bragg diffraction grating that spatially separates the diffracted and undiffracted beams in the waveguide plane. A number of such gratings may be cascaded on a single planar waveguide and used to perform a complex of switching functions. Using this approach, RCA Laboratories has developed devices based on a new and simple waveguide made by diffusing metallic niobium into LiTaO₃ substrates.¹⁰ With them, the power and voltage required to switch or modulate large fractions (over 80%) of the input light at frequencies close to the microwave range are low enough to be compatible with the outputs of electronic circuits.¹¹

These waveguide films are formed by evaporating approximately 500 Å of metallic Nb onto polished LiTaO₃ substrates. These are placed in a weakly oxidizing atmosphere for 6 to 10 hours at about 1100°C so that the Nb diffuses into the LiTaO₃, forming a thin film of LiNb_xTa_{1-x}O₃ (LNT). This process produces maximum index differentials between the film and substrate on the order of 1%, which makes for a relatively strong optical waveguide with losses on the order of 1 dB/cm.

Potential for light-on-light operations

We later discuss how even more efficient switches than the Bragg grating switch have been made using the waveguide properties. In addition to producing active waveguide devices with these guides, we have made detailed studies of the physical properties of the LNT films.¹⁰ The particular system of diffusing niobium into LiTaO₃ has further interesting potential in that both LiTaO₃ and LiNbO₃ can be doped with iron in such a way that they become light-sensitive. Since refractive-index changes can be thus induced in the waveguide by the action of light, this property opens the possibility for using these waveguides to perform light-on-light operations.

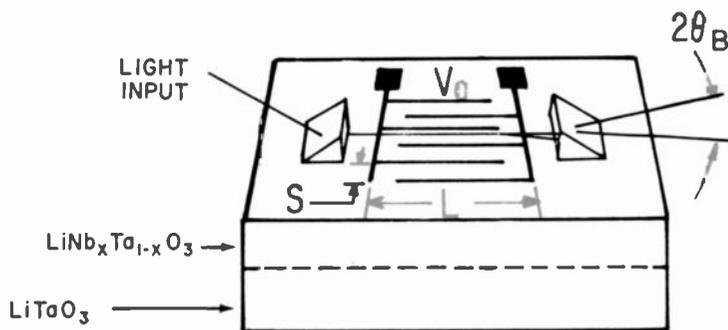


Fig. 1

Grating modulator on LNT waveguide. Guided light is diffracted through an angle $2\theta_B$ when a voltage is applied to the interdigital electrodes. S is 7.6 micrometers and L is 0.3 cm.

As shown in Fig. 1, laser light is coupled into the film with SrTiO₃ prism couplers. The effective index for the guided light may be calculated from the coupling angle and is found to be consistent with values expected for this system. Single-mode operation is obtained; this implies that the thickness at which the graded index difference between film and substrate falls to 10% of its maximum value is between 0.7 and 1.6 μm .¹²

The waveguide modulator is produced by applying a voltage to an interdigital electrode pattern deposited on the waveguide surface as shown in Fig. 1. Applying the voltage to the electrodes results in a periodic variation in the refractive index; this variation acts as a phase diffraction grating. For L sufficiently large the grating may be considered a "thick" or Bragg grating. Light entering the grating at an angle θ_B is diffracted through an angle $2\theta_B$ in the waveguide plane, where

$$\sin\theta_B = \lambda_0 / 4Sn_r$$

and n_r is the effective index for the guided mode being considered. The fraction of the entering light diffracted is

$$I/I_0 = \sin^2 \Delta\phi / 2$$

and to the first order in r' , the phase shift is

$$\Delta\phi = -r'(n_r)^3(L/\lambda_0)E$$

E is the average in-plane field component caused by the voltage V_0 , and r' is the effective electro-optic coefficient. Thus, I/I_0 has the form $\sin^2 BV_0$.

Experimental waveguide results

Fig. 2 shows the percentage diffracted as a function of voltage for three laser

wavelengths. The solid curves are plots of $\sin^2 BV_0$, normalized to the data at $I/I_0 = 75\%$. The functional agreement is good. We observe no variation in these percentages from dc up to pulses with rise times below 3 ns, which is the limit of our measurement capability. The measured capacitance is 20 pF, which indicates capacitive power requirements below 0.2 mW/MHz.

The pulse response is shown in Fig. 3. The apparent 3-ns rise time of the light is limited by the response of the photomultiplier. Since the Pockels effect is capable of responding to frequencies of about 10^{14} Hz, the real limitation of this device is set by circuit-match considerations.

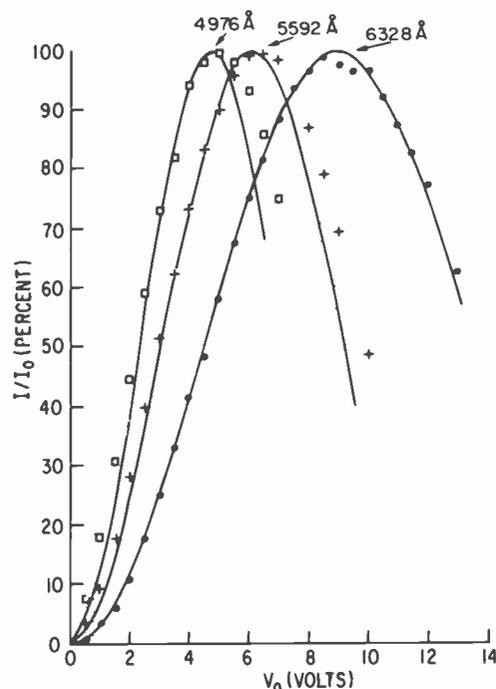


Fig. 2

Percentage of light diffracted in waveguide modulator as a function of voltage.

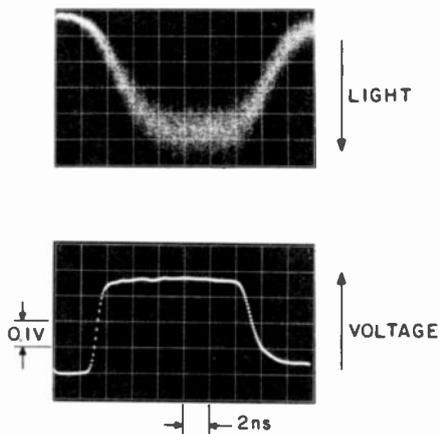


Fig. 3
Pulse response of waveguide modulator. The apparent rise time is limited by the response of the photomultiplier.

The $\text{LiNb}_x\text{Ta}_{1-x}\text{O}_3$ optical waveguides described here are relatively simple to make, have excellent and controllable waveguide properties, and can be oriented to make optimal use of the strong electro-optic effect of both LiNbO_3 and LiTaO_3 . The high efficiency and low voltage and power requirements of this modulator represent an order-of-magnitude improvement over bulk devices and earlier waveguide grating modulators.

Fig. 4 illustrates the waveguide modulator's size advantage over the bulk modulator. Here a commercial state-of-the-art bulk electro-optic modulator is contrasted with a waveguide modulator hypothetically packaged to include input and output grating couplers and a protective case. The waveguide modulator operates at approximately 10V and requires powers below 0.5 mW/MHz; the bulk modulator requires at least ten times its power and voltage.

Although the waveguide modulator could be packaged to act as a bench modulator, it is more likely to be used in conjunction with fiber-optic transmission lines. Proper coupling devices to connect the waveguide modulator directly to fiber-optic waveguides are under development at RCA Laboratories.

Overcoming coupling problems

There are many problems to be overcome in making direct couplers between fibers and planar electro-optic waveguide devices. Direct coupling by end-firing techniques is impractical because of both size and symmetry differences between the

fiber and the planar guide. In addition, the beam spread of light from the end of a multi-mode fiber is too large to be accepted by the modulator structures. Even with a single-mode fiber, the beam spread from an end-fired coupler will be too large to accept because of diffraction from the very small core of the fiber.

An alternative approach is to attempt to couple the fiber to the film guide using their evanescent fields. Ideally, simply bringing the core of the fiber close to the film should allow some coupling between the two systems. Here, the difficulty is the phase-velocity mismatch between light traveling on the fiber, which has a refractive index of approximately 1.5, and light on the film, with refractive indexes on the order of 2.2.

Coupling via evanescent fields and a diffraction grating

Nonetheless, our approach is indeed based on the use of evanescent-field coupling. We allow the evanescent fields of the two guides to overlap in a planar region containing a phase diffraction grating, which serves as the required phase-matching element (Fig. 5). Coupling then takes place over a controllable interaction length. The effective coupling aperture, H , may therefore be designed to reduce the diffractive beam spread in the film waveguide to values that the switch can accept. With this method, theoretical coupling efficiencies on the order of 90% have been predicted¹³ between single-mode fibers and film guides using blazed backward wave gratings.

Multi-mode fibers may also be coupled to single-mode films. This converts the spread in velocity among the fiber modes to an

angular spread in the film, and so reduces coupling efficiency. The spread, however, may be rather small ($<0.2^\circ$) for the low-loss, low-optical-aperture fibers that are most attractive for optical communications. Nevertheless, coupling efficiencies on the order of 50% may still be realized.

Since the coupler is located entirely on the surface plane of the film waveguide, its fabrication and packaging characteristics are compatible with the requirements expected for the mass production of planar optical waveguide devices and systems.

To understand the operation of the coupler, consider the core of an optical fiber in intimate contact with a grating formed in the surface of a planar (film) optical waveguide, as in Fig. 5.

If a mode with propagation constant $|\beta_1| = 2\pi n_c/\lambda_0$ can propagate in the fiber and a mode with propagation constant $|\beta_2| = 2\pi n_f/\lambda_0$ can propagate in the film, strong coupling may take place if the fields of the waves overlap in the grating region and if the grating vector $|k_g| = 2\pi/d$ closes the vector triangle illustrated in Fig. 5. This occurs when

$$d/\lambda_0 = (n_2^2 + n_c^2 - 2n_2n_c\cos\theta)^{-1/2} \quad (1)$$

and

$$\phi = \arcsin(n_c d \sin\theta/\lambda_0).$$

Here, n_c and n_f are the effective refractive indexes in the fiber and film, respectively. The critical coupling length, L , for which there is maximum transfer between the guides depends on the grating characteristics and conditions of field overlap. For this geometry it is clear that the effective coupling aperture, H , depends

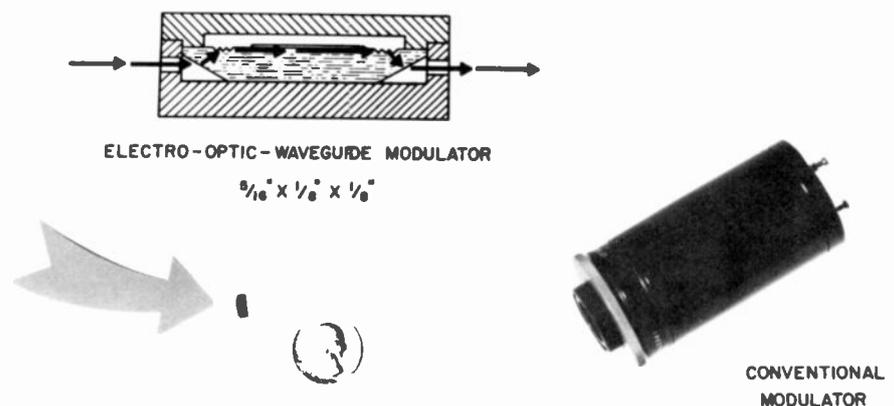


Fig. 4
Waveguide modulator has a size advantage over the bulk modulator. The waveguide modulator is pictured as it would appear if provided with grating couplers and packaged.

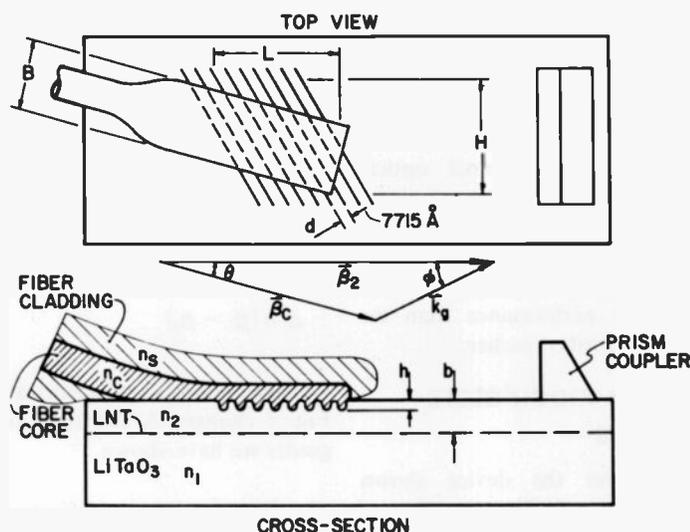


Fig. 5 Coupling between fiber and film waveguide is done with a diffraction-grating system.

on L and the angle θ . In addition, by choosing $L \gg B$, H becomes essentially independent of B for $\theta > 0$.

Without the grating, a fiber mode would couple by refraction to an unbound substrate wave.¹⁴ The grating may be viewed as scattering unbound waves into guided directions, and the efficiency of this scattering can be made high by using blazed gratings.¹⁵ If a forward wave grating is chosen ($\theta < 90^\circ$), light will be lost to higher grating orders, restricting the ideal efficiency to 50% or less. This restriction can be avoided with backward wave coupling ($\theta > 90^\circ$).¹⁶

Working with multimode fibers

Our discussion has, to this point, been based on coupling a single-mode fiber or a single mode of a multimode fiber to a single-mode planar guide. The in-plane propagation direction of a single-mode planar guide is unrestricted, so we might expect that many of the fiber modes could be coupled to the planar guide by assigning a different direction in the planar guide for each fiber mode. We shall see that the grating coupler can perform this function.

A multi-mode fiber, with core and cladding refractive index of n_c and n_s , respectively, will sustain modes with a effective refractive index n_{eff} ranging from $n_{eff} \approx n_c$ for the lowest-order to $n_{eff} \approx n_s$ for the highest-order mode. If the coupling grating is designed to couple the lowest-order (slowest) mode, higher modes may also

couple with varying efficiencies, depending on the angular selectivity of the grating. A grating coupler with a long critical coupling length will have high angular selectivity, and vice versa. Long critical coupling lengths are associated with a combination of weak gratings and small field overlaps.

What happens is that as the angle ϕ changes by an amount $\Delta\phi$, the diffraction efficiency falls off as¹⁷

$$I(\Delta\phi) = I_0 \frac{\sin^2(k_s L \Delta\phi / 2)}{(k_s L \Delta\phi / 2)^2} \quad (2)$$

and the coupling efficiency for the modes that do not exactly satisfy the condition of Eq. 1 are reduced. Thus, the coupling process causes a beam spread, which converts the velocity spread in the multi-mode fiber to an angular spread in the film. In addition, in both the multi-mode and single-mode cases, light coupled from fiber to film will have an angular spread $(\Delta\theta)_L$ caused by the finite size of the effective optical aperture H . For a single mode the spread will be

$$(\Delta\theta)_L = 2n_2\lambda_0 / L \cos\phi \quad (3)$$

In the multi-mode case this will be increased by adding $(\Delta\theta)_m$. Clearly, a compromise between coupling efficiency and angular spread will have to be made for any multi-mode coupler intended for use with waveguide devices having limited acceptance angles.

We may summarize our discussion to this point as follows. In using a grating to

couple a multi-mode fiber to a single-mode film waveguide, the spread in phase velocities among the fiber modes is converted to an angular spread in the planar guide. In addition to the coupling spread, there is an additional spread due to the finite aperture (Eq. 3).

If a single planar-guide mode with minimized angular spread, such as one that might be produced by coupling a TE_{00} laser beam into the film guide with a prism, is grating-coupled out into a multi-mode fiber, the coupling will be unrestricted by the angular characteristics. In this case we would expect the coupling efficiency to be determined entirely by the characteristics of the grating and field overlaps. This expectation was used to experimentally characterize the developmental coupler by measuring the output coupling from planar-film waveguide to fiber when the input to the film is a well-collimated laser beam.

Coupling experiments

In our experiment an unclad multi-mode (diameter 0.002 in.) plastic (PMMA) fiber is embossed onto a grating etched into the film-waveguide surface. The film waveguide is a single-mode (TE_{00}) diffused LNT waveguide with $n_2 = 2.195$ and $b \approx 1.0 \mu\text{m}$. The fiber index is close to 1.5 and, since there is no cladding, $\Delta n = 0.5$. The grating space d is 7715 \AA , which requires θ to be 13.8° and ϕ to be 25.8° .

The grating coupling characteristic is first determined by coupling 6328-Å HeNe laser light into the LNT film with a prism

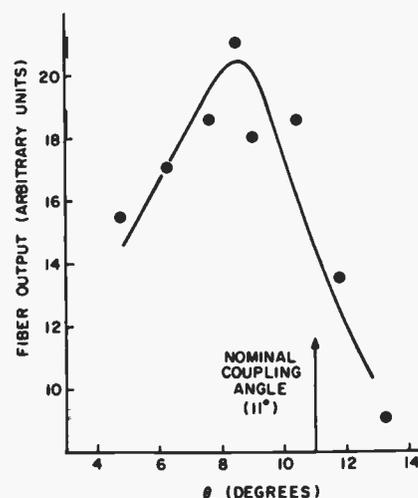


Fig. 6 Push-pull stripe guide switch uses phase matching to couple the two guides.

coupler using a lens to produce a well-collimated (essentially parallel) guided beam that enters the grating coupling region at the nominal coupling angle. When this is done, we find a coupling efficiency of 6% by measuring the ratio of light emerging from the fiber to light entering the prism and correcting for the prism input coupling, waveguide loss and fiber loss, which are all independently measured. The angular dependence of the coupling output can be verified by rotating the entry angle of the beam in the planar guide; this is plotted in Fig. 6. Obtaining the expected angular dependence verifies the physical basis of the coupling.

The multi-mode fiber-to-film coupling is determined by focusing the 6328-Å laser into the free end of the plastic fiber, using a 40-power microscope objective. This overfills the acceptance angle of the fiber, ensuring a broad excitation of the modes. The modes are additionally mixed by using at least a meter of fiber bent in a ten-cm radius. The light intensity coupled out of the film waveguide at the prism coupler is then measured and compared to the input light, correcting for the other losses, and we find a fiber-to-planar coupling of 0.11%. This low number is readily understood in terms of the theory outlined earlier, since most of the modes fall in a region where the $(\sin x/x)^2$ dependence is small.

These experiments demonstrate the coupler principle and give promise that as we learn to make improved gratings and increase the quality of the physical contact between fiber and film, the theoretical values of coupling will be obtained.

System expectations

We thus briefly consider what might be expected using a low-loss (<20 dB/km) multi-mode fiber of the type now commercially available. We assume the cladding is left in place to avoid the large Δn of an unclad fiber. The cladding might appear as indicated by the dotted region in Fig. 5. For these fibers $\Delta n \approx 0.006$ and $n_c = 1.457$ at 6328 Å.

Using a critical coupling length $L = 100 \mu\text{m}$, and requiring $\Delta\theta = 0.2^\circ$, we predict $I/I_0 \approx 0.46$ for an ideal correctly-blazed grating. This must be reduced by 50% if we choose to work in the forward-wave region, giving a predicted 34% coupling efficiency into a high-index planar guide. If we had chosen to work in the backward-wave region and were willing to accept 0.5° spread, approximately 50% coupling

would be predicted. A single-mode fiber in a similar coupler would produce approximately 90% coupling.

The grating-coupler approach can also be applied to stripe or channel optical waveguides. In the next section we give a brief description of stripe guide switches that make use of the unique properties of the LNT waveguide system and promise to give even better performance than the grating switch described earlier.

Stripe guide modulators and switches

We now consider the device shown schematically in Fig. 7. The operation of this type of coupler-modulator has been described in great detail in a number of places,¹⁸ so we will only give the briefest outline. If β_1 and β_2 are the propagation constants in guides 1 and 2, under phase-match conditions $\beta_1 = \beta_2$ and light will be coupled between the guides. If initially all the light is in guide 1 after traveling a distance, L , equal to the critical coupling length, all the light will be transferred to guide 2. If the interaction continues over a long length, light will be coupled back and forth between the guides.

If $\beta_1 \neq \beta_2$, the amount of light coupled will be reduced and the coupling period shortened. The coupling is described by¹⁸

$$\begin{aligned} I_2/I_0 &= 1 - (I_1/I_0) \\ &= [\kappa^2 / (\kappa^2 + (\Delta/2)^2)] \cdot \\ &\quad \sin^2[(\kappa^2 + (\Delta/2)^2)^{1/2} x] \end{aligned} \quad (4)$$

where I_0 is the intensity initially entering the coupling region and, at $x=0$, $I_2=0$ and $I_1=I_0$. The coupling coefficient κ is related to the critical coupling length through

$$\kappa = 2/\pi L. \quad (5)$$

The phase mismatch is measured by Δ , given by

$$\Delta = |\beta_1 - \beta_2| \quad (6)$$

Voltage applied to the electrodes shown in Fig. 7 changes Δ and for our geometry guides we have shown

$$\Delta \approx (2\pi/\lambda_0) r' n'^3 (V/a) \quad (7)$$

where r' is the effective electro-optic coefficient and n' is the effective refractive index as defined in Ref. 18. Thus, a voltage can vary the amount of light coupled between the two guides.

The extinction ratio is defined by

$$\eta = |I_1 - I_2| / (I_1 + I_2) = |I_1 - I_2| / I_0 \quad (8)$$

where the values of I_1 and I_2 are taken after the switch and a lossless system is assumed. Extinction ratios close to 1 (100%) are desired.

For the type of device illustrated in Fig. 7, η will depend strongly on how closely the actual device length approaches the critical coupling length. The critical coupling

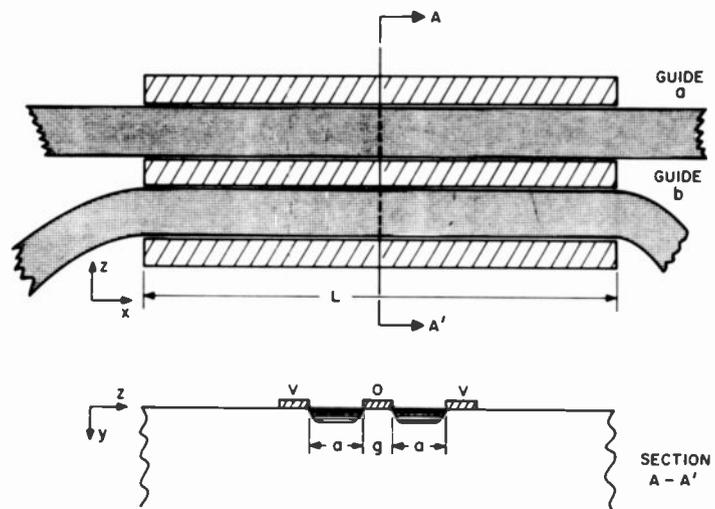


Fig. 7 Push-pull stripe guide switch uses phase matching to couple between the two guides.

length is proportional to the reciprocal of the coupling constant, which, in turn, is extremely sensitive to the device dimensions and propagation characteristics.

Conventional photolithography is used to fabricate the stripe guide couplers; Nb is deposited on a LiTaO₃ substrate, covered with photoresist and exposed through a mask. After development the Nb is etched, leaving the desired pattern.

Manufacturing the coupled stripe waveguides

This relative simplicity also applies to how the coupled stripe waveguides are formed. After guide formation, the switching electrodes are deposited using evaporated chrome-gold and a second photolithographic step. Fig. 8 is a photomicrograph of this stripe guide showing an entry horn on the upper guide. When the stripe guides are correctly formed, light coupled into the upper guide through the entry horn emerges from the lower guide through a second horn (not visible in the figure). The dimensions are (referring to Fig. 5) $a = 5 \mu\text{m}$, $g = 2 \mu\text{m}$, and $L = 2 \text{mm}$. The center electrode is $2 \mu\text{m}$ wide. Thus, it is obvious that careful photolithographic techniques must be used, although such techniques are well within present manufacturing capabilities. The results of our initial experiments with this type of switch¹⁹ are shown in Fig. 9—the voltage behaves as expected theoretically.

Future versions of these devices are predicted to require drive voltages on the order of only 1V and drive power less than $10 \mu\text{W}/\text{MHz}$. These very low values hold promise for system flexibility that would make these types of switches extremely

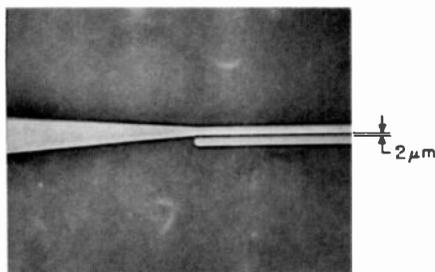


Fig. 8 **Stripe guide directional coupler.** Light entering the upper guide through the entry horn emerges from a similar exit horn (not shown in this photomicrograph) in the lower guide.

