VOLUME 6 □ NUMBER 9 IEEE spectrum SEPTEMBER 1969

IEEE spectrum

## features



THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

# Three new reasons to specify

# DIGIEC


Model 262

## Digital Multimeter

**(AC, DC, Ohms)**

A .1% Multimeter for everyday usage. A necessity for design and development, production, quality control, or anywhere that DC volts and current, AC volts, and ohms are measured. Battery pack available.

**$375**


Model 311

## Precision Calibrator

**(Voltage & current source)**

A .01% precision voltage and current Calibrator that serves as a working standard. In addition, the high current capability may be used as a lab source for developing those critical circuits so essential in electronics today.

**$650**

Model 691

## 21 Column Printer

**(3 lines per second)**

A Drum Printer which is expandable from 4 to 21 columns to satisfy your specific requirements. This versatile printer accepts all standard BCD inputs and provides 38 symbols along with "floating" decimal point.

starting at **$770**

DIGIEC® by **UNITED SYSTEMS CORPORATION**

918 Woodley Road • Dayton, Ohio 45403 • (513) 254-6251

Representatives Throughout the world

For complete specifications request new catalog D69A

Circle No. 2 on Reader Service Card.

**the cover**

*Two little people move about in a house of light and shadow. The cover photo was made by taking multiple exposures of a statically deflected (rather than vibrating) varifocal mirror. This simulates the vibrating varifocal mirror described in the article beginning on page 37.*

## departments

**IEEE SPECTRUM** is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: 345 East 47 Street, New York, N.Y. 10017. Cable address: ITRIPLEE. Telephone: 212—752-6800. Published at 20th and Northampton Sts., Easton, Pa. 18042. **Change of address** can be made effective with any particular issue if received by IEEE Headquarters by the 10th of the preceding month. **Annual subscription:** IEEE members, first subscription $3.00 included in dues. Single copies $1.00. Prices for nonmember subscriptions available on request. **Editorial correspondence** should be addressed to IEEE SPECTRUM at IEEE Headquarters. **Advertising correspondence** should be addressed to IEEE Advertising Department, at IEEE Headquarters. Telephone: 212—752-6800.




**OTHER IEEE PUBLICATIONS:** IEEE also publishes the PROCEEDINGS OF THE IEEE, the IEEE STUDENT JOURNAL, and more than 30 Transactions for IEEE Groups with specialized interests within the electrical and electronics field. Manuscripts for any IEEE publication should be sent to the editor of that publication whose name and address are shown on page 10A of the January 1969 issue. When in doubt, send the manuscript to E. K. Gannett, Editorial Services, at IEEE Headquarters, for forwarding to the correct party.

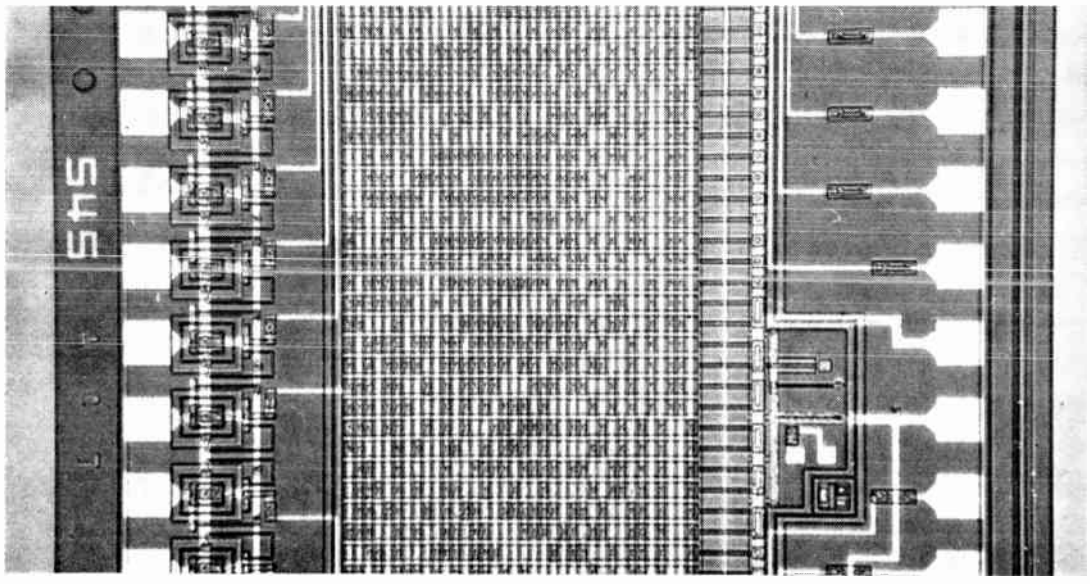


**FIGURE 1. Portion of an LSI digital word translator that contains more than 12 650 active components per square centimeter. The integrated subsystem translates a 6-bit input word to 64 24-bit output words.**

# The art of building LSIs

***Large-scale integrated circuits can be made in substantial quantities once it is recognized that the success of the final product depends on the proper application of the interrelated factors of design, fabrication, and testing***

**Herschel T. Hochman** *Honeywell Inc.*

*Today solid-state technology has improved to such a degree that very large IC chips are being made with yields up to 30 percent. This article highlights some of the main elements that determine whether LSIs can be manufactured in large quantities. The author claims that obtaining reasonable yields is readily achievable once the factors that stand out as deterrents to success are recognized. Previously developed manufacturing techniques are used for LSI production. In the future, however, lasers and the computer will be applied to enable engineers to produce components faster with greater reliability.*

Although large-scale integrated (LSI) circuits are still in their infancy, many good, sophisticated circuits can be made using some previously developed solid-state manufacturing techniques. Properly applied, these well-tested methods can be adapted to standard circuits to increase yields and lower costs. In the future however, the laser and computer, and new alignment equipment, will make many of today's rules and methods obsolete.

Confusion and a number of misconceptions fog the meaning of the term "LSI." Therefore, before discussing circuit designs or manufacturing problems I shall attempt to clear the air by giving what I believe is a workable definition of large-scale integrated circuits: A large-scale integrated circuit is a silicon chip on which there have been deposited a large number of active and/or passive components connected in such a way as to perform a multitude of circuit functions. In the past, criteria such as the number of gates, size of the chip, and/or components per square centimeter have been used to classify LSI devices. However, when used to describe compactness or complexity, these terms are misleading and nondescriptive. When one examines an LSI chip, he finds that variations in component density can occur depending upon the percentage of silicon used by the passive and active components.

For example, Fig. 1 is a circuit containing approximately 2400 components in an active area (excluding bonding pads) of about 3000 × 5000 micrometers. This gives a count of slightly more than 16 000 components per cm². If the bonding area is included, the count is still more than 12 650 components per cm². The ten-stage counter shift register shown in Fig. 2 contains more than 7000 components per cm². About 65 percent of the chip area is taken up by passive resistors.

## Design considerations

In the early days of solid-state circuit design, packaging, artwork, and testing were of little importance during the initial design phase. At the start of an LSI circuit design, however, these considerations are every bit as important as the calculation of the device parameters and tolerances. A simple lack of attention to packaging details can add thousands of dollars and months of effort to a project. It is imperative that testing receive early attention. But because there are no simple methods to check LSI circuits, testing at the wafer stage is minimal. Attempting to trace a circuit containing over 2400

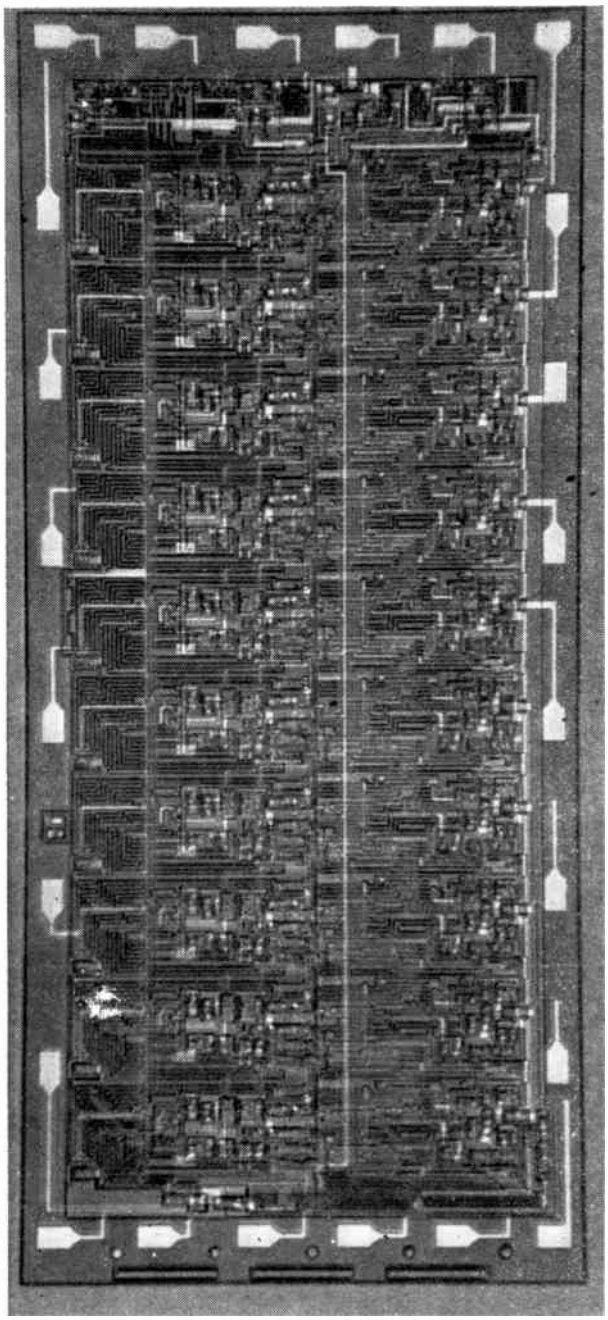


**FIGURE 2. This monolithic subsystem can be used as a ten-stage counter or shift register. About 65 percent of the chip area is taken up by passive resistors.**

active diffused components with more than 4000 junctions is an almost impossible task. But with proper forethought, pads can be brought out so that certain portions of the circuit can be tested and components added for parameter evaluation. In addition, space for test probe points, or special test interconnect patterns, can be added to the device's metal mask.

Since large-scale integration implies maximum density of components in minimum area, the circuit designer is presented with a challenge to engineer the circuit operation with devices that may not be optimized for the particular current or power they will be required to handle. In other words, given a choice of components, those designed into the LSI would not represent those used if the circuit were breadboarded of discrete elements.

What then should be the criteria for the design in terms of component parameters? First, let us look at the resistors. Even though only diffused resistors are used, total circuit resistance can be very high. For example, the resistance in the circuit of Fig. 2 is greater than 7.8 megohms. Three methods are used to achieve high resistance values: sheet resistivity of the base diffusion; very long and narrow geometric configurations; and buried resistors.

Whether or not sheet resistivity is used depends, to a degree, upon the tolerance that must be maintained over the wafer. The author has found that with 200 ohms/square an overall 10 percent tolerance is possible. In a group of wafers whose sheet resistances varied from 125 ohms/square to 150 ohms/square, it was possible to achieve a 5 percent tolerance from run to run and wafer to wafer. Generally, the lower the resistivity, the tighter the tolerance spread.

As far as long, narrow geometric configurations are concerned, some highly sophisticated mask-making equipment that is available today can easily make 6-$\mu$m line widths with consistency. However, it is difficult to control the accuracy and thus the resistance of large numbers of these lines over a 6000- by 6000-$\mu$m area.

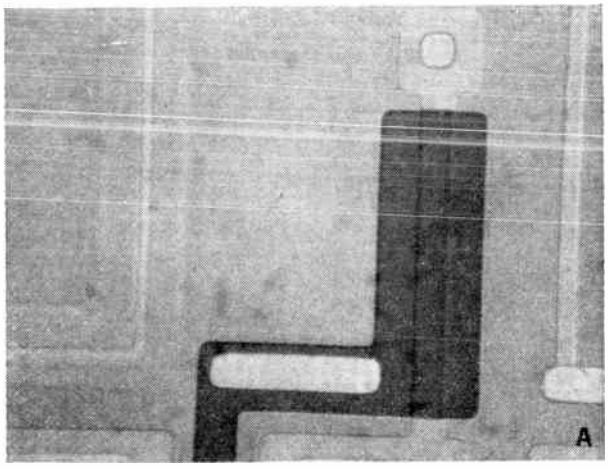
Because up to 10 000 ohms/square can be obtained, it is sometimes advantageous to use buried resistors in an LSI circuit design. But although the high ohmic capability is desirable, there are also disadvantages. It is extremely difficult to maintain close control of resistor tolerances and keep a desired transistor beta. Also, transistor parameters continuously change as the emitter diffusion process acts to establish the proper resistance value. When using this technique, the designer must carefully take into account worst-case conditions. The structure of a typical buried resistor is shown in Fig. 3. It is similar to the junction field-effect transistor and produces the nonlinear resistance characteristic shown in Fig. 3(C). This limits its usefulness.

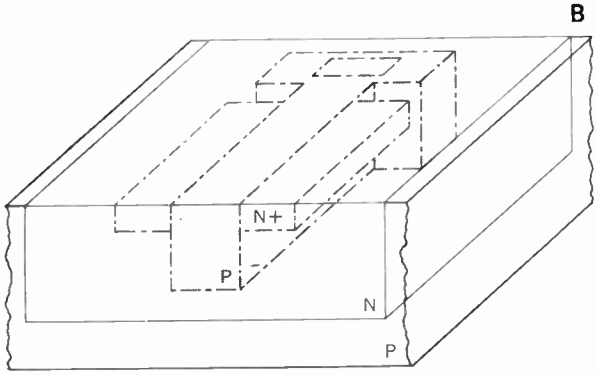
Test transistors can be used as an aid in the development of custom or prototype LSI circuits. These can be used for beta adjustment, which aids fine tuning of the fabrication process. It is almost impossible to use meter probes to check an LSI transistor, but if a fairly large standard transistor is used, its beta can be correlated with those of the small internal transistors. Figure 4 shows the types of transistors used in most LSIs. In addition to the test transistor, the engineer should also include a resistor tapped at 10 000 and 25 000 ohms in his test kit. This resistor can be used to verify earlier sheet-resistance measurements and to check the effectiveness of the metal-to-silicon alloying. Contact problems can be seen quickly.

One more point should be mentioned. The designer will probably find it necessary to utilize underpasses or tunnels for the LSI conductor circuit. With the high component density of LSI wafers, it's impossible to lay out a circuit without using this technique. More on this will be discussed later in the article.

## Mask fabrication

The two items that probably determine the success or failure of an LSI circuit are the masks and epitaxial layer. A book could be written about each, but only the



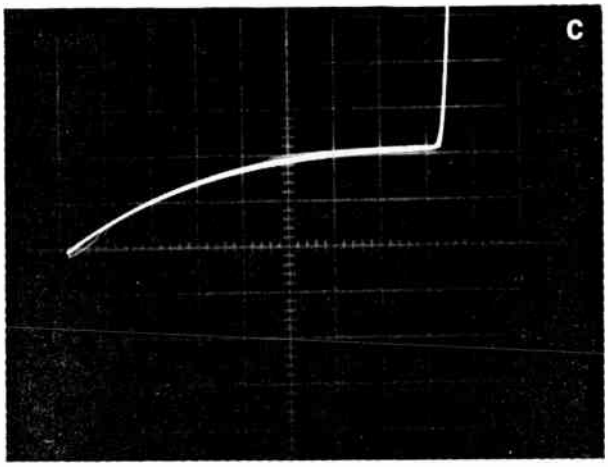




**FIGURE 3. Photomicrograph (A) of an LSI buried resistor (n+ region), cross section (B) of the buried resistor area, and nonlinear resistance characteristics (C).**

important points will be discussed here. Generally the discussion will center around mask fabrication.

The size of the artwork is not fixed and is usually determined by several factors, such as chip size, drafting skill, copy board size, first reduction capability, and the step and repeat reduction capability. The ratio of the circuit artwork (layout cuts) to final chip size can be 200, 250, 400, or 500. For example, a 7620- by 7620-$\mu$m final chip cut 500 times would need a copy board and artwork approximately 4 by 4 meters.

Standard rubylith film comes in 4-foot widths; so obviously if the cutting ratio is 500, standard techniques cannot be used. If, on the other hand, the cutting ratio is 300, the accuracy is reduced and the minimum final line width is approximately 6.35 $\mu$m. (Most draftsmen have difficulty cutting less than a 1270-$\mu$m width on the ruby.) It is obvious that new techniques have to be employed. These can take the form of photocomposing or paste-on.

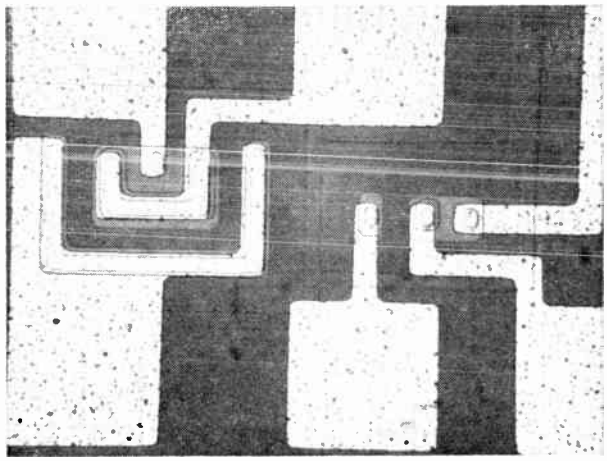


**FIGURE 4. Typical transistor geometries used in LSIs.**

Photocomposing is a technique of splitting the artwork into sections, making a first reduction of each section, and then piecing each section together on a master plate during the step and repeat operation. This requires the utmost precision; but this precision is obtainable with currently available equipment. Designers should anticipate the use of this technique when constructing the original layout since special considerations have to be made.

Paste-on artwork is art (usually photographs of a repetitive circuit) that has been pieced together on the copy board. After piecing, the paste-on is reduced and stepped. The success of this procedure depends upon the dimensional stability of the film as well as the accuracy maintained while piecing the parts together. When the artwork is ready for the first reduction, it must be handled with care. Stretching or bulging result in poorly registered masks and additional shootings.