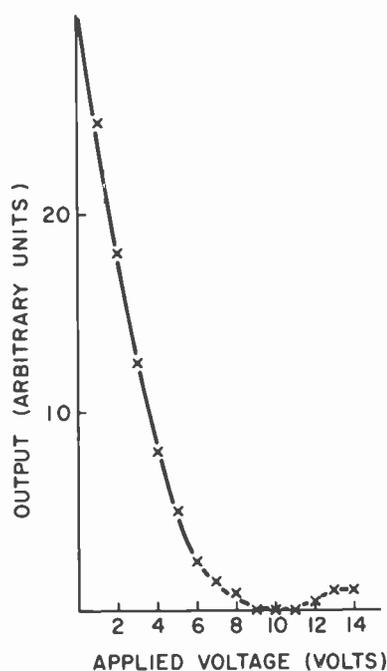


Fig. 9 **Experimental variation** of output of stripe guide coupling modulator as function of applied voltage. The small bump at 14 V is significant because it confirms the theoretically predicted behavior.

attractive in applying integrated optics to telephone switching, data processing, and repeaters for optical-fiber communication systems.

Summary and future work

The work at RCA Laboratories on thin-film dielectric optical waveguides has resulted in the production of optical switches and modulators operating with voltages below 10V and powers below $0.5 \text{mW}/\text{MHz}$. These values are more than an order of magnitude lower than bulk modulators and are compatible with the outputs of conventional integrated-circuit technology. While the chief use for these devices lies in the field of optical communications, applications to optical data-processing systems (such as the modulator element in optical video disc recording) may also be visualized.

An important missing link for using "integrated optics" elements is the coupler to optical fibers. Work at RCA Laboratories on the development of grating couplers has progressed to the point that the basic correctness of this method has been demonstrated. More

work is required, however, to improve the coupling efficiency to useful values (losses less than 3 dB per coupler pair).

Optical waveguide modulators for video-disc-type applications require improvements in the materials and coupling methods to produce lower insertion losses, higher optical power handling capacity, and better beam quality in the blue-green spectral region than is now available.

In the next few years, the optical waveguide program at RCA Laboratories will concentrate on obtaining improved fiber/thin-film couplers and on improving the blue-green performance of the waveguide modulators.

Acknowledgments

The research at RCA Laboratories reviewed in this paper is the result of the joint efforts of W. Phillips, C.C. Neil, R.A. Bartolini, A. Miller, D.J. Channin, M.T. Duffy, and the author. Our efforts have been encouraged by the informed support of B.F. Williams. Free use has been made of much joint published material. The author wishes to express his appreciation for the long and fruitful association he has enjoyed with these friends and colleagues.

References

1. Hammer, J.M.; Channin, D.J.; Duffy, M.T.; and Neil, C.C.; *IEEE J. Quantum Electron.* Vol. 11 (1975) p. 138.
2. Reinhart, F.K.; and Miller, B.I.; *Appl. Phys. Letters* Vol. 20 (1972) p. 36.
3. Klein, M.B.; and Abrams, R.L.; *IEEE J. Quantum Electron.* Vol. 11 (1975) p. 609.
4. Zeidler, G.; *J. Appl. Phys.* Vol. 42 (1971) p. 884.
5. Ramo, S.; and Whinnery, J.R.; *Fields and Waves in Modern Radiot.* 2nd. ed., New York, John Wiley and Sons, 1953 pp. 388-393.
6. See, for example, Cohen, L.G.; Kaiser, P.; MacChesney, J.B.; O'Connor, P.B.; and Presby, H.M.; *Appl. Phys. Letters* Vol. 26 (1975) p. 472.
7. Hammer, J.M.; *Proc. Soc. Photo-Optical Inst. Engineers* Vol. 53 (1975) p. 60.
8. St. Ledger, J.F.; and Ash, E.A.; *Electron. Letters*, Vol. 4 (1968) p. 99.
9. Hammer, J.M.; *Appl. Phys. Letters* Vol. 18 (1971) p. 147.
10. Phillips, W.; and Hammaer, J.M.; *J. Electronics Materials* Vol. 24 (1974) p. 545.
11. Marcuse, D.; *IEEE J. Quantum Electron.* Vol. 9 (1973) p. 1000.
12. Hammer, J.M.; Bartolini, R.A.; Miller, A.; and Neil, C.C.; *Appl. Phys. Letters* Vol. 28 (1976) p. 192.
13. Peng, S.T.; and Tamir, T.; *Opt. Commun.* Vol. 11 (1974) p. 405.
14. Dalgoutte, D.G.; *Opt. Commun.* Vol. 8 (1973) p. 124; and Ulrich, R.; *J. Opt. Soc. Am.* Vol. 63 (1973) p. 1419.
15. Dakas, M.L.; Kuhn, L.; Heidrich, P.F.; and Scott, B.A.; *Appl. Phys. Letters*, Vol. 16 (1970) p. 523.
16. Gordon, E.I.; *Proc. IEEE* Vol. 54 (1966) p. 1391.
17. Tamir, T., ed.; *Integrated Optics*, New York, Springer Verlag, 1975, see Hammer, J.M.; Ch. 4, pp. 185-189.
18. Phillips, W.; and Hammer, J.M.; *Digest, Topical Meeting on Integrated Optics*, Salt Lake City, Jan 12-14, Paper Tu A5.
19. Witke, J.; *RCA Review* Vol. 35, (1974) p. 198.

Digital filters

L. Shapiro

This third article of the series introduces the concept of the digital filter and lays the groundwork necessary to design infinite-impulse-response filters.

In digital signal processing, we need a filter for the same reasons that we need one in analog signal processing, namely, to remove noise or to select or reject a particular frequency or band of frequencies. Thus, a noisy analog signal $x(t)$, as shown in Fig. 1, may be passed through a lowpass analog RC filter to obtain a "cleaned up" analog output signal $y(t)$. The digital process is essentially the same. We sample the noisy analog signal (Fig. 2a) and obtain noisy samples (Fig. 2b). The problem is to remove the overlay of noise from the digital samples and obtain a cleaned up digital¹ signal (Fig. 2c).

Although the general objective—to remove noise from a noisy signal—is the same in both analog and digital cases, the filtering problem is quite different. Thus, the simple analog low-pass filter of Fig. 1 would be quite inappropriate and actually destroy our sample pulse, leaving only a remnant consisting of a small piece of exponentially rising signal followed by a long exponential tail-off. Evidently, we need a different approach to obtain our cleaned up digital signal.

If we compare our cleaned-up analog signal (Fig. 1c) with its noisy parent (Fig. 1a), we come to the conclusion that its amplitude at any moment depends quite substantially upon the previous amplitude history of the noisy signal. In addition, although not quite so obvious, the amplitude of the cleaned-up analog signal (at any moment) also depends quite substantially upon the previous amplitude history of the cleaned-up signal. (Of course, all values of the cleaned-up analog signal naturally depend upon the action, or transfer function, of the analog lowpass filter.)

¹There is a bit of a semantic problem here since our signal has not yet been quantized and, hence, should be classified as discrete rather than digital. However, to simplify this development we will ignore this distinction and continue to refer to all pulse-type signals as digital since, in practice, such signals would necessarily be quantized at some point in the process before being acted upon by the digital filter.

Reprint RE-22-4-14
Final manuscript received September 1, 1976.

In carrying these relationships over to the requirements of our digital filter, we necessarily conclude that our digital filter must have a memory (e.g., a storage and/or delay capability) as well as a computing capability. This ability is necessary so that the digital filter will be able to adjust the amplitudes of the samples on an individual basis while preserving the discrete nature of the information.

From a mathematical standpoint, in which the memory aspects of the digital filter are taken into account, the relationship between input and output pulse trains may be expressed as a difference equation

$$y(n) + b_1y(n-1) + b_2y(n-2) = a_0x(n) + a_1x(n-1) + a_2x(n-2) \quad (1)$$

Here, a_i and b_i are the weighting coefficients and n is an arbitrarily selected reference pulse along the time axis as

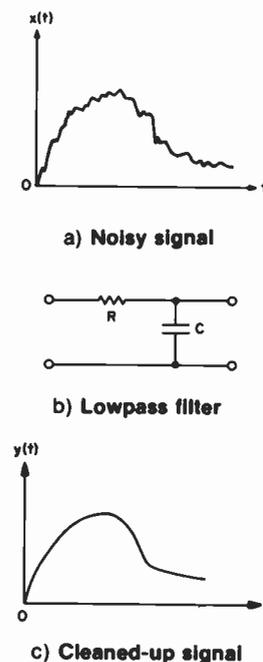


Fig. 1 Filtering action needed to clean up a noisy analog signal can be provided by a simple lowpass RC circuit.

shown in Fig. 3, which compares the noisy input $x(nT)$ and the filtered output $y(nT)$. [In accordance with previous articles of this series we will continue to use the notation T for the (sampling) time interval between adjacent pulses. For clarity, however, we have dropped the constant T from the arguments.]

Digital computer as a digital filter

Actually, our digital filter must take the form of a digital computer. This computer may be either a general all-purpose computer, so that only software is needed, or else a dedicated, or "hard-wired," computer designed to perform only a specific filtering function.

This introduces a very interesting aspect of our digital filter. A digital computer is

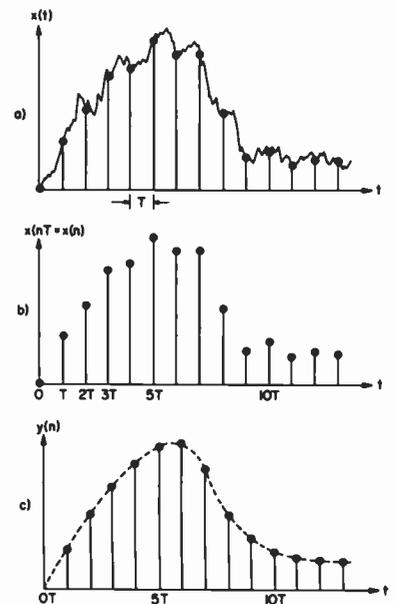


Fig. 2 Digital filtering process starts with a sampled noisy analog signal; thus, the samples represent a mixture of information and noise. The problem is then to remove the noise from the digital samples.

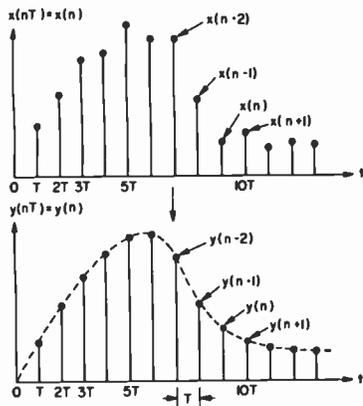


Fig. 3
Numbered samples along the time axis correspond to a difference equation that relates input $x(n)$ to output $y(n)$.

presently capable of doing very much more than simply performing a routine filtering operation. The available signal manipulations can be made very complex and highly sophisticated. Hence, our digital filter can now perform processing on the input signal far beyond the capabilities of any array of analog components. Thus, we have suddenly progressed from a simple RC filtering operation into a new world in which almost anything is possible. We would like to refer the interested reader to the first article of this series for a brief account of what can be done with our innocent-looking digital filter. For the present, however, we will proceed in a modest way and only gradually extend our digital filter capabilities.

Our initial approach will deal with situations where, at least from an analog standpoint, a need is recognized that may be met by a conventional analog filter. This need is then expressed as a set of (analog) filter specifications from which both analog filter and desired digital filter are readily obtained. Although this process may sound artificial and of limited scope, such is not really the case, and it is representative of a great deal of the digital filter design work being done at the present time. First, however, we should note that with this approach we are really dealing with the design of a so-called "infinite-impulse-response" (IIR) digital filter, as described below.

IIR vs FIR

The design approach for infinite-impulse-response filters is quite different than that for the finite-impulse type.

At this point, we want to use the impulse function, $\delta(t)$, which appeared in the first

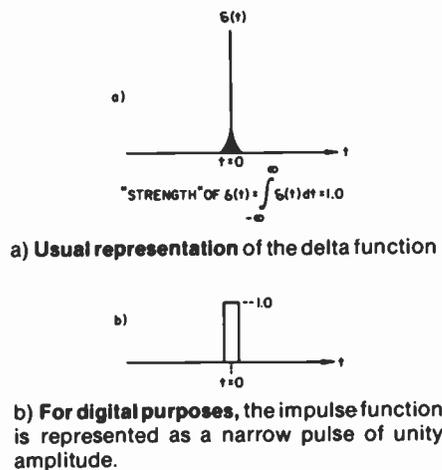


Fig. 4
Strength of the impulse function (delta function) is unity as normally used in digital calculations.

article of this series as a sampling operator (see Fig. 4a). We will now use this function^b as the input to a digital filter so that we may obtain its impulse response. If this response never decays exactly to zero, no matter how long a period of time elapses, the filter is classified as an infinite-impulse-response (IIR) filter. If, however, the output of the digital filter does fall exactly to zero after a finite period of time, we classify the filter as a finite-impulse-response (FIR) filter.

The design of IIR and FIR filters is usually treated separately since the design approach is quite different for each. As indicated above, in this article we will deal with the IIR filter and illustrate the principal methods of its design. In each case, we will assume that we are dealing with a so-called "DTLTI" (discrete-time linear-time-invariant) system so that the filter behavior will not change during the time it is operating on our input pulse.

The z-transform can be quite advantageous in the design of digital filters.

The z-transform is a modified Laplace transform that deals directly with pulse trains or, perhaps more exactly, with time series (of numbers). We will now take note of the z-transform approach to the extent that it will be needed in our design examples. Additional information about the z-transform will be found in Appendix A.

The z-transform of a discrete function $x(nT)$ may be defined as follows:

$$X(z) = \sum_{n=0}^{\infty} x(n)z^{-n} \quad (2)$$

^bSubsequent calculations are based on the assignment of unity to the "strength" of this function. It is sometimes convenient to represent the delta function as a narrow rectangular pulse of unity amplitude as shown in Fig. 5b.

where z is related to the Laplace variable s by the relation $z = e^{sT}$, and s has the usual value $\sigma + j\omega$.

A very interesting aspect of Eq. 2 is that the transform contains, within itself, the original (untransformed) time series. It follows then, that whenever we are able to represent the z-transform of a function in the form of a series in negative powers of z , the associated time series will appear naturally as the coefficients.^c As shown in Appendix A, the application of the z-transform to our difference expression (Eq. 1) gives

$$Y(z) + b_1z^{-1}Y(z) + b_2z^{-2}Y(z) = a_0X(z) + a_1z^{-1}X(z) + a_2z^{-2}X(z) \quad (3)$$

We can now obtain our transfer function $H(z)$, which is simply the ratio of $Y(z)$ to $X(z)$, e.g.

$$H(z) = \frac{Y(z)}{X(z)} = \frac{a_0 + a_1z^{-1} + a_2z^{-2}}{1 + b_1z^{-1} + b_2z^{-2}} \quad (4)$$

Our digital transfer function $H(z)$ is a key quantity in digital signal processing and, in general, can be used as follows:

1) We can apply $H(z)$ to a given input $X(z)$ and obtain the output $Y(z)$, e.g.,

$$Y(z) = H(z)X(z) \quad (5)$$

The output time series, $y(n)$, is then obtained by taking the inverse z-transform of $Y(z)$, e.g.,

$$y(n) = Z^{-1}[Y(z)]$$

2) Given a particular $H(z)$, we can implement it in terms of hardware, or suitable software (for a general-purpose digital computer).

3) Given a desired prototype analog filtering function, or set of filtering specifications, we can develop the corresponding $H(z)$.

4) Given a particular $H(z)$, we can examine it critically for its performance parameters and stability.

Space limitations do not permit a practical discussion on the physical realization of digital filters. Interested readers are referred to texts by Stanley⁹ and by

^cAnother exceedingly interesting situation is that when $X(z)$ represents a transfer function, the time series embodied in the z-transform represents the impulse response of the associated system.

Rabiner and Gold.¹¹ A model for such a realization, however, is given in the problem at the end of this article.

Design examples

We will design a digital filter that has the frequency response of a simple low-pass filter.

The design approaches examined in this article are based on the development of the desired digital transfer function from a prototype analog filter or corresponding set of filter specifications. We generally follow the development in Stanley.⁹

The type of desired digital filter may be any of the following three:

- 1) A digital filter having the same *frequency response* as the prototype analog filter over the portion of the frequency spectrum of interest. This objective can, in general, be only approximately achieved so that a frequency tolerance specification is necessary.
- 2) A digital filter having the same *impulse response* as the prototype analog filter.
- 3) A digital filter having the same *step function response* as the prototype analog filter.

The above three digital filters, although based on the same prototype analog filter, will, in general, have substantially different *frequency responses*. Therefore, we should investigate each of these three responses for any given digital filter. In our design examples, we will begin with a digital filter

Notation:

These symbols are used throughout the paper, but extensively in the design examples. The lower case letters apply to signals in the time (t , nT , or n) domain; upper-case letters refer to signals in the frequency (s or z) domain.

$x(t)$, $X(s)$	Analog input
$y(t)$, $Y(s)$	Analog output
$G(s)$	Analog transfer function = $Y(s)/X(s)$
$x(n)$, $X(z)$	Digital input
$y(n)$, $Y(z)$	Digital output
$H(z)$	Digital transfer function = $Y(z)/X(z)$

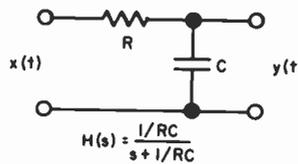


Fig. 5 This low-pass analog filter can be the starting point from which we develop the equivalent digital filter.

that is to have the *frequency response* of a simple low-pass analog filter.

Given the low-pass analog filter of Fig. 5, we want to develop a digital filter that has a similar frequency response over a frequency range extending from dc to cutoff.

What do we mean by a "similar" frequency response on the part of a digital filter?

If a sinewave input signal $x(t)$ is applied to our analog filter (as in Fig. 6a), the output will be a sinewave of the same frequency but of different amplitude and phase as shown by $y(t)$ in Fig. 6b. For the digital filter, the corresponding input and output are shown in Figs. 6c and 6d, where our sample pulses possess amplitudes corresponding to the envelope of the associated analog signal. In each case, the argument of the digital-filter signal function is given as nT indicating that these signals are defined only for multiples of the sampling interval T . As previously indicated, the symbol for the sampling interval, T , is often dropped for reasons of clarity, and the argument is given simply as n . (See the Notation table.)

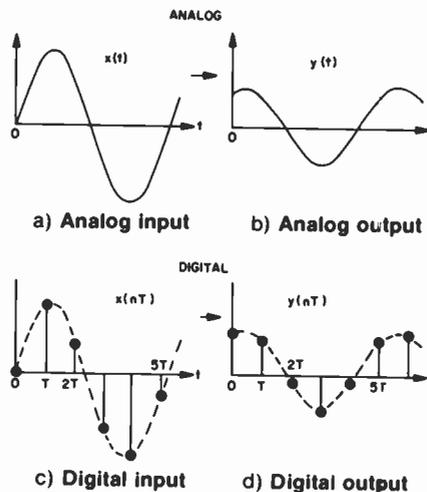


Fig. 6 Frequency response approach. Analog filtering changes phase and amplitude but not the frequency of a signal; our digital filter is designed to match the analog response.

We can now find the transfer function.

The procedure for obtaining the transfer function of the desired digital filter (for a frequency response match) is straightforward:

- 1) Obtain the transfer function of the prototype analog filter.
- 2) Obtain and plot the frequency response of the prototype analog filter.
- 3) Apply the bilinear transformation to the prototype analog transfer function to obtain the corresponding digital transfer function
- 4) Obtain and plot the frequency response of the digital transfer function and compare it with the frequency response of the prototype analog filter. Comparison of the two plots will show the spectral region of acceptable correspondence.

We organize our work in accordance with the above sequence of steps:

- 1) The transfer function of our analog filter of Fig. 5 is easily obtained.

$$G(s) = \frac{1/sC}{R + 1/sC} = \frac{1}{1 + sRC} \quad (6)$$

- 2) The frequency response, $A(f)$, is obtained by replacing the Laplace variable s by its imaginary part $j\omega$ and then taking the absolute value of the result

$$A(f) = |G(j\omega)| = [1 + (\omega/\omega_c)^2]^{-1/2} \quad (7)$$

where $\omega_c = 2\pi f_c = 1/RC$.

This frequency response is plotted on Fig. 7 for $f_c = 400$ Hz and on Fig. 8 for $f_c = 400$ Hz and 357.13 Hz.

- 3) We now develop the transfer function for our digital filter. This is accomplished by means of the bilinear transformation^d

$$s = \left(\frac{2}{T} \right) \left(\frac{1 - z^{-1}}{1 + z^{-1}} \right) \quad (8)$$

Application of this transformation to Eq. 6 gives us the digital transfer function which is conveniently placed in the following form:

$$H(z) = \frac{a_0 + a_1 z^{-1}}{1 + b_1 z^{-1}} = \frac{Y(z)}{X(z)} \quad (9)$$

^dIn complex-variable theory, a bilinear transformation is linear in each of the variables separately (the other one being held fixed). The transformation of Eq. 8 satisfies this condition for the variables s and z . Geometrically speaking, a bilinear transformation has the property that it transforms lines into lines and circles into circles [see, for example, "Complex Variables and Applications," R.V. Churchill, McGraw-Hill, 1960.]

where

$$a_0 = a_1 = \frac{T}{T + 2RC}$$

$$b_1 = \frac{T - 2RC}{T + 2RC}$$

The difference equation connecting the input and output of our digital filter is readily obtained from Eq. 9 by working back from the above transfer function, see Appendix A.

$$y(n) + b_1 y(n-1) = a_0 x(n) + a_1 x(n-1) \quad (10)$$

4) The frequency response of the digital transfer function of Eq. (9) is obtained by replacing the Laplace variable s by its imaginary part $j\omega$, noting that $z = e^{sT} \rightarrow e^{j\omega T} = \cos \omega T + j \sin \omega T$. With some minor manipulation we obtain

$$H(e^{j\omega T}) = \frac{a_0 + a_1 \cos \omega T - j a_1 \sin \omega T}{1 + b_1 \cos \omega T - j b_1 \sin \omega T}$$

We take the absolute amplitude of the above expression to obtain its frequency response^{*}

$$A(f) = |H(e^{j\omega T})| \quad (11)$$

$$= \sqrt{\frac{a_0^2 + a_1^2 + 2a_0 a_1 \cos \omega T}{1 + b_1^2 + 2 b_1 \cos \omega T}}$$

At this point, the spectral region that is valid for our development can be evaluated numerically. For the sake of our example, we establish the following set of specifications based on the prototype filter of Fig. 5.

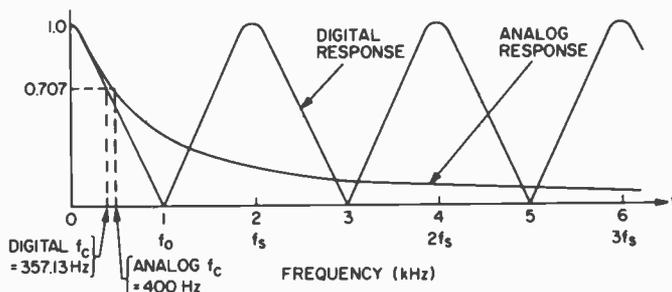


Fig. 7 Comparison of frequency responses. Note that the steeper fall-off of the digital filter response has resulted in a lowering of its cutoff frequency from 400 Hz to 357.13 Hz.

Cutoff frequency for analog filter, $f_c = 400$ Hz

Sampling frequency, $f_s = 2000$ Hz

Folding frequency, $\frac{1}{2} f_s = 1000$ Hz

Sampling interval, $T = 1/f_s = 0.0005$ s

Results are plotted on Fig. 7 for both the analog and digital responses. The analog response tails off monotonically to zero with increasing frequency. The digital response, however, is periodic and repeats at multiples of the sampling frequency, f_s . For a region in the neighborhood of dc, however, there is a fair agreement between these two responses. This region is shown on an expanded scale on Fig. 8. We may note the following:

- 1) Frequency responses converge to unity at dc.
- 2) The digital filter exhibits a cutoff frequency of only 357.13 Hz, although the cutoff frequency of the analog filter from which it was derived was 400 Hz, which may be a significant deviation.
- 3) For comparison, the frequency response of a similar analog filter with a cutoff frequency of 357.13 Hz (instead of 400 Hz) is also given on Fig. 8. In this case, although frequency responses match exactly at dc and cutoff, the match at other points in the spectrum may not be as good.

The impulse responses of our filters are shown in Fig. 9; the step-function responses, in Fig. 10. For the analog filter, these responses were obtained in the usual way using Laplace transform techniques. The digital filter responses were obtained by straightforward use of the difference equation [Eq. 10], where the values of the constants were computed as follows:

*There is a problem in semantics here since the quantity $A(f)$ is usually called the amplitude response to distinguish it from the phase response of the transfer function.

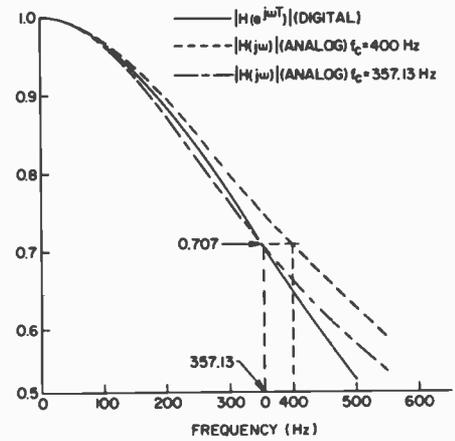


Fig. 8 Expanded comparison of frequency responses of analog and digital filters.

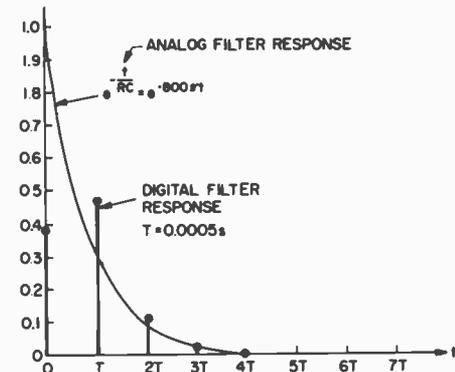


Fig. 9 Comparison of impulse responses of analog and digital filters designed for a frequency response match.

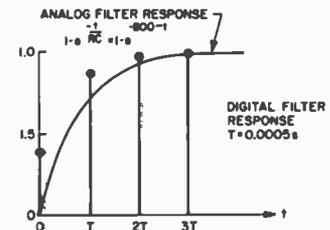


Fig. 10 Comparison of step-function responses of analog and digital filters designed for a frequency response match.

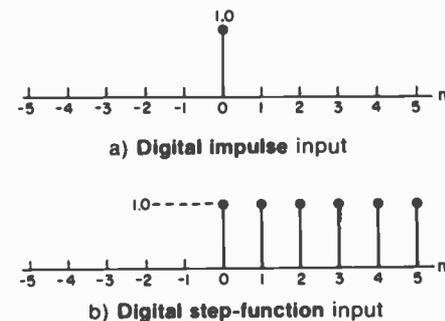


Fig. 11 Standard digital inputs. These types of inputs were used to obtain the results shown in Figs. 9 and 10.

$$RC = 1/\omega_c = 1/[2\pi(400)]$$

$$a_0 = a_1 = \frac{T}{T + 2RC}$$

$$= \frac{0.0005}{0.0005 + 2/(800\pi)}$$

$$= 0.386$$

and

$$b_1 = \frac{T - 2RC}{T + 2RC} = -0.228$$

The impulse and step function digital inputs were used as defined in Fig. 11.

Figs. 9 and 10 show substantial differences between the prototype analog filter responses for the impulse and step-function inputs as compared to the digital filter responses for the same inputs. Apparently, the bilinear transformation can give only very approximate results when we seek to design a filter to obtain a given impulse or step function response. We must seek another method.

Impulse-invariance approach to digital filter design

In this section we explore the design of an IIR digital filter to give the same impulse response as that of a given analog filter. To make the filtering problem more interesting, we have chosen a second-order Butterworth analog filter having a 3-dB cutoff frequency of 50 Hz. The sampling rate f_s is 500 Hz. In this case, the design procedure is somewhat different from that used to obtain the digital filter for a frequency response match. We proceed as follows:

- 1) Obtain the transfer function for the prototype analog filter.
- 2) Obtain the *impulse* response of the prototype analog filter. This is conveniently done by applying the impulse response to the transfer function (using the Laplace transform) and then converting back to the time domain.
- 3) Establish the desired digital response by replacing the variable in the (continuous) analog response (t) by its digital equivalent (nT).
- 4) Apply the z-transform to the desired digital response to obtain this response in the z domain.

5) Using $Y(z) = H(z)X(z)$, obtain the desired digital filter transfer function by solving for $H(z)$, e.g., $H(z) = Y(z)/X(z)$. In the present case, the computation is simplified by the fact that the z-transform of the input impulse, $X(z)$, is unity.

1) Establish the transfer function of the prototype analog filter.

The basic form for the "amplitude-squared" function of the Butterworth maximally flat filter is^f

$$A^2(f) = \frac{1}{1 + (\omega/\omega_c)^{2k}} \quad (12)$$

$$= \frac{1}{1 + (\omega/\omega_c)^4}$$

where k is the order of the filter and ω_c is the radial cutoff frequency = 100π rad/s. We immediately obtain the amplitude frequency response (plotted in Fig. 12). However, to obtain the impulse response we first must obtain the analog transfer function in the Laplace "s domain" formulation. We begin by normalizing our radial frequency ω to the cutoff frequency ω_c .

$$A^2(f) = \frac{1}{1 + \omega^4} \quad (13)$$

and then replace ω^2 by $-s^2$. Since s is complex, we write^g

^fSee Stanley⁷ for a good account of the amplitude-squared function. This is the same amplitude function $A(f)$ that we called the frequency response in the previous section of this article.

^gSince s is a complex number and the entire right-hand side of Eq. 14 represents the square of a complex quantity, the left-hand side necessarily becomes the product of this quantity and its complex conjugate, e.g., $H(s)$ and $H(-s)$.

$$A^2(f) \rightarrow H(s)H(-s) = \frac{1}{1 + s^4} \quad (14)$$

We obtain the poles of Eq. 14 using the conventional method of obtaining the fourth root of a complex number.^h

$$s^4 = -1 = \exp [j(180^\circ + q360^\circ)]$$

$$s = \exp [j(180^\circ + q360^\circ)/4]$$

$$= s_1, s_2, s_3, s_4 \quad (15)$$

(for $q = 0, 1, 2, 3$).

This operation leads to poles at the following four locations:

$$s_1 = \frac{1}{\sqrt{2}} + j \frac{1}{\sqrt{2}}$$

$$s_2 = -\frac{1}{\sqrt{2}} + j \frac{1}{\sqrt{2}}$$

$$s_3 = -\frac{1}{\sqrt{2}} - j \frac{1}{\sqrt{2}}$$

$$s_4 = \frac{1}{\sqrt{2}} - j \frac{1}{\sqrt{2}}$$

To obtain a stable transfer function, we associate only the poles in the left half-plane with $H(s)$ in Eq. 14.ⁱ This gives us

$$G(s) = \frac{1}{(s - s_2)(s - s_3)} \quad (16)$$

$$= \frac{1}{s^2 + \sqrt{2}s + 1}$$

We revert to our unnormalized form of the transfer function with the simple transformation

^hSee Churchill, R.V.; *Complex variables and applications*; 2nd Ed. (McGraw-Hill Book Co.; New York; 1960) p. 14.

ⁱAn excellent account of these manipulations is given in chapter 6 of Ref. 9.

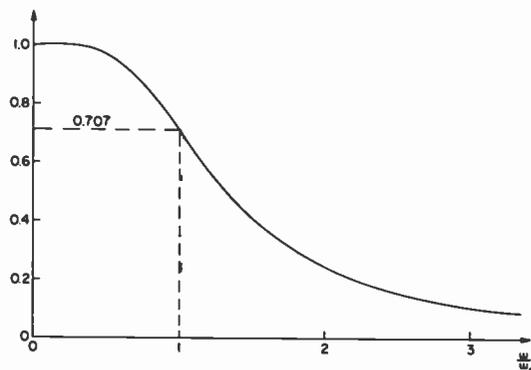


Fig. 12 Amplitude frequency response, $A(f)$, of second-order Butterworth analog filter.

$$s \rightarrow \frac{s}{\omega_c} = \frac{s}{100\pi} \quad (17)$$

which leads to

$$G(s) = \frac{10^4 \pi^2}{s^2 + 100\pi \sqrt{2} s + 10^4 \pi^2} \quad (18)$$

2) Obtain the impulse response of the (analog) transfer function.

Since the Laplace transform of the input function, $X(s)$, is unity (see Table 1), we have only to obtain the inverse transform of our transfer function of Eq. 18 to obtain our impulse response in the time domain, e.g.,

$$\begin{aligned} Y(s) &= G(s)X(s) \\ &= G(s) [1] \\ &= G(s) \end{aligned}$$

We begin by completing the square in the denominator

$$Y(s) = \frac{10^4 \pi^2}{(s + 50\pi \sqrt{2})^2 + 5000\pi^2}$$

and then make use of the Laplace transform pair:

$$\frac{1}{(s + \alpha)^2 + \beta^2} \longleftrightarrow \frac{e^{-\alpha t} \sin \beta t}{\beta}$$

and obtain our expression for the output, $y(t)$,

$$y(t) = 100 \sqrt{2} \pi \exp(-50 \sqrt{2} \pi t) \sin(50 \sqrt{2} \pi t) \quad (19)$$

This impulse response for our prototype analog filter is shown in Fig. 13.

3) Establish the actual impulse response of the digital lowpass filter.

We replace the argument, t , in Eq. 19 with nT .

$$y(nT) = 100 \sqrt{2} \pi \exp(-50 \sqrt{2} \pi nT) \sin(50 \sqrt{2} \pi nT) \quad (20)$$

This desired response is also shown on Fig. 13 (for pulses occurring at moments nT).

4) Transform the digital impulse response to the z domain.

We do this by applying the z -transform to Eq. 20, making use of the z -transform pair (see Table 1):

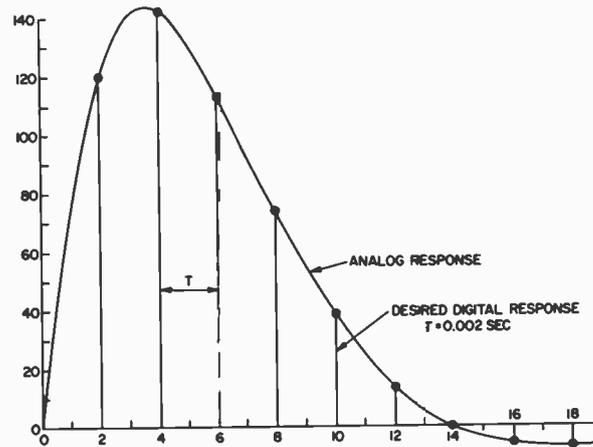


Fig. 13 Impulse response of second-order Butterworth filter. This analog filter response information will help us establish the actual impulse response of the digital lowpass filter.

$$\sin naT \longleftrightarrow \frac{z \sin aT}{z^2 - 2z \cos aT + 1}$$

which evaluates to

$$Z[\sin(222.14nT)] = \frac{0.4298 z}{z^2 - 1.8058 z + 1} \quad (21)$$

Our actual expression requiring transformation to the z domain (Eq. 20), however, includes an exponential factor multiplying the sine function. Fortunately, there is a z -transform which takes care of this contingency

$$e^{-naT} x(n) \longleftrightarrow X(e^{aT} z) \quad (22)$$

Hence, we need only alter the argument, z , of Eqs. 20 and 21. Direct evaluation of our modified argument of Eq. 22 gives

$$z \rightarrow 1.5594z$$

With some manipulation, we may now write

$$\begin{aligned} Z[\exp(-222.14nT) \sin(222.14nT)] \\ = \frac{0.2756 z^{-1}}{1 - 1.1580z^{-1} + 0.4113z^{-2}} \quad (23) \end{aligned}$$

We have so far ignored the coefficient $100 \sqrt{2} \pi$ in Eqs. 19 and 20. To check the impulse response of our digital filter against the impulse of the prototype analog filter (Fig. 13), it may be convenient to carry this multiplying constant along without change. However, it is more usual to multiply our z -transform transfer function by the sampling interval T at this stage

in order to obtain a match of frequency response (rather than impulse response) between digital and analog versions. We will not burden the reader with the rigorous justification for this step. However, it may be inferred from a study of Appendix A, Eqs. A9 and A10.

Our final expression for the digital filter, therefore, is obtained by multiplying the result of Eq. 23 by $100 \sqrt{2} \pi (0.002)$. The result is

$$Y(z) = \frac{0.2449 z^{-1}}{1 - 1.1580z^{-1} + 0.4113z^{-2}} \quad (24)$$

Table 1 Z-Transform of various function pairs and operation pairs used in this paper.

Function pairs	$x(n)$	$X(z)$	$X(s)$ (Laplace transform)
	$\delta(n)$	1.0	1
	Step function [1.0 or $u(n)$]	$\frac{z}{z-1}$	$\frac{1}{s}$
	nT	$\frac{Tz}{(z-1)^2}$	$\frac{1}{s^2}$
	$\sin naT$	$\frac{z \sin aT}{z^2 - 2z \cos aT + 1}$	$\frac{a}{s^2 + a^2}$
	$\cos naT$	$\frac{z^2 - z \cos aT}{z^2 - 2z \cos aT + 1}$	$\frac{s}{s^2 + a^2}$
Operations pairs:			
	$x(n) \rightarrow X(z)$		
	$e^{-naT} x(n) \rightarrow X(e^{aT} z)$		

5) Obtain the desired digital-filter transfer function.

It may be noted that, since the z -transform of the delta function is unity, the output of the system $Y(z)$ is equal to the z -transform of the transfer function itself, e.g.

$$Y(z) = H(z) X(z) = H(z)$$

We have therefore obtained our desired digital filter transfer function in Eq. 24.

Our result is easily checked

The difference equation can be reconstructed from the transfer function

$$H(z) = \frac{Y(z)}{X(z)} = \frac{0.2449z^{-1}}{1 - 1.580z^{-1} + 0.4113z^{-2}}$$

We obtain

$$Y(z) - 1.1580z^{-1}Y(z) + 0.4113z^{-2}Y(z) = 0.2449z^{-1}X(z)$$

Applying the inverse z -transform to the above (see Appendix A, Eq. A16) we obtain the difference equation

$$y(n) - 1.1580y(n-1) + 0.4113y(n-2) = 0.2449x(n-1)$$

The impulse response of the digital filter is obtained by successive solutions of the difference equation, beginning with the digital-impulse input of Fig. 11, at $n = 0$. After dividing by $T (= 0.002)$ we find that, within round-off error, the match is exact (see Table II).

Step-invariance method

The design of a digital filter having the same step-function response as that of a given analog filter follows closely the method used above for the impulse-

invariance approach. The steps are

1) Obtain the transfer function for the prototype analog filter.

2) Obtain the step-function response of the prototype analog filter. This is conveniently done utilizing Laplace transforms.

3) Establish the desired digital response by replacing the time variable in the continuous analog response (t), by its digital equivalent (nT).

4) Apply the z -transform to the desired digital response to obtain this response in the z domain.

5) Using the relationship $Y(z) = H(z)X(z)$, obtain the desired digital transfer function by solving for $H(z)$, e.g. $H(z) = Y(z)/X(z)$, where $X(z) = z/(z-1)$ (Appendix A).

Space does not permit detailing the various steps in the development of the step-function-invariant digital transfer function for the second-order Butterworth filter used in the impulse-invariant example above. The result, however, is

$$H(z) = \frac{0.1454z^{-1} + 0.1078z^{-2}}{1 - 1.1580z^{-1} + 0.4112z^{-2}}$$

and the associated difference equation is

$$y(n) - 1.1580y(n-1) + 0.4112y(n-2) = 0.1454x(n-1) + 0.1078x(n-2)$$

We can check the results.

The digital step-function response may be calculated directly by repeated use of the above difference equation starting with the digital step-function input of Fig. 12. We find that the correspondence between the two filters is exact.

It becomes interesting to compare the frequency responses of our various digital filters based on the same second-order

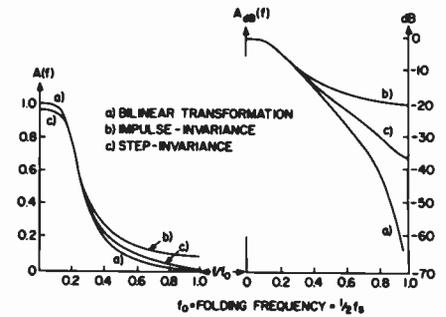


Fig. 14 Comparison of frequency responses of digital filters designed from the same prototype analog (second-order Butterworth) filter.

Butterworth analog filter when designed either by the bilinear transformation for a matched frequency response, or by the above methods for a matched impulse response or a matched step-function response. These responses, on both a linear and logarithmic scale, are shown on Fig. 14 where it is evident that there are wide differences in each case, particularly in the cutoff region.

Concluding remarks

This article treats the design of digital filters based upon the action—or transfer function—of a given analog filter. However, this approach (although widely used) represents only a very small part of the tremendous flexibility and capability of a computer-based digital filter. For example, it is a simple matter to take the (discrete) Fourier transform of a time series and obtain a frequency series. In either case, the information is in series form and, insofar as the filter (e.g., the digital computer) is concerned, it does not (and, indeed, cannot) distinguish between the two. It blithely goes ahead and modifies the various terms in the series in accordance with whatever program has been entered. Hence, we have the alternative of operating in either one or the other of the time and frequency domains.

Table II Impulse-invariance-designed digital filter. Comparison of analog and digital amplitudes of impulse response. The response match is exact within roundoff errors.

$n =$	0	1	2	3	4	5	6	7	8	9	10
Analog	0	122.46	141.82	113.87	73.55	38.34	14.16	0.63	-5.10	-6.16	-5.04
Dig. $\times 500$	0	122.45	141.8	113.8	73.5	38.3	14.1	0.6	-5.1	-6.15	-5.0

The possibilities here are breath-taking and will be treated further in the next article. In effect—within certain limits—we will be able to write our own bill as to the nature of the filtering action we desire, quite independently of what has traditionally been considered possible for conventional analog filters.

A future article will deal with the Fast Fourier Transform (FFT) which, by virtue of its speed, has made possible many of the totally unexpected signal-processing approaches that have appeared in the digital domain.

References

The following list represents only a small portion of the available literature in this field. However, it contains material which, in the opinion of the writer, would be particularly suited for readers planning to become further acquainted with the design of digital filters.

Laplace transform: CEE courses offer a good introduction.

We highly recommend the CEE course M52, Fourier Analysis and the Laplace and

Z-transforms, for an entrance into this field. Alternatively, CEE course M2, Engineering Mathematics II, treats the Laplace transform. Readers wishing to begin with the general subject of Fourier analysis would be advised to arrange for the CEE course M1, Engineering Mathematics I, before registering for M2.

For an excellent introduction:

1. Skilling, H.H.; *Electrical engineering circuits*, Second Edition (John Wiley & Sons, Inc.; 1965).

For further work on a slightly higher level:

2. Holbrook, J.G.; *Laplace transforms for electronic engineers*, Second (Revised) Edition (Pergamon Press; 1966).
3. Cheng, D.K.; *Analysis of linear systems*, (Addison-Wesley; 1959).

The z-transform—Appendix A summarizes the subject.

4. "Fourier analysis and the Laplace and z-transforms," CEE course M52.

For a first introduction:

5. Cheng, D.K.; *Analysis of linear systems*, (Addison-Wesley; 1959).
6. Cadzow, J.A., *Discrete-time systems, an introduction with interdisciplinary applications*, (Prentice-Hall, Inc.; 1973).

Advanced treatment:

7. Doetsch, G.; *Guide to the applications of the Laplace and z-transforms*, (Van Nostrand Reinhold Co.; 1971).
8. Jury, E.I.; *Theory and application of the z-transform method*, (Robert E. Krieger Publishing Co.; 1964).

Direct design of digital filters. Start with Stanley.

9. Stanley, W.D.; *Digital signal processing* (Reston Publishing Co., Inc.; Prentice-Hall co.; 1975).

On a somewhat higher level:

10. Oppenheim, A.V.; and Schaefer, R.W.; *Digital signal processing* (Prentice-Hall, Inc.; 1975).
11. Rabiner, L.R.; and Gold, B.; *Theory and application of digital signal processing*, (Prentice-Hall, Inc.; 1975).

Series manipulations.

The following paperbacks represent a good accumulation of material on power series.

12. Knopp, K.; *Infinite sequences and series*, (Dover Publications, Inc.; 1956).
13. Jolley, L.B.W.; *Summation of series*, (Dover Publications, Inc.; 1961).
14. Moore, C.N.; *Summable series and convergence factors*, (Dover Publications, Inc.; 1966).

The tutorial treatment of series in the standard mathematical texts seems to have fallen into disrepute. The following texts, however, do have chapters on the treatment and analysis of series:

15. Boas, M.L.; *Mathematical methods in the physical sciences*, (John Wiley & Sons, Inc.; 1966).
16. Pipes, L.A.; *Applied mathematics for engineers and physicists*, Second Edition (McGraw-Hill Book Co.; 1958).
17. Sokolnikoff, I.S.; and Sokolnikoff, E.S.; *Higher mathematics for engineers and physicists*, Second Edition (McGraw-Hill Book Co., Inc.; 1941).
18. Reddick, H.W.; and Miller, F.H.; *Advanced mathematics*, Third Edition (John Wiley & Sons, Inc.; 1955).

Glossary—a collection of terms from parts 1, 2, and 3 of this series.

Aliasing: A phenomenon arising as a result of the sampling process in which high frequency components of the original analog signal (whether information or noise) appear as lower frequencies in the sampled signal. Aliasing occurs when the sampling rate is less than twice the highest frequency existing in the original analog signal.

Amplitude function: As used in filter design, the amplitude function represents the absolute value of the frequency response. A filter is sometimes specified by the expression for the square of this function, e.g., $A^2(f)$.

Analog/digital converter: A circuit which samples an analog signal at specified periods of time to produce a discrete signal which is then quantized.

Analog signal: A signal that is continuous in both time and amplitude.

Bilinear transformation: A transformation in complex variable theory which is linear in each of the variables separately (the other one being held fixed). The usual variables in our work are s and z . Geometrically speaking, a bilinear transformation has the property that it transforms lines into lines and circles into circles; see, for example, *Complex Variables and Applications* by R.V. Churchill (McGraw-Hill, 1960).

Binary code: A language in which each symbol (or pulse) has only one of two possible meanings

(or levels). Each symbol (or pulse) then represents one bit.

Binary signal: A digital signal with only two available amplitudes or levels, variously called on/off, one/zero, or high/low. A binary signal may be "positive" in the sense that the "one" level may be a positive voltage, or it may be "negative" in the sense that the "one" level may be a negative voltage. In either case the "zero" level is ground. A binary signal may also be "bipolar", in which case the "one" is usually a positive voltage while the "zero" is a negative voltage of the same amplitude.

Bipolar pulse: A two-level pulse with both levels being equal in magnitude but opposite in polarity. The "one" is usually assigned to the positive level and the "zero" to the negative level.

Bit: A unit of information corresponding to the selection of one of two equally likely possible alternatives. It appears when the logarithm inherent in the definition of information (H), below, is taken to the base 2.

$$H = (T/\tau) \log_2 n$$

where T is message duration (seconds); τ is the width of slot assigned to each pulse (seconds); and n is the number of information-bearing levels in each pulse.

Capacity: The capacity (C) of a system is the maximum number of bits it can process, or transmit, per second, e.g., $C = H/T$ where T is time required for the processing or transmission

and H is information content (in units of bits). **Componder.** A device consisting of a compressor at the transmitting end and an expander at the receiving end which operate as nonlinear amplifiers to obtain a more advantageous amplitude-quantizing relationship for the reduction of noise. The process of companding is particularly important with audio signals. The principles have also been effectively used in the processing of picture information.

Continuous amplitude signal: A signal that is able to assume any amplitude value, usually between certain prescribed limits.

Continuous time signal: A signal that is defined for all values of time, usually between certain prescribed limits.

Dirac delta function: A function defined by the following relationships:

$$\delta(t) = \begin{cases} 0 & , \text{ for } t \neq 0 \\ \text{arbitrarily large} & , \text{ for } t = 0 \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(t) dt = 1.0$$

As a consequence,

$$\int_{-\infty}^{\infty} f(t) \delta(t) dt = f(0)$$

The delta function is also called the impulse function.

Digital/analog converter: A circuit which transforms a digital signal into an analog signal, usually by some type of filtering action.

Digital filter: A discrete system (as defined below) which usually operates to remove noise or separate out a particular frequency band for separate processing.

Digital signal: A discrete signal in which the available amplitude values constitute a discrete series, each member of which can be represented by a number having a finite number of digits. The terms digital and discrete are sometimes loosely used interchangeably.

Discrete signal: A signal defined only at a particular set of time values. Between these values the amplitude may be zero or have an amplitude of no additional information value.

Discrete system: A system which operates on an input discrete signal $x(n)$, to generate another discrete signal $y(n)$ according to some well-defined transfer function.

Encoding: The process of translating a message from one language to another; e.g., we may translate a message of M possible meanings conveyed by a signal pulse of M possible levels, into two pulses such that these M possible meanings may be represented by various combinations of these two pulses. That is, $M = m^2$. Here, m represents the required number of levels in each of the two pulses involved in the translation. The value of m is selected as required to establish the equality.

Finite-impulse-response filter: A filter in which the response to an impulse input reaches exactly zero after a finite period of time.

Fourier methods: See the first article of this series including the Appendix

Frequency-division multiplexing (FDM): A system in which a number of separate channels are assigned to different frequency bands within an overall channel bandwidth. This is conveniently done by appropriate choice of carrier frequencies. The individual channels are then separated out at the receiving end with filtering techniques.

Frequency domain: A graphical way of representing signals in which the horizontal axis is calibrated in units of frequency. Alternatively

it may be applied to a mathematical representation of signals in which the variable is in units of frequency.

Holding circuit: A circuit used at the receiving terminal in time division multiplexing which lengthens the individual pulses (after separation and routing to their respective channels) to the full sampling time interval in order to aid in the demodulation process.

Impulse function response (impulse response): The response of a system to an impulse input.

Impulse-invariance: The condition in which two systems have identical impulse responses. This is sometimes applied to a pair of systems in which one is continuous and the other discrete. The response match is then taken only at the sampling points of the discrete system.

Infinite-impulse-response (IIR) filter: A filter in which the response to an impulse input never quite reaches zero after a finite period of time.

Impulse function: See delta function.