The LSI wafers contain many circuits and perform a multitude of functions. It is obvious, therefore, that each mask must have an extremely low defect level. The most critical phase of the LSI manufacturing process is that of making the working plate. Most circuits are in the 2500- by 2500-$\mu$m category. On a 3.8-cm wafer (disregarding edge devices) there are approximately 165 usable devices; on a 7500- by 7500-$\mu$m device there are about 18 devices. In the first case a defect level of 20 percent on each mask is tolerable; but in the second case, this level of defect would mean a yield of almost zero. The matter is further complicated by the fact that six different masks (each with its own defect level) are used throughout the fabrication process.

Here is a list of possible defects:

**1. Breaks.** These are particularly prevalent on the narrow lines used for circuit isolation and narrow resistors. They cause devices to be short-circuited to each other as well as open resistors.

**2. Pinched or necked-down lines.** These lines develop

during the photoresist operation and result in the aforementioned defects.

**3. Image size variation.** Since tolerances are tight an oversize image could cause two diffused regions to overlap, such as the emitter into the collector.

**4. Poor image density.** This becomes a problem as lines become more narrow. On masks in a series that contain both 5-$\mu$m lines as well as wider, 25-$\mu$m lines, improper spacing of the images on the artwork may cause poor density or image deformation. This occurs because the light surrounding a small image can be scattered in the camera as well as in the photo repeater. As an analogy, it is very easy to see light through a pinhole in a dark room, but if light suddenly surrounds the pinhole, it obscures the small opening. Since these conditions can change with variables such as lens, lighting, and processing, experimental test patterns develop the proper rules to follow for the particular facility doing the work. These are not difficult to formalize, but they must be taken into consideration when developing an LSI technology.

**5. Contamination.** Under this heading falls a variety of particles. They can be caused by poor emulsion on the plate, improper handling in processing, and poor water and/or chemicals. A wafer containing only 18 to 20 circuits can tolerate very few contaminants. Depending on the mask used, these can create pinholes and open and short circuits. Point-of-use filters can be used to reduce particle size to less than 0.45 $\mu$m. They will remove resin particles as well as bacteria from the water system.

Additional contaminants in the form of photoresist are picked up each time the mask is used on a wafer. Projection printing methods in the future may eliminate this problem. Contamination is acceptable if the density is minimized and the particle size is less than 0.4 $\mu$m.

**6. Interference rings.** These are created by improper contact printing and result in open and short circuits. Figure 5 illustrates this condition and Fig. 6 shows the result on the die. With LSIs particularly, the closely spaced metal covering the die requires a mask that, unfortunately, is susceptible to this condition.

Due to the above conditions there are occasions when only one master plate is used per wafer. This is expensive, but is cheap when the value of the large integrated circuit is considered. We have made no distinction between emulsion and chrome masks since the above defects apply equally to both.

## Fabrication

Fabrication is the process of bringing together quality epitaxial layers into which precisely controlled impurities are diffused. Building the LSIs is not a "shotgun" technique; each fabrication step must be planned from the design stage, and each individual step interacts closely with all others.

**Material.** The quality of the epitaxial layer shares the spotlight with the mask as the key to successful fabrication. As with masks, the number of defects acceptable for an LSI wafer is far below that of the ordinary integrated circuit. Epitaxy today is generally good; the resistivity and thickness variation can be held to a tight tolerance. For example, resistivity below 1 ohm·cm can be held to $\pm 10$ percent whereas a thickness between 5 and 9 $\mu$m can be held to $\pm 5$ percent. Normally greater variations can be tolerated in resistivity. Other tolerable conditions include washout of the buried region or a slight lemon-peel finish. The one condition that must be eliminated is shown in Fig. 7. Here an epitaxial defect has formed in an emitter site due to a contaminant left on the wafer during epitaxial growth. This contamination can result from many sources including (1) initial wafer cleaning, (2) the buried slug diffusion process, (3) inadequate oxide removal, (4) the cleaning process, and (5) from within the reactor prior to epitaxial deposit. Source (5) can be eliminated if the etching (with hydrochloric acid) is done prior to deposition of the epitaxial material.

Contaminants come from unexpected sources. Wafers have been received from the manufacturer with polishing compound so heavily embedded in them that repeated washing with hard cotton swabs was required to remove the dirt. The cotton, of course, can work into the cleaning solution and eventually become imbedded on the surface of the wafer. To ensure that the wafers are truly clean, they should be carefully observed under both light and dark field magnification. Then the epitaxial growth can be assumed to be free of defects.

**Diffusion.** With the tremendous emphasis placed on

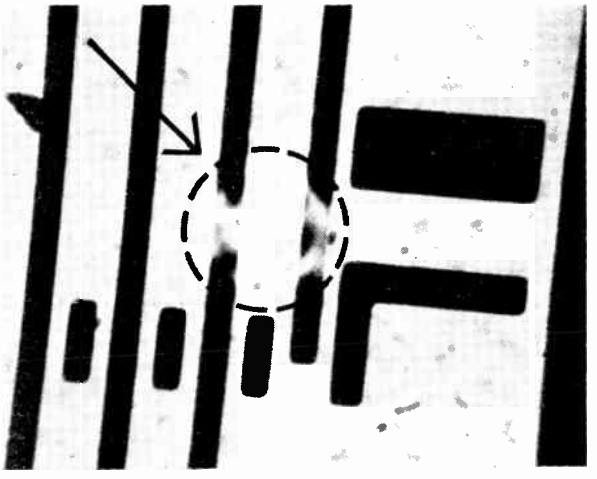
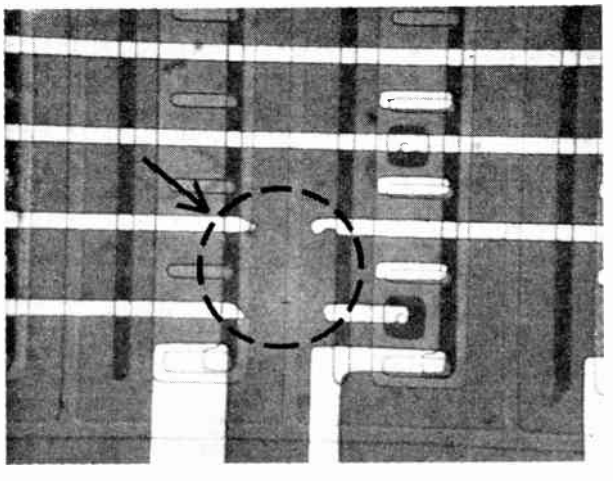


**FIGURE 5. Interference rings due to poor printing.**

**FIGURE 6. Open aluminum conductors caused by interference rings on the emulsion mask shown in Fig. 5.**



diffusion in the past few years, it is no surprise that this part of the fabrication process should give the least amount of trouble. Still, processing LSI wafers does require much care or in-line process control. Since each die covers such a large area or "length" of wafer, large beta and resistor variations can exist within the die. New types of flow systems, tube designs, and wafer-positioning equipment permit greater diffusion control and less parameter variation than ever before.

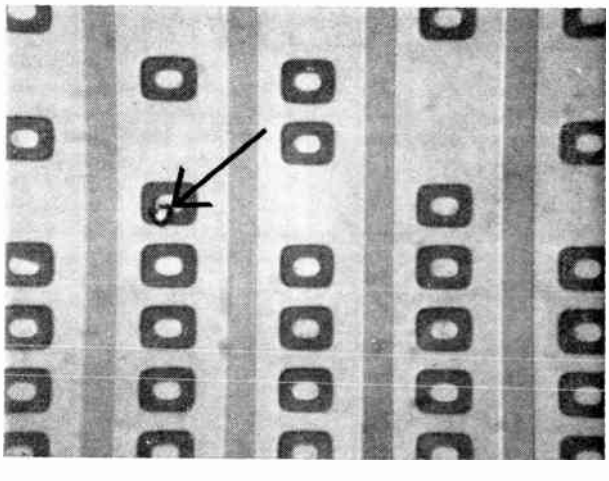
The first of the various diffusion processes that we will examine in detail is the isolation process. Isolation oxide openings are very close to the base region and impurities must be diffused only to a depth necessary to isolate the components, and no further. Too long a time at high temperature can cause a lateral diffusion, which may short-circuit the base region to the isolation region. This implies, of course, that the epitaxial thickness is uniform over the entire wafer surface. If, after epitaxial deposition, washout is severe, the isolation pattern will be misaligned over the buried regions and isolation will not take place since the impurities will cease penetration when it meets the heavily doped buried region.

With low sheet resistivity, it is easy to reproduce transistors with low base resistance ($r_b'$), high cutoff frequency ($f_t$), and matched $V_{BE}$. But the amount of resistance that the designer can put in the circuit is limited. Conversely, using high sheet resistance reduces circuit capacitance and power dissipation. Therefore, as with other integrated circuits, compromises[1] must be made. The differences lies in the reproducibility and control needed over large areas of silicon. For example, in some applications there is a need for accurate Zener voltages. For a given sheet resistivity, precise control of impurity depth is needed to accomplish this.

Underpasses or n+ tunnels (Fig. 8) are formed during the emitter diffusion process. Standard design rules should require the tunnels be placed in the base region since short-circuiting the n+ to the base during metalization will reduce the capacitance of the region. Some LSIs may have 10 to 20 of these tunnels.

As previously stated, buried resistors formed during the emitter diffusion process can be used as a means of obtaining high resistance values. If a wide range of resistor tolerances and transistor beta (10 to 25) is allowed the processing problem does not represent a major issue. As the required beta reaches the 40 or above level, the resistor becomes uncontrollable in terms of absolute value. In other words, the transistor beta cannot be used as a means to assess the value of the buried resistor. Even though base width is a prime factor in determining beta, other factors such as lifetime of the charge carriers, emitter efficiency, and surface effects can alter beta and give a false indication of what the resistor's value will be upon completion of the processing.

**FIGURE 7. Defect (arrow) in emitter site caused by contamination on wafer before growth of epitaxial layer.**



## Photolithography

The photoresist operation is a good focal point from which to view the integrated circuit operation. From this vantage point, the quality of the entire fabrication process can be evaluated. The photoresist process is extremely sensitive, and success depends heavily upon previous wafer treatment. Many wafers are rejected at this stage. Causes for rejection are warpage, misalignment, and contamination acquired during diffusion.

As devices get smaller, chip size gets larger, and wafers increase in diameter, the photoresist operation becomes the weak link in the chain. Photoresist must be thin enough to resolve 2.5-$\mu$m-width lines, yet not be subject to pinholes. Some new positive acting resists, although having better resolution capability, are limited in their resistance to certain types of etch solutions.

Pinholes can be reduced by double-exposing the film with two masks of the same layer; or by exposing once, shifting the mask to a new position, and exposing again. Any defect in the mask that would have otherwise allowed the resist to develop will shift, and the unexposed resist will be polymerized during the second exposure. Care must be exercised when performing this operation since the area or line width size will be reduced if the alignment is not accurate.

Thinner oxide can be used to shorten etch times. But if this is done, careful attention must be paid to contamination and the alloying process. A contaminated thin oxide can cause inversion layer problems due to the proximity of the metal and the silicon surface. Voltage potentials developed across oxide layers can induce channel currents.

The present-day equipment and photoresist techniques are well able to produce circuits such as those in Fig. 1 with 6.0-$\mu$m minimum line widths and 3-$\mu$m spacings between one diffused region and another. For example, the spacing between emitter oxide opening and base oxide opening is about 3 $\mu$m. Under laboratory conditions, some components with simple circuits have been produced that have 1.0-$\mu$m line widths. Although this is certainly a step in the right direction, producing thousands of these line widths over a large die, time and time again, is another matter.

As part of the photoresist operation, measurements of line widths are made on the mask, on the wafer after photoresist development, and after etching. In this way, process control is maintained and resistors can be accurately brought into specification (within 5 percent). Few alignment machines are available that can satisfactorily maintain dimensional control (in terms of exposure) on 3.8- to 5.0-cm LSI wafers. Depth of field is not satisfactory and major modifications must be made to adapt mechanically to the specific processing methods.

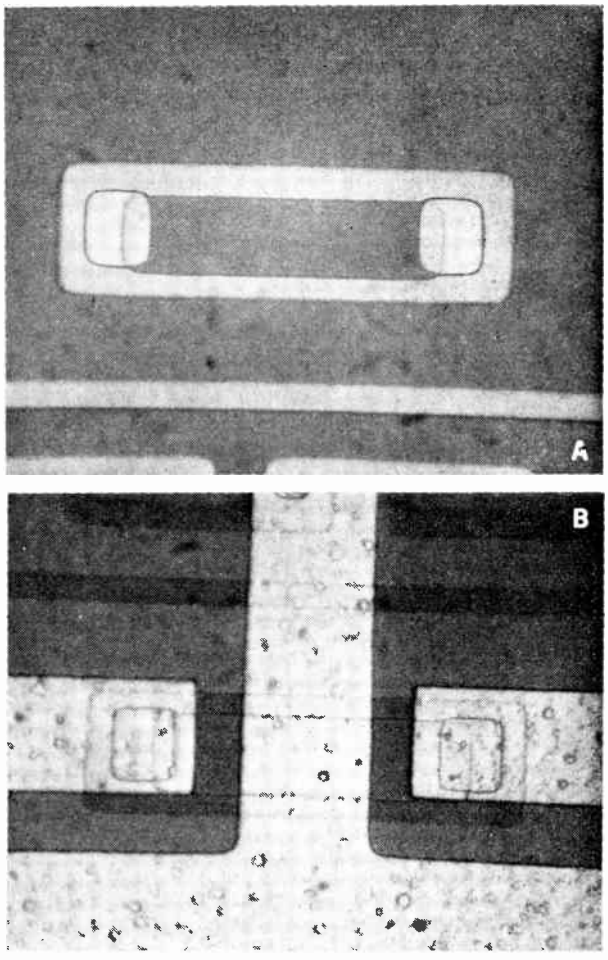



**FIGURE 8. Typical n tunnel shown (A) before metalization and (B) after the metalization has been completed.**

Chemicals and photoresist must also be tailored; and filtering and centrifuging become integral parts of the process. The operators also have to know the circuit as well as its topography. This is necessary because compensation is often made to remedy a slight misregistration or defect in the mask. Even though alignment marks are put on the mask, they do not necessarily imply perfect alignment of the circuit images over the wafer.

It is of utmost importance to standardize the photoresist process, test the process, and assign necessary correction factors in the early design phase. If, for example, during base oxide removal, the bases are slightly undersized due to improper photoresist techniques, the alignment tolerance for subsequent emitter oxide removal is no longer available. Or, if the isolation oxide opening is slightly enlarged and the base oxide opening follows the same trend, short-circuiting of base regions to isolation regions can easily occur.

In terms of the photoresist operation the metal widths and spacings are extremely critical. A basic problem is created during the etching of closely spaced lines. A small amount of aluminum may be left short-circuiting one line to another. This is caused either by hydrogen bubbles in the etch adhering to the aluminum and inhibiting the etch rate, or incomplete removal of emulsion during development. Leaving the wafer in the etch to remove the short circuits may mean severe undercutting of the balance of the aluminum on the wafer; whenever practical, aluminum lines and spacings should be wide.

Wafer rework almost invariably leads to yield loss. Tests can and should be designed at every step of the process so that repetition of the fabrication steps may be avoided.

## Assembly

Many wafers are lost due to careless handling during assembly. Scribing and breaking operations should not be a source of trouble if proper wafer thickness is maintained, sharp diamond points are used for scribing, and the scribe line is properly oriented to the crystal. However, the scribe lines must be kept clear of oxide and metal.

A major problem with using present-day die mounting equipment for LSI assembly is that the operator is unable to observe the entire operation as it is being performed. He may be unaware that the die is skewed in the package, or gold has splattered on top of the die.

Again, it's most important to consider packaging early. The yield, as well as the device's ability to function, will seriously be affected by all packaging decisions.

Earlier, we listed packaging as part of the design phase. This recommendation is all too often completely disregarded; the result is usually a poorly designed component. For example, major problems associated with die bonding are, in most cases, due to the package. The package's cavity must allow room for the die collect, thereby necessitating an oversized cavity. After assembly, the circuit operation may be limited by the package's ability to dissipate heat. At Honeywell an ultrasonic die attachment is used to bond the die in the package, and die collects are ordered the day the artwork is cut. Figure 9 shows some LSI packages used at Honeywell's Aero-Space Facility. These range from a standard 22-lead flat pack with a 0.64-cm diameter to a 60-lead flat pack with a 1.4- by 1.4-cm lead frame.

The flat-pack lead frame design is very flexible. By simply modifying the lead frame, it's easy to change the location or number of bonding pads. This costs about $500 as compared with the $2500 to $3500 tooling charge for a new package. Another nicety is that it can be done to almost any package the engineer wants—provided, of course, that the die fits.

Figure 10 illustrates undesirable pad placement. Crossing wires (circle) can be seen in the corner of the chip. However, changing the pads around would change the size of the die and make the package unusable. But sometimes a compromise between pad placement and design can be made if quality control and other cognizant personnel are aware of the problem at an early stage. Otherwise, packaged circuits may be rejected due to flying leads, wires crossing, improper pad positioning relative to active area, and incorrect pad size in relation to ball bond size.