Information content: The information content of a message \mathcal{H} is defined in terms of the probability of the occurrence of the event in question, e.g., $\mathcal{H} = \log(1/p_k)$ where p_k is the probability for the occurrence of the particular event. For the case where M events are equally likely, the probability for the occurrence of any one of these events is $1/M$ and we obtain $\mathcal{H} = \log M$. If M has been encoded in the form of n^m , we can write $\mathcal{H} = m \log n$. For the commonly occurring case where the logarithm is taken to the base 2, we obtain the information \mathcal{H} in bits.

Linear system: One in which the behavior of the system is not dependent upon the amplitude of the input signal, or upon the simultaneous presence of other signals.

Linear time-invariant system (LTI system): A system in which the transfer function does not change during the processing of the pertinent information.

Negative pulse: A two-level pulse in which the "one" level is a negative voltage and the "zero" level is at ground potential.

Normalized power: The power that would be developed by the signal across a one-ohm resistor.

Positive pulse: A two-level pulse in which the "one" level is a positive voltage and the "zero" level is at ground potential.

Quantization: The process of constraining the values of a signal, whether continuous or discrete, to assume one or another of a discrete set of values. By quantizing a discrete signal we obtain a digital signal.

Quantization noise: A type of noise inherent in the quantizing process due to the difference in amplitude between the original analog signal and the quantized signal. This implies an uncertainty regarding the original information, which is then construed as quantization noise.

Single-polarity pulse: A two-level pulse in which one of the levels is at ground potential. The ground potential level is normally designated as the "zero" level.

Step-function response (step response): The response of a system to a step-function input.

Step-invariance: The condition in which two systems have identical step-function responses. This is sometimes applied to a pair of systems in which one is continuous and the other discrete. The response match is then taken only at the sampling points of the discrete system.

Time-division multiplexing (TDM): A system in which the sample pulses are very narrow as compared to the sampling time. It is hence feasible to insert a number of such pulses, corresponding to different information channels, into the time period between the successive samples of any one channel. We hence obtain a sequence of pulses during each sampling period. The pulses are routed to their respective channels at the receiving end.

Time domain: A graphical way of representing signals in which the horizontal axis is calibrated in units of time; alternatively, it may be applied to a mathematical representation of signals in which the variable is in units of time.

Transfer function: The quantitatively expressed ratio of the response at one point in a system to an input at another point in the system.

Z-transform: A modification of the Fourier transform for use with digital signals in which the Laplace variable s is replaced by $z = e^{sT}$. The meaning and utilization of this transform is developed in appendix A.

Appendix A: The z-transform

We continue to use the notation of our first article so that the sampled function is indicated by an asterisk. Thus, given a train of delta functions (or impulses), we represent the sampling process of our original continuous analog time function, $x(t)$, as

$$x^*(t) = x(t) \sum_{n=0}^{\infty} \delta(t - nT) \quad (\text{A1})$$

$$x^*(t) = \sum_{n=0}^{\infty} x(nT) \delta(t - nT) \quad (\text{A2})$$

Eqs. A1 and A2 represent the same information except that Eq. A2 is in more convenient form for further manipulations. It might be argued that $x(nT)$ is exactly equal to $x^*(t)$. This may be true, but we can also argue that multiplication by the delta-function factor insures that we wipe out any signal occurring between integer multiples of the sampling interval T

and makes for a more precisely defined pulse train.

We now take the Laplace transform of Eq. A2

$$X^*(s) = \sum_{n=0}^{\infty} x(nT) e^{-nsT} \quad (\text{A3})$$

Our result on the right-hand side is obtained by noting that our variable is t , and hence, only the delta function need be transformed. The transform of the undelayed delta function, $\delta(t)$, is unity. The second term in the argument of $\delta(t - nT)$ is nT and represents a delaying operation. In Laplace transform language, this delay is given by the exponential factor in Eq. A3.^a

^aAn excellent source for review of the Laplace transform as well as a comprehensive table of Laplace transform pairs is *Laplace Transforms for Electronic Engineers* by James G. Holbrook (Macmillan Co., N.Y., 1959).

We now make the substitution

$$z = e^{sT} \quad (\text{A4})$$

Eq. A4 defines the relationship between the Laplace variable s and the z -transform variable z . With this substitution, Eq. A3 becomes

$$X(z) = \sum_{n=0}^{\infty} x(nT)z^{-n} \quad (\text{A5})$$

Eq. A5 defines the z -transform of the time series $x(nT)$. We no longer need the asterisk for $X(z)$ since the variable z , by itself, indicates that we are dealing with a time series. Another extremely important property of the relationship in Eq. A5 is the fact that the original time series appears in the summation as an inherent part of the z -transform. Hence, at any time that we can write the z -transform of a function in the form of Eq. A5, e.g., as a summation in negative powers of z , the associated time series automatically appears in our result. A simple example may be helpful. For example, suppose that we are given the following z -transform

$$G(z) = Tz/(z-1)^2 \quad (\text{A6})$$

We can immediately extract the associated time series by expanding the denominator and performing a straightforward long division.

$$G(z) = Tz/(z^2 - 2z + 1) = Tz^{-1} + 2Tz^{-2} + 3Tz^{-3} + \dots \quad (\text{A7})$$

Our time series, then, is

$$T, 2T, 3T, \dots \quad (\text{A8})$$

This implies that we can rewrite our expression for $G(z)$ as

$$G(z) = \sum_{n=0}^{\infty} (nT) z^{-n} \quad (\text{A9})$$

A further important implication is that, in accordance with the definition of the z -transform in Eq. A5, we can say that the z -transform of the series corresponding to nT is the right-hand side of Eq. A6 or

$$Z\{nT\} = Tz/(z-1)^2 \quad (\text{A10})$$

This brings up the question of z -transform pairs for use in our work. There are, unfortunately, only a few sources available (to the knowledge of the writer) of z -transform tables. Probably the most complete has been assembled by Jury.⁶

Fortunately, use of the z -transform is quite analogous to that of the Laplace transform so that the same approach may be used in either case. We transform to the z representation as convenient during the course of the problem and then transform our final answer back to the time domain when we want to determine the appearance of our result along a time axis.

A few additional remarks about the z -transform may be in order. For example, since the z -transform represents the transform of a time series (which exists within the z -transform itself) it often becomes convenient to be able to move easily from a series representation to a closed-form representation of a function. Thus, readers planning to engage in such calculations may find it advantageous to become conversant with the standard forms for series expansions, such as the binomial theorem and other forms of the power series.⁷

⁶Application of the z -Transform Method by E.I. Jury (Krieger Publishing Co.; Huntington, N.Y.; 1964; reprinted with corrections 1973).

⁷See the references for source material on series expansions.

As another remark in the handling of the z -transform, due to the rather incomplete nature of available z -transform pairs, the writer has found it quite advantageous to make extensive use of the reduction of z -transform functions to a number of simply transformable terms using the partial-fraction approach exactly as in the Laplace case.⁸ One precaution should be noted, however. Practically every z -transform appearing in the tables of z -transform pairs contains z to at least the first power in the numerator. Therefore, in obtaining partial-fraction expansions of a function, the manipulation should be so arranged as to save at least one power of z for the numerator in each term. There are various devices for accomplishing this but space, unfortunately, does not permit us to expand upon these tricks of the trade. A good deal of interesting and well written information about the z -transform may be found in Cadzow.⁹

The z -transform in difference equations

We have found that the operation of a digital filter may, in general, be described by a difference equation of the type given in Eq. 1. Let us now apply the z -transform to the solution of this equation. We may desire the solution in either of two ways, i.e., with, or without, the capability of incorporating initial conditions. Let us consider, first, the case in which we wish to incorporate initial conditions into the solution. For this case we make use of the so-called "left-shifting" z -transform

$$f(n+m) \longleftrightarrow z^m F(z) - \sum_{i=0}^{m-1} f(i)z^{m-i} \quad (\text{A11})$$

When we go back only two time intervals from the reference moment, n , we need only the following three versions of the above transform pair

$$\begin{aligned} f(n) &\longleftrightarrow F(z) \\ f(n+1) &\longleftrightarrow zF(z) - zf(0) \\ f(n+2) &\longleftrightarrow z^2F(z) - z^2f(0) - zf(1) \end{aligned} \quad (\text{A12})$$

It is now convenient to rewrite our difference equation as follows:

$$y(n+2) + b_1y(n+1) + b_2y(n) = a_0x(n+2) + a_1x(n+1) + a_2x(n) \quad (\text{A13})$$

We transform term by term and obtain the following:

$$\begin{aligned} [z^2Y(z) - z^2y(0) - zy(1)] + b_1[zY(z) - zy(0)] \\ + b_2Y(z) = a_0[z^2X(z) - z^2x(0) - zx(1)] \\ + a_1[zX(z) - zx(0)] + a_2X(z) \end{aligned} \quad (\text{A14})$$

We solve for $Y(z)$ and write

$$Y(z) = \frac{a_0z^2 + a_1z + a_2}{z^2 + b_1z + b_2} X(z) + \frac{z^2[y(0) - a_0x(0)] + z[y(1) + b_1y(0) - x(1) - a_1x(0)]}{z^2 + b_1z + b_2} \quad (\text{A15})$$

The separation of terms in the solution for $Y(z)$ is quite significant. Thus, the coefficient of $X(z)$ is independent of initial conditions and constitutes the transfer function connecting the output to the input (initial conditions

⁸See CEE course M52, Session 5, for the Heaviside Expansion Theorem in z -transform notation. See Cheng for the Laplace version.

⁹Discrete Time Systems, James A. Cadzow (Prentice-Hall, 1973).

are not considered). The second term on the right represents the contribution of the initial conditions and modifies the output accordingly. Notice the direct analogy to the Laplace formulation where, in a similar situation, the transfer function appears separately from the initial conditions.

If we are interested in obtaining only the transfer function, things are somewhat simpler and we may make use of the "right-shifting" z-transform

$$\begin{aligned}
 f(n-m) &\longleftrightarrow z^{-m} F(z) \\
 \text{or} \\
 f(n) &\longrightarrow F(z) \\
 f(n-1) &\longrightarrow z^{-1} F(z) \\
 f(n-2) &\longrightarrow z^{-2} F(z)
 \end{aligned}
 \tag{A16}$$

We operate directly on our difference equation as it appears in Eq. 1 and obtain

$$\begin{aligned}
 Y(z) + b_1 z^{-1} Y(z) + b_2 z^{-2} Y(z) \\
 = a_0 X(z) + a_1 z^{-1} X(z) + a_2 z^{-2} X(z)
 \end{aligned}
 \tag{A17}$$

which gives us our transfer function with much less fuss

$$Y(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2}}{1 + b_1 z^{-1} + b_2 z^{-2}} X(z)
 \tag{A18}$$

The transfer functions as developed in Eqs. A15 and A18 are easily transformed from one form to the other by multiplication of numerator and denominator by the proper power of z . At this stage, however, it may be of interest to take a quick look at the transfer function and draw a rapid conclusion regarding its behavior over the frequency spectrum. Those readers familiar with the Laplace transform will recall that the frequency response of a function in the Laplace transform representation is obtained by letting the Laplace variable s go directly to its imaginary part $j\omega$, e.g.,

$$s = \sigma + j\omega
 \tag{A19}$$

The resultant expression is then evaluated for various values of ω in the spectral regions of interest. We do exactly the same thing with our transfer function in the z-transform representation. Now, however, s occurs in the exponent of Napierian e as follows:

$$s = \sigma + j\omega \longrightarrow j\omega
 \tag{A20}$$

We substitute the last expression on the right for z wherever it appears in our expression for the transfer function. We then examine the behavior of the transfer function, again, in the spectral regions of interest. Now, however, things are a bit different. Our basic quantity is an exponential function with an imaginary exponent. The result is that it is periodic and repeats every time the argument (e.g., the factor multiplying the imaginary quantity j) increases by a multiple of 2π , or, in effect,

$$\omega T = 2\pi p \quad p = 0, 1, 2, \dots
 \tag{A21}$$

The above relationship tells us that the frequency response of our transfer function repeats exactly for the various integer values of p . If we make use of the fact that our sampling interval, T , is the reciprocal of the sampling frequency, f_s , we easily arrive at the result that the frequency response of our transfer function repeats exactly at integer multiples of the sampling frequency, a result that we had previously discovered in the first article of this series.

Transfer function in the z-transform domain

As usual, we define the transfer function as the ratio of the output to the input of system, in which initial conditions have been separated out and do not appear,

$$H(z) = \frac{Y(z)}{X(z)}
 \tag{A22}$$

For our previous case, such as indicated in Eq. 1, our transfer function followed immediately from application of the z-transform to the difference equation giving the results of Eqs. A15 and A18. We may now write

$$H(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2}}{1 + b_1 z^{-1} + b_2 z^{-2}}$$

In the general case, where input and output relations involve pulses positioned over many points along the time axis, Eq. A23 may be generalized,

$$H(z) = \frac{\sum_{k=0}^M a_k z^{-k}}{\sum_{k=0}^N b_k z^{-k}}
 \tag{A24}$$

A final problem to chew on

The mathematical tools explored and used in this article have value far beyond the field of digital signal processing. In fact, they may be resorted to whenever information is available in discrete form. The following example should illustrate this point.

The national income of this country is calculated quarterly and consists of the sum of the following:

- 1) Consumer expenditures (purchase of consumer goods).
- 2) Induced private investments (leading directly or indirectly to the purchase of capital equipment for increasing production).
- 3) Government expenditures.

We let $y(k)$ = National Income
 $i(k)$ = Induced private investment
 $c(k)$ = Consumer expenditures
 $u(k)$ = Government expenditures

Paul A. Samuelson, a noted economist,* postulated the following in 1939:

- 1) Consumer expenditures in the k -th accounting quarter are proportional to the national income in the preceding $[(k - 1)$ th]; quarter, e.g., $c(k) = a y(k - 1)$.
- 2) Induced private investment in any accounting period is proportional to the increase in consumer expenditure of that period over the preceding period, e.g.,

$$i(k) = b [c(k) - c(k - 1)]$$

- 3) Government expenditure is the same for all accounting periods.

Problem

Show that we may express the national income for the k -th quarter as

$$y(k) = u(k) + a(1 + b)y(k - 1) - aby(k - 2)$$

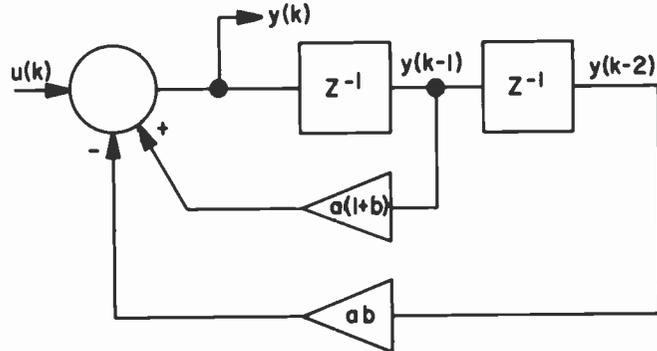
Answers to review questions from Part 2 of this series

- 1) The inherent properties of the human ear which favor the companding of voice signals are:

- The relationship between the loudness sensation experienced by the human ear and the actual physical sound intensity in decibels (e.g., logarithmically) is very nearly linear, see next page. [from *Acoustical Engineering*, Harry F. Olson (Van Nostrand, 1957)].
- Normal speech makes much more use of soft, rather than loud sounds. An approximate relationship based on

Problem

Show that the national income of this country can be modeled as shown below.



where z^{-1} represents a delay operator of one-quarter of a year (e.g., $T = 3$ months).

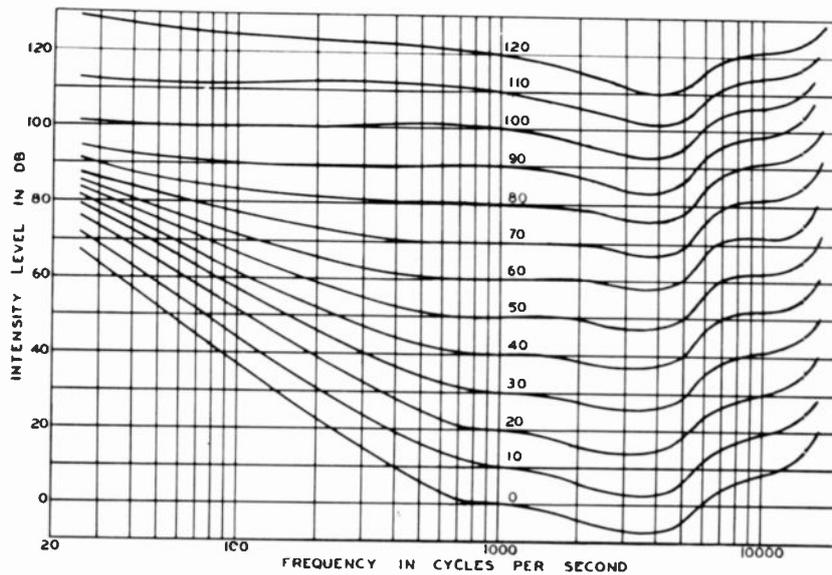
Although this model of our economy is admittedly drastically oversimplified, there is yet enough substance in it to raise some valid points. The interested reader, for example, may experiment with various values for the constants (assuming reasonable initial conditions). He may be surprised at the government expenditure required to keep the economy on an even keel.

*Samuelson, P.A.: "Interactions Between the Multiplier Analysis and the Principle of Acceleration," *Review of Economic Statistics*, 21 (1939), pp. 75-78.

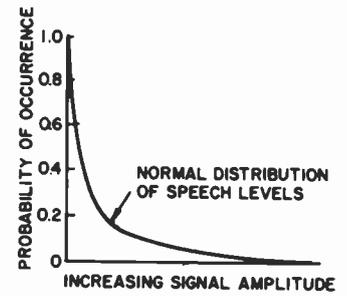
Ed. note: The answers to these questions will be given in the next issue. If you answer them yourself and don't want to wait, write to Dr. L. Shapiro, Continuing Engineering Education, RCA Bldg. 204-2, Camden, N.J. 08101 for the solutions.

a linear amplitude scale is shown on the next page [from *Communication Systems Engineering Handbook*, Donald H. Hamsher, Editor, (McGraw-Hill, 1967) p. 10-12].

Hence, by use of companding we perform the dual functions of favoring the weaker speech sounds to which the ear is so much more sensitive while at the same time favoring the (same) low levels which occur much more frequently in normal human speech.



Loudness sensation of the human ear related to sound intensity.



Relationship of soft to loud sounds in normal human speech.

2) Cross-talk occurs when transmission lines couple their signals into each other causing the spurious appearance of voice signals on lines that should not be transmitting such signals. By the process of companding we raise the amplitude level of the weaker noise signals (at the expense of the reduced gain for the stronger signals) thus making these (originally) weaker voice signals relatively immune to the risk of being swamped by spurious competing cross-talk signals. [For further discussion see *Communication Systems and Techniques*, Schwartz, Bennett, and Stein (McGraw-Hill, 1966)].

Every man should get married. If he marries a good woman, he will be very happy. If he marries a bad woman, he will become a philosopher...

(This statement by Socrates was probably motivated by a concern about the low marriage rate in Athens of the fifth century B.C.)

We will skip the obvious male chauvinistic implications of the above statement and consider only the two alternatives, e.g., the probability of the male member of the union becoming a very happy individual or else becoming a philosopher. Almost the only objective data available is the divorce rate versus the marriage rate. In recent years, with the easing of the divorce laws, these two rates have approached each other, with the divorce rate, in the case of a certain west coast state, even exceeding the marriage rate. We may then say, very approximately, that with an equal probability for the occurrence of either of the two alternatives, the informational aspects of the two alternatives is close to maximum since we have a maximum uncertainty regarding the outcome of a union between any two individuals of opposite sex.

*My candle burns at both ends;
It will not last the night;
But ah my foes, and, oh my friends—
It gives a lovely light.*

Edna St. Vincent Millay seems to be quite happy with her approach to the consumption of candle wax. The writer, however, doubts that in general such excessive consumption would actually produce the "lovely light" that so entranced our poetess. If we accept this premise we would then say that the above poem does indeed contain a great deal of information, since it states that, despite the improbability for the production of a "lovely light" in the case of excessive candle wax consumption (by the average individual) such "lovely light" did, in fact occur for Edna—or at least, so she wrote.

Operation of the RCA Frequency Bureau

R.E. Simonds
N.B. Mills
J.F. Eagan

This RCA service agency assists divisions of the Corporation with equipment approval, station licenses, frequency allocation, and other matters involving contact with the FCC, governmental agencies, and international organizations.

During the forty years since its establishment, the basic functions of the RCA Frequency Bureau have remained much the same, although its scope has increased substantially. Modes of radiocommunication have increased from simple telegraphy and telephony to complex satellite multi-channel digital pulse transmission. The useful radio spectrum has expanded from 30 MHz to above 30 GHz, and the number of users has grown materially. At the present time, the FCC has authorized the use of over 5 million radio stations, so the administrative and technical problems associated with maintaining orderly use of the radio spectrum have necessarily become more intricate. This is reflected in an increase in volume and complexity of national and international rules and technical standards governing the use of the radio spectrum.

Radio station licenses

Most RCA licenses are obtained from the Federal Communications Commission, but from time to time operating authorities from other Government agencies have been acquired for work in connection with Government contracts. In addition, RCA maintains ship licenses issued by Liberia and Panama. The range of licenses processed through the Bureau includes such varied operations as Space and Earth stations and associated terrestrial microwave facilities, Ship stations, Coast stations, fixed HF, VHF and microwave Point-to-Point, AM, FM and TV Broadcast stations, including their auxiliary Broadcast stations, Rural Radio and Domestic Public Radio, Aircraft, Business, Manufacturers', and Experimental Radio stations.

Radio station licensing and operation must be in accordance with the FCC Rules and Regulations and the Regulations of international treaties to which the United States is a signatory. As a result, the Frequency Bureau must concern itself with the details of both domestic and international rules and regulations that govern the variety of services in which RCA operates stations or sells equipment.

At the present time the Frequency Bureau is responsible for maintaining records on well over 1300 licenses issued



Norman B. Mills has been associated with the Frequency Bureau since 1941, dealing with radio station licenses as well as use of frequencies throughout the radio spectrum. He is presently Manager of the New York office of the Bureau and also serves on the Executive Committee of the Radio Technical Commission for Marine Services (RTCM) and is a member of the National Industries Advisory Committee (NIAC).

Contact him at:
RCA Frequency Bureau
50 Broad St.
New York, N.Y.
Ext. 5003



John F. Eagan, Jr. had completed thirty-five years of service with the RCA Frequency Bureau when he died this April. He had been Manager of the Cherry Hill Frequency Bureau since 1953. He represented RCA on several committees, including the International Electrotechnical Commission and the Radio Technical Commission for Aeronautics.

Raymond E. Simonds' biography and photo appear with his other article in this issue.

by the FCC and 490 ship licenses issued by foreign governments. License information must be reviewed and updated from time to time to conform to various changes in the requirements and to accommodate advances in communications technology that were not available when the original application was submitted. As a result, there is a continuing need to file information with the FCC. All licenses have a specific renewal period that varies according to service, so the Bureau maintains computerized records to insure timely renewal filing. License applications vary in degree of complexity from two or three pages to more complex forms, including supporting documents. The Frequency Bureau also provides data for National Broadcasting Company requests for operating authority for on-the-spot coverage of news stories and sports events.

Equipment approval

The greater demand for and increased use of the radio spectrum has made frequency management increasingly important. Over the years, and especially recently, the Commission has therefore developed more stringent technical requirements for the radio frequency equipment over which it has jurisdiction. To minimize the possibility of interference to the various radio services it regulates, the Commission has adopted technical standards and evaluation procedures for equipments that emit rf energy. These procedures consist of type approval, type acceptance and certification of equipment. The rules also provide that equipment requiring any of these three approvals may not be sold or leased until the approval has been granted.

Type approval calls for submitting one or more sample units of the equipment to the FCC Laboratory Division in Laurel, Md. for examination and measurement. All units subsequently marketed by the grantee must be identical in all respects to the sample tested or include only changes authorized by the Commission. This applies to such equipment as Broadcast Modulation Monitors and Class I Television Devices. Class I Television equipment must operate within a channel allocated for television broadcasting; its output signal is connected to a receiving device by either wires or coaxial cable in addition to other requirements. RCA "SelectaVision" and the various ping pong, tennis and hockey games using the television screen as the play area are examples of this classification.

Type acceptance requires advance approval of licensed radio transmitting equipment by the Commission. It determines that the equipment complies with the rf-interference standards based on tests conducted by the manufacturer or his agent. Again, all units subsequently marketed must be identical to the sample tested except for variations authorized by the Commission. Type acceptance is required for transmitters in the Land Mobile,

Marine, Aviation, Class A and D Citizens, and Broadcast Radio Services.

Under *equipment certification*, the Commission acknowledges officially that the units tested by the manufacturer or his agent are designed to meet applicable technical standards. Certification is mandatory for any radio receiver which is capable of tuning in the band from 30 to 890 MHz, and therefore encompasses FM and TV Broadcast Receivers. (RCA has a tv broadcast receiver certification facility at Bloomington, Ind.) It also governs the use of Wireless Microphones, Radio-controlled Garage Door Openers, Field Disturbance Sensors and similar equipment. At the present time, most of the non-Government rf equipments manufactured or marketed by RCA are subject to these approvals. The Frequency Bureau processes all RCA applications for equipment approval and the resulting Commission grants, assuring efficient handling on a cost-effective basis. The Bureau also administers RCA plant certification, an additional Commission procedure whereby all manufacturing facilities using industrial heating or ultrasonic equipments are certified to meet prescribed rf radiation limits.

Committee work

Members of the RCA Frequency Bureau have assisted or acted as the RCA representative on a number of domestic and international technical committees concerned with radio regulations, frequency allocations and equipment specifications. The Bureau participates in International World Administrative Radio Conferences and in FCC proceedings where there may be a joint interest by more than one part of RCA. Since the Bureau has a prime responsibility for World Administrative International Radio Conference work of the International Telecommunication Union (ITU), it coordinates the necessary preparatory work. This work is particularly important inasmuch as the agreements reached by international conferences become the law of the United States when ratified by Congress.

The Bureau also actively participates in the work of the International Radio Consultative Committee (CCIR), and passively in the International Telegraph and Telephone Consultative Committee (CCITT) and the International Electrotechnical Commission (IEC). Usually this participation is supplemented by assistance furnished by appropriate experts within the Corporation. RCA is represented on all government-industry committees concerned with the allocation of radio frequencies.

Members of the Bureau have served as technical consultants to such groups as the Joint Technical Advisory Council of the IEEE and EIA (JTAC), the Electronic Industry Association (EIA), the Frequency Management Advisory Council of the President's Office of Telecommunication Policy, and the National Association of Broadcasters (NAB) FCC De-regulation committee.

Technical liaison

As part of its technical liaison, the RCA Frequency Bureau is routinely called upon to render interpretations of the domestic and international rules and regulations that are of a technical and operational nature. This may involve in-depth study of existing rules and case histories or consultation with the staffs of the FCC or Office of Telecommunications Policy.

Current rules and knowledge of changes being proposed are of importance to the RCA operating companies, as well as manufacturing and sales personnel. The Bureau supplies appropriate parties with the sections of these rules and regulations of interest to them and furnishes proposed changes, amendments and associated documents on a continuous basis. This service is supplied to over 600 people throughout RCA.

Many Frequency Bureau activities require engineering studies, in addition to frequency coordination with other radio systems. Prior frequency coordination is required by FCC Rules and must be completed before filing applications for earth stations in the space communication services or terrestrial microwave stations. This involves notifying other system operators of the proposed use of the desired frequencies. The other operators are then required to respond to our notification within 30 days. A continuing analysis is also carried out to protect existing RCA systems. In order to provide this frequency coordination service and rapid distribution of information, the Frequency Bureau performs many of its functions using data-processing techniques.

The records of the Frequency Bureau are a source of information to many parts of RCA. The Frequency Bureau maintains the most recent International Radio Consultative Committee (CCIR) and Consultative Committee International Telegraph and Telephone (CCITT) documents and provides copies of excerpts from them in response to a large number of requests from many companies and divisions of the Corporation. Through its knowledge of the records, functions and organization of the FCC and other federal agencies, the Bureau is able to assist in advance coordination and preparation of applications and engineering and marketing studies.

Conclusions

For more than forty years, the RCA Frequency Bureau has contributed to RCA's manufacturing, service, and communications entities with its concepts of centralized accumulation and dispersal of information, cost-effective support in licensing and equipment approval, and Corporate representation before U.S. Government and international organizations. Effective participation in important growth areas in the future will be a continuing goal of the Frequency Bureau.

U.S. Government and international agencies with whom the Frequency Bureau deals.

<i>International organizations</i>	<i>National organizations</i>
International Telecommunication Union	Federal Communications Commission
CCIR CCITT	Office of Telecommunications Policy
International Electrotechnical Commission	Department of Transportation
International Maritime Consultative Organization	U.S. Coast Guard Federal Aviation Administration
Other international	National Oceanographic and Atmospheric Adm. Department of State Other national

RCA entities assisted by the Frequency Bureau.

RCA Global Communications	Commercial Communications Systems Division
RCA Alaska Communications	Government Systems Division
RCA American Communications	Avionics Systems Missile and Surface Radar
National Broadcasting Co.	Astro Electronics Automated Systems
Picture Tube Division	
RCA Service Company	Consumer Electronics
Distributor and Special Products Div.	Solid State Division
David Sarnoff Research Center	Other RCA

How the Communications Act affects you

R.E. Simonds

If you design any equipment that transmits or receives rf energy, FCC Rules and Regulations affect your work in some way. The FCC's power comes from the Communications Act of 1934 and its amendments.

Everyone recognizes that laws have an important role in society. However, the laws governing use of the radio spectrum—and the engineer's role in writing, implementing, interpreting, and enforcing these laws—may not be widely recognized. The communications services and products of RCA are subject to FCC Rules and Regulations, which are based on the Communications Act of 1934 and its subsequent amendments.

General powers of the FCC

Many of the FCC Rules and Regulations are of a technical nature—about half of the general powers of the FCC listed in Section 303 of the Communications Act have engineering aspects. Consider those relating to the assignment of frequencies to the several classes of stations used for broadcasting and radiocommunication:

Sec 303.

GENERAL POWERS OF THE COMMISSION

Sec. 303. Except as otherwise provided in this Act, the Commission from time to time, as public convenience, interest, or necessity requires shall—

(c) Assign bands of frequencies to the various classes of stations, and assign frequencies for each individual station and determine the power which each station shall use and the time during which it may operate;

(d) Determine the location of classes of stations or individual stations;

(f) Make such regulations not inconsistent with law as it may deem necessary to prevent interference between stations and to carry out the provisions of this Act: *Provided, however, that changes in the frequencies, authorized power, or in the times of operation of any station, shall not be made without the consent of the station licensee unless, after a public hearing, the Commission shall determine that such changes will promote public convenience or interest or will serve public necessity, or the provisions of this Act will be more fully complied with;*

(h) Have authority to establish areas or zones to be served by any station;

The Radio Spectrum between 10 kHz and 275 GHz is divided into bands of frequencies allocated to the various radio services such as Broadcasting, Aeronautical, Land and Maritime Mobile, Radionavigation and Radio-

location, Fixed (Point-to-Point) and Radio Services using Space techniques. Knowledge of propagation characteristics over the radio spectrum is necessary to make adequate and suitable frequency assignments to stations that must meet the operational needs of these diverse services. For example, in the United States, standard AM Broadcast Service frequencies are identified as Local, Regional or Clear-channel, and assignments are made depending upon the size of the geographical area to be served. In some instances directional antennas are employed to obtain the most effective coverage or to reduce interference in certain undesired directions. Similarly, FM Broadcast stations are classified according to power and coverage area as Class A, B or C.

The frequency range allocated to Television Broadcasting extends from 54 MHz to 890 MHz. In order to compensate for this wide difference in frequency and variation in propagation characteristics between the lowest and highest frequencies, stations operating on the lower vhf channels are authorized a maximum effective radiated power of 100 kW at an antenna height of 500 ft, stations on channels 7 through 13 have a maximum power of 316 kW at 500 ft, and in order to provide equal coverage, uhf stations employ a maximum power of 5 MW.

Television Broadcast Stations are assigned in accordance with a Plan contained in the FCC Rules and Regulations, taking into account the Commission's obligation given in Section 307 (b):

Sec. 307

(b)⁴⁸ In considering applications for licenses, and modifications and renewals thereof, when and insofar as there is demand for the same, the Commission shall make such distribution of licenses, frequencies, hours of operation, and of power among the several States and communities as to provide a fair, efficient, and equitable distribution of radio service to each of the same.

The Plan also provides for assignment of stations in such a way as to prevent interference by taking into account the so-called "taboos," whereby assignments made to uhf television stations consider the deficiencies of many home tv receivers in their ability to suppress local oscillator radiation, images, and intermodulation products.

Similarly, additional frequencies have been made available to Land Mobile Services in some of the larger

OMEGA • SHIP-TO-SHORE • SONAR • DECCA • LORAN C • MARITIME TELEGRAPHY • AM RADIO • LORAN A • DISTRESS AIRBORNE • AMATEUR • INTERNATIONAL BROADCAST • SHIP-TO-SHORE • CITIZEN'S BAND • RADIO ASTRONOMY • VHF-TV • FM BROADCAST • AIRCRAFT • AIR TRAFFIC CONTROL • RECREATIONAL BOATS • TELEMETRY • EMERGENCY AIRCRAFT • SURVIVAL • SATELLITE • CITIZEN'S BAND • UHF-TV • MOBILE • AIR-TRAFFIC RADAR • MICROWAVE OVENS • SATELLITE UPLINK-DOWNLINK • MICROWAVE RELAY • AIRBORNE DOPPLER RADAR • RADAR ALTIMETERS • RADAR • SATELLITE

cities using uhf-tv channels 14 through 20 on a shared basis in a manner designed to avoid interference to the television broadcast service.

Many of the frequencies used for space communications are shared with terrestrial services. In order to make such sharing effective without mutual interference, it is necessary to coordinate the frequencies in planning the locations of earth stations and terrestrial microwave stations. The coordination calculations are complex, but each applicant for a station license must perform them.

It is obvious from the foregoing that in order to make assignments to stations so that these broadcasting and communications services can be provided without mutual interference, it is necessary to take into account the technical characteristics of the various parts of the radio spectrum together with terrain factors, areas to be served and powers of transmitters. These objectives of the Act can only be achieved by engineering solutions.



Raymond E. Simonds, Director of the RCA Frequency Bureau, has been associated with the Bureau since 1941, dealing with the allocation, assignment and use of frequencies throughout the radio spectrum. Mr. Simonds has been a member of the U.S. Delegation to a number of international radio conferences and is currently an industry advisor to the FCC Steering Committee preparing for the 1979 ITU General World Administrative Radio Conference.

Contact him at:

RCA Frequency Bureau, Washington, D.C., Ext. 4236

Equipment standards

Sec. 303

(e) Regulate the kind of apparatus to be used with respect to its external effects and the purity and sharpness of the emissions from each station and from the apparatus therein;

This is the authority under which the FCC prescribes technical specifications that must be met by transmitters used in the various radio services. The FCC Rules and Regulations governing each radio service contain specifications of frequency stability that must be maintained by transmitters, the level of suppression of spurious emissions, and the modulation characteristics of the transmitters. The FCC has a procedure of equipment type-acceptance that requires manufacturers to submit measurement data to the FCC on each model of transmitter offered for sale, attesting that it meets the requirements of the rules. Transmitters meeting the rules are subsequently accepted by the FCC for licensing. Type-acceptance recently became a newsworthy item when the FCC did not approve a large number of citizen's band transceivers.

Sec. 303

(g) Study new uses for radio, provide for experimental uses of frequencies, and generally encourage the larger and more effective use of radio in the public interest;

By way of encouraging the larger and more effective use of radio, the FCC issues experimental licenses when it is necessary to radiate a signal in just about any part of the radio spectrum for any purpose found by the FCC to be in the public interest. Perhaps the only proviso is that no interference be caused to regular radio services.

Sec. 303

(r) Make such rules and regulations and prescribe such restrictions and conditions, not inconsistent with law, as may be necessary to carry out the provisions of this Act, or any international radio or wire communications treaty or convention, or regulations annexed thereto, including any treaty or convention insofar as it relates to the use of radio, to which the United States is or may hereafter become a party.

This section of the Act gives the FCC authority to incorporate provisions of international treaties in its Rules and Regulations. Most international treaty matters of concern to the FCC are those developed within the International Telecommunication Union (ITU), a specialized agency of the United Nations. Its purposes, as

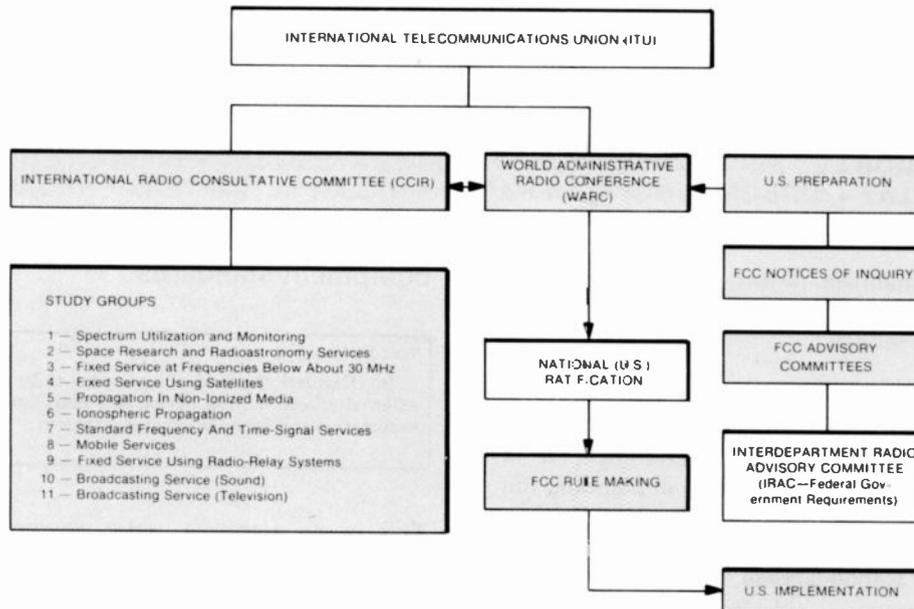


Fig. 1
How international radio regulations are made. Shaded blocks indicate areas in which industry engineers normally participate.

contained in the ITU Convention, are quite similar to those contained in the Communications Act: To promote development of technical facilities with a view to improving the efficiency of telecommunications services, to effect allocation of the radio frequency spectrum, and to register radio frequency assignments in order to avoid interference between radio stations of different countries.

Technical matters within the responsibility of the ITU are dealt with by two technical arms referred to as the CCIs. The CCITT is the consultative committee for telegraph and telephone matters and the CCIR is for radiocommunication matters.* Technical questions having international implications are normally studied by the CCI's and their recommendations are then developed for international adoption.

Article 1 of both the ITU Telegraph Regulations and Telephone Regulations are indicative of the importance of the recommendations of the CCITT.

Article 12 of the Radio Regulations, entitled Technical Characteristics of Equipment and Emissions, relies heavily upon Recommendations developed by the CCIR for guidance in selecting transmitting and receiving equipment. Registration of frequencies by the International Frequency Registration Board (IFRB) requires technical evaluation as to the probability of producing interference with other systems. The criteria employed in evaluating interference levels are based upon Recommendations of the CCIR.

The development of CCIR Recommendations is an area that is the province of the radiocommunications engineer. Where such Recommendations become a matter of inter-

*The initials represent, in French, Comité Consultatif International Télégraph et Téléphone, and Comité Consultatif International des Radiocommunications

Telegraph Regulations

Article I

Purpose of the Telegraph Regulations

1 1.1) The Telegraph Regulations lay down the general principles to be observed in the international telegraph service.

(2) In implementing the principles of the Regulations, Administrations* should comply with the C.C.I.T.T. Recommendations, including any Instructions forming part of those Recommendations, on any matters not covered by the Regulations.

2 2. These Regulations shall apply regardless of the means of transmission used, so far as the Radio Regulations and the Additional Radio Regulations do not provide otherwise.

*) or recognized private operating agency(ies)

Telephone Regulations

Article I

Purpose of the Telephone Regulations

1 1.1) The Telephone Regulations lay down the general principles to be observed in the international telephone service.

(2) In implementing the principles of the Regulations, Administrations* should comply with the C.C.I.T.T. Recommendations, including any Instructions forming part of those Recommendations, on any matters not covered by the Regulations.

2 2. These Regulations shall apply regardless of the means of transmission used, so far as the Radio Regulations and the Additional Radio Regulations do not provide otherwise.

*) or recognized private operating agency(ies)

national regulation, they are almost invariably incorporated into the Rules and Regulations of the FCC. The procedure for developing international Radio Regulations, including the role of the CCIR, is shown in Fig. 1.

Television receivers

Sec 303
 (g) Have authority to require that apparatus designed to receive television pictures broadcast simultaneously with sound be capable of adequately receiving all frequencies allocated by the Commission to television broadcasting when such apparatus is shipped in interstate commerce, or is imported from any foreign country into the United States, for sale or resale to the public.

Generally speaking, FCC jurisdiction is limited to transmitters; its jurisdiction over receivers is limited to restricting (radio) radiation from them. A 1962 amendment to the Communications Act, however, authorized the FCC to require television receivers to be able to tune all channels, both vhf and uhf. When the FCC did adopt rules requiring tv receivers to have comparable vhf and uhf tuning capability, it produced a challenge for engineers to devise effective methods of achieving this comparability.

Inspections

Sec 303
 (n) Have authority to inspect all radio installations associated with stations required to be licensed by any Act or which are subject to the provisions of any Act, treaty, or convention binding on the United States, to ascertain whether in construction, installation, and operation they conform to the requirements of the rules and regulations of the Commission, the provisions of any Act, the terms of any treaty or convention binding on the United States, and the conditions of the license or other instrument of authorization under which they are constructed, installed, or operated.

The FCC inspects radio installations, and in so doing it may be necessary to perform technical measurements to insure compliance with the terms of the station's license. Such measurements are normally carried out by an engineer. In the event that violations of the regulations are observed or detected during such inspections, a notice of violation is served upon the licensee. It is therefore necessary in such cases to take corrective measures to bring the station into compliance again and provide assurance that such corrective measures are effective.

Technical requirements for ships

Sec 355
 (e) The main and reserve installations shall, when connected to the main antenna, have a minimum normal range of two hundred nautical miles and one hundred nautical miles, respectively; that is, they must be capable of transmitting and receiving clearly perceptible signals from ship to ship by day and under normal conditions and circumstances over the specified ranges.

In order to implement this provision of the Act, the FCC had to determine what constituted "clearly perceptible signals from ship to ship by day and under normal

conditions and circumstances over the specified ranges" on the international distress frequency of 500 kHz. Several engineers sailed on ships and measured the signal-to-noise ratios under various atmospheric conditions. They took into account the average configuration of the vessels and the typical antennas in use and determined that a field intensity of 82.5 microvolts per meter was required at 200 miles in order to produce a "clearly perceptible signal." They calculated this to be the equivalent to 30 millivolts per meter at one nautical mile (over sea water) and determined that a transmitter having 200 watts output into a typical ship antenna could produce the required field strength.

Administrative procedure act

The Administrative Procedure Act requires that government agencies give prior notice of proposed rule making and provide opportunity for interested parties to file comments. As has been pointed out, many FCC Rules and Regulations are of a technical nature, and although comments are filed in a legal format, the substance of comments on technical matters is normally prepared by engineers. Fig. 2 shows the procedures that are usually followed to develop FCC Rules and Regulations.

Conclusion

The foregoing examples demonstrate the important role of the radiocommunication engineer in developing and implementing national and international radio regulations. The orderly administration of the radio spectrum requires engineers having knowledge of the state of the art.

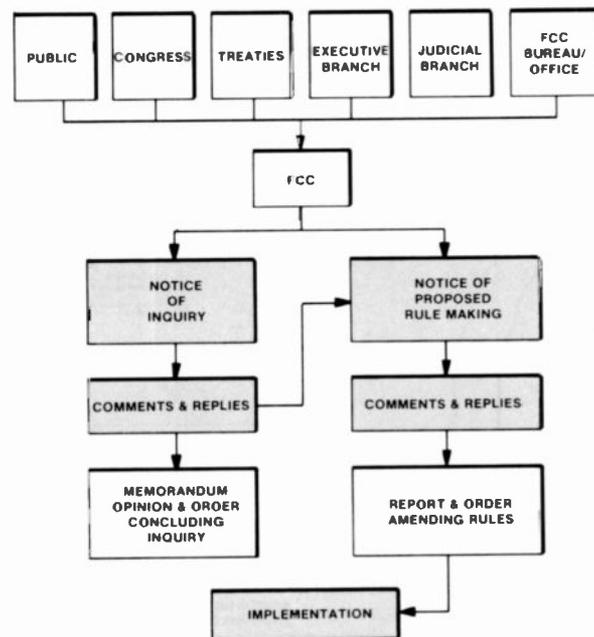


Fig. 2
 How FCC rules are made. Shaded blocks indicate areas in which industry engineers normally participate in technical rule-making proceedings.

Ground-control system for Satcom satellites

J. Lewin

RCA American Communications' two Satcom satellites, presently in U.S. domestic communications service, are supported by a highly reliable ground-control system to insure uninterrupted communications coverage. Accordingly, the Satcom ground system embodies a high degree of automation, redundancy, and autonomy of operation.

The key mission requirements of the Satcom ground segment are to control three satellites in geostationary orbit; provide continuous monitoring of health and status, and command and control; and keep the satellites on station within $\pm 0.1^\circ$ in longitude and latitude. The ground segment must also support transfer-orbit operations for each spacecraft, with an expected peak loading of two spacecraft in geostationary orbits and one in a transfer orbit. Fig. 1 represents the ground complex, set up for transfer-orbit operations.

The basic design considerations for the ground segment were operational reliability and economy of operation, which dictated a solution relying on a high degree of autonomy and automation. However, human decision is called for in the system at critical junctures.

High-reliability system

This emphasis on high reliability, which included the consideration of catastrophic ground-station failures, led to the concept of two redundant control stations. Each station is capable of performing the required Tracking, Telemetry, and Control functions (TT&C) in the orbital arc designated for domestic communications (99° to 132.5° W Longitude).

Locating these TT&C stations at Vernon Valley, N.J. and Moorpark, California, provides even wider coverage of the orbital path for the support of transfer-orbit operations. The addition of leased Intelsat facilities at Carnarvon, Australia, and Fucino, Italy produced global coverage. (See Fig. 2). The Intelsat stations were

instrumented by RCA for compatibility with the Satcom spacecraft.

Operating economics require the Satellite Operations Control Center (SOCC) and TT&C to be located together. The combined TT&C/SOCC is then collocated with a major satellite-communications earth station. This set-up means that a single operator can use a central control console to monitor the performance, ascertain the orbital position, and command each of the three spacecraft.

Dual but autonomous control

Normally, the Vernon Valley TT&C/SOCC is the primary Satcom Systems Controller for all spacecraft, with

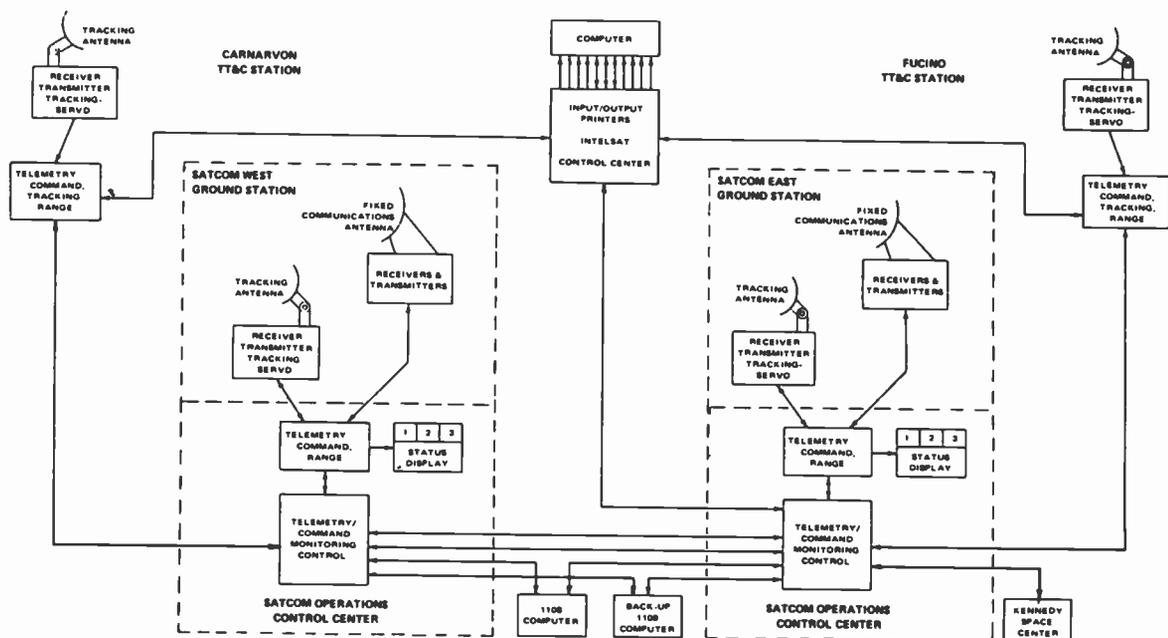


Fig. 1
Ground-station setup for a transfer-orbit operation. The two Intelsat stations are leased for such operations only; the two U.S. stations alone are sufficient for covering the arc needed for domestic communications.