As the die becomes larger, the circuitry more complex, and the number of pads increases from 14 to 40 or 60, the wire-bonding operation becomes more critical. A die may have to spend as much as 15 or 20 minutes on the heat column at temperatures between 320° to 360°C. This certainly will not enhance the yield. Device parameters can change and previously tested circuits may fail due to leakage, beta changes, or purple plague after bonding. These defects can be minimized if ultrasonic aluminum wire bonding is performed at room or rela-

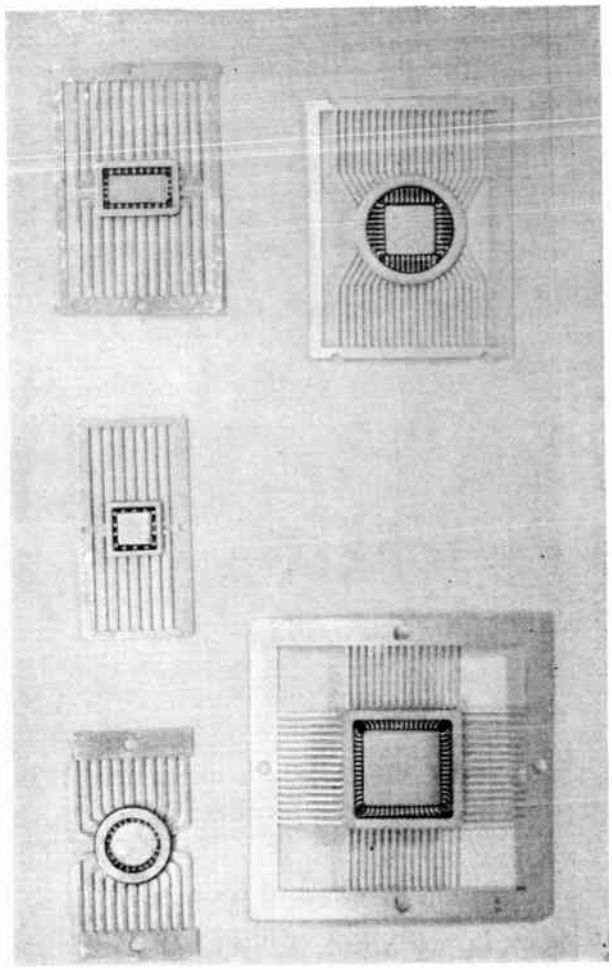


FIGURE 9. Some typical LSI packages.

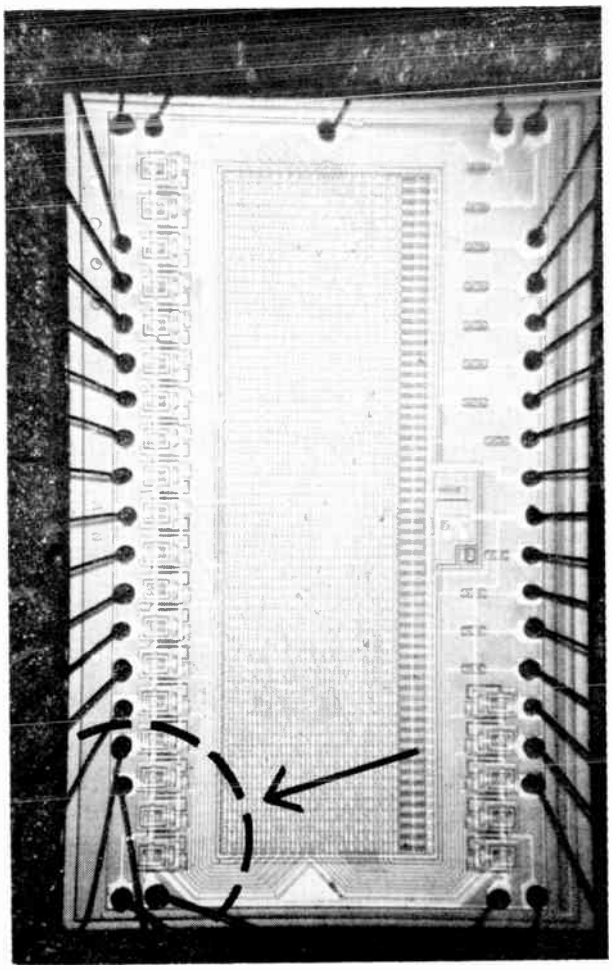


FIGURE 10. Wired digital word translator. The arrow indicates crossing wires caused by bad pad placement.

tively low temperatures. A defect is shown in Fig. 11.

The sealing operation is the last critical phase in assembly. Experience has shown that the number of perfectly hermetically sealed packages decreases as the package size increases. The result is a decrease in a yield. Solder-clad lids improve hermetic sealing and yield is higher. This is because solder balls and solder flow inside the package are eliminated. A flat-pack perimeter sealer has a cycle time of only 1.5 minutes, with the device at the sealing temperature only a matter of seconds. This is an advantage in terms of device yield, but a disadvantage in that tooling and control settings must be changed for each type of package.

## Testing

Although much has been written about the new LSI testers and the methods of functionally testing the final product, there is a need to explore in greater detail in-process testing and test circuitry on the die.

In-process tests can be used to check component isolation. To take full advantage of LSI density, chip components must be close together; therefore the isolation diffusion process is kept to a minimum. Once the base oxide removal is complete a quick electrical test will determine if isolation is achieved. If the components are not isolated, a short time in the furnace will complete the operation. In addition, any breakdown exhibited by the collector to substrate junctions will yield information as to range of the epitaxial resistivity.

In-process tests can also be used to measure beta and Zener voltage. If the parameter tolerances are very wide, there may be no need for this measurement. However, with the extra effort and time needed for this test, the component parameters can be made fairly tight, giving the designer a more-optimized and higher-performance circuit. In-process control is most critical if buried resistors are used.

Transistor beta is measured after contacts are opened. It should be recognized that due to the manner in which contact is made by a probe point on bare silicon, the beta measurement is not absolute. In fact, leakage due to induced inversion layers could go undetected. However, once the metal is put on and alloyed, measurements of the test resistors and transistors are invaluable for predicting the probable success of the circuit. Measurements such as beta normal, beta inverse, saturation voltage, junction breakdowns, p-n-p beta, leakage, and resistance values can be made. This is important information that can readily be correlated to circuit operation and can save countless hours of toil and frustration in trying to track down an error source in a malfunctioning circuit.

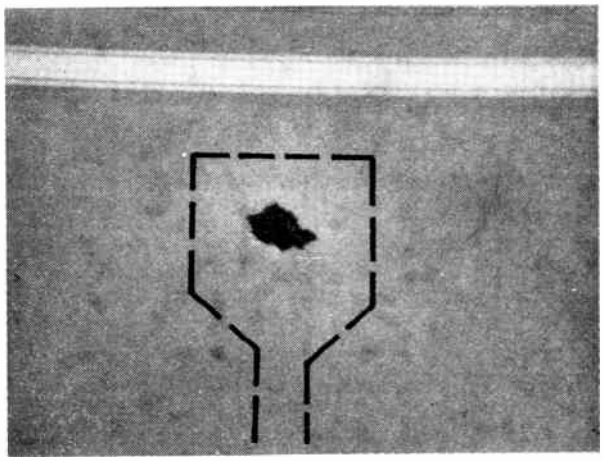
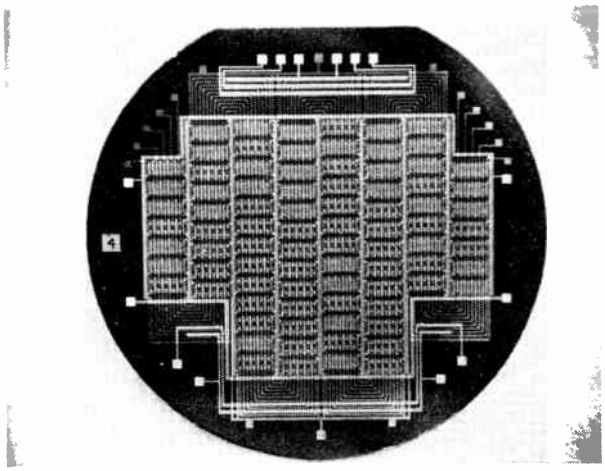


**FIGURE 11. Breakdown of oxide due to ultrasonic bonding.**

**FIGURE 12. This total monolithic system was fabricated using discretionary wiring techniques. The multilayered metal is separated by silicon dioxide dielectric.**



The process engineer should, in turn, use this information to monitor the fabrication processes. Jointly, the design and process engineers will be able to improve the circuit's performance. This technique cannot be overemphasized as an aid in troubleshooting the LSI.

If a calculated, logical approach is not used to detect malfunctions, rework may be done unnecessarily on the wafers; this may require removal of metal and repetition of many processing steps. As mentioned previously, the result is low yield due to breakage, leakage, and change in parameters. Test transistors and resistors are invaluable troubleshooting aids.

There is another desirable test vehicle that can be designed into the LSI wafer. When the metal mask is stepped, a new reticle can be inserted and stepped onto the mask. This reticle will be a test pattern with pads brought out from portions of the circuit, which may function as a gate, a bit, or buffer section. This test circuit uses up only one die on the wafer.

It is advisable that the initial quantity of wafers for a new, complex circuit be limited, and the completed wafers moved through the test process with the highest priority. This allows early identification of mask errors and gives the design engineer time to calculate and determine what changes are necessary to assure success on the subsequent wafer runs.

In summation, it can be seen that successful large-scale integrated circuit production is the result of the complex interactions of design, fabrication, and testing. Close cooperation between the circuit designer and the process engineer is essential, as this new technology necessitates some new approaches. Even though each area has been discussed, the types and numbers of problems that can be encountered have not been exhausted. Obviously, many other conditions will exist as a function of a particular company's location, policy, or personnel. The circuits discussed in this article have been built successfully, meeting product assurance criteria and customer requirements over the temperature range.

What does the future hold? Whereas Figs. 1 and 2 present LSIs with fixed interconnect patterns, Fig. 12 shows a total monolithic system utilizing discretionary wiring and multilayer metal connecting over 60 repetitive circuits or cells that had probed good out of a possible 84. Each cell on the 3.2-cm wafer contains 143 components and measures 3175 by 1400 $\mu$m. More than 325 bits are utilized on this particular wafer. A low-temperature silicon dioxide separates two metal layers. The first layer is partially customized, with a bus-bar arrangement surrounding the standard cell pattern.

Ultimately, the computer will design the necessary interconnect masks based on test data received by wafer testing. The advantage will be the short turnaround time and optimum layout of routing connections. As complexity increases, additional layers will be added specifically to handle signal interconnections and power distribution. This is unquestionably the future trend of the large-scale integrated circuit, and with this trend will come cost savings and high reliability. It is this type of system that, in the not-too-distant future, will utilize the full potential of the semiconductor technology.

The comments and suggestions of C. J. Cerulli in the preparation of this article are appreciated. Appreciation is also expressed to Lloyd Horton and Less Schubeck for their assistance with the photographs.

REFERENCE

1. Hochman, H. T., and Hjelle, D., "Optimizing transistor parameters in integrated circuits," *Solid State Technol.*, Sept. 1966.

**Herschel T. Hochman** (M) received the B.S.E.E. degree from Johns Hopkins University in 1962, and has done graduate work at the University of Dayton. For the past ten years he has been involved in all phases of semiconductor technology including dielectric isolation, thin film, and radiation hardening of digital and linear bipolar and MOS circuitry. Mr. Hochman was a charter member of the Westinghouse R&D Microcircuit Group in Baltimore, Md. He set up the National Cash Register Microcircuit Group Laboratory in Dayton, Ohio, where he was manager from 1962 to 1966. In 1966 he joined the Norden Division of United Aircraft Corporation as chief of the R&D Microelectronics Group. In 1967, the production activities of the Minuteman program were made a part of this activity. He joined Honeywell in 1968 as a staff engineer and is now project engineer and supervisor of LSI circuit processing.

# Vibrating varifocal mirrors for 3-D imaging

*To relieve some of the complexity that exists between man and machines, a three-dimensional interface is needed. A practical 3-D system is not available, but here is a technique that satisfies many autostereoscopic requirements*

**Eric G. Rawson** *Bell Telephone Laboratories, Inc.*

*As technology advances, the interactions between man and machine become more complex. A reliable three-dimensional man–machine interface could help alleviate some of the complexity; but an effective 3-D display has, so far, eluded technologists. Thus, we have been forced to make unnatural compromises when dealing with data that are essentially three-dimensional in nature. This article describes a system that may meet the autostereoscopic display requirements in many situations.*

A miniature television camera held by an astronaut made the beauty of the moon and earth vivid for millions of viewers. This same camera, when focused on instruments and controls within the Apollo 10 cabin, testified to the complexity of man–machine interactions.

Much of the data involved in these interactions are three-dimensional in nature; and technologists have continually stressed the need for a good three-dimensional man–machine interface. The lack of such a device has forced us into unnatural compromises in data handling. For example, the three-dimensional positions of aircraft in the vicinity of an airport are presented on a two-dimensional interface—a cathode-ray tube (CRT). If the aircrafts' altitudes are shown, they are represented by numbers painted beside the radar echo marks. Like the air traffic controller, the submarine commander who operates in a three-dimensional environment would probably be happier if he could replace his two-dimensional CRT sonar display with an equivalent three-dimensional man–machine interface. In short, he would like an autostereoscopic projector. Recently, the use of vibrating varifocal mirrors for stereoscopic display[1-3] have made important additions to 3-D projection techniques.

## Autostereoscopic displays

Stereoscopic display systems can be divided into two broad classes: those that are autostereoscopic and those of the stereo pair type.[4] Stereo-pair-type devices use two slightly different images, and an optical system that directs one image to each eye [Fig. 1(A)]. Home stereo viewers are of this class, as were the ill-fated 3-D movies of a few years ago. Autostereoscopic display systems (holograms are perhaps the best known example) project the light rays that emanate from a reconstructed image over a relatively wide solid angle; and can therefore be viewed by several observers, over a range of distances and from any direction within the solid angle [Fig. 1(B)]. Another example of an autostereoscopic process is integral photography (Fig. 2), which was invented in 1908 by Gabriel Lippmann.[5] Since then, interest in integral photography has languished due to the lack of suitable "fly's-eye" lenses. However, the availability of high-quality plastic lens arrays has recently stimulated activity in this area.[6-9] A modified form of the integral photograph, in which the fly's-eye array of spherical lenslets is replaced by an array of plastic cylindrical lenses, is known as a parallax panoramogram.[10]

A third example of an autostereoscopic projection technique, and the one from which the vibrating varifocal mirror display has evolved, makes use of a rapidly vibrating or rotating screen on which is projected a sequence of images.[11-15] Figure 3(A) shows one form of the device.[12] A rotating projection screen is illuminated by a high-brightness CRT and a projection lens positioned on the rotation axis. The motion of the screen spreads these images throughout the three-dimensional volume swept by the screen; if the periodic motion of the screen is at a high enough frequency (about 15 Hz or higher), persistence of vision creates the impression of a three-dimensional image. A related display device[13] [Fig. 3(B)] makes use of a flat projection screen that oscillates, piston-like, toward and away from the viewer. The mechanics of the system limit the oscillation amplitude, and hence the depth of the three-dimensional image volume swept out by the screen.

A vibrating varifocal mirror display system is basically an oscillating screen that relieves several mechanical problems by oscillating the image of the screen instead of the screen itself.

## Vibrating varifocal mirror display

In 1961, Muirhead[16] noted that a thin sheet of aluminized Mylar plastic film stretched taut over an airtight circular frame could be pneumatically distorted to form a good-quality concave or convex mirror, and that the curvature, and hence the focal length, of the mirror could be conveniently varied by decreasing or increasing the static air pressure on the Mylar's back surface. These "varifocal" mirrors were constructed with diameters up to 3.66 meters. A few years later, Dr. Alan Traub at Mitre Corporation recognized the potentialities of the varifocal

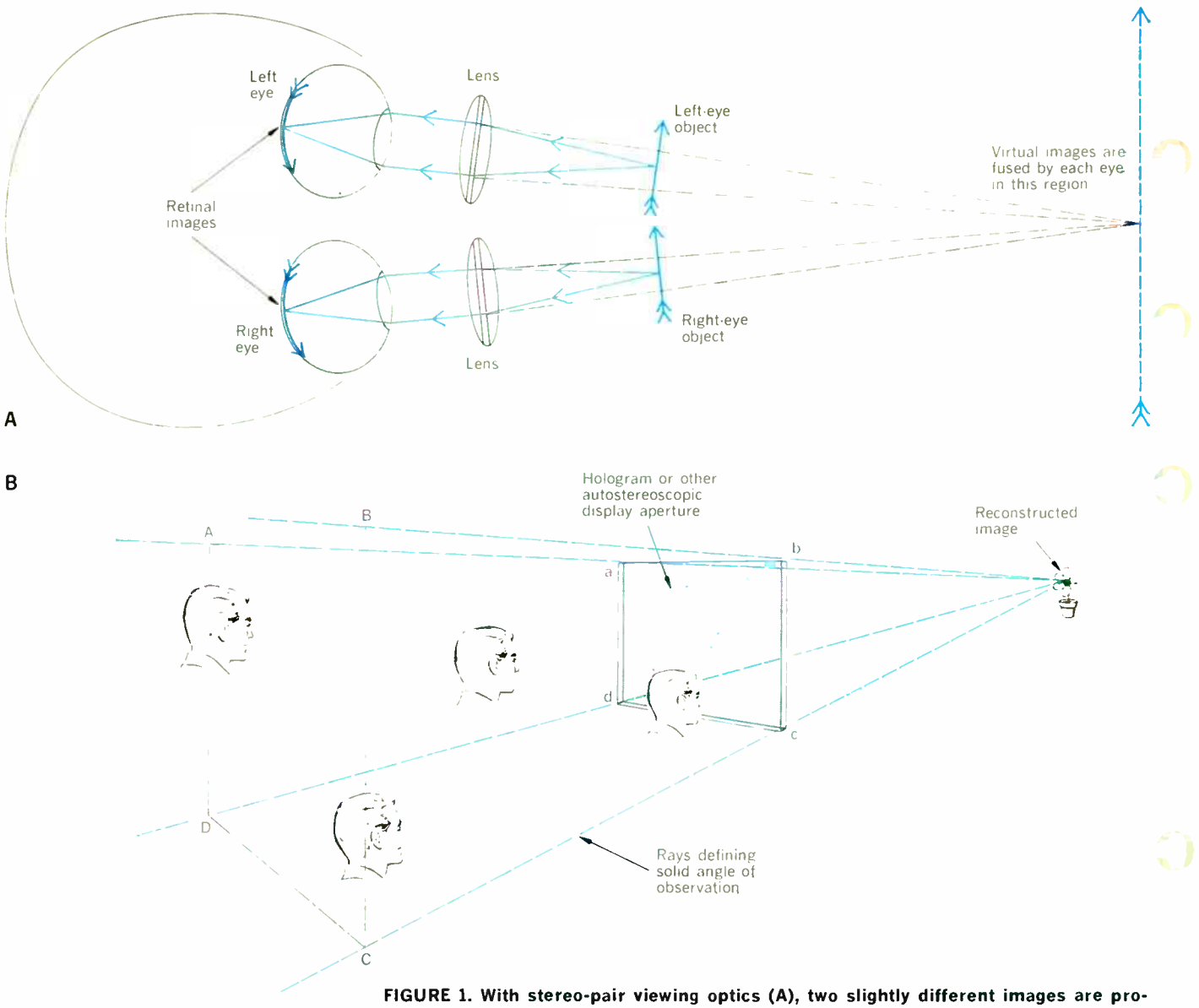
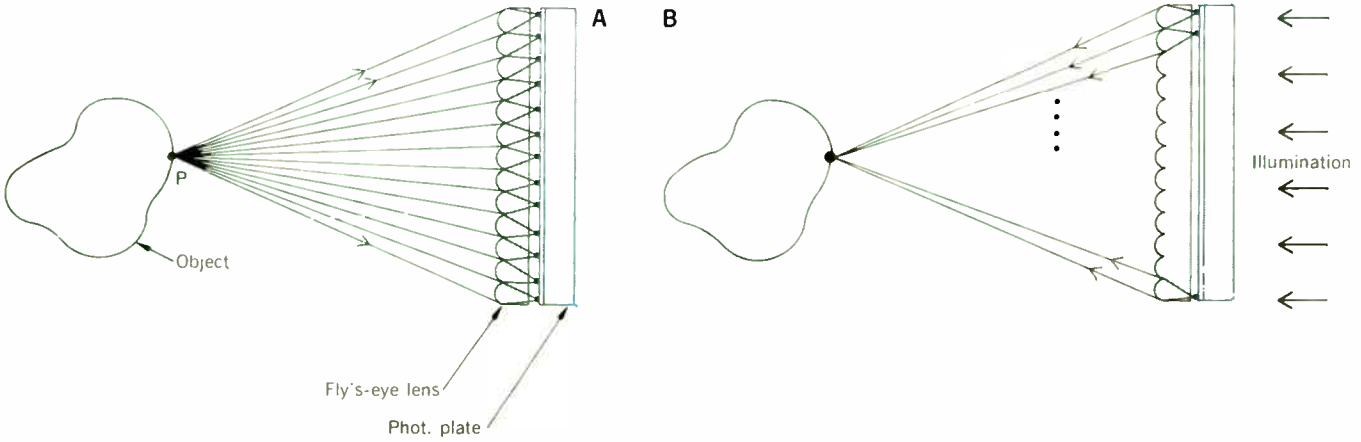