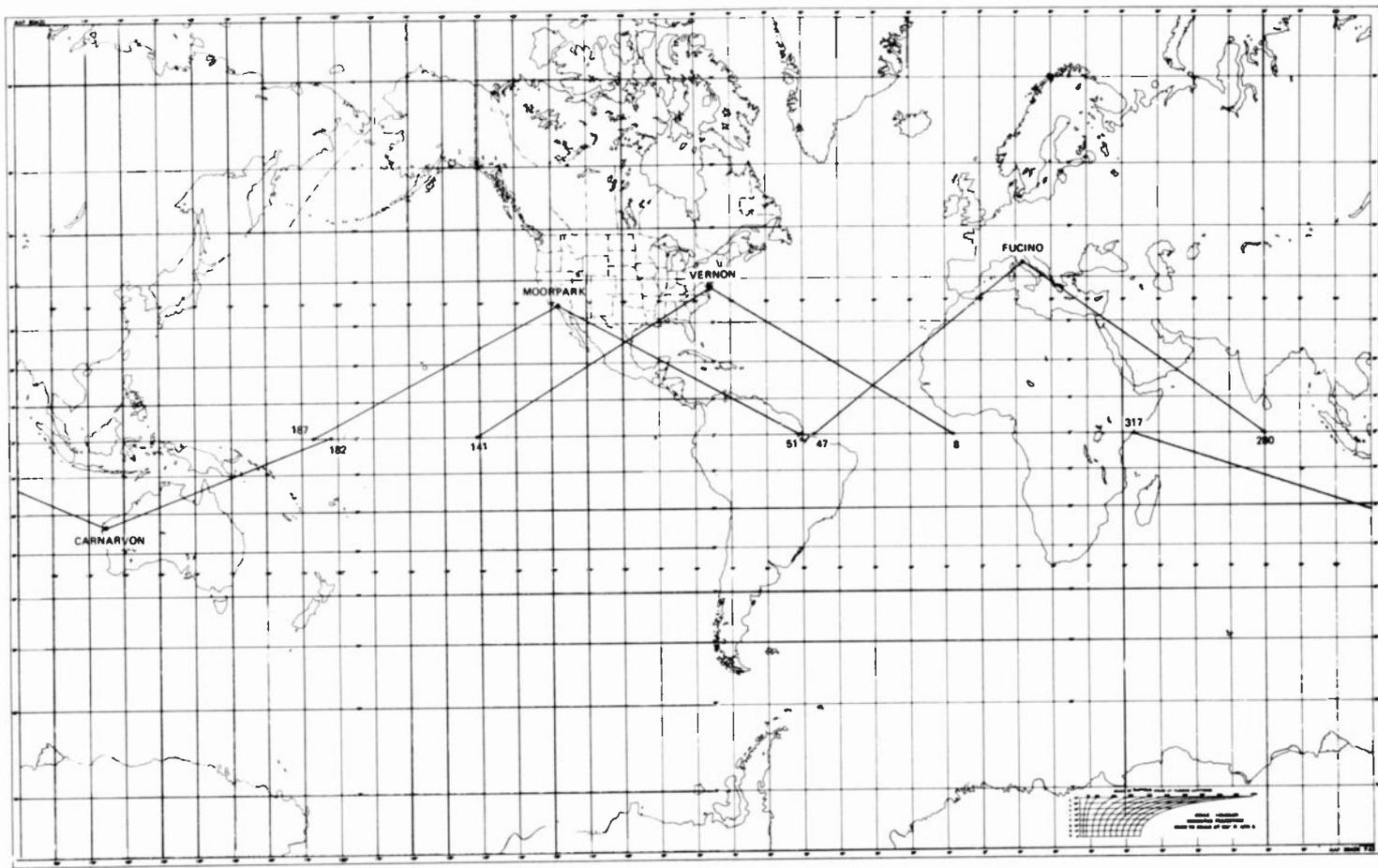


Fig. 2
Four stations provide worldwide coverage for transfer-orbit operations when their antennas are at 10° elevation and the satellites are at synchronous altitude. Numbers on chart are longitude expressed on a 360° basis.

Moorpark operating in a "hot" standby mode. At a moment's notice, however, the Moorpark site can assume the role of Systems Controller should the need arise, as in case of equipment malfunction or scheduled maintenance downtime, for example. This concept yields additional operating flexibility by permitting sharing of control, where one site controls one spacecraft and the other site controls the remaining spacecraft. This kind of load-sharing was practiced during the launch of Satcom F-2. In that case, Moorpark acted as controller for RCA Satcom F-1, in geostationary orbit, while Vernon Valley acted as Mission Controller for F-2, in transfer orbit. After the successful injection of F-2 into its geosynchronous (drift) orbit, control of F-1 reverted back to Vernon Valley.

The requirement of making each TT&C station autonomous has been achieved, except for orbit-related computations. The high precision and speed required there, particularly during transfer-orbit

operations, requires the use of large-scale computers. Their relatively low utilization, coupled with the high cost of purchase, led to the leasing of Univac 1108s from computer utilities. Again, reliability requirements, including the possibility of catastrophic failures, dictated the use of two physically separated computers, one prime and one backup. During transfer-orbit operations the backup computer was used as a "hot" standby: i.e., it was continually updated to keep abreast of the mission, as well as periodically exercised to cross-check with the primary computer.

State-of-the-art permitting, (i.e., the availability of low-cost, high-precision, high-speed minicomputer) a self-contained orbit-computing capability is under consideration for the TT&C/SOCC.

Computer complex

Two Hewlett-Packard 2100 minicomputers, designated "data" and "control," are at the core of each TT&C/SOCC.

These minicomputers and the two leased Univac 1108 computers, which are both accessible from each U.S. ground station, produce a system, and individual TT&C/SOCCs, with a high degree of automation. An HP 2100 is also in use at each of the two transfer-orbit stations at Carnarvon and Fucino.

The eight-computer complex is interconnected via a communications network that enables computer-to-computer transfer of data and commands (Fig. 3).

Two dedicated, full-duplex, 4800-baud lines interconnect the Moorpark and Vernon Valley TT&C/SOCCs: one line interconnects the data computers, transferring spacecraft telemetry received by one computer to the other; the second interconnects the two control computers. Command-list transfers and station coordination are done using this link. Finally, a 300-baud line is

Reprint RE-22-4-18
 Final manuscript received September 27, 1976.

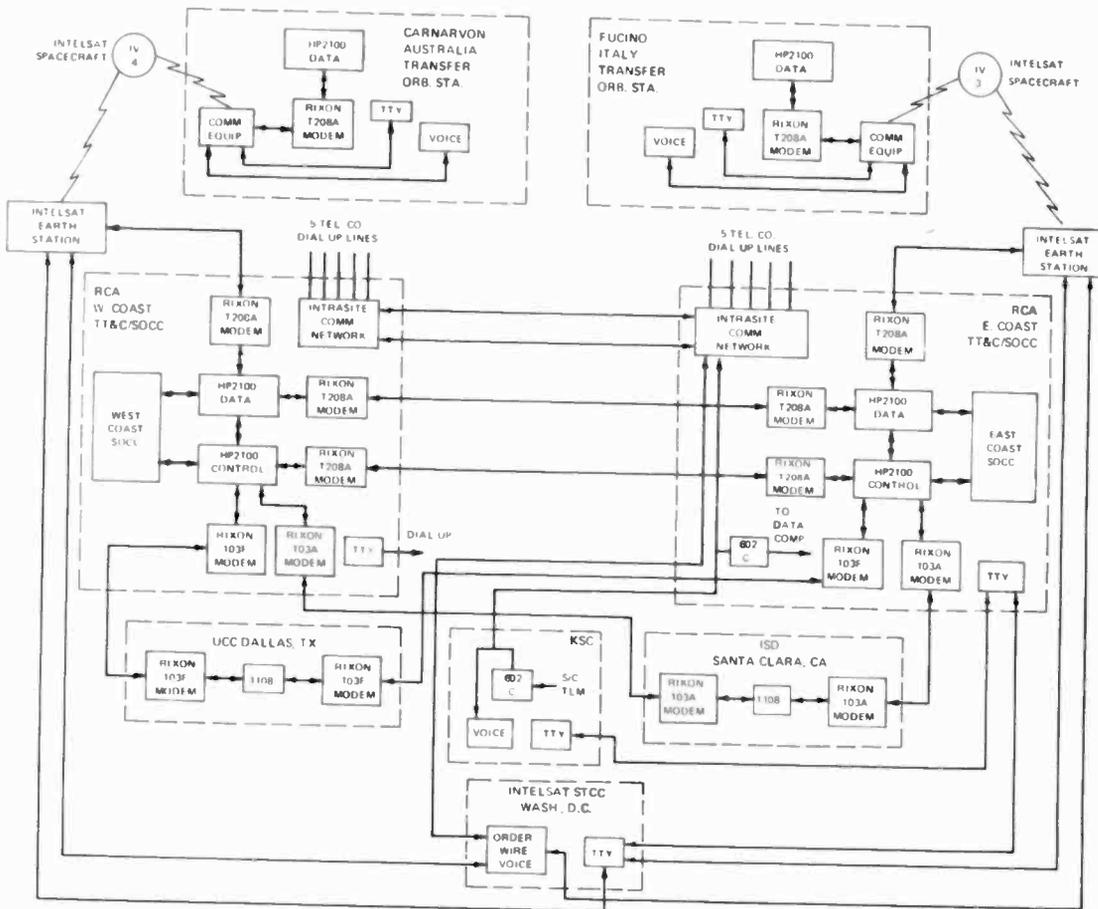


Fig. 3
Data and commands are transferred back and forth between eight computers in the system. Data computers deal with telemetry, ranging and attitude data; control computers deal with command sequences.

provided for data transfer to and from the Univac 1108 computers.

TT&C/SOCC ground station

The TT&C/SOCC has three basic functional areas: the rf receiving and transmitting equipment, the signal-processing equipment, and the spacecraft operations control center.

Both TT&C/SOCC ground stations are identical in capability and hardware. Within each station a high degree of redundancy and "cross-strapping" is provided, minimizing the effects of single-point failures. For example, should the computer-controlled automated commanding fail, either one of two redundant "command and range tone generators" provides manual command generation (see Fig. 4).

Each ground station can command or range with each of the three spacecraft, one at a time; telemetry can be received and processed simultaneously from all three spacecraft.

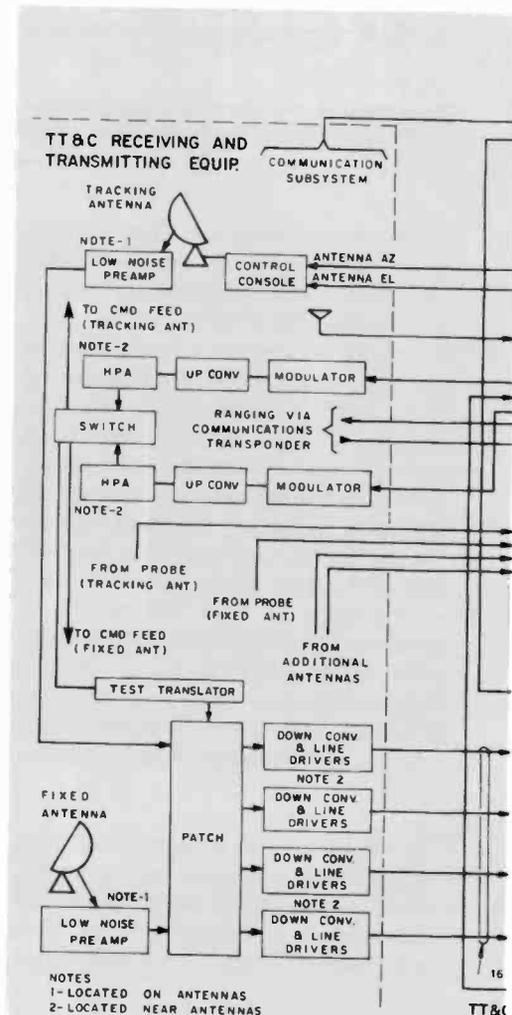
Antennas

Each spacecraft in orbit has a dedicated antenna, which performs two functions during commercial operations—communications and command and control.

All antennas have 13-meter dishes and a G/T of 32.4 dB/°K. The rf feeds are of dual orthogonal polarization, capable of receiving and transmitting simultaneously in both polarizations. The feeds are designed for 500 MHz of bandwidth, with 6 GHz uplink and 4 GHz downlink. The gain at 6 and 4 GHz is 56 dB and 53 dB, respectively. When driven by a 3 kW transmitter, the antenna system is capable of producing an effective isotropic radiated power (EIRP) of 89 dBw, which is ample for ranging and commanding during both transfer-orbit and on-station commercial operations.

Tracking system

To facilitate transfer-orbit operations, one antenna at each TT&C/SOCC is equipped



with auto-track capability. This antenna has an extended range ($\pm 270^\circ$ in azimuth and 90° in elevation) and higher angular velocity ($3^\circ/s$) when compared with the standard communications antenna. Four tracking elements comprise the monopulse feed system.

The pointing accuracy of the antenna is 0.025° rms and the tracking errors in the auto-track mode are less than 0.008° rms. The shaft position readout for azimuth and elevation has a resolution of 0.01° ; it is transmitted to the data computer for further processing, in conjunction with spacecraft ranging.

In addition to its auto-track mode, the antenna can also be controlled via computer-driven "programmed-track" and manual-track modes. Both tracking modes were successfully used at each TT&C/SOCC site during the transfer orbit of Satcom F-1 and F-2 when attempting an initial acquisition (the first time the

spacecraft appears over the horizon). Predictions for azimuth and elevation generated by the Univac 1108s off-line software are used to position the antenna.

When on-station, the auto-track antenna can track either of the two orthogonally polarized spacecraft beacon frequencies: 3700.5 MHz, horizontally polarized (east-west); and 4199.5 MHz, vertically polarized (north-south), originating at the spacecraft communications antenna. The commanding frequency is 6423.5 MHz and is horizontally polarized.

When in transfer orbit, the spacecraft's omni antenna is used, with both beacons copolarized. The command signal is orthogonal to that of the beacons. The beacon polarization is linear and parallel to the spacecraft's spin-axis, which changes in the course of the transfer mission as the spacecraft spin axis precesses. Consequently, manual polarization adjustments have to be made during this phase of tracking.

The tracking antenna has a polarization adjustment over $\pm 45^\circ$ for both orthogonally positioned rf feeds. Additionally, the polarization of the uplink feeds can be adjusted relative to the downlink feeds by $\pm 10^\circ$ in order to optimize polarization alignment.

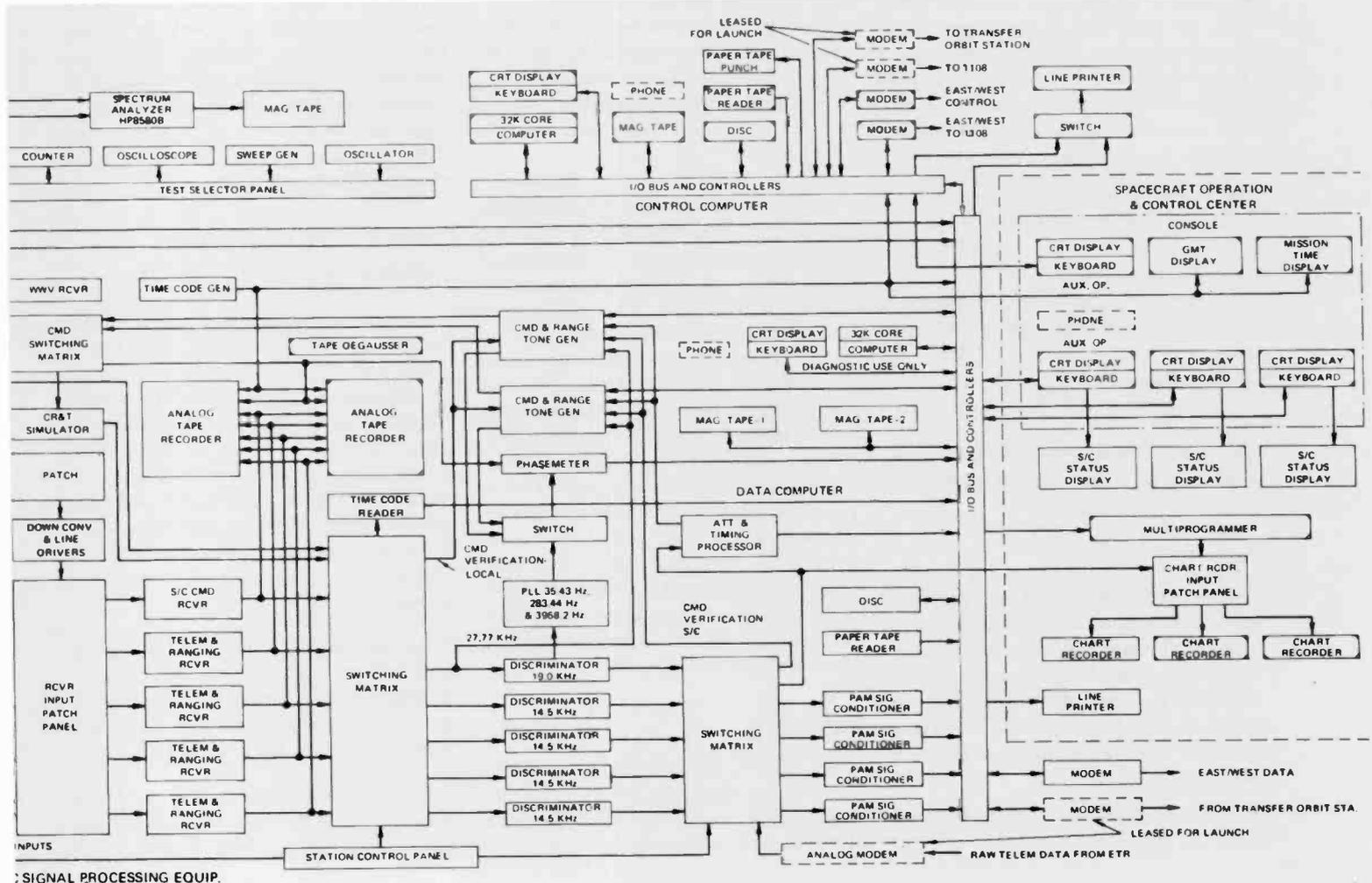
Redundant low-noise receivers (LNRs), provided for each polarization, enhance the reliability of the telemetry downlink. The primary pair is thermoelectrically cooled and has a maximum noise temperature of 55 K. The standby LNRs are uncooled, with a noise temperature of 62 K. Fig. 4 is a block diagram of the TT&C receiving and transmitting system.

Commanding system

The command system is designed to minimize errors on several levels: coding, transmission, and command generation.

Under routine operating conditions, com-

Fig. 4 Combined TT&C/SOCC ground station. Each of the two such Satcom stations is almost totally autonomous.



manding is performed using "canned" command lists designed to execute the desired functions. Generating and executing each list entails authorization. The control computer provides command-list generation, storage, and initiation. Commands are time-tagged and, once authorized, will be executed in accordance with the predetermined time sequence.

Spacecraft commanding entails the concept of load, then verify, then execute. The received command is retransmitted via the spacecraft beacon to the TT&C/SOCC, where it is compared to the original command. Upon verification, an "execute" command is sent to the spacecraft. In addition to the space-to-ground link, verification is also provided via a ground-antenna feedback loop, facilitating rapid fault-isolation to the ground segment.

Finally, protection is provided in the command format and code. A Hamming distance of two exists between the spacecraft address and operation codes (op-codes), so a single-bit error at the spacecraft will not result in acceptance of the command.

Beyond the preceding, commands which could adversely affect revenue-producing capability or spacecraft life are classified as "hazardous" commands. These cannot be transmitted without activating the "hazardous command key," located on the operations control console.

The format allows for 256 command op-codes, of which 210 have been implemented in the spacecraft. Command transmission is at 100 bits per second. The modulation is PCM/FSK/FM, with ternary FSK; the execute tone is a different frequency from either the "mark" or "space," giving added protection against false command execution.

Given the above described system, the probability for false command reception by the spacecraft is less than 10^{-22} .

Telemetry

Telemetry is updated from each spacecraft every 2 seconds; there are 128 channels per frame. One hundred and ten channels require processing by the data computer in real time. A total of 145 points of information (155 with both beacons) are relayed from the spacecraft either as status or analog data.

Each Satcom spacecraft can transmit two telemetry signals simultaneously, one on

each beacon. PAM/FM/PM is used for "housekeeping" telemetry; analog/FM/PM is used with both attitude telemetry during transfer orbit, and "dwell" telemetry where a given channel is examined for closer evaluation of a given parameter. A 14.5-kHz subcarrier is frequency-modulated by either PAM or analog data. The modulated subcarrier in turn phase-modulates the beacon rf carrier.

After down-conversion (see Fig. 4), telemetry is patched to one of the four telemetry and ranging receivers, where it is demodulated and switched to one of the four 14.5-kHz discriminators. After recovery of the baseband signal, PAM telemetry is switched to any one of four PAM signal conditioners, where analog-to-digital (A/D) conversion is made. The digitized data is received by the data computer for processing, logging on disc and digital tape, and display on CRTs, stripcharts, or line-printer hard copy.

After it is received, attitude data is routed to the attitude and timing processor, where it is quantized and transmitted to the data computer for display on stripcharts or for pre-processing into an "attitude data file," which is transmitted to the off-line 1108 for subsequent determination of the spacecraft's spin-axis orientation. Dwell data is routed directly for display on one of the 18 available stripchart recorder channels.

Telemetry is backed up on several levels, providing for greater system operating reliability and flexibility. Two seven-track analog tape recorders can continually store received telemetry (including ranging and command verification data). Should the data computer be unable to process in real time because of a failure or scheduled maintenance, the stored analog tape may be played back to recover telemetry data. Analog tapes may be maintained for a period of several days as a backup to the processed data logged on the digital tape. The digital tapes are stored for several months for use in intermediate term trend analysis. Subsequently, "data compression" is performed on these tapes. Compressed tapes are then stored for long-term trend analyses and posterity.

Ranging

Slant range to the spacecraft is measured by using the spacecraft command receiver and beacon transmitter in transponder fashion. Four coherently generated sine-

wave tones are transmitted one at a time via the command uplink, then received via the beacon downlink. The range is determined by measuring the shift in phase angle between each transmitted and received tone.

The four tones, generated by the command and range tone generator under data-computer control, are 35.4 Hz, 283.4 Hz, 3968.1 Hz, and 27.777 kHz. One cycle of the low tone measures to within 4241 km; the next two higher tones resolve ambiguities to within 530 and 37 km; and the high-frequency tone resolves the range to within 5.409 km.

Using all four tones together, the resultant range measurement is typically accurate to within 25 meters (1 sigma), when using one-station ranging, and 12 meters (1 sigma) when using two-station ranging.

The ranging system employs FM/FM modulation on the uplink and FM/PM on the downlink.

Typically, ranging to each spacecraft is performed subsequent to a maneuver. Data is collected hourly, alternating between Vernon Valley and Moorpark, to achieve greater accuracy. Ranging tones are applied for a period of approximately 15 minutes, each time. Accumulated tracking-data files are transmitted to the off-line 1108, via the control computer, for orbit-determination processing.

Spacecraft operations control center (SOCC)

A six-cabinet wraparound console is the focal point for Satcom spacecraft system operations. Four CRT/keyboards are the prime monitoring and controlling devices. Three of these keyboard terminals and their associated character-mode CRT screens interface with the data computer and also request and display telemetry data. All three CRTs can display telemetry from one spacecraft or each CRT can display telemetry from a different spacecraft.

Three additional 23-inch black-and-white tv monitors suspended from the ceiling act as repeaters, enabling operations analysts to monitor spacecraft health and status without crowding the console area.

A fourth character-mode CRT/keyboard interfaces with the control computer and acts as the executive-command input to the

system. Command generation, authorization, and spacecraft control are done here.

An "operator control panel" allows the console operator to monitor and control the TT&C equipment remotely: e.g., computer status monitoring, stripchart recorder, and analog tape-recorder control. The panel also contains the hazardous-command lock and alarms.

Intersite voice and data communications control is performed at the control console, where the operator can control data transmission to the off-line 1108 or to the other TT&C/SOCC, for example. A mission-time clock and a GMT clock round out the console equipment. Both clocks are driven from a time-code generator that is synchronized with a WWV receiver.

In addition to the console and its equipment, a line printer, three six-channel stripchart recorders connected to the data computer, and a second printer, switchable to either the data or control computer, give the control center area the required hard-copy capability.

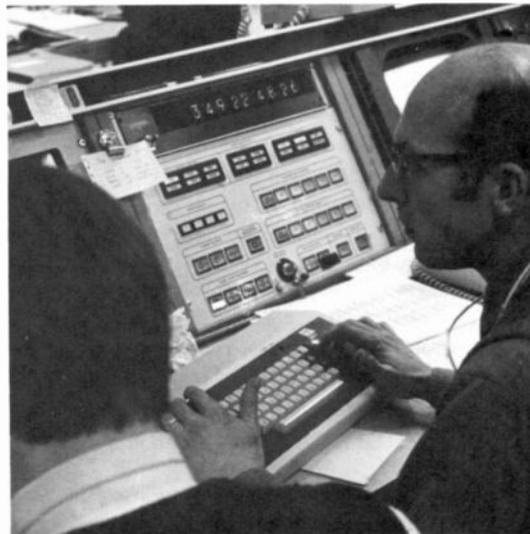
Computer software

Development of new computer software was a major undertaking needed to accommodate a new spacecraft.

The software system consists of two major parts: real-time software, needed to accept and operate on telemetry and provide spacecraft command and control; and off-line software, associated with orbital computations, establishing orbital position, planning maneuvers to effect the desired transfer orbit and on-orbit station-keeping.

All the software except the orbit determination and prediction software are new; the latter was based on the "space 360" program used by the U.S. Government.

The off-line software performs the key function of orbit determination and prediction, spin-axis attitude determinations, spin-axis precession planning, apogee kick



Clockwise, from top:
Control console at Vernon Valley during launch of Satcom F-1.
TT&C/SOCC operations room at Vernon Valley.
TT&C antenna, Vernon Valley.

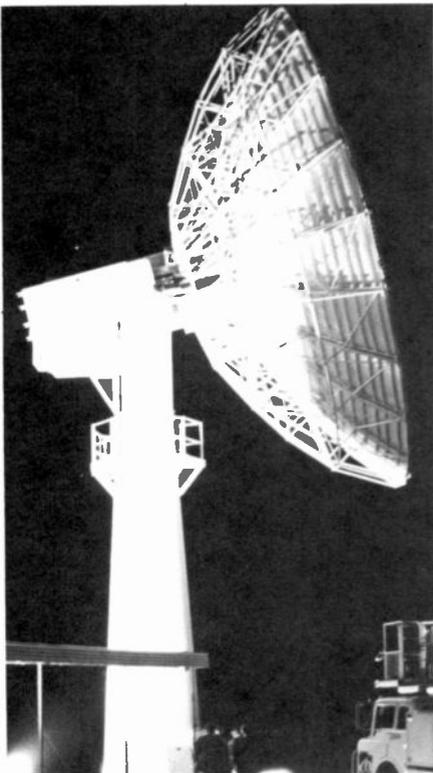


Table I
Real-time software resident in data and control computers.

<i>Key modules of the real-time software system</i>	
<i>Data computer</i>	<i>Control computer</i>
Telemetry calibration processing	Command-list generation
Telemetry logging	S, C Command-list execution
Telemetry data intersite tele-communications	Station Command-list execution
Video display	East coast/ West coast system coordination
Chart recorder display	Intrasite communications
Attitude data processing	1108 file processing
Antenna (Az&E1) data collection	1108 execution control
TT&C antenna control	1108 file transmission
Ranging data processor	

Jack Lewin is Principal Member of Americom's Space Systems Engineering staff. He has over 19 years of experience in system design, analysis, and testing in digital communications, data processing, command and control, and computer programming and applications. He joined RCA in 1959 and has worked on numerous programs in missile- and spacecraft-related projects in these fields. His most recent assignment was as project engineer on the Satcom spacecraft command, ranging, and telemetry subsystem, the ground TT&C/SOCC subsystem, and mission operations. His present assignment is as group leader on mission and systems software.

Contact him at:
Space Systems Engineering
RCA American Communications
Piscataway, N.J.
201-885-4189



motor firing planning, maintaining and updating the thruster performance parameters used in maneuver-planning, hydrazine fuel status, station acquisition planning, station-keeping planning, and data-base management.

Table I lists the key software modules comprising the "real-time" software system resident in the data and control computers.

Performance testing

The common thread throughout the Satcom ground system development was the newness of its constituent parts. This meant that meeting the launch date with the desired degree of confidence became a challenging exercise in testing and rehearsals.

To expedite testing, the orbit-determination and prediction software, for example, were tested against spacecraft data supplied by NASA and Telesat, Canada.

The ultimate way of testing a satellite tracking system, however, is with a satellite in orbit. This enables testing of not only software, but the antennas, the uplink, the downlink, and data and control computers, as well as the personnel.

In order to have this testing done prior to the launch of the Satcom spacecraft, RCA arranged to have NASA's ATS-1, ATS-5 and Anik II exercise the ranging subsystem and the orbit-determination and prediction software.

This accelerated system debugging immeasurably. Also, incorporating live ranging into the mission rehearsals lent a high degree of reality to the exercises, and led to the ultimate achievement of two successful missions.

Acknowledgments

A number of technical groups and organizations each contributed in their specialized way to the success of the Satcom ground segment development and realization.

At RCA American Communications, Inc., the TT&C and Mission Operations Engineering, Satcom Spacecraft Control, Earth Station Engineering, and the Earth Station Implementation activities each contributed significantly to the Satcom mission success by insuring the timely availability of the required systems, facilities, and personnel, all of which met or exceeded performance requirements.

RCA Astro-Electronics delivered hardware, computer software, and manpower that helped produce two flawless Satcom launch and transfer-orbit operations.

Telesat Canada's Computer Software System and Mission Analysis activity contributed invaluable advice toward a well-functioning timely system.

NASA allowed RCA to use the ATS-1 and ATS-5, which helped debug the Satcom TT&C/SOCC system, and contributed to a highly reliable operating ground system prior to the launch of the Satcom satellites.

References

1. Napoli, J.; and Christopher, J.; "RCA Satcom system," *RCA Engineer*, Vol. 21 No. 1 (Jun Jul 1975) pp. 23-29.
2. Napoli, J.; and Greenspan, J.; "RCA Satcom, the next generation domestic communications satellite system," (Sep 16-19, 1975) Wescon Communication Satellite Systems.
3. Becken, E.D.; "Satellite communications," *RCA Engineer*, Vol. 22 No. 1 (Jun Jul 1976) pp. 39-41.
4. Brook, A.W.; "RCA Satcom system," *RCA Engineer*, Vol. 22 No. 1 (Jun Jul 1976) pp. 42-49.
5. Keigler, J.E.; "RCA Satcom - maximum communications capacity per unit cost," *RCA Engineer*, Vol. 22 No. 1 (Jun Jul 1976) pp. 50-55.
6. Cuddihy, J.; and Wash, J.M.; "RCA Satcom earth station facilities," *RCA Engineer*, Vol. 22 No. 1 (Jun, Jul 1976) pp. 58-63.
7. Christopher, J.; Greenspan, D.; and Plush, P.; "The launch and in-orbit test elements of the Satcom system," *RCA Engineer* Vol. 22 No. 1 (Jun Jul 1976) pp. 64-70.
8. Bell, J.R.; and Bell, C.G.; editors, *Mini-Computer Software*, North Holland Publishing Co., 1976. See Smith, R.D.; and Mills, R.W.; "RCA Satcom programming technology."

Pen and Podium

Recent RCA technical papers and presentations

To obtain copies of papers, check your library or contact the author or his divisional Technical Publications Administrator (listed on back cover) for a reprint. For additional assistance in locating RCA technical literature, contact RCA Technical Communications, Bldg. 204-2, Cherry Hill, N.J., extension PY-4156.

Astro Electronics

G.T. Tseng|K.J. Phillips

Attitude stability of a flexible dual spin spacecraft with active nutation damping using products of inertia, *J. Astronomical Sc.*, Vol. 24, No. 3 (7-9/76)

Missile and Surface Radar

M.W. Buckley

Project management—planning, scheduling and control—AMA Seminar, Chicago, IL (10/11-14/76)

M.W. Buckley

Project/program management—Drexel University, Phila., PA (11/8-10/76)

J.R. Fogleboch|F.I. Palmer

J.O. Taylor|R.W. Sudbury

Solid state transceiver time sidelobe performance—*Digest GOMAC*, Lake Buena Vista, FL (11/9-11/76)

M.W. Buckley

Project/program management—PERT/CPM Workshop, The Behrend College, Erie, PA (Pennsylvania State U. Continuing Education) (9/30-10/2/76)

B.M. Fell

Electron mobility for a two-dimensional semiconductor at low temperatures—Condensed Matter Symp., SUNY, Stony Brook, L.I., NY (11/76)

J.C. Volpe

RCA and radar—evolving technology in the real world—*Signal*, Vol. 31, No. 2 (10/76) pp. 40-45

M.W. Buckley

Management planning, scheduling and control, AMA Meeting, Baltimore, MD (9/27-28/76)

M.W. Buckley

Management by objectives—Philadelphia Naval Base Officers' Club, Phila., PA (9/15/76)

D.L. Pruitt

Thyristor switches for super power intermittent duty operation—*Proc. Pulse Power Systems Workshop*, Naval Surface Weapons Center, White Oak, MD (9/21-23/76)

J.T. Threston

Enhanced carrier operations through use of AN/SPY-1 radar system—*Proc. Amer. Soc. Naval Engineers*, San Diego, CA, (10/7-8/76)

D. Shore

RPV: the future is here—*National Defense*, vol. 61, No. 339 (11/76)

Automated Systems

J.C. Willett

The soldering process and solder joint inspection—Printed Circuit Technology QC Seminar, Burlington, MA (12/3/76)

M.L. Johnson|C.F. Matthews

Transferring usable information to the maintenance man—11th Intl. Logistics Symp., Valley Forge, PA (8/19/76)

M.J. Cantella

Observations of sky brightness spatial variations—SPIE Technical Symp., San Diego, CA (8/27/76).

M.J. Cantella

High resolution camera tubes—1976 Special Summer Program on Photoelectronic Imaging Devices, San Diego, CA (8/21/76)

C.A. Asbrand

Surveillance with night vision equipment—Dept. of Criminal Justice Crime Lab, Clinton C.C., Plattsburgh, NY (9/76).

M.J. Cantella

Electrical input-output storage tubes—1976 Special Summer Program on Photoelectronics Imaging Devices, San Diego, CA (8/18/76)

O.T. Carver

EQUATE—New technology ATE—Defense Expo '76, Wiesbaden, Germany (10/6-8/76)

T.E. Kupfrian

Commercial integrated circuits in selected military applications—IEEE Boston Section Reliability Chapter, Burlington, MA (1/12/77)

E.W. Richter|V.D. Holaday

Designing microwave ATE to meet UUT requirements—AUTOTESTCON, Arlington, TX (11/10-12/76)

Corporate Engineering

H. Kleinberg|F.M. Oberlander

Standardization in a multidivision company—*Standards Engineering*, Vol. 27, No. 4 (8/75) pp. 62-64

Advanced Technology Laboratories

E.P. Herrmann|D.A. Gandolfo

A. Boornard|D.B. Stepps

CCD programmable correlator—*Proc. 3rd Intl. Conf. on Applications of Charge Coupled Devices*, Edinburgh, Scotland (9/28-30/76)

E.P. Herrmann|D.A. Gandolfo

A. Boornard|D.B. Stepps

A high-speed, CCD-scanned photosensor for gigabit recording applications—*Proc. 1976 Electro-Optical Sys. Design Conf./Intl. Laser Exposition*, New York, NY (9/14-16/76)

Government Communications Systems

O.E. Bessette

Recent advances in magnetic recording of digital data—Intl. Telemetry Conf., Los Angeles, CA (9/28-30/76)

A.M. Burke|F.L. Putzrath

NASA standard spacecraft tape recorder—Int. Telemetry Conf., Los Angeles, CA (9/28-30/76)

O. Black

Water soluble residue fluxes—Practical Soldering Technology, Anaheim, CA (9/21-23/76)

D. Hampel|J. Rothweiler

An LSI FFT signal detection and demodulation processor—Government Microcircuit Conf., FL (11/10/76)

Broadcast Systems

M. S. Siukola

New multi-station top mounted fm

antenna—26th Annual IEEE Broadcast Symp., Washington, DC (9/23-24/76)

O. Ben-Dov

Radiation patterns of broadcast antennas from aperture illumination—26th Annual IEEE Broadcast Symp., Washington, DC (9/23-24/76)

A.J. Fandozzi

Noise measuring techniques for television transmitters—26th Annual IEEE Broadcast Symp. Washington, DC (9/23-24/76)

Picture Tube Division

R. Marshall

Design criteria for platinum-rhodium alloy sheath thermocouples for stable operation above 1300C—*Proc. Instrument Society of America*, Houston, TX (10/11-14/76). Also in *Proc. ISA-76 Annual Conf.*, Vol. 31, Part 3

Laboratories

W.C. Stewart

On differential phase contrast with an extended illumination source—*J. Optical Soc. America*, Vol. 66, No. 8 (8/76) p. 813

J.E. Carnes|R.L. Rodgers III

Recent results on CCD imagers—*Laser 75 Opto-Electronics*, p. 78

H. Kressel|M. Ettenberg

Low-threshold double-heterojunction AlGaAs/GaAs laser diodes: theory and experiment—*J. Appl. Phys.*, Vol. 47, No. 8 (8/76) p. 3533

G.W. Hughes|R.J. Powell
M.H. Woods

Oxide thickness dependence of high-energy electron, VUV, and corona-induced charge in MOS capacitors—*Appl. Phys. Lett.*, Vol. 29, No. 6 (9/76) pp. 377-79

G.L. Schnable

Applications of electrochemistry to fabrication of semiconductor devices—*Fabrication of Semiconductor Devices*, Vol. 123, (9/76) p. 310-C

L.C. Upadhyayula

Quasi-enhancement mode of operation of transferred electron logic devices (TELDs)—*Electronics Lett.*, Vol. 2, No. 10 (10/13/76)

C.R. Wronski|D.E. Carlson
R.E. Daniel|A.R. Triano

Electrical properties of a-Si solar cells—*Digest*, 1976 IEEE Intl. Electron Devices Mtg., Washington, D.C. (12/6-8/76).

D.O. North

Theory of model characters, field structure, and losses for semiconductor laser—*IEEE J. Quantum Electronics*, Vol. QE-12, No. 10 (10/76) pp. 616-624.

A. Rosen|P.T. Ho|J.B. Klatskin

Fabrication and thermal performance of a

novel trapatt diode structure—*IEEE Trans. Electron Devices* (2/77)

C.J. Nuese

Diode sources for 1.0 to 1.2 um emission—*Technical Digest*, 1976 Intl. Electron Devices Mtg., Washington, D.C. (12/6/76) pp. 125-128

R.J. Himics|M. Kaplan
N.V. Desai|E.S. Poliniak

Poly (cyclopentene sulfones) as electron beam resists—SPSE Business Graphics Abstracts, Society of Photographic Scientists and Engineers Symp., Washington, DC (11/9-13/76) pp. 82-86

D.R. Carter|K.G. Hernqvist

Advances in helium-cadmium lasers—*Proc. of Electro-Optical Systems Design Conf.*, New York, NY (9/14-16/76)

E.S. Kohn

IR imaging with Schottky barrier CCD arrays—Solid State Circuits Committee of the IEEE Circuits and Systems Society (10/15/76)

M. Skolnin|L.C. Thanh
F. Levy|G. Harbeke

High-field magneto-absorption investigations of exciton states in Pb₂—*Int. Conf. on Layered Semiconductors and Metals*, Bari, (9/6-10/76)

D. Meyerhofer

New technique of aligning liquid crystals on surfaces—*Appl. Phys. Lett.*, Vol. 9, No. 11 (12/1/76) pp. 691-2

M.S. Abrahams|C.J. Buiocchi|R.T. Smith
C.W. Cullen|J.F. Corboy, Jr.|J. Blanc

Early growth of silicon on sapphire—I: transmission electron microscopy—*Electrochemical Soc. Fall '76 Mtg.*, Las Vegas, NV (10/76)

J. Blanc|M.S. Abrahams|G.W. Cullen
J.F. Corboy, Jr.|C.J. Buiocchi|R.T. Smith
Early growth of silicon on sapphire—II: models—*Electrochemical Soc. Fall '76 Mtg.*, Las Vegas, NV (10/76)

P.H. Robinson|R.V. D'Aiello|H. Kressel
Dichlorosilane: a silicon source for epitaxial growth—*Electrochemical Soc. Fall '76 Mtg.*, Las Vegas, NV (10/76)

R.J. Himics|D.L. Ross

The examination of poly (5-hexene-2-one sulfone) as a positive-working photoresist—*Technical Papers*, Soc. Plastics Engineers Conf., Ellenville, NY (10/13-15/76), pp. 26-33

R.J. Himics|M. Kaplan
N.Y. Desai|E.S. Poliniak

The synthesis, characterization, evaluation, and processing of selected sulfone copolymers as electron beam resists—*Technical Papers*, Soc. Plastics Engineers Conf., Ellenville, NY (10/13-15/76) pp. 269-283

P.T. Ho|A. Rosen|J. Klatskin

Power combination of broadband trapatt

amplifiers—*Electronics Lett.*, Vol. 12, No. 1 (1/8/76) p. 16

M.T. Gale|K. Knop

Color-encoded focused image holograms—*Appl. Optics*, Vol. 15, No. 9, (9/76) pp. 2189-2198

A.G.F. Dingwall|R.E. Stricker

C²L: A new high speed, high density bulk CMOS technology—IEEE Intl. Electron Device Conf., Washington, DC (12/7/76) p.3

J.I. Pankove

New materials for electroluminescent devices—1976 Intl. Conf. Solid State Devices, Tokyo (9/2/76)

R.A. Geshner

LSI photolithography—I.G.C. Conf. on Microphotolithography, Ipswich, MA (9/13/76)

A.C. Ipri|J.C. Sarace

CMOS/SOS process evaluation techniques—GOMAC, Buena Vista, FL (11/9-11/76)

A. Bloom|R.A. Bartolini
L.K. Hung|D.L. Ross

A non-polymeric host for recording volume phase holograms—*Appl. Phys. Lett.*, Vol. 29, No. 8 (10/15/76) pp. 483-4

D. Vilkomerson|R. Mezrich|M.F. Etzold

An improved system for visualizing and measuring ultrasonic wavefronts—VII Intl. Symp. on Acoustical Imaging and Holography

A. Bloom|R.A. Bartolini|P.L.T. Hung

The effect of host on volume phase holographic recording properties—SPE 4th Technical Conf. on Photopolymers, Ellenville, NY (10/13/76)

R.A. Bartolini|A. Bloom

Review of organic holographic recording media—*Optical Soc. Annual Mtg.*, Tuscon, AZ (10/76)

A. Bloom|P.L.K. Hang

Effect of dye structure on order parameter in a nematic liquid crystalline host—6th Intl. Liquid Crystal Conf., Kent, OH (8/24/76)

C.R. Wronski|D.E. Carlson|R.E. Daniel
Schottky-barrier characteristics of metal-amorphous-silicon diodes—*Appl. Phys. Lett.*, Vol. 28, No. 9 (11/1/76) pp. 602-605

R. Williams|R.S. Crandall|P.J. Wojtowicz
Melting of crystalline suspensions of polystyrene spheres—*Physical Review Lett.*, Vol. 37, No. 6 (8/9/76) pp. 348-351

D.A. deWolf

Waves in random media: weak scattering revisited—1976 Intl. IEEE/APS Symp. & USNC/URSI Mtg., Amherst, MA (10/11-15/76)

G.L. Schnable

Reliability of MOS devices in plastic packages—*International Micro-electronics Conf.* (6/9/76)

J.R. Sandercock

Simple stabilization scheme for maintenance of mirror alignment in a scanning Fabry-Perot interferometer—*J. Physics E: Scientific Instrument*, Vol. 9 (1976) p. 567

E. Sichel|R. Miller

Thermal conductivity of highly oriented pyrolytic boron nitride—*Thermal Conductivity* 14 (1976)

RCA Records

A. Devarajan|D. Mishra
Computerized technique for analyzing inventory control problems—American Inst. of Industrial Engineers Annual Systems Engineering Conf., Boston, MA (12/3/76)

D. Mishra

Computer-aided work sampling—Work-sampling workshop, U. of Wisc. Ext., Milwaukee, WI (11/17/76)

Patents

Automated Systems

S.C. Hadden|L.R. Hulls
P.J. Staney|E.M. Sutphin, Jr.

Tachometer without physical connection to internal combustion engine—3978719

Government Communications Systems

S.P. Clurman

Reduction of hunting in synchronous motor—3988653

E.J. Nossen|E.R. Starnier

Accurate digital phase/frequency extractor—3984771

Advanced Technology Laboratories

B.W. Siryj

Object orientation apparatus—3986604

K.C. Hudson|R.F. Kenville

Optical scanner with large depth of focus—3989348

B.W. Siryj

Labeling apparatus—3984279

Solid State Division

A.A. Ahmed

Equalization of base current flow in two interconnected transistor amplifiers—3987368

A.A. Ahmed

Current amplifier—3990017

D.R. Carley

Method of making a semiconductor device—3980507

A.A. Ahmed

Absolute-value circuit—3989997

A.J. Pikor

Multiple mesa semiconductor structure—3988765

R. Denning|W.G. Einthoven

Thermally balanced PN junction—3988759

H.A. Wittlinger

Series energized transistor amplifier—3986132

S.W. Kessler, Jr.

Transcendent semiconductor device—3984861

C.F. Wheatley, Jr.

Bridge-output amplifier with direct-coupled differential-mode feedback—3983502

A.F. Arnold

Nickel-gold-cobalt contact for silicon devices—3982908

O.H. Schade, Jr.

Radiation responsive voltage dividing circuit—3982197

S. Berkman|J.G. Marton

Graphite susceptor structure for inductively heating semiconductor wafers—3980854

M. Glogolja

Protection circuit—3980930

H. Arnoldi

Transistor testing circuit—3979672

O.H. Schade, Jr.

Differential amplifier circuit—3979689

M.I. Payne

Information storage circuit—3979735

M.I. Payne|B.S. Dalal

Transistor switching circuit—3979611

A.A. Ahmed

Current level detector—3979606

H.W. Justice

Etchant for silicon nitride and borosilicate glasses and method—3979238

A.D. Checki|A.G. Frey

Automatic assembly of semiconductor devices—3978579

Consumer Electronics

J. Avins

Transient suppression in television video systems—3984865

D.H. Willis

Gating circuit for a video driver including a clamping circuit—3984864

B.W. Beyers, Jr.

Character generator for television channel number display with edging provision—3984828

S.A. Steckler|A.R. Balaban

Deflection waveform correction signal generator—3984729

J.B. George

High power remote control ultrasonic transmitter—3984705

R.L. Shanley, II|J.M. Yongue

Apparatus for accentuating amplitude transients—3983576

M.E. Miller|J.G. Amery

Velocity correction system with damping means—3983318

J.C. Schopp

Turntable speed control system—3983316

L.A. Cochran

Switching arrangement for flesh tone correction and chrominance overload control circuits—3982273

J.C. Peer|D.W. Luz

Power supply for a television receiver—3980821

K.R. Woolling, Jr.

Counter type remote control receiver including noise immunity system—3980956

J.B. George

Tuner bandswitching system for a television tuning system—3980959

Laboratories

A.V. Tuma

Digital remote control for electronic signal receivers—3987414

C.R. Wronski|A.D. Cope|B. Abeles
Low dark current photoconductive device—
3987327

L.S. Napoli
Method of forming conductive coatings of predetermined thickness by vacuum depositing conductive coating on a measuring body—3987214

E.C. Giaino, Jr.
Method of increasing the image exposure and developing sensitivity of magneto-electric printing system—3986872

M.Ettenberg|H.Kressel
Solar cell device having two heterojunctions—3990101

F.C. Duigon|S. Liu
Planar trapatt diode—3990099

C.J. Busanovich|R.M. Moore
Selenium rectifier having hexagonal polycrystalline selenium layer—3990095

C.T. Wu
Display device utilizing magnetic storage—
3988738

H. Kressel|V.L. Dalal
Solar cell device having improved efficiency—3988167

W.F. Kosonocky
Introducing signal at low noise level to charge-coupled circuit—3986198

P.K. Weimer
Charge transfer memories—3986176

J.I. Pankove
Insulating nitride compounds as electron emitters—3986065

H.L. Pinch|S.T. Opreko
Vapor deposition of cermet layers—3985919

J.L. Vossen, Jr.|F.R. Nyman
G.F. Nichols
Adherence of metal films to polymeric materials—3984907

G.S. Kaplan|A.D. Ritzie
Homodyne communication system—
3984835

J. Rosen|E.J. Denlinger
Two-inductor varactor tunable solid-state microwave oscillator—3984787

J. Gross|W.H. Barkow
Display system utilizing beam shape correction—3984723

S. Larach
Apparatus and method for analyzing biological cells for malignancy—3984683

I. Gorog
Flying spot scanner unaffected by ambient light—3984629

N. Feldstein
Method for electrolessly depositing metals

using improved sensitizer composition—
3982054

P.N. Yocom|J.P. Dismukes
Luminescent sulfides of monovalent and trivalent cations—3981819

P.E. Haferl
Deflection circuit—3980927

H.R. Beelitz|D.R. Preslar
Electrical circuit—3979607

S.A. Lipp
Chemical vapor deposition of luminescent films—3984587

F.Z. Hawrylo
Ohmic contact—3984261

P.M. Russo
Electron image identifying system—
3983571

H.E. Schade
Vacuum electron device having directly-heated matrix-cathode-heater assembly—
3983443

D.J. Channin
Liquid crystal display—3981559

P.M. Heyman|R.L. Quinn|I. Gorog
Electrochromic display device—3981560

Distributor and Special Products Division

P.J. Mikulich, Jr.|J.D. Callaghan
Broadband antenna system with the feed line conductors spaced on one side of a support boom—3984841

Broadcast Systems

L.J. Bazin
High efficiency deflection circuit—3983452

A.M. Goldschmidt|W.A. Dischert|J.R. West
System for controlling tension of magnetic tape—3982160

Missile and Surface Radar

J.L. Christensen
RF cross correlation radar—3981013
(assigned to U.S. Gov't)

R.P. Perry
Data processor reorder shift register memory—3988601

R.P. Perry
Digital matched filtering using a step transform process—3987285

J.R. Ford
Apparatus for erecting a true vertical axis—
3985033

J.O. Horsley
High-speed counter with reliable count extraction system—3982108

D. Olivieri|R.J. Socci
Heat spreader and low parasitic transistor mounting—3982271

Picture Tube Division

M.H. Wardell, Jr.
Cathode ray tube assembly fixture—
3989233

B.K. Smith
Yoke mount assembly—3986156

T.A. Saulnier
Photograph method employing organic light-scattering particles for producing a viewing-screen structure—3981729

D.D. VanOrmer
Shadow mask color picture tube having non-reflective material between elongated phosphor areas and positive tolerance—
3979630

L.J. Dimattio
Electron tube base—3979157

RCA Ltd. Canada

C.H. Todd
Method of air letting an evacuated cathode ray tube—3981554

G.J. Lo|F.L. Papworth
N.E. Tenne-Sens|M.V. O'Donovan|G. Dziub
Metal plated body composed of graphite fibre epoxy composite—3982215

Avionics Systems

J.E. Miller
Tracking gate servoed by relative range—
3987441

SelectaVision Project

M.L. McNeeley|H. Rees
Apparatus for producing injection molded and centrally apertured disc records—
3989436

F.R. Nyman|J.L. Vossen, Jr.
D.G. Fisher|G.F. Nichols
Metal coating for video disc—3982066

RCA Ltd. Belgium

H.J. Digneffe
Alternating current control system—
3990000

Dates and Deadlines

Upcoming meetings

Ed. Note: Meetings are listed chronologically. Listed after the meeting title (in bold type) are the sponsor(s), the location, and the person to contact for more information.