**FIGURE 1. With stereo-pair viewing optics (A), two slightly different images are projected by two lens into the eyes that fuse the images at some point in space. Autostereoscopic display (B) allows viewers to see the reconstructed image from anywhere within the viewing "cone."**

**FIGURE 2. Integral photography. A—A fly's-eye lens forms multiple images on a photographic plate. The rays show the imaging of a point "P" on the object. B.—After the plate is reverse-processed, repositioned, and illuminated from behind, the light rays are refocused to point "P" and the image of the original object is reconstructed.**



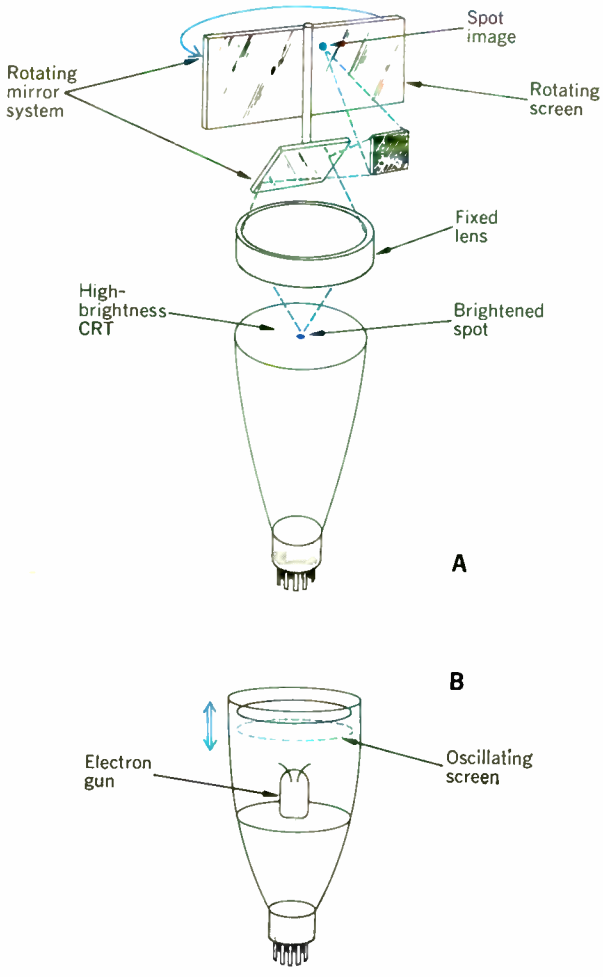



**FIGURE 3. Autostereoscopic display system using a rotating screen (A),[12] and using an oscillating screen (B).[15]**

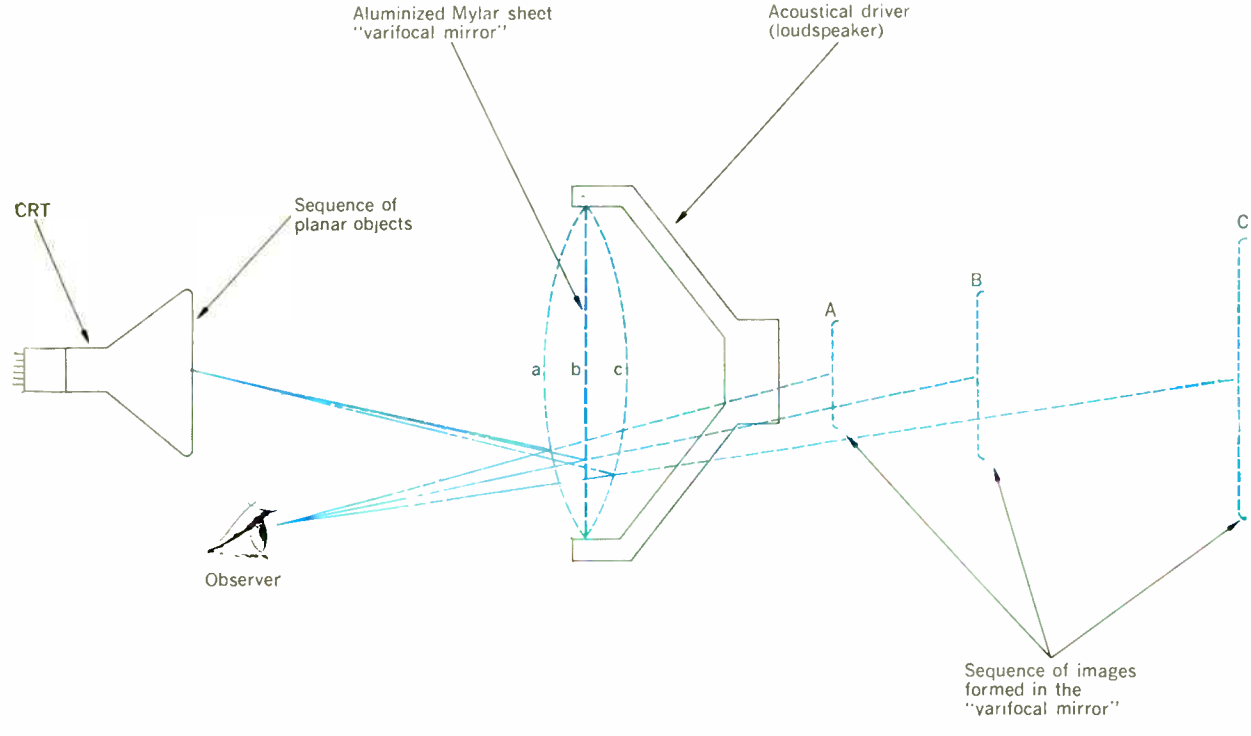
mirror for autostereoscopic imaging.[1] The essentials of the method are illustrated in Fig. 4. The thin aluminized Mylar film is stretched taut and driven sinusoidally by a 15- to 60-Hz tone from a loudspeaker. If the Mylar mirror is taut enough and the amplitude of the oscillation not too large, the mirror's surface is essentially a sphere of continuously changing curvature. Thus, when an observer views an object (such as the face of a CRT) by reflection in the mirror, the changes in curvature cause a corresponding change in the position of the reflected image. In a typical operation a rapid sequence of perhaps 20 or 30 two-dimensional images appears on the object screen. During this time the loudspeaker causes the aluminized Mylar mirror to change curvature smoothly from one extreme (a) to the other (c). As a result the image position sweeps from A to C and the sequence of images is spread out more or less evenly between the two extremes. This display sequence is repeated cyclically at a frequency of 15 Hz or higher. Due to persistence-of-vision effects, the result is an autostereoscopic image that is essentially a transparent stack of two-dimensional images viewed in the varifocal mirror.

The nature of the imaging process is governed by the spherical mirror equation,[17] which says that the amplitude of the image position motion, AC, is typically 15 to 30 times larger than the corresponding mirror oscillation amplitude, ac. By increasing the ratio of the object distance to the mirror diameter, or by increasing the mirror oscillation amplitude, the distance to the farthest image plane C can be easily increased to infinity. This allows wide flexibility in the image depth range.

## Two laboratory applications

Traub demonstrated his discovery in a variety of configurations, one of the most interesting of which was a real-time display of a computer-generated autostereoscopic image.[1] A computer-controlled CRT display con-

**FIGURE 4. Principle of varifocal mirror autostereoscopy.**



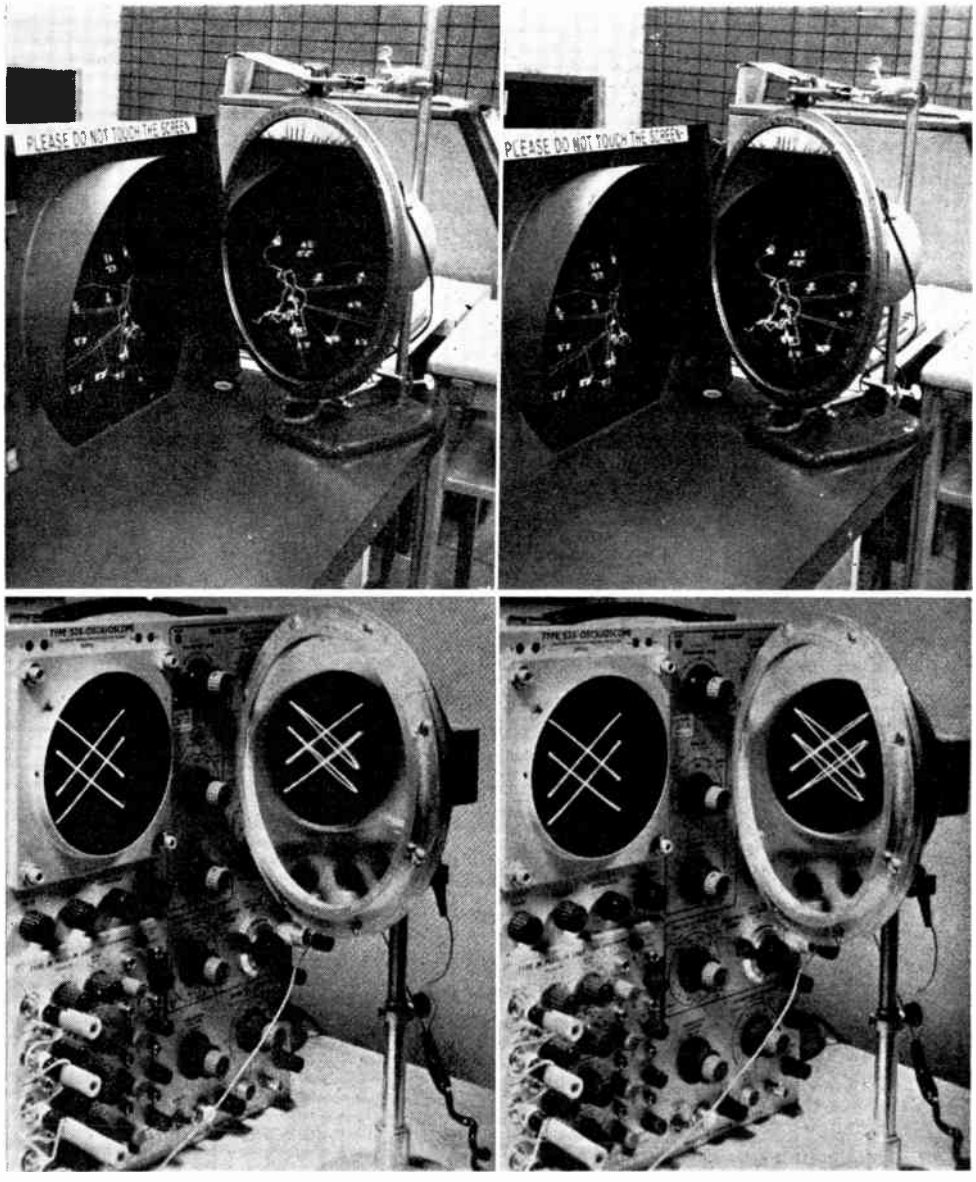



**FIGURE 5. Stereo pairs (top) showing varifocal mirror displays of Traub's computer simulation of a 3-D radar display, and (bottom) a 3-D Lissajous figure. The stereo pairs can be viewed by hand holding a pair of lenses (of 10- to 30-cm focal lengths) in front of each eye.**

sole was used to generate the required sequence of two-dimensional source images. This system was used to simulate an air traffic controller's three-dimensional radar console, in which the altitudes as well as the positions of aircraft were easily perceivable. Figure 5(A) shows a stereo pair of the radar display; also shown is a stereo pair of a three-dimensional Lissajous figure [Fig. 5(B)].

Another application of varifocal mirrors, a computer-generated autostereoscopic movie projection system,[3] was made in the author's laboratory. Figure 6 shows the system schematically. A special, high-speed, 16-mm movie projector casts a sequence of 15 movie frames onto a rear projection screen, during which time the image plane advances toward the observer. This is followed by 15 opaque frames (during which time the image plane retreats to its starting point). Thus, a single three-dimensional image volume is assembled from a spatially distributed sequence of 15 planar images. To accomplish this, the projector runs at 450 frames per second.

The 16-mm movie film was generated using a Stromberg-Carlson 4020 microfilm recorder under the control of a GE 645 computer. In order to synchronize the mirror oscillations to the free-running movie film, the computer was programmed to draw sync marks (small transparent areas) in the corners of appropriate movie frames. During projection, the resulting light pulses are photoelectrically detected and used to generate the sine wave required to drive the loudspeaker.

Figure 7 illustrates the autostereoscopic nature of the movie image. This movie consists of a line drawing of a three-dimensional house with a front yard and two front doors, through which a boy and a girl move back and forth. The top two photographs are oblique views of the image from different directions within the solid angle, and the third is a single frame that is included to assist in interpreting the first two. The blurring in these pictures

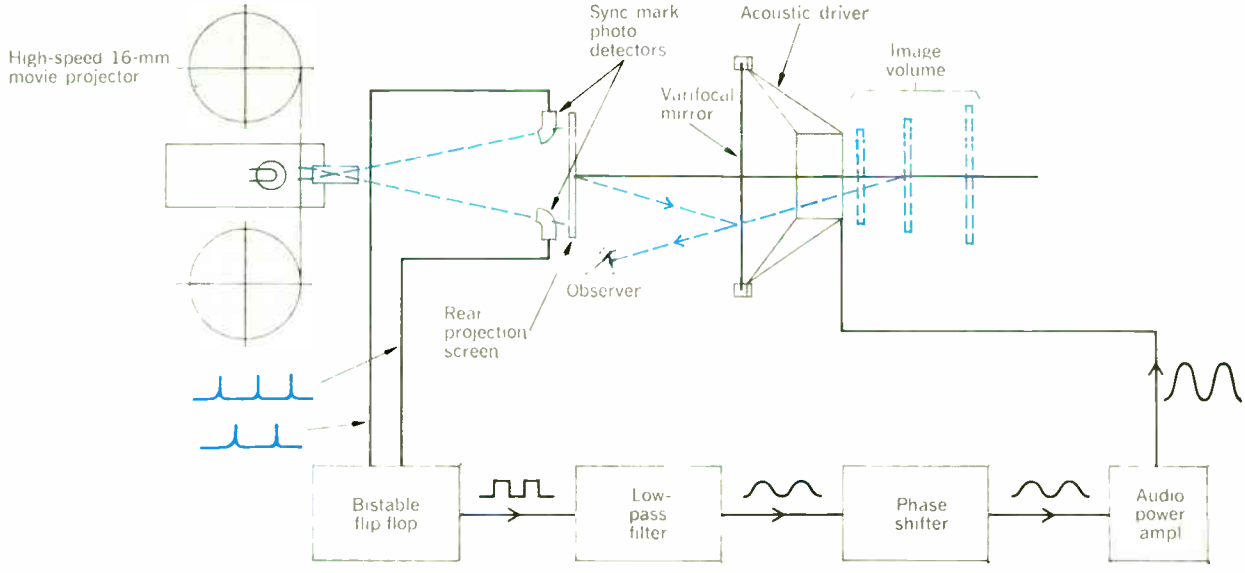
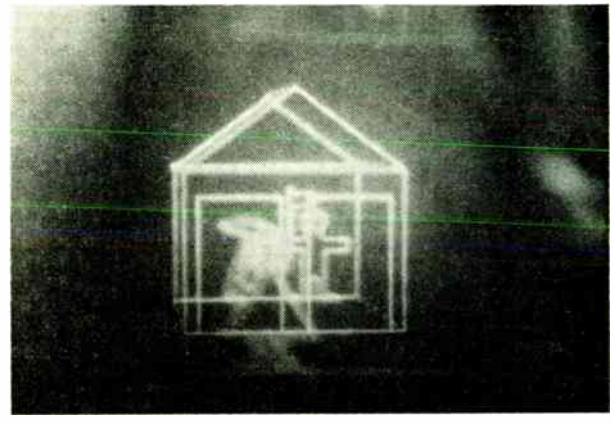



**FIGURE 6. Schematic diagram of a 3-D computer-generated movie projection system. Photodetectors are used to detect sync marks and generate the mirror driving voltage.**



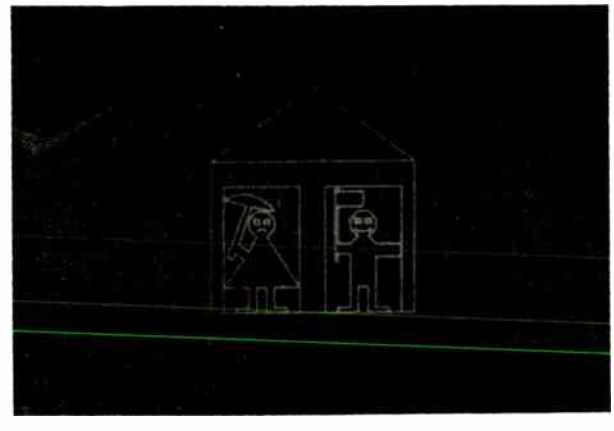


**FIGURE 7. The top two photographs show two oblique views of the 3-D computer-generated movie. Blurring is due to figure movement and image jitter during exposure. The bottom photo is a single frame to assist in the visual interpretation of the top two photos.**

is considerably greater than that noticed in direct viewing, and is due to the problem of photographing a low-brightness moving image.

### Peculiarities of varifocal mirror systems

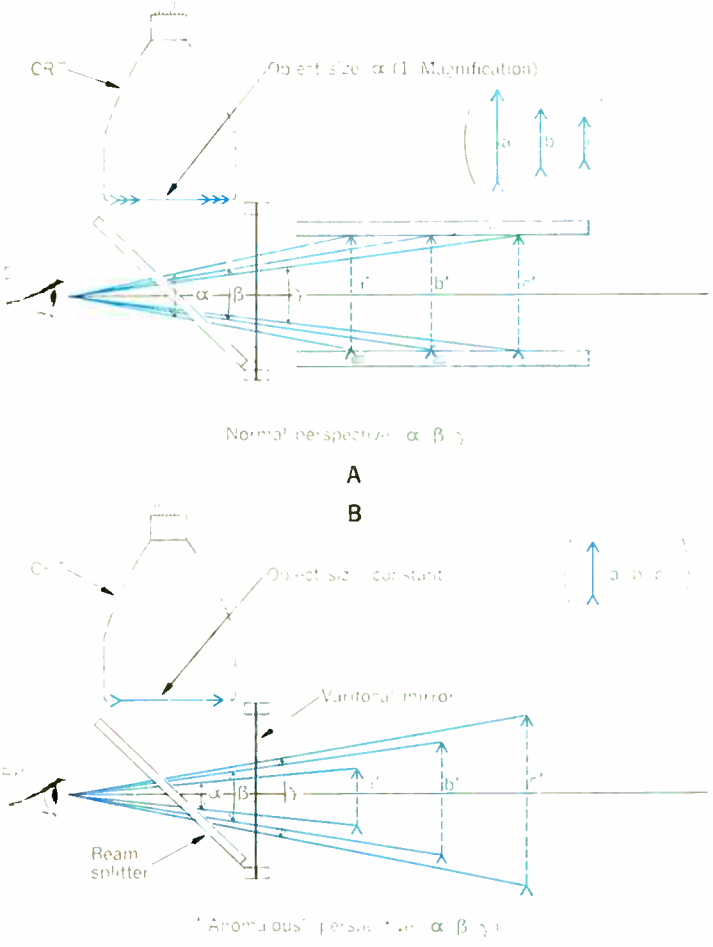
It is apparent that this technique has its own peculiar limitations and shortcomings. First, it generates a transparent or "phantom" image. That is, the important depth cue of interposition—the obscuring of farther portions of a scene by nearer portions—is missing. This suggests that varifocal mirror displays may find their most successful applications where symbolic data (such as three-dimensional position coordinates) rather than realistic images (such as scenes or people) are being displayed.

Another peculiarity of the varifocal process is that, as the image moves along the depth axis toward the observer, the image size diminishes. This is shown by Fig. 8(A). Traub has called this effect "anomalous perspective,"[1] since objects of equal size are imaged in such a way that distant images subtend larger angles at the observer's eye than do near images. Figure 8(B) suggests a simple cure for anomalous perspective. The scale of the object pictures is modulated so that the size of the object is inversely proportional to the instantaneous magnification. The result is a constant lateral scale throughout the image volume.

Other peculiarities of varifocal mirror systems come to light when one considers what is the best distribution of image planes along the depth axis. Figure 9 illustrates two such image distributions: the first (A) involves an even spacing of planes along the depth axis; and the

second (B) is an uneven distribution in which each plane subtends an equal angular increment at the observer. This latter distribution is suitable for displays requiring infinite or near-infinite depth ranges.

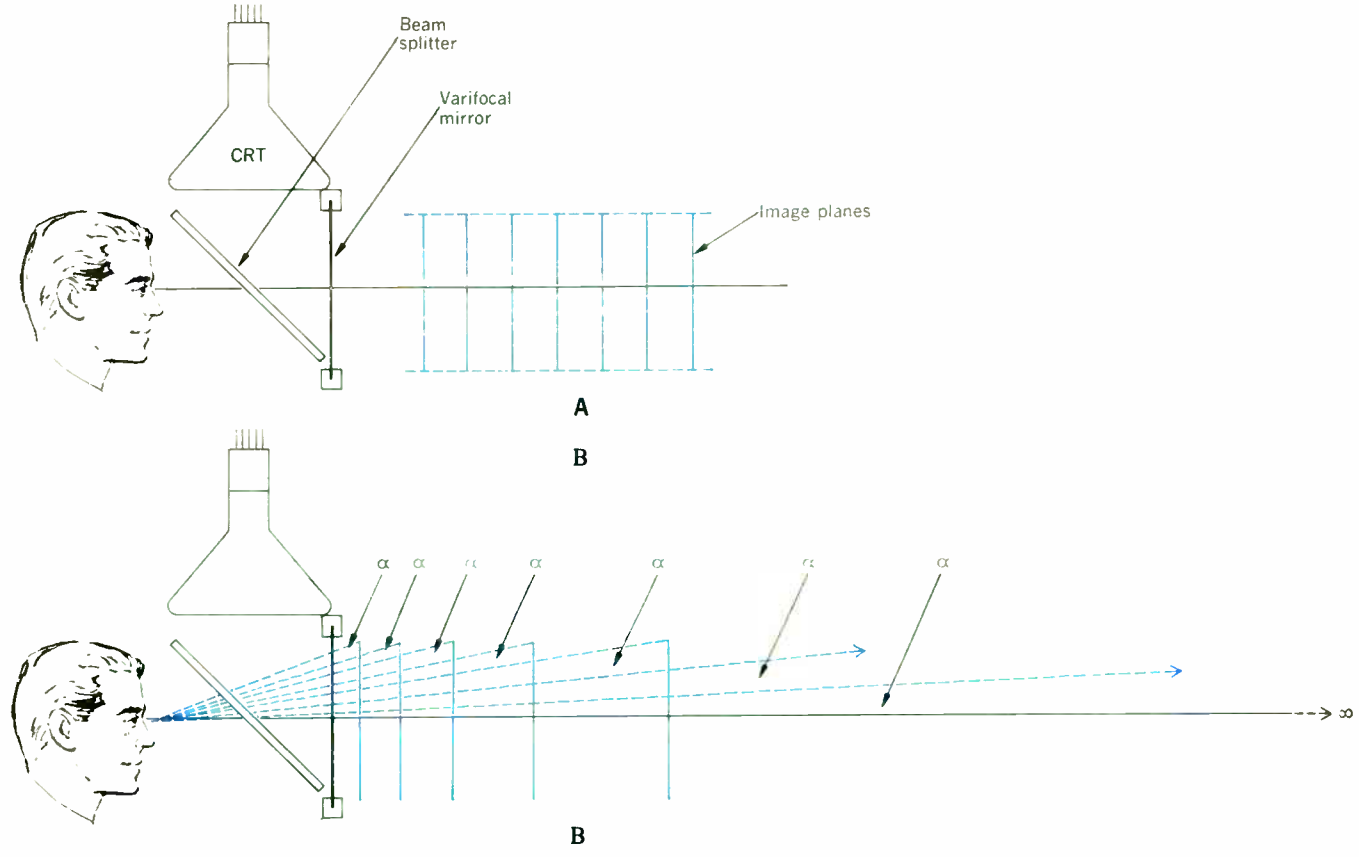
In order to achieve an even distribution of images along the depth axis, a linear time sweep of the depth axis is often desirable. This is true because the patterns usually appear on the object screen at a constant rate (generally, the fastest possible rate). Therefore, to minimize the retrace time, a sawtooth-like motion of the image plane along the depth axis is desirable. It is here that two more peculiarities appear. The first is that the image position is not a linear function of the mirror displacement.[3] Thus the sawtooth image motion requires a more complex mirror motion and loudspeaker driving voltage waveform. The second peculiarity is that the speaker–mirror combination will usually have a highly nonlinear frequency response, as illustrated in Fig. 10. In this figure, salt granules collect along nodal lines and show the nature of the mode of oscillation. It can be




A

B

**Figure 8. "Anomalous" perspective (A) in which the angle subtended at the observers eye of normally equal sized subjects increase with distance. If the object varies inversely as the magnification (B), the subtended angle decreases normally with distance.**




A

B

B

**FIGURE 9. Optimum distribution of images along the depth axis depends upon the application. A—A linear distribution is suitable for displaying 3-D functions in a rectilinear coordinate system. B—The nonlinear system is suitable for "scenic" displays spanning great depth ranges.**

seen that the desired zero-order mode of oscillation is attained only at frequencies up to about 150 Hz with this particular 20-cm-diameter varifocal mirror. At higher frequencies, higher-order modes of oscillation appear. It is apparent that a waveform such as a sawtooth wave, which is rich in high-frequency harmonics, will excite these undesired higher-order modes of oscillation in the Mylar. However, a filtered sawtooth wave in which the harmonic components above about 150 Hz are strongly attenuated has been successfully used to drive the mirrors in a quasi-sawtooth manner, achieving a scanning duty cycle of about 90 percent.[3]

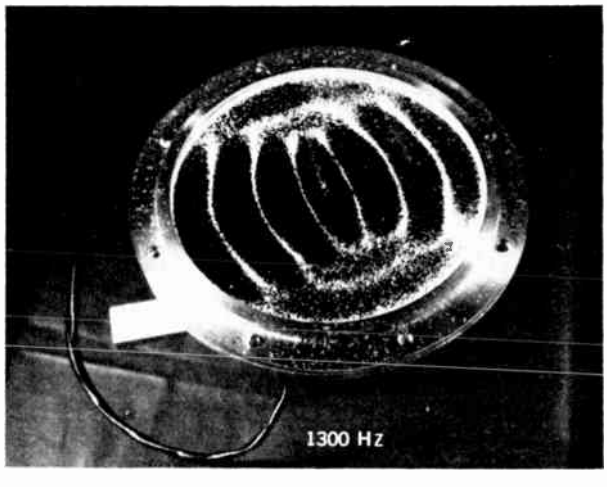
## Varifocal mirrors in light sculpture

As an interesting sidelight, vibrating varifocal mirrors have recently been used in light sculptures by New York artist Robert Whitman, working in collaboration with the author and under the auspices of Experiments in Art and Technology, a nonprofit organization whose aim is to encourage the collaboration of artists and engineers. In these works, several large varifocal mirrors, some of them 1.2 meters in diameter and others 1.2 meters square, are each acoustically driven by four, 38-cm-diameter loudspeakers. The mirrors are driven for a random time interval ranging from 1 to 30 seconds with one of five randomly selected waveforms (sine waves and sawtooth waves of various frequencies); then the mirror is quiescent for a random time interval ranging from 1 to 30 seconds. The cycle then begins again with the random selection of another waveform. Certain waveforms on the mirrors are accompanied by stroboscopic illumination of the observers in the vicinity of the mirror. The strobe light frequency is adjusted to within about 1 Hz of the fundamental mirror frequency, resulting in the observer's reflected image moving back and forth along the depth axis at the difference frequency, about 1 Hz. Due to the large size of the mirrors, many high-order vibrational modes are excited, resulting in complex undulations of the reflected images.

## Conclusion

Do vibrating varifocal mirrors provide the much needed autostereoscopic display device discussed earlier? In many respects they do. The hardware is simple, inexpensive, and reliable. However, a cost increase does appear in the form of additional system bandwidth requirements as compared with its two-dimensional equivalent. Ideally one should have almost as many stacked $z$-axis images as one has resolved spots along the $x$- and $y$-axes of the corresponding two-dimensional image. This suggests bandwidth increases of 100 to 1000 times. On the other hand, visually acceptable systems have been demonstrated using as few as 15 resolved depth planes. For specialized display systems, the cost of additional bandwidth may not be prohibitive. Furthermore, when used to display data in which most of each image plane is dark, it may be preferable to draw only the bright portions of each frame rather than raster-scan the frame completely, thus achieving a saving in bandwidth. But despite the bandwidth requirements and their other limitations, vibrating varifocal mirrors may provide an important new display technique.

**FIGURE 10. Oscillation modes of the Mylar film are indicated by salt granules collected along nodal lines.**





REFERENCES

1. Traub, A. C., *Appl. Opt.*, vol. 6, p. 1085, 1967.
2. Traub, A. C., Document No. M68-4, Mitre Corp., 1968.
3. Rawson, E. G., *Appl. Opt.*, vol. 7, p. 1505, 1968.
4. Valyus, N. A., *Stereoscopy*. New York: Focal Press, 1966.
5. Lippmann, G., *J. de Phys.*, ser. 4, vol. 7, p. 821, 1908.
6. Ives, H. E., *J. Opt. Soc. Am.*, vol. 21, p. 171, 1931.
7. Pole, R. V., *Appl. Phys. Letters*, vol. 10, p. 20, 1967.
8. Chutjian, A., and Collier, R. J., *Appl. Opt.*, vol. 7, p. 99, 1968.
9. Burckhardt, C. B., *J. Opt. Soc. Am.*, vol. 58, p. 71, 1968.
10. Valyus, N. A., *op. cit*, p. 108.
11. "CRT provides three-dimensional displays," *Electronics*, pp. 54–57, Nov. 2, 1962.
12. "New display gives realistic 3-D effect," *Aviation Week*, pp. 66–67, Oct. 21, 1960.
13. "3-D display," *Space/Aeronautics*, pp. 60–67, Sept. 1962.
14. Goldberg, A. A., "3-D display system," *Proc. IRE*, vol. 50, p. 2521(L), Dec. 1962.
15. Withey, E. L., "Cathode-ray tube adds third dimension," *Electronics*, vol. 31, p. 21, May 23, 1958.
16. Muirhead, J. C., *Rev. Sci. Inst.*, vol. 32, p. 210, 1961.
17. Born, M., and Wolf, E., *Principles of Optics*. New York: Pergamon, 1959.



**Eric G. Rawson** received the B.A. and M.M.A. degrees in physics from the University of Saskatchewan in 1959 and 1960, and the Ph.D. degree in physics from the University of Toronto, Canada, in 1966. At the University of Saskatchewan, he was active in nuclear spectroscopy and in upper atmospheric physics. At Toronto he carried out spectroscopic studies of Brillouin scattering of light by gases, studies of relaxation phenomena in gases, and the measurement of ultrasonic velocities in gases. In the course of this work, he discovered the phenomenon of the propulsion and orientation stabilization of certain dust particles through air within a laser cavity, which became known as the "runners and bouncers" phenomenon. Since 1966 he has been with Bell Telephone Laboratories, Murray Hill, N.J., where he has been working on optical information processing techniques. These include optical memories, autostereoscopic displays involving integral photography and vibrating varifocal mirrors, and the computer-automated design of complex lenses. For the past three years he has been active within "Experiments in Art and Technology," collaborating with artists on a variety of projects. He is a member of the Optical Society of America.


# The future of UHF transmission lines

*In predicting the future of UHV transmission lines, the author has followed a twofold approach. First, a quantitative analysis is made of the characteristics of future UHV lines, as conceived in the light of present-day techniques. Second, new solutions considered particularly suitable for UHV lines are discussed. Two basic assumptions are made: (1) that the need to transmit ever-greater quantities of electric energy by overhead lines will continue and (2) that society will become increasingly concerned over esthetic considerations.*

## The future in the light of present-day techniques

Existing EHV lines are characterized, from the structural standpoint, by bundle conductors, V strings, and steel lattice towers. The conductors are placed in a horizontal configuration, and the line is shielded by means of two ground wires. A typical tower design for these lines is shown in Fig. 1(A); this design is sometimes replaced by guyed towers, of the $\pi$ or V type. The bundles are made of steel–aluminum subconductors, with outside diameters of about 30–35 mm; the increase of the overall section per phase to meet the requirements of larger transmitted load is generally obtained by increasing the number of subconductors per bundle. The lengths of the spans are limited by economical rather than by technical reasons, and remain approximately constant around 400–500 meters; there is only a slight tendency to increase the span length with the voltage.