JAN 10-14, 1977 — **AIAA 13th Annual Mtg. and Technical Display**, Wash. DC **Prog Info:** Martin Newman, AIAA, 1290 Sixth Ave., New York, NY 10019

JAN 18-20, 1977 — **Reliability and Maintainability Conf.** (IEEE) Marriott, Phila., PA **Prog Info:** J.M. Wiesen, Dept. 1220, Sandia Labs, Albuquerque, NM 87115

FEB 7-9, 1977 — **Aerospace and Electronic Systems Winter Convention (WINCON)** (IEEE) Sheraton-Universal, North Hollywood, CA **Prog Info:** H.S. Abrams, Litton Systems, Inc., G&CS Div., 5500 Canoga Ave., Woodland Hills, CA 91364

FEB 16-18, 1977 — **Intl. Solid State Circuits Conf.** (IEEE) Sheraton, Phila., PA **Prog Info:** John H. Wuorinen, Bell Labs, Whippany, NJ 07981

FEB 22-24, 1977 — **Optical Fiber Transmission Conf.** (IEEE, OSA) Williamsburg Lodge, Williamsburg, VA **Prog Info:** R.D. Maurer, Sullivan Park, Corning Glass Works, Corning, NY 14830

FEB 28-MAR 3, 1977 — **COMPCON Spring** (IEEE) Jack Tar, San Fran., CA **Prog Info:** L.C. Hobbs, Hobbs Assoc., POB 686, Corona Del Mar, CA 92625

MAR 16-18, 1977 — **Vehicular Technology Conf.** (IEEE) Orlando Hyatt House, Orlando, FL **Prog Info:** G.F. McClure, 1730 Shiloh Lane, Winter Park, FL 32789

MAR 17-19, 1977 — **Simulation Symp.** (IEEE) Tampa, FL **Prog Info:** Ira M. Kay, POB 22573, Tampa, FL 33622

MAR 21-23, 1977 — **Industrial Applications of Microprocessors** (IEEE) Sheraton, Phila., PA **Prog Info:** S.J. Vahaviolos, Eng. Research Ctr., Western Electric Corp., POB 900, Princeton, NJ 08540

MAR 21-23, 1977 — **American National Metric Council, Third Annual Conf. and Exposition**, McCormick Inn, Chicago, IL **Prog Info:** ANMC, Conference Committee, 1625 Massachusetts Ave., N.W., Suite 501, Wash., DC 20036

MAR 23-25, 1977 — **Computer Architecture Symp.** (IEEE, ACM) College Park, MD **Prog Info:** Bruce Wald, Naval Research Lab., 4555 Overlook Ave., Washington, DC 20390

MAR 18-21, 1977 — **Semiconductor Power Converter Intl. Conf.** (IEEE) Walt Disney Contemporary, Orlando, FL **Prog Info:**

Eberhart Reimers, USAMERDC, AMXFB-EA, Electrical Equip. Div., Fort Belvoir, VA 22060

APR 6-8, 1977 — **Digital Processing of Signals in Communication** (IERE, IEEE) Loughborough, UK **Prog Info:** IERE, 8-9 Bedford Square, London WC1B 3RG, England

APR 18-20, 1977 — **Fifth Conf. on Chemical and Molecular Lasers**, Stouffer's Riverfront Inn, St. Louis, MO **Prog Info:** Dr. W.Q. Jeffers, Helios Inc., POB 2190, Boulder, CO 80302

APR 18-21, 1977 — **Design Engineering Conf. and Show**, New York, NY **Prog Info:** Technical Affairs Dept., ASME, United Engineering Center, 345 East 47th Street, New York, NY 10017

APR 19-21, 1977 — **Society for Information Display Intl. Symp. and Exhibition**, Sheraton-Boston Hotel, Boston, MA **Prog Info:** Lewis Winner, 152 West 42nd St., New York, NY 10036

APR 21-22, 1977 — **Picture Data Description and Management** (IEEE) U. of Ill., Chicago, IL **Prog Info:** K.S. Du, Purdue Univ., School of EE, Lafayette, IN 47907

APR 23-28, 1977 — **American Ceramic Soc., Electronics Div., 179th Annual Mtg. & Exposition**, Conrad Hilton Hotel, Chicago, IL **Prog Info:** Dr. Richard M. Rosenberg, E.I. duPont de Nemours & Co., Inc., Photo Products, Bldg. 428, Buffalo Ave., Niagara Falls, NY 14302

APR 25-27, 1977 — **Circuits & Systems Intl. Symp.** (IEEE) Del Webb's Towne House, Phoenix, AZ **Prog Info:** W.G. Howard, Motorola Integrated Circuits Center, Mail Stop MR, POB 20906, Phoenix, AZ 85036

MAY 2-5, 1977 — **Offshore Technology Conf.** (IEEE et al) Astrodome, Albert

Thomas Convention Ctr., Houston, TX **Prog Info:** OTC, 6200 N. Central Expressway, Dallas, TX 75206

MAY 3-6, 1977 — **EUROCON 77** (IEEE et al) Venice, Italy **Prog Info:** Alberto Vandind Buti, c/o AEI, Viale Monza 259, 20126 Milan, Italy

JUN 1-3, 1977 — **Conf. on Laser Engineering and Applications** (IEEE/OSA) Wash., DC **Prog Info:** Conf. Mgr.: Anne J. Morandiere, Courtesy Assoc., Suite 700, 1629 K Street, N.W. Wash., DC 20006

JUN 20-22, 1977 — **14th Design Automation Conf.** (ACM/IEEE) International Hotel, New Orleans, LA **Prog Info:** David W. Hightower, MS907, Texas Instruments, Inc., POB 5012, Dallas, TX 75222

Calls for papers

Ed. Note: Calls are listed chronologically by meeting date. Listed after the meeting (in bold type) are the sponsor(s), the location, and deadline information for submittals.

APR 26-28, 1977 — **Microwave Power Tube Conf.**, Naval Postgraduate School, Monterey, CA **Deadline Info:** (abst) 1/30, Lynwood Cosby, U.S. Naval Research Laboratory, 4555 Overlook Ave., S.W., Wash., DC 20390

JUN 20-24, 1977 — **Intl. Electromagnetic Symp.** (IEEE, URSI) Stanford U., Palo Alto, CA **Deadline Info:** 3/1, Prof. K.K. Mei, Dept. of EE and CS U. of California, Berkeley, CA 94720

JUL 26-29, 1977 — **Intl. Conf. on Crime Countermeasure Science and Engineering**, (IEEE et al) Oxford University, Oxford, England **Deadline Info:** (Abst) 1/14, John S. Jackson, *Proceedings* Editor, Dept. of Electrical Engineering, U. of Kentucky, Lexington, KY 40506

Clip out and mail to Editor, *RCA Engineer*, 204-2, Cherry Hill, N.J.

RCA Engineer

Have we your correct address?
If not indicate the change below:

Name _____ Code # _____

Street or Bldg. _____

City and State or Plant _____

*Please indicate the code letter(s) that appear next to your name on the envelope.

Engineering News and Highlights

Automated Systems team of nine cited for laser rangefinder work



Design to Unit Production Cost of the AN/GVS-5 Hand-Held Laser Rangefinder brought this team a TE award. From left to right are: Anthony Amato, Manager Products Engineering; team members Fred Pratt, Sam Waldstein, Jay Woodward, Walt Radcliffe; Dr. Harry Woll, Div. Vice President and General Manager; team members Bob Guyer, Ferd Martin, Larry Blundell, Norm Roberts, and Jim Quinn; Gene Stockton, Chief Engineer; and Bill Hannan, Manager Radiation Systems Engineering.

Hilal wins award at Government Communications Systems



Design and Development of TENLEY Dedicated Loop Encryption Device won a Technical Excellence Award for George Hilal. Left to right are Hilal; J.V. Fayer, Manager; N. VanDelft, Leader; and C.A. Schmidt, Manager.

Bosselaers cited for Equate accomplishments at Automated Systems

Robert Bosselaers received the Technical Excellence Individual Award for his accomplishments in the advancement of the Equate RF Synthesizer design. RF Synthesizers, up to 18 GHz with varying modulation types, have been a key factor in maintaining RCA's front running position in the ATE product line.

Left to right: Dr. Harry Woll, Div. Vice President and General Manager; Bob Bosselaers; and Gene Stockton, Chief Engineer.



Behnen



Bronecke



Landry



Pschunder



Socci



Udicious

Six receive technical excellence awards at Moorestown

Steve Behnen—for his outstanding achievements in developing the weapon selection function for the AEGIS ship Combat Direction System.

Pete Bronecke—for his work in transforming a standard automatic integrated-circuit tester (DATRON 4400) into a special high-speed device for testing complex hybrid circuits.

Norm Landry—for his outstanding performance in the development of an extremely short waveguide horn radiating element for phased array radars.

Ralph Pschunder—for further advancements in the DYNA series of structural analysis computer programs. His updated DYNA-3 provides substantial cost reduction and greater ease of modeling of complex structures, and further includes a shock analysis subrouting incorporating specified Navy standards for shock analysis.

Bob Socci—for his contributions in the redesign of the AN/SPY-1 RF receiver in a new version that provides a 70% reduction in production cost with improved performance, accessibility, and reliability and maintainability.

Ralph Udicious—for his extraordinary achievement in the installation, checkout, and certification of the CMS-2Y compile capability in the AN/TPQ-27 program support facility.

Licensed engineers

When you receive a professional license, send your name, PE number (and state in which registered), RCA division, location, and telephone number to: *RCA Engineer*, Bldg. 204-2, RCA, Cherry Hill, N.J. New listings (and corrections or changes to previous listings) will be published in each issue.

RCA Records Division

August Skele, Indianapolis, Ind.; PE-16489 Ind.

Melvin K. Martin, Indianapolis, Ind.; PE-16685 Ind.

Robert E. Fuller, Indianapolis, Ind.; PE-16697 Ind.

Picture Tube Division

Paul Justus, Circleville, Ohio; PE-41191 Ohio.

Obituary

Frank J. Bingley, who retired from RCA in 1971 after nine years as an engineer with Astro Electronics, died on November 16. He was 70. Mr. Bingley was a graduate of the University of London in physics and mathematics. He came to this country in 1931 to work for the Philco Corp. after developing some of the earliest scanning systems for television for the Baird Television Co. of London.

As chief television engineer for Philco, he designed and helped construct one of the world's first all-electric experimental television stations.

In 1962, Mr. Bingley joined RCA Astro Electronics where he helped in the development of a high resolution tv camera tube for the Earth Resources Satellite and the RCA Apollo color television camera used to provide live telecasts from the moon.

He published numerous technical papers in his field, and in 1956 was awarded the Vladimir K. Zworykin prize of the Institute of Radio Engineers as outstanding television engineer of the year.



Promotions

Missile and Surface Radar

C. Laible from Mbr. Engrg. Staff to Sr. Mbr. Engrg. Staff (M. Korsen, Pedestals/RF/Trans.)

R. Wasson from Mbr. Engrg. Staff to Sr. Mbr. Engrg. Staff (M. Korsen, Electrical Integration)

A. Dennison from Princ. Mbr. Engrg. Staff to Ldr. Engrg. Sys. Projects (R. Howery, Comp Prog AN/SPY1A)

H. Millett from Mbr. Engrg. Staff to Sr. Mbr. Engrg. Staff (C. Martin, Install. Documentation)

E. Britt from Sr. Mbr. Engrg. Staff to Ldr. Sys. Engrg. (B. Inman, Ship Integration)

R. Creedon from Sr. Mbr. Engrg. Staff to Ldr. Syst. Engrg. (B. Inman, Ship Integration)

N. Ricciardi from Mbr. Engrg. Staff to Sr. Mbr. Engrg. Staff (J. Nessmith Hybrid Radar/CAMEL)

C. Tipple from Sr. Mbr. Engrg. Staff to Ldr. Engrg. Sys. Projects (J. Nessmith, Man/Machine)

L. Clayton from Sr. Mbr. Engrg. Staff to Princ. Mbr. Engrg. Staff (S. Steele, Comb. Software Systems)

D. Haimowitz from Mbr. Engrg. Staff to Sr. Mbr. Engrg. Staff (S. Steele, Comp. Prog. Integ/Simu.)

R. Rader from Sr. Mbr. Engrg. Staff to Ldr. Engrg. Sys. Projects (S. Steele, Weap. Sys. Soft Devel.)

J. Schwaninger from Princ. Mbr. Engrg. Staff to Ldr. Engrg. Sys. Projects (S. Steele, Dist. Data Proc. Sys.)

C. White from Princ. Mbr. Engrg. Staff to Ldr. Engrg. Sys. Projects (S. Steele, Weap. Sys. Soft Devel.)

C. Yeisley from Sr. Mbr. Engrg. Staff to Ldr.

Engrg. Sys. Projects (S. Steele, Adv. Software Engrg.)

H. Bharmal from Mbr. Engrg. Staff to Sr. Mbr. Engrg. Staff (S. Steele, Weap. Sys. Soft Devel.)

Astro Electronics

S. Ravner from Staff Systems Scientist to Mgr., Speciality Engineering (T. Aukstikalnis)

W. Lindorfer from Staff Systems Scientist to Mgr., Speciality Engineering (H. Curtis)

J. Balcewicz from Engineer to Mgr., Speciality Engineering (H. Curtis)

H. Curtis from Staff Systems Scientist to Mgr., Engineering (W. Manger)

R. DeBastos from Staff Systems Scientist to Mgr. Project (A. Schnapf)

A. Garman from Mgr. Project to Mgr. Communications Satellite Programs (C. Constantino)

Consumer Electronics

L.A. Cochran from Sr. Mbr. Engineering Staff to Mgr., Signal Processing Engineering (J.P. Bingham, Engineering Development)

Solid State Division

J. Sundburg from Ldr., Technical Staff to Mgr., Bipolar Assembly and Test, Manufacturing

J. Arico from Mbr., Technical Staff to Ldr., Technical Staff.

Staff announcements

RCA Laboratories

Kerns H. Powers, Director of the Communications Research Laboratory, has announced the organization as follows: **Guy W. Beakley**, Head, Image Processing Research; **Jon K. Clemens**, Head, Signal Systems Research; **James J. Gibson**, Fellow Technical Staff; **H. Nelson Crooks**, Manager, High-Density Recording Project; **William D. Houghton**, Head, Special Projects Research; **Eugene O. Keizer**, Head, Video Systems Research; **Charles B. Oakley**, Head, Broadcast Systems Research; **Robert E. Flory**, Fellow, Technical Staff; **J. Guy Woodward**, Fellow, Technical Staff; and **Daniel A. Walters**, Head, Satellite Systems Research.

Paul Rappaport, Director of the Process and Applied Materials Research Laboratory, has announced the organization as follows: **Glenn W. Cullen**, Head, Materials Synthesis

Research; **Chih C. Wang**, Fellow, Technical Staff; **Leonard P. Fox**, Head, "SelectaVision" Processing Research; **Bernard Hershenov**, Staff Advisor; **Richard E. Honig**, Head, Materials Characterization Research; **D. Alex Ross**, Manager, Division Liaison; **Daniel L. Ross**, Head, Organic Materials and Devices Research; **George L. Schnable**, Head, Process Research; **Charles W. Mueller**, Fellow, Technical Staff; **Brown F. Williams**, Head, Optical Materials and Devices Research; and **Richard Williams**, Fellow, Technical Staff.

Commercial Communications Systems Division

Andrew F. Inglis, Division Vice President and General Manager of the Commercial Communications Systems Division, has announced the organization as follows: **William L. Firestone**, Division Vice President and General Manager, Avionics Systems; **Joseph P. Ulasewicz**, Division Vice President and General Manager, Mobile Communications Systems; **Jack F. Underwood**, Division Vice President, Plans and Programs; and **Neil Vander Dussen**, Division Vice President and General Manager, Broadcast Systems.

Distributor and Special Products Division

Paul B. Garver, Division Vice President and General Manager of Distributor and Special Products Division, has announced the following appointments: **Gene W. Duckworth**, Division Vice President, Business Development and International Operations; and **Fred G. Wenger**, Division Vice President, Sales.

Solid State Division

Thomas T. Lewis, Director of Electro-Optics Operations, has announced the organization as follows: **Thaddeus J. Grabowski**, Manager, Market Planning—Displays & Emitters; **Clarence H. Groah**, Manager, Operations Control; **Leonard W. Grove**, Manager, Manufacturing—Electro-optics; **N. Richard Hangen**, Manager, Market Development—Research & Development; **Fred A. Helvy**, Manager, Applications Engineering—Electro-Optics; **Edward F. McDonough**, Manager, Market Planning—Photodetectors; **Ronald G. Powers**, Manager, Solid State Detectors—Canada; **Carl L. Rintz**, Manager, Market Planning—Imaging Devices; and **Eugene D. Savoye**, Manager, Engineering—Electro-Optics.

Harold R. Krall, Manager of Electro-Optics Equipment Operations, has announced that the responsibility for the manufacturing,

marketing, and engineering of Charge Coupled Device (CCD) Cameras will be assumed by Electro-Optics Equipment Operations. He announced the organization of Electro-Optics Equipment Operations as follows: **Carl F. Adams, Jr.**, Administrator, Operations Control; **David E. Bowser**, Leader, Engineering—Subassemblies & Equipment; **Joseph T. Gote**, Manager, Manufacturing—Electro-Optics Equipment; **Victor C. Houk**, Manager, Market Planning—Closed-Circuit Video Equipment; and **Robert L. Rodgers**, Manager, Engineering—Closed-Circuit Video Equipment.

Leonard W. Grove, Manager of Manufacturing—Electro-Optics, has appointed **William N. Henry** as Manager, Manufacturing—Silicon Devices (CCD and Si targets).

Carl R. Turner, Division Vice President of Solid State Power Devices, has announced the organization as follows: **Dale M. Baugher**, Director, Power Engineering; **Melvin Bondy**, Manager, Operations Planning & Administration—Power; **Ralph S. Hartz**, Manager, Mountaintop Engineering; **John E. Mainzer**, Director, Power Manufacturing Operations; and **Donald Watson**, Director, Product Marketing—Power.

Ralph S. Hartz, Manager of Mountaintop Engineering, has announced the organization as follows: **Donald E. Burke**, Manager, Device Development Engineering; **Robert J. Satriano**, Manager, Equipment Engineering; and **Louis V. Zampetti**, Manager, Type Engineering.

Dale M. Baugher, Director of Power Engineering, has announced the organization as follows: **Jacques M. Assour**, Leader, Process Engineering; **Dale M. Baugher**, Acting Administrator, New Product Planning; **Leonard H. Gibbons**, Manager, Applications Engineering—Power; and **Wallace D. Williams**, Leader, Test Engineering.

Norman C. Turner, Manager of Somerville MOS Operations, has announced the organization of Photomask Operations as follows: **David S. Jacobson**, Manager, Photomask Engineering & Tooling; and **Evan P. Zlock**, Manager, Photomask Production.

Picture Tube Division

D. Joseph Donahue, Division Vice President of Engineering, has appointed **James C. Miller**, Manager, Technical Projects and Engineering Administration.

Matthew M. Bell, Manager of Equipment Engineering, Circleville Glass Operations, has announced the following appointments: **F.L. Armstrong**, Manager Equipment Design Engineering—Forming; **B.B. LeMay**, Manager, Equipment Design Engineering—Finishing; and **R.W. Marshall**, Manager, Instrumentation Engineering.



Tannenbaum is TPA for Government Communications Systems

Dan Tannenbaum has been appointed Technical Publications Administrator for Government Communications Systems. In this capacity, Mr. Tannenbaum is responsible for reviewing and approving technical papers; for coordinating the technical reporting program; and for promoting the preparation of papers for the *RCA Engineer* and other internal and external journals.

Mr. Tannenbaum is currently Staff Engineer for the Government Communications Systems Division in Camden; his more recent responsibilities as Manager of Digital Equipment Engineering have been in the design and development of digital equipment, tactical transmission processing equipment, and telephone switching terminals. Mr. Tannenbaum has been with RCA for 27 years after receiving the BS in Engineering Physics from the University of Michigan in 1949. He is a licensed professional engineer in New Jersey and is also a senior member of the IEEE.

Mobile Communications Systems

George J. Mitchell, Manager of Mobile Product Operations, has appointed **Lee F. Crowley**, Manager, Communications Products Engineering and Product Management.

Service Company

Raymond J. Sokolowski, Division Vice President, Consumer Services, has appointed **J. William McGee** Director, Technical Support.

Editorial Representatives

The Editorial Representative in your group is the one you should contact in scheduling technical papers and announcements of your professional activities.

Commercial Communications Systems Division

Broadcast Systems W.S. SEPICH* Broadcast Systems Engineering, Camden, N.J.
K. PRABA Broadcast Systems Antenna Equip. Eng., Gibbsboro, N.J.
A.C. BILLIE Broadcast Engineering, Meadow Lands, Pa.

Mobile Communications Systems F.A. BARTON* Advanced Development, Meadow Lands, Pa.

Avionics Systems C.S. METCHETTE* Engineering, Van Nuys, Calif.
J.McDONOUGH Equipment Engineering, Van Nuys, Calif.

Government Systems Division

Astro-Electronics E.A. GOLDBERG* Engineering, Princeton, N.J.

Automated Systems K.E. PALM* Engineering, Burlington, Mass.
A.J. SKAVICUS Engineering, Burlington, Mass.
L.B. SMITH Engineering, Burlington, Mass.

Government Communications Systems D.A. TANNENBAUM* Engineering, Camden, N.J.
H.R. KETCHAM Engineering, Camden, N.J.

Government Engineering M.G. PIETZ* Advanced Technology Laboratories, Camden, N.J.

Missile and Surface Radar D.R. HIGGS* Engineering, Moorestown, N.J.

Research and Engineering

Corporate Engineering H.K. JENNY* Technical Information Programs, Cherry Hill, N.J.

Laboratories C.W. SALL* Research, Princeton, N.J.

Solid State Division

J.E. SCHOEN* Engineering Publications, Somerville, N.J.
H.R. RONAN Power Devices, Mountaintop, Pa.
S. SILVERSTEIN Power Transistors, Somerville, N.J.
A.J. BIANCULLI Integrated Circuits and Special Devices, Somerville, N.J.
J.D. YOUNG IC Manufacturing, Findlay, Ohio
R.W. ENGSTROM Electro-Optics and Devices, Lancaster, Pa.

Consumer Electronics

C.W. HOYT* Engineering, Indianapolis, Ind.
R.J. BUTH Engineering, Indianapolis, Ind.
P.E. CROOKSHANKS Television Engineering, Indianapolis, Ind.
C.P. HILL Manufacturing Technology, Indianapolis, Ind.

SelectaVision Project

F.R. HOLT SelectaVision VideoDisc Engineering, Indianapolis, Ind.

RCA Service Company

J.E. STEOGER* Consumer Services Engineering, Cherry Hill, N.J.
R. MacWILLIAMS Marketing Services, Government Services Division, Cherry Hill, N.J.
R.M. DOMBROSKY Technical Support, Cherry Hill, N.J.

Distributor and Special Products Division

C.C. REARICK* Product Development Engineering, Deptford, N.J.

Picture Tube Division

E.K. MADENFORD* Engineering, Lancaster, Pa.
N. MEENA Glass Operations, Circleville, Ohio
J.I. NUBANI Television Picture Tube Operations, Scranton, Pa.
C.W. BELL Engineering, Marion, Ind.

RCA Communications

Alascom P.WEST* RCA Alaska Communications, Inc., Anchorage, Alaska
Americom D.L. LUNDGREN* RCA American Communications, Kingsbridge Campus, N.J.
Globcom W.S. LEIS* RCA Global Communications, Inc., New York, N.Y.

RCA Records J.F. WELLS* Records Eng., Indianapolis, Ind.

NBC

W.A. HOWARD* Staff Eng., Technical Development, New York, N.Y.

RCA Ltd

H.A. LINKE* Research & Eng., Montreal, Canada

Patent Operations

J.S. TRIPOLI Patent Plans and Services, Princeton, N.J.

Electronic Industrial Engineering

J. OVNICK* Engineering, N. Hollywood, Calif.

*Technical Publications Administrators (asterisked * above) are responsible for review and approval of papers and presentations.

RCA Engineer

A TECHNICAL JOURNAL PUBLISHED BY CORPORATE TECHNICAL COMMUNICATIONS
"BY AND FOR THE RCA ENGINEER"

Printed in U.S.A.

Form No. RE-22-4