Let us consider now what evolution in line design would be determined by these present-day techniques in relation to the increase of the transmission-voltage levels. For this purpose, we have decided to produce a quantitative analysis; in our opinion this analysis, although it may involve arbitrary choices that many people will not perhaps share, will be useful for providing homogeneous

**I. Characteristics of lines at various system voltages**

| Highest system voltage ($U_m$), kV: | 420 | 525 | 765 | 1000 | 1300 | 1500 |
|---|---|---|---|---|---|---|
| Overall aluminum section per phase (S), mm² | 1240 | 1660 | 2680 | 3780 | 5250 | 6300 |
| Number of subconductors per phase (n) | 2 | 3 | 4 | 6 | 8 | 8 |
| Subconductor diameter ($\phi$), mm | 34.5 | 32.4 | 35.8 | 34.7 | 35.5 | 38.8 |
| Conductor-tower clearance (d), meters | 3.00 | 3.90 | 5.60 | 7.20 | 8.50 | 9.40 |
| Switching impulse 50% discharge voltage of tower insulation ($V_{50}$), per unit | 3.2 | 2.95 | 2.60 | 2.25 | 1.95 | 1.80 |
| Admissible 1% switching overvoltage ($U_{1\%}$), per unit | 2.65 | 2.45 | 2.15 | 1.85 | 1.60 | 1.50 |
| Conductor-ground clearance at midspan (C), meters | 7.2 | 8.45 | 10.8 | 13.1 | 15.0 | 16.2 |
| Span length (L), meters | 400 | 420 | 445 | 475 | 500 | 515 |
| Midspan sag (s), meters | 12 | 13.5 | 15 | 17 | 19 | 20 |
| Conductor height at the tower (H), meters | 19.2 | 21.7 | 25.8 | 30.1 | 34.0 | 36.2 |
| Interphase distance (D), meters | 7.30 | 9.20 | 12.8 | 16.1 | 19.0 | 20.8 |
| Tower width (A), meters | 20.0 | 25.4 | 35.6 | 45.2 | 53.3 | 58.4 |
| Tower height (B), meters | 24.6 | 28.2 | 35.5 | 42.25 | 47.9 | 51.5 |
| Line-size parameter (right of way) ($S_L$), meters | 35.5 | 42.3 | 52.0 | 62.5 | 72.0 | 76.5 |
| Tower-size parameter ($S_T = 1000AB/L$), m²/km | 1230 | 1700 | 2840 | 4020 | 5110 | 5840 |
| Voltage gradient of lateral phase conductor ($g_m$), kV/cm | 15.15 | 14.3 | 14.45 | 13.8 | 13.6 | 14.1 |
| RI limit gradient of lateral phase conductor ($g_{admissible}$), kV/cm | 15.8 | 15.7 | 15.35 | 15.5 | 15.25 | 14.85 |
| Voltage gradient at ground (G), kV/m | 7.35 | 9.50 | 11.4 | 13.1 | 16.55 | 17.55 |
| Surge impedance ($Z_S$), ohms | 284 | 268 | 264 | 249 | 240 | 245 |
| Surge impedance loading ($P_S$), MW | 560 | 925 | 1970 | 3615 | 6335 | 8265 |
| Specific line-size parameter ($S_L/P_S$), m/GW | 64 | 47 | 26 | 17 | 11 | 8 |
| Specific tower-size parameter ($S_T/P_S$), m³/km·GW | 2200 | 1830 | 1440 | 1110 | 810 | 700 |

*Although ultrahigh-voltage lines up to 1500 kV can be conceived on the basis of traditional techniques, significant reductions in right-of-way width and in the dimensions of tower structures can be achieved through the adoption of new methods*

**Luigi Paris** *Ente Nazionale per l'Energia Ellettrica*

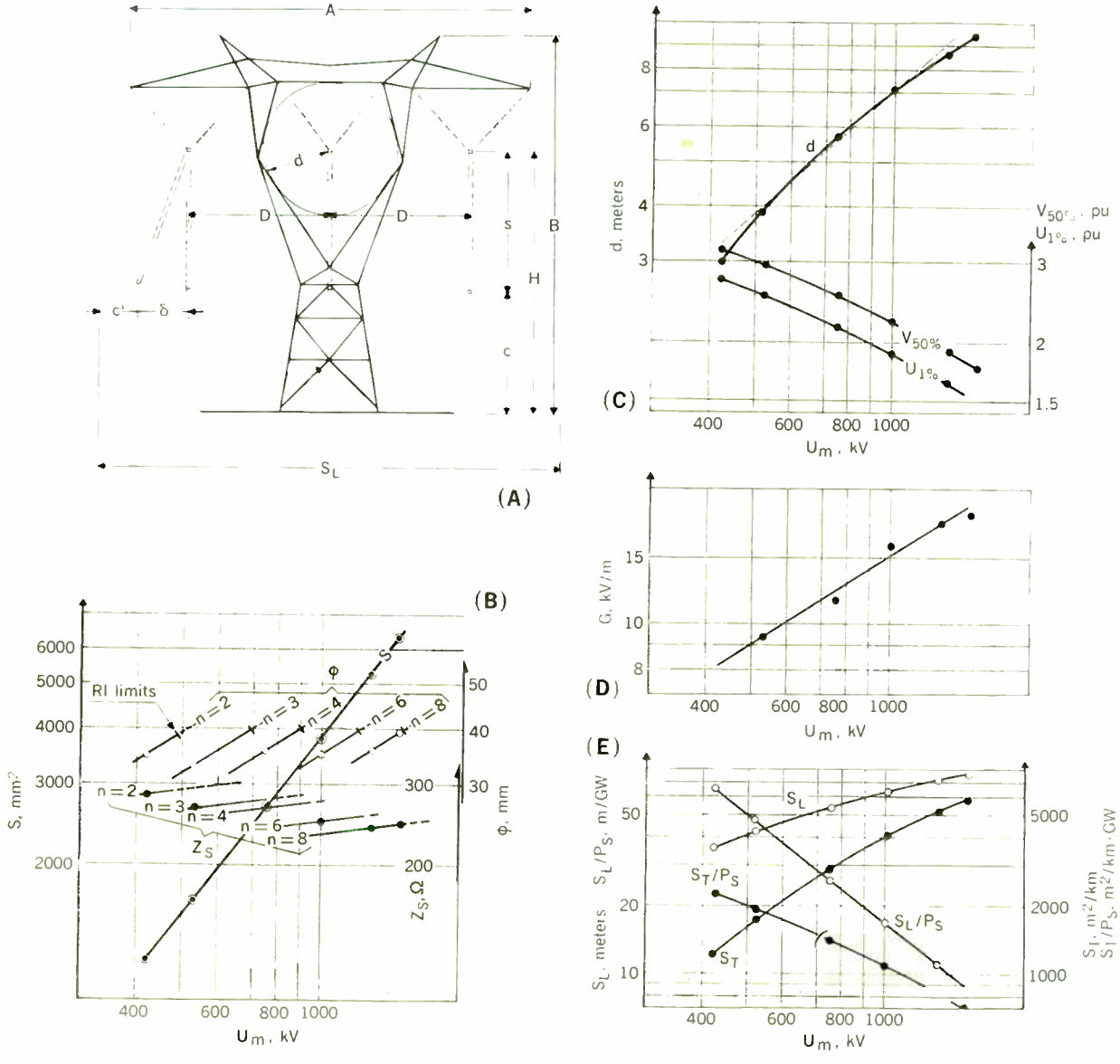



**FIGURE 1.** Basic characteristics of UHV lines of the traditional type vs. highest system voltage, $U_m$. (A) Tower geometry. (B) Conductor characteristics. (C) Insulation characteristics. (D) Voltage gradient G at ground. (E) Size parameters.

- **S** = overall aluminum section per phase
- **φ** = subconductor diameter
- **n** = number of subconductors per phase
- **$Z_S$** = surge impedance of line
- **$V_{50\%}$** = switching impulse 50% discharge voltage of tower insulation
- **$U_{1\%}$** = admissible 1% switching overvoltage
- **d** = conductor-tower clearance
- **$S_L$** = line-size parameter (right of way)
- **$S_T$** = tower-size parameter
- **$S_L/P_S$** = specific line-size parameter
- **$S_T/P_S$** = specific tower-size parameter

terms of reference for the main quantities with which the designer is concerned.

A series of overhead lines in the 420–1500-kV range have been designed, following uniform criteria.[1] The main characteristics of lines having highest system voltages 420, 525, 765, 1000, 1300, and 1500 kV are reported in Table I. These characteristics are shown in Fig. 1(B), (C), and (E) as a continuous function of voltage. The design criteria were chosen in such a way that for the lines in the 420–765-kV range the characteristics would on an average be the same as for the actual lines built in recent years.

The main assumptions used for this quantitative analysis are the following:

1. The aluminum cross sections of conductors have been assumed to increase more than proportionally with the voltage [$S \approx kU_m^{1.3}$; see Fig. 1(B)]. This assumption is based on the experience so far gained with lower voltages; on the other hand, it has been borne in mind that the number of subconductors increases as voltage increases, and that therefore the surge impedance of the line is reduced, thus causing an increase in the power rating more than proportional to the square of the voltage. Moreover, it has been taken into account that the economical density is reduced as voltage increases owing to the increased importance of corona problems, which tend to move economical density toward the lower values.

2. The number of subconductors per phase, for a given cross-sectional area, is the minimum needed to meet the radio-interference (RI) requirements for residential areas.[2] Figure 1(B) shows the subconductor diameter as a function of voltage when these criteria are applied.

3. Insulation levels, and therefore tower dimensions, are chosen in such a way as to obtain a constant ratio (equal to approximately 0.8) between the cost of the "inactive" components of the line (towers, foundations, insulators) and the cost of the "active" components (conductors). It should be observed that only the cost of the inactive components is affected by the insulation level. This criterion leads to a progressive reduction in the insulation levels and therefore in the allowable switching overvoltages. The values of these overvoltages, as given in Table I, were determined on the basis of a 2 percent failure probability of line insulation. It should be observed that the overvoltage value found for 1500-kV systems (1.5 pu) is the same as the one obtained by others.[3,4]

4. All the structural design rules conform to the criteria reported in Ref. 1 and are such as to offer comparable safety levels. Span length and consequently conductor sag and tower height have been chosen so as to ensure the maximum saving.

Line dimensions have been carefully considered and have been defined by means of two parameters: (1) the "line-size" parameter, which essentially corresponds to the width of the right of way; and (2) the "tower-size" parameter, which gives an indication of the space taken up by the towers[3] and which is the product of the tower width, the tower height, and the number of towers per kilometer, and is expressed in square meters per kilometer.

To give an idea of the space utilization in power transmission, specific figures can be obtained by dividing the line-size and tower-size parameters by the surge impedance loading of the line.

From the results of the analysis one can deduce the following:

1. The relative values of the maximum allowable switching overvoltages are to be drastically reduced as voltage increases [see Fig. 1(C)] if it is desired to keep the cost of the "inactive" components of the line within the usual limits. Between 1200 and 1500 kV (maximum switching overvoltages lower than 1.7 pu), the control of switching overvoltages, though involving technical difficulties, appears to be solvable at the present stage of technological development.[5,6] Above 1500 kV (maximum switching overvoltages lower than 1.5 pu), switching overvoltages will hardly be contained within the required limits and therefore the size and cost of the line would increase with voltage much more rapidly.

2. The insulation distance $d$ in Fig. 1(C) and, consequently, the insulator string length, increase proportionally with voltage up to 1000 kV; above this value, the increase of the string length with voltage is slightly reduced. Therefore, with regard to power frequency insulation under polluted conditions, serious problems would arise above 1000 kV only if a possible nonlinear behavior of very long insulator strings is assumed. Under this assumption, unless the insulator characteristics improve, an increase in the insulation distances will be needed, which in turn will involve an increase in the cost of towers.

3. Bundles having up to eight subconductors and diameters up to one meter are needed for the highest voltage levels.

4. The forces exerted on insulator strings and tower members, which increase about proportionally with the diameter and number of subconductors, increase more than proportionally with voltage. From 420 kV to 1500 kV, these forces increase by approximately five times, whereas the average dimensions of the tower framework increase less than proportionally with the voltage. It can then be stated that the traditional single- or double-angle members may be used in tangent towers up to 1500 kV; for special towers above 1000 kV, composite members will probably be needed.

5. The line-size parameter (width of the right of way) and, in general, the linear dimensions of towers increase less than proportionally with the voltage (the line-size parameter doubles between 420 kV and 1500 kV), whereas the tower-size parameter increases more than proportionally (it increases by five times over the range of 420 kV to 1500 kV). Correspondingly, considerable reductions in the specific dimensions of the line (i.e., dimensions per unit of the transmitted power) are obtained; for instance, the specific line-size parameter is reduced to an eighth of its value at the lower voltage.

6. The electric field at ground increases considerably (almost in proportion to the voltage) when the clearance to ground is fixed according to the traditional criteria. For the present time, it is difficult to evaluate the actual risks deriving from higher values of the electric field at ground; these, however, could lead to larger clearances to ground and therefore to higher cost and larger towers. To overcome these disadvantages, the introduction of earth wires located under the conductors has been proposed.[3]

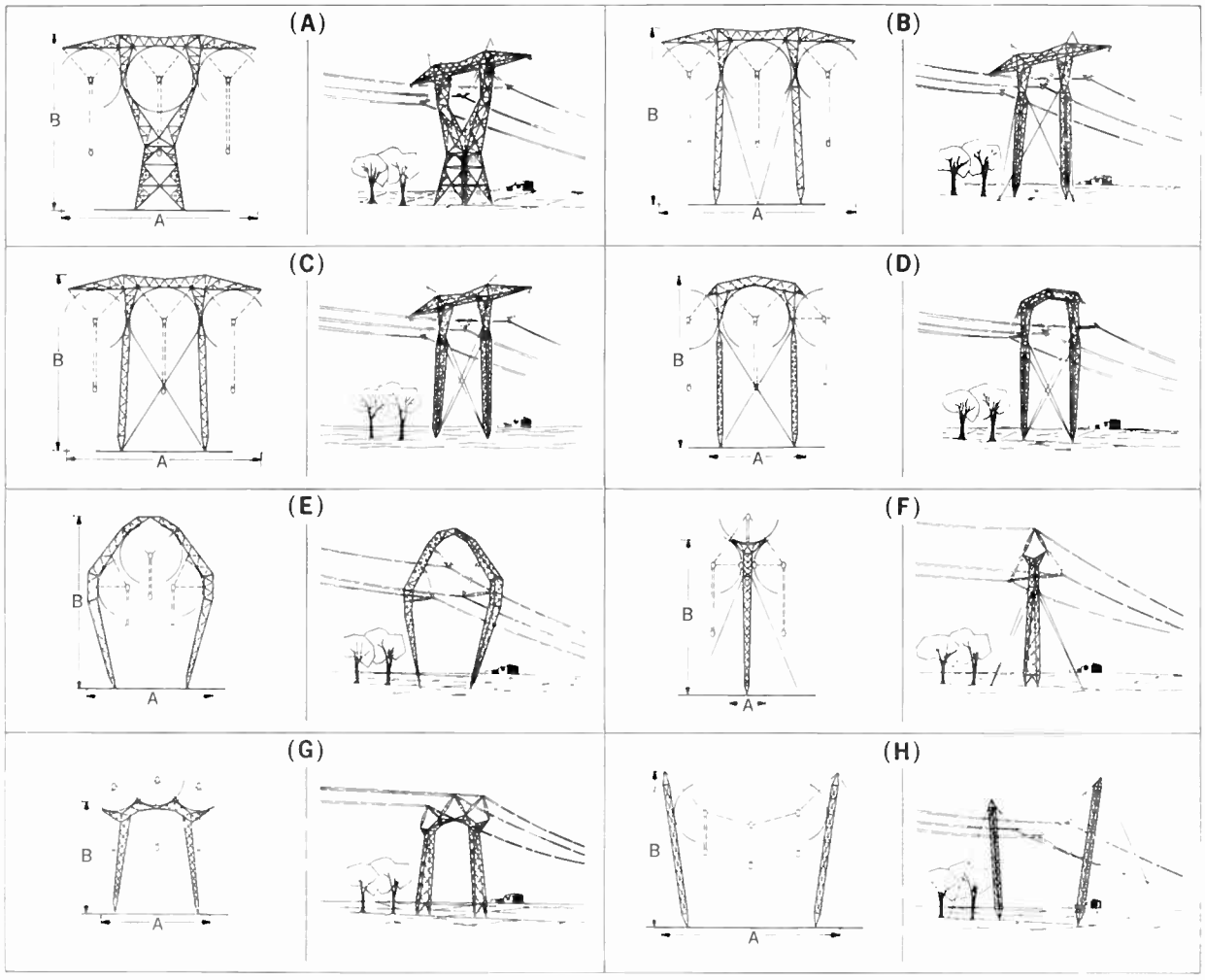
From the result of this analysis it may be concluded that the line engineer, even if working on the basis of present techniques, will not actually encounter significant technical difficulties in designing UHV lines up to

1500 kV, provided the system engineer can control switching overvoltages up to 1.5 pu. Only a few problems could arise in connection with bundles of six or eight subconductors. From the economical point of view, it should be pointed out that the overall cost of the line increases with voltage in a rather reasonable way, since it actually increases in proportion to the cost of the active components.

Optimistic people might actually be quite satisfied with the reductions in the specific dimensions of the line obtainable as voltage increases. However, it should be considered that higher voltage levels are introduced when the load density is increased—with the result that the annual requirement for new lines remains practically constant. It turns out that a reduction in the specific dimensions cannot by itself be considered a real success; actually, the efforts of the designer should be directed to limiting the increase in the absolute dimensions as far as possible.

To obtain this aim, traditional solutions have to be given up and other solutions found. To be precise, these should (1) reduce the metallic structures, both for esthetic reasons (the efforts of designers and engineers

**FIGURE 2. New types of towers for 1500-kV lines, (C) to (H), compared with towers, (A) and (B), of the traditional type.**




| Solutions | | | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) |
|---|---|---|---|---|---|---|---|---|---|---|
| Conductor-tower clearances, meters | central phase | $d_c$ | 9.50 | 9.00 | 9.00 | 9.00 | 8.50 | 8.00 | 8.00 | – |
| | lateral phase | $d_e$ | 9.00 | 9.00 | 9.00 | 8.50 | 8.50 | 8.50 | 8.00 | 8.50 |
| Tower dimensions, meters | width × height | A × B | 58.4 × 49 | 58.4 × 49 | 58.4 × 49 | 27 × 49 | 36 × 49 | 10 × 45 | 31 × 39 | 49 × 45 |
| Size parameters | line width | $S_L$, meters | 75 | 75 | 75 | 74 | 48 | 55 | 59 | 61 |
| | tower area | $S_T$, $m^2/km$ | 5556 | 5556 | 5556 | 2610 | 3420 | 870 | 1920 | 4282 |
| | tower volume | $S_V$, $m^3/km$ | 6880 | 1774 | 1440 | 800 | 610 | 550 | 660 | 630 |
| | tower area at ground | $S_g$, $m^2/km$ | 630 | 1090 | 220 | 200 | 190 | 310 | 500 | 790 |
| Voltage gradient at ground | | $G_0$, kV/m | 17.5 | 17.5 | 17.5 | 17.5 | 13 | 10 | 10 | 10 |

Span length is about 500 meters for solutions (A)–(D), (F), and (G), and about 400 meters for solutions (E) and (H).

who seek to preserve the beauty of the landscape are directed toward this end[7]) and to improve the switching-surge performance of air insulation[8,9] and (2) reduce the line width and consequently the right of way.

## The future in the light of new techniques

Figure 2 shows the geometry of towers based on traditional techniques as compared with the geometry of towers based on new techniques, for 1500-kV lines. This figure gives also some quantitative data that show the main characteristics of the different solutions. In particular, consideration is given to two other "size" parameters necessary to evidence the differences between the various structures: (1) the "tower volume" parameter, which is given by the volume of the body of the lattice structure of one tower multiplied by the number of towers per kilometer; and (2) the "tower area at ground" parameter, which is given by the area of the quadrilateral circumscribing at a 2-meter distance the supporting points of the tower, multiplied by the number of towers per kilometer.

All the towers considered in Fig. 2 have the same switching-surge insulation level; the tower–conductor clearances are different because of the air insulation performance, which varies in relation to the structure shape.

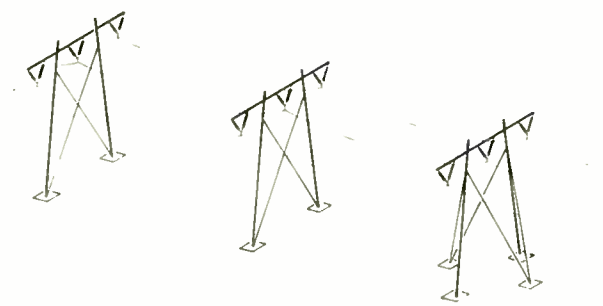
In the illustrations, the earth wires have not been drawn in, since they have minor structural importance and since some of the solutions do not admit of them.

The traditional structure already dealt with in the first part of the article is illustrated in Fig. 2(A). In Fig. 2(B) is shown the conventional guyed solution, which is widely used for voltages up to 500 kV. This solution allows for considerable reduction in the tower volumetric dimensions, although it somewhat increases the dimensions at ground. Moreover, it has the advantage of a lower "sensitivity" of the cost in respect of insulation, if compared with the solution of Fig. 2(A).[1]

The other solutions shown in Fig. 2 make use of new technological possibilities. These solutions, already studied and in part experimented on 420-kV lines in Italy, spring mainly from the consideration that the essential function of the tower is to withstand the vertical and transverse forces exerted by the conductors, whereas the function of withstanding longitudinal forces, which is necessary in order to ensure the stability of the whole line, may be left to only a few special towers.

It is therefore possible to conceive a line in which most towers lack longitudinal strength (bidimensional towers). As an example, Fig. 3 shows the solution experimentally

**FIGURE 3. Line with bidimensional towers. (Photograph shows the bidimensional tower of a 420-kV Italian line.)**



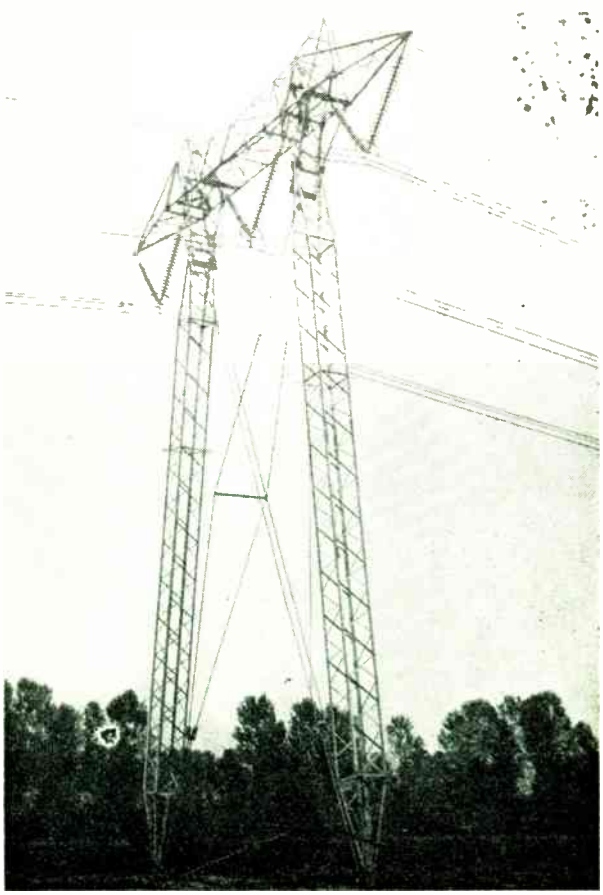
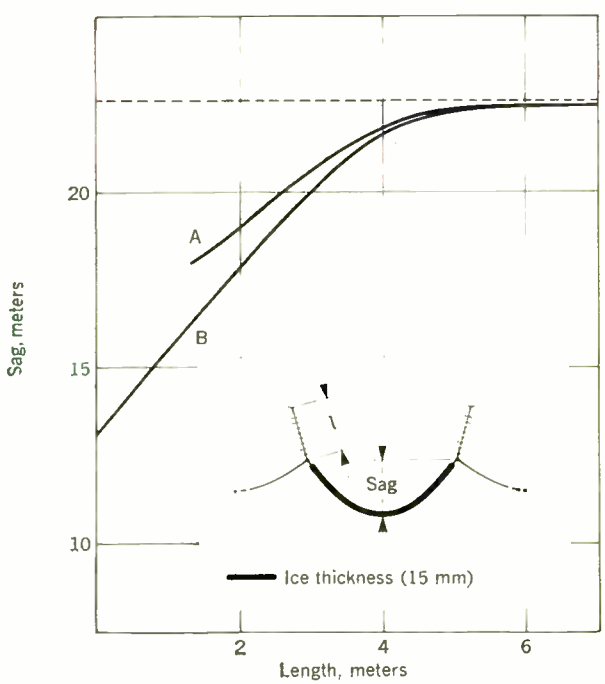


**FIGURE 4. Sag in an overloaded span as a function of the length of the insulator string. The other spans are assumed to be unloaded. (A) Line shown in Fig. 3. (B) Corresponding traditional line.**



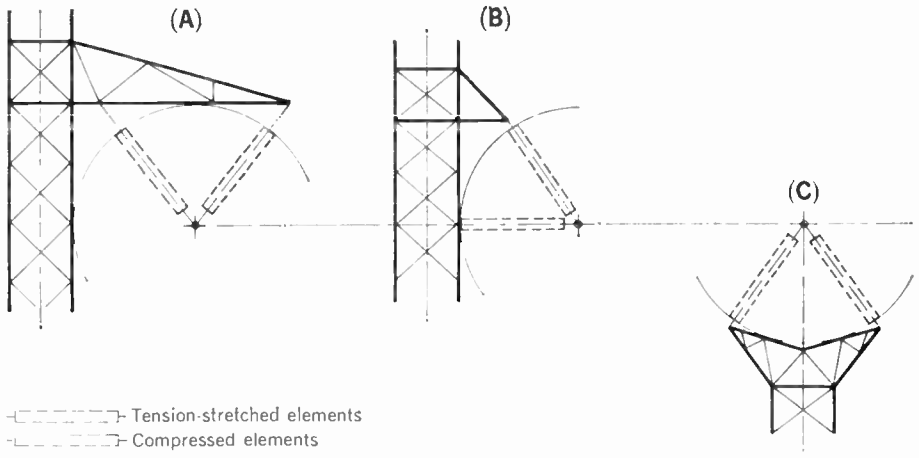



**FIGURE 5. Bidimensional insulating structures. (A) V string. (B) Insulating crossarm. (C) Reverse V crossarm.**

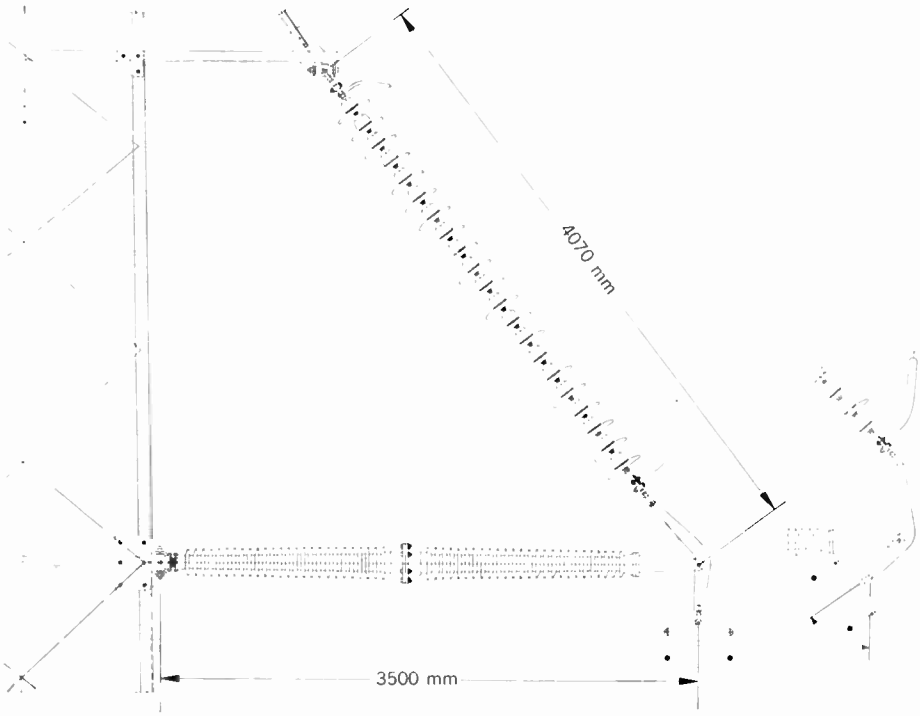



**FIGURE 6. Insulating crossarm for a double-circuit 420-kV line under construction in Italy, under normal and windy conditions.**

adopted in Italy for a 420-kV line, in which bidimensional guyed towers, as in Fig. 2(C), alternate with conventional guyed towers. It might be feared that a solution of this kind would increase the risk of the conductors making contact with objects crossed over by the line, because of the considerable longitudinal mobility of the conductor suspension points in the bidimensional towers. Figure 4 shows, in concise form, the reduction in clearance caused by a formation of ice concentrated on all the conductors and earth wires of one span, as a function of the length of the insulator string; this is shown in respect both of the traditional types of line and of the bidimensional tower line (Fig. 3). As can easily be seen, when the insulator strings exceed 4 meters in length (as is normal in UHV lines), the difference in behavior between the two lines is no longer appreciable; the longitudinal mobility of the conductor connection points, which is ensured by the long suspension strings, is already so great that any further increase in that mobility has little effect.

The main problem involved in designing lines that make use of these techniques is represented by the research to be done into the type and the frequency of the towers that have to be used in order to ensure the stability of the line.[10] In the case of the line in Fig. 3, the stabilizing tower is a conventional guyed suspension tower; one tower of this type for every five bidimensional ones would appear to be more than sufficient to make the line stable.

The use of bidimensional structures makes it much easier to achieve "insulating structures" to take the place of insulator strings and also of parts of the steel structures. The simplest and most direct insulating structure is a rotating crossarm consisting of a tensioned and a compressed insulating element, shown in Fig. 5(B), which can replace the traditional type of crossarm equipped with V strings, shown in Fig. 5(A). The "reverse V," shown in Fig. 5(C), is another insulating structure. Since the structure is formed by two elements that are generally com-

pressed, it is possible to support the conductors from below instead of suspending them, with a consequent considerable saving in the metallic structure. Future research will have to be devoted to finding the best materials for insulating structures; however, it is conceivable that they could be constructed with traditional materials.

Figure 6 shows, as an example, the insulating crossarm adopted in Italy on a double-circuit 420-kV line. The tension rod consists of a string of suspension insulators and the compressed element consists of two porcelain post insulators. Good compression performance can be obtained with porcelain in insulator lengths up to those suitable for UHV systems. The low bending strength of porcelain, however, makes it impossible for the compressed element to withstand even secondary bending stresses. In the case of bundled conductors, avoiding secondary stresses involves complex design solutions. Some design complications arise even with twin conductors, as can be seen from Fig. 6.

Going back now to the detailed analysis of the types of towers illustrated in Fig. 2, we note that the bidimensional guyed tower shown in Fig. 2(C) corresponds to a type that has been already used on an experimental basis on 420-kV Italian lines. Its advantage, as compared with the three-dimensional guyed tower in Fig. 2(B), is essentially in the cost; furthermore, this tower presents a slight decrease in the volumetric dimensions and a more important decrease in the dimensions at ground, while the lattice structures remain basically the same.

In the tower shown in Fig. 2(D), the external crossarms of the bidimensional guyed tower are replaced by insulating crossarms. Thus a satisfactory reduction of the tower size and, in part, of the volumetric dimensions is achieved.

The tower in Fig. 2(E), which can be derived from that in Fig. 2(D) by raising the central conductor and turning inside the insulating crossarms, is characterized by an extremely reduced line width (less than that of the existing 765-kV lines). The tower volume and the tower area at ground are also considerably reduced; on the other hand, this solution implies an increase in the tower area. To avoid excessive heights of the tower in connection with this special shape, it is advisable to reduce the span length from about 500 to 400 meters. Due to the reduction in the interphase distances, the surface gradient at the conductor increases by 15 percent in respect to the Fig. 2(D) tower. This makes it necessary to take steps to reduce the occurrence of corona—for instance, by increasing the number of subconductors per phase or the subconductor diameter. On the other hand, the electric field at ground level is reduced to values corresponding to those of existing 765-kV lines.

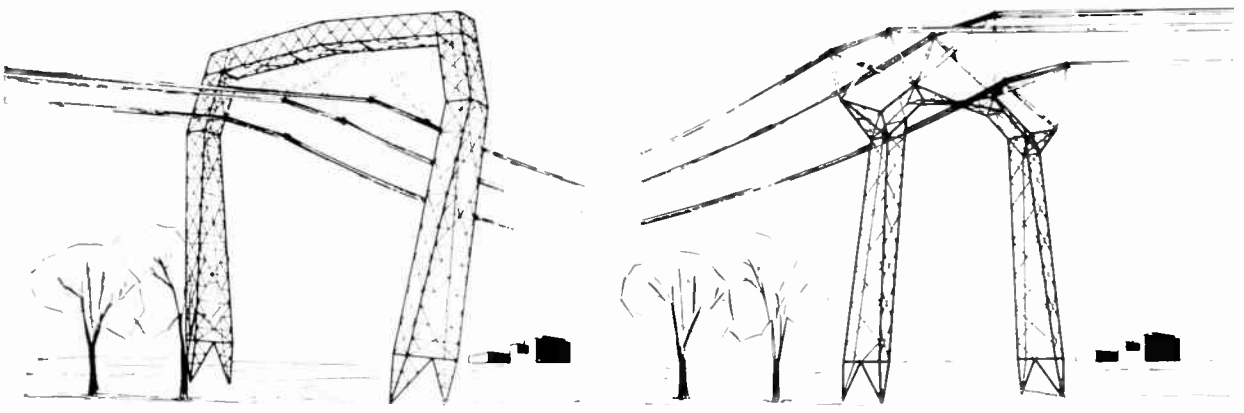
The designer who makes use of the solution offered by Fig. 2(E) is confronted with the problem of fixing a limit for the interphase clearance of the line. This problem has never arisen with traditional structures, since they are designed for a clearance between phases of more than double the clearance to ground. In choosing the interphase distance for this solution (12 meters), we have followed the rules discussed in Ref. 1, which seem to be fairly acceptable given the present state of knowledge. The possibility of adopting tower solutions of this type makes it advisable to go deeper into the problem of the switching impulse behavior of interphase distances, and in particular into the behavior of extremely long electrodes, such as the conductors of a line.

The solution in Fig. 2(F), which makes use of one reverse V string and of two insulating crossarms, achieves the minimum tower area together with a considerable reduction in the line width; moreover, a considerable reduction in the electric-field strength at ground is obtained. The tower is three-dimensional and only the insulating structures are bidimensional. The structure does not provide for earth wires, unless some important modifications are introduced.

The solution in Fig. 2(G), which is derived from the traditional solution by replacing the V strings with "reverse V" strings, also makes it possible to reduce considerably both the tower dimensions and the line width. The tower is three-dimensional and only the insulating structures are bidimensional. In this case too, corona problems in connection with increase in the conductor gradient, as well as switching-surge problems, arise as a consequence of the reduction in the interphase clearances. On the other hand, also in this case the electric field at ground is relatively low (equal to that of the existing 500-kV lines). The installation of earth wires is possible only if the structure is considerably modified.

Finally, the solution illustrated by Fig. 2(H) reduces the

**FIGURE 7. Two possible solutions for stabilizing towers for a line equipped with the tower shown in Fig. 2(G).**



metallic structure to very small proportions. This is possible if high-ultimate-strength insulators are available and if the power-frequency strength of the insulator strings can be ensured under extremely high phase-to-phase voltage stresses.

For all the solutions shown in Figs. 2(C) to 2(H), only the characteristics of bidimensional towers have been considered. These towers, which are by far the most numerous in the line, have to alternate with other towers—whose function, as mentioned, is a stabilizing one and which, in general, do not possess the characteristics of lightness of those that have already been described. There are many possible solutions for stabilizing towers both as to type and frequency, but to suggest any solutions here would perhaps be not only premature, but also needlessly lengthy. However, just to give an idea, Fig. 7 indicates two possible solutions for stabilizing towers corresponding to the structural solution in Fig. 2(G); obviously, another possibility consists in using normal dead-end towers.

In describing all the new solutions, we have avoided speaking of the considerable saving that can be achieved in tower costs. Actually, we feel that these solutions are primarily of interest not because of the saving in cost, but because they offer considerable reduction in the dimensions. On the other hand, the cost advantage could be partly overshadowed by the difficulties in line erection that the new solutions present. These difficulties, which seem to increase progressively from Fig. 2(C) to 2(H), are hard to evaluate without any direct experience. However, we can definitely say, on the basis of our own experience, that the increased burden of the erection operations does not turn to be particularly significant in the case of the Fig. 2(C) configuration.

## Conclusions

1. Ultrahigh-voltage lines up to 1500 kV can be conceived on the basis of traditional techniques, keeping about constant the ratio between the cost of the "inactive" components of the line (towers, foundations, and insulators) and the total cost of the line. For this purpose:
   (a) The switching overvoltages are to be controlled to values as low as 1.5 pu.
   (b) The power-frequency voltage strength per unit length of insulator strings as long as 10 meters, should not be lower than the strength of shorter strings now used.
   (c) Increases of the gradient at ground up to 18 kV/m are to be accepted.

Under these conditions, the linear dimensions of towers increase less than proportionally with the voltage (in particular the width of the right of way doubles over the range of 420 kV to 1500 kV), while the tower area increases more than proportionally (it becomes about five times larger as the voltage increases from 420 kV to 1500 kV).

2. A significant reduction in the right-of-way width, as well as in the dimensions of the tower structures, can be achieved if new techniques are adopted that are based essentially on the use of bidimensional towers and insulating structures. By applying these methods, 1500-kV lines with dimensions comparable to those of 765-kV (and even of 500-kV) lines can be conceived.

3. For the design of future UHV lines, research is needed particularly in the following fields:
   (a) Control of switching overvoltages to values below 1.7 pu.
   (b) Pollution performance of long insulator strings, up to 10 meters.
   (c) Possibility of accepting high gradients at ground.
   (d) Electrical and mechanical performance of bundles with a high number of subconductors.

In particular, to adopt the new techniques that are suggested here for UHV lines, research is needed in the following fields:
   (a) Development of long insulating elements that can withstand compression stresses combined with rather high secondary bending stresses.
   (b) Establishment of requirements for minimum interphase distances between line conductors.
   (c) Stabilization of lines equipped with bidimensional towers.
   (d) Erection of lines with bidimensional towers and insulating structures.

Revised text of a paper presented at the American Power Conference, Chicago, Ill., April 22–24, 1969. Scheduled for publication in the proceedings of the conference.

REFERENCES

1. Paris, L., and Comellini, E., "Cost reduction of EHV lines," Rept. 422, CIGRE, 1966.

2. Paris, L., and Sforzini, M., "RI problems in HV-line design," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-87, pp. 940–946, Apr. 1968.

3. Anderson, J. G., and Barthold, L. O., "Design challenges of transmission lines above 765 kV," Rept. 10, IEEE EHV Transmission Conf., Montreal, Sept. 1968.

4. Catenacci, G., Carrara, G., Furioli, G., and Dellera, L., "A first look on the main electrical design problems," Rept. 11, IEEE EHV Transmission Conf., Montreal, Sept. 1968.

5. Kimbark, E. W., and Legate, A. C., "Fault surge versus switching surge—a study of transient overvoltage caused by line-to-ground faults," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-87, pp. 1762–1769, Sept. 1968.

6. Colclaser, R. G., Wagner, C. L., and Donohue, E. P., "Multistep resistor control of switching surges," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-88, July 1969.

7. "Electric transmission structures," EEI Pub. 67-61.

8. Paris, L., "Influence of air gap characteristics on line-to-ground switching surge strength," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-86, pp. 936–947, Aug. 1967.

9. Paris, L., and Cortina, R., "Switching and lightning impulse discharge characteristics of large air gaps and long insulator strings," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-87, pp. 947–957, Apr. 1968.

10. Paris, L., and Comellini, E., "Bidimensional structures for transmission lines," presented at the 1969 IEEE Summer Power Meeting, Dallas, Tex., June 22–27.

**Luigi Paris** (M) was born and educated in Pisa, Italy, where he received the Ph.D. degree in electrical engineering in 1950. Since he started his career in the electric utility field he has been concerned mainly with the design and construction of EHV lines. He was a member of the staff who designed the Italian 420-kV system. He is now with ENEL (the Italian National Electric Agency), where for some years he directed the Electrical Research Centre, being particularly concerned with the problems related to the planning and engineering of electric power generation and transmission systems. Since 1967 he has been vice manager of the Hydro and Electric Engineering Centre of ENEL, in charge of transmission engineering. He is also professor of power system analysis at the University of Pisa.

# An introduction to synthetic-aperture radar

*Radar beams can be narrowed without being narrowed. It's all a matter of sifting information from shifted frequencies by processing data on an optical bench*

**William M. Brown, Leonard J. Porcello**
*The University of Michigan*

*One way of achieving fine-resolution terrain imagery using airborne, side-looking radar is to boost frequency; another is to decrease along-track tracking. Neither is attractive. But they would have had to do were it not for coherent wave radar—upon which synthetic-aperture radar is premised. Using coherent radar, resolution can depend, not on the width of the beam, but on Doppler frequency shift. The azimuthal resolution of side-looking radar can therefore be of the same order of magnitude as that for range resolution. The key to converting the theoretical groundwork into a "full-bodied" system is an appropriate data-processing scheme. And the simplest scheme for working, processing, and deciphering the data is optical.*

Although the amount of information carried by an electromagnetic signal is constrained, depending on the technicians' resourcefulness, there are limits and *limits*. Were it not for some electronic "trickery," therefore, the range resolution in the direction of flight of present-day, side-looking, airborne radar would be relatively poor—and the future, despite increasingly sophisticated and costly components, would not hold much promise for improvement.

One useful technique for overcoming the physical constraints that limit the more "straightforward" type of radar systems is called synthetic-aperture radar. Using it, along-track resolution does not depend exclusively on how narrow the beam is. Rather, resolution is provided by Doppler processing the received echoes. In this way, greatly improved performance—an order of magnitude or more—is possible.

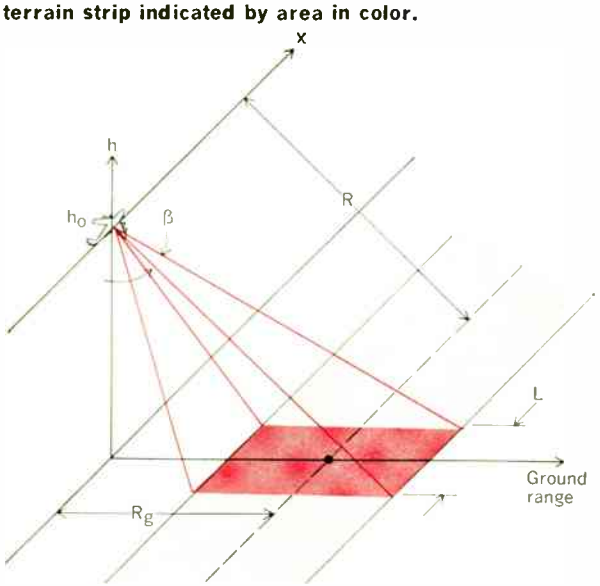
## The synthetic-aperture concept

The genesis of the synthetic-aperture concept appears to have been the work of Carl Wiley of the Goodyear Aircraft Corp. in the early 1950s; the development of his ideas is retraced in a paper by Sherwin *et al.*[1] Wiley observed that a one-to-one correspondence exists between the along-track coordinate of a reflecting object (being linearly traversed by a radar beam) and the *instantaneous Doppler shift* of the signal reflected to the radar by that object. He concluded that a frequency analysis of the reflected signals could enable finer along-track resolution than that permitted by the along-track width of the physical beam itself.

This "Doppler beam-sharpening" concept was not only exploited by Goodyear, but by a group at the University of Illinois. An experimental demonstration of the beam-sharpening concept was carried out by the Illinois group in 1953 through use of airborne coherent X-band pulsed radar, "boxcar" circuitry, a tape recorder, and a frequency analyzer.

One major problem, recognized quite early, was implementation of a practical data processor that could accept wide-band signals from a storage device and carry out the necessary Doppler-frequency analysis at each

**FIGURE 1. Antenna of side-looking radar illuminates terrain strip indicated by area in color.**



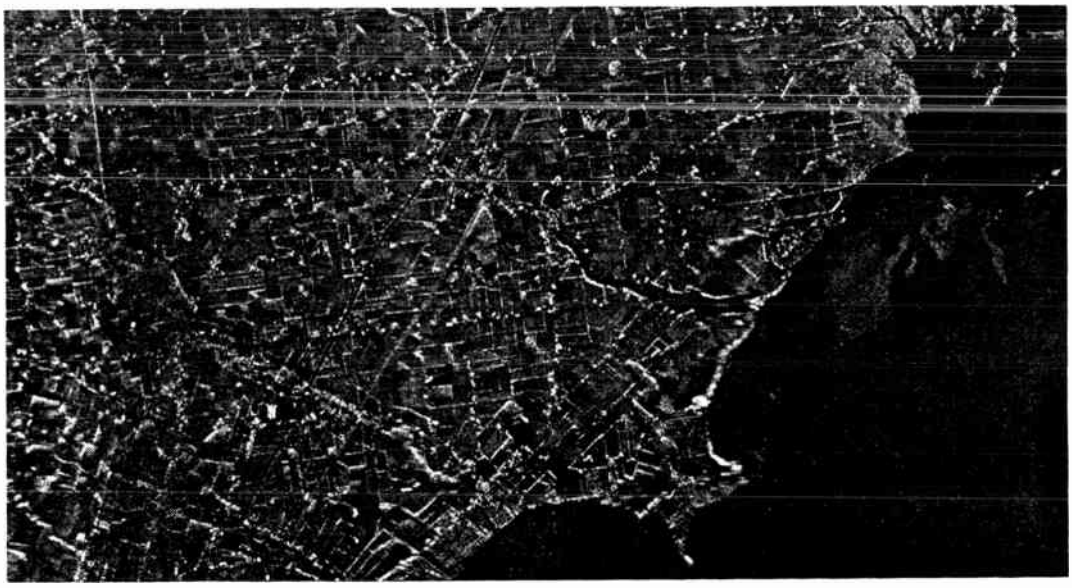


**FIGURE 2. Here is an example of side-looking radar's imaging capabilities.**

resolvable slant range. During 1953 a summer study was held at the University of Michigan. Known as Project Wolverine, it served as a point of departure for expanded efforts in, and diverse approaches to, fine-resolution, terrain-imaging radar.

The basic geometry of the side-looking, terrain-imaging radar is shown in Fig. 1. The airborne radar radiates through an antenna whose vertical pattern is chosen to illuminate a terrain strip properly—the colored region in Fig. 1—parallel to the flight path. The antenna beam, usually narrow in the along-track (azimuthal) dimension, is not rotated, but scans the terrain as a result of the forward motion of the aircraft. The radar system generates a two-dimensional image of the reflectivity distribution of the terrain strip, with one dimension proportional to slant range and the second to along-track coordinate position. Some examples of side-looking radar imagery are shown in Fig. 2. The slant-range dimension is vertical, whereas along-track is horizontal.

A side-looking radar achieves slant-range resolution through the use of pulsing and time-delay sorting, as illustrated in Fig. 3. If the radar transmits a very short pulse, reflected by a target at slant range $R$, then the round-trip propagation time between the radar and this target is given by

$$\Delta T = \frac{2R}{c} \tag{1}$$

where $c$ is the propagation speed of the radar wave. The reflections from targets at different ranges will, naturally, arrive at the receiver with different time delays. If the transmitter pulse is very short, say of duration $\tau$, then the returns from targets at sufficiently different ranges will be nonoverlapping in time. Specifically, the required separation is

$$\Delta R_s \geq \frac{c\tau}{2}$$

Accepting this as a measure of slant-range resolution $\rho_R$, then

$$\rho_R \approx \frac{c\tau}{2} \tag{2}$$

and the corresponding ground-range resolution is

$$\frac{c\tau}{2} \sec \psi$$

where $\psi$ is the depression angle of the line of sight to the target with respect to local horizontal, as shown in Fig. 3.

In a simple, pulsed radar, generation of a pulse of duration $\tau$ requires a transmitter bandwidth of the order of

$$W \approx \frac{1}{\tau}$$

and preservation of the range resolution $c\tau/2$ requires that the receiver and display also have bandwidth $W \approx 1/\tau$. Obviously, the key to achieving fine range resolution is wide-band radar transmitters and receivers—as amply demonstrated in pulse-compression technology.* The bandwidth sets the fundamental constraint on range resolution:

$$\rho_R \approx \frac{c}{2W} \tag{3}$$

By way of example, a radar with a bandwidth of 1 GHz provides a theoretical range resolution of

$$\rho_R \approx 15 \text{ cm}$$

Now the along-track or azimuthal cross section of the

* In a pulse compression radar, we radiate "dispersed" pulses with time–bandwidth product $\tau W \gg 1$ (rather than $\tau W \approx 1$ as in the case above) and use a data processor to "compress" the received pulses to a time duration $1/W$.

antenna pattern has a half-power angular width of $\beta$ radians. The corresponding along-track beam width at range $R$ is

$$L \approx \beta R \tag{4}$$

If the distance $L$ is accepted as measure of the along-track resolution $\rho_x$ of this radar, then the only recourse for achieving fine resolution at long range is to make $\beta$ very narrow. But an antenna aperture with along-track dimension $D$, operating at its diffraction limit at wavelength $\lambda$, yields a half-power angular beam width of

$$\beta \approx \frac{\lambda}{D} \text{ radians} \tag{5}$$

therefore,

$$\rho_x \approx \frac{\lambda R}{D} \tag{6}$$

To keep $\rho_x$ small as $R$ increases, $D$ must be increased and/or $\lambda$ decreased. Each of these options becomes unattractive beyond certain limits. Large-$D$ antennas are incompatible with airborne operation; operation at very short wavelength leads to weather limitations; and, finally, construction of a diffraction-limited airborne antenna with $D \gtrsim 10^3\lambda$ can be costly as well as difficult. By way of example, a radar with $D \approx 10^3\lambda$ realizes, at slant range $R = 10$ km, an along-track resolution of

$$\rho_x \approx 10 \text{ meters}$$

The simple radar system of Fig. 1, therefore, has the potential to achieve fine range resolution, but appears to be constrained to relatively poor azimuthal resolution at long operating ranges. The synthetic-aperture technique can improve azimuthal resolution to the point where $\rho_x$ is commensurate with $\rho_R$.

If the radar emits a sinusoid of frequency $\nu_0$ Hz, and if the angle $\gamma$ (see Fig. 4) changes sufficiently slowly, then the reflected signal reaching the receiver is a sinusoid with slowly varying frequency $\nu = \nu_D + \nu_0$ where the instantaneous Doppler shift $\nu_D$ is given by

$$\nu_D = \frac{2v}{c} \nu_0 \cos \gamma \tag{7}$$

When $\theta = \frac{\pi}{2} - \gamma$ is small compared with one radian:

$$\nu_D \approx \frac{2v}{\lambda R} (x_0 - x) \tag{8}$$

where $\lambda = c/\nu_0$. Therefore, at any range $R$, the Doppler shift $\nu_0$ is a linear function of $(x - x_0)$, and a frequency analysis of the return displays the radar reflectivity of the terrain as a function of $(x_0 - x)$.

The synthetic-aperture technique, although first conceived from the Doppler viewpoint, can also be approached from an entirely different, but mathematically equivalent, viewpoint. Following from Fig. 1, observe that a reflector at range $R$ is illuminated by the radar while the latter moves through a distance of

$$L = \beta R$$

If the physical antenna is regarded as one element of a linear array (extending in the direction of flight), occupying in time sequence all the elemental positions germane to the array, then, intuitively, it should be possible to "synthesize" an aperture of length $L$ by suitably storing the received signal before processing the data. At first glance, one would expect this synthetic aperture to have a "synthetic angular beam width" of approximately

$$\beta' \approx \frac{\lambda}{L} \text{ radians} \tag{9}$$

and a corresponding along-track width $R\beta'$. Substituting for $\beta'$, $R\beta'$ equals $D$.

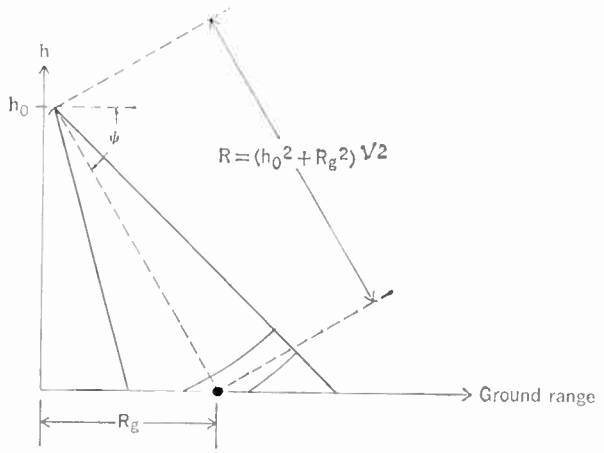
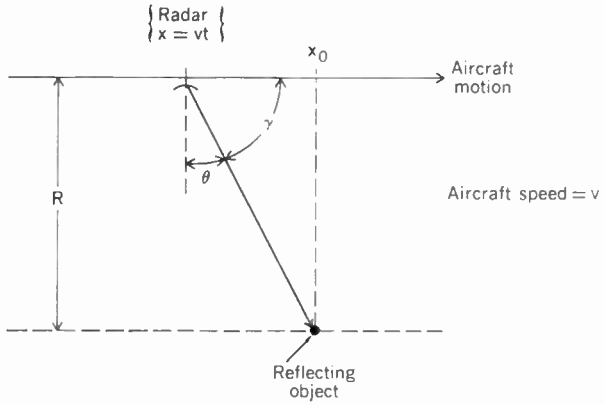



**FIGURE 3. Behind side-looking radar resolution are pulsing and time-delay sorting.**

**FIGURE 4. Coherent radiation enables the radar beam, through processing, to be effectively narrowed—hence the appellative, synthetic-aperture radar.**




## Radar observation of a rotating target

The motion of a terrain-imaging, synthetic-aperture radar past a target gives essentially arbitrary resolution in the along-track dimension (perpendicular to the range dimension). It should be expected that other forms of relative motion between radar and rigid target should lead to similarly fine resolution capability. In fact, if a stationary, coherent radar illuminates a target field consisting of a rotating rigid body, as in Fig. 5, it is possible to form a two-dimensional image of the object's radar reflectivity distribution by carrying out a delay-Doppler (or range-Doppler) analysis of the reflected signals.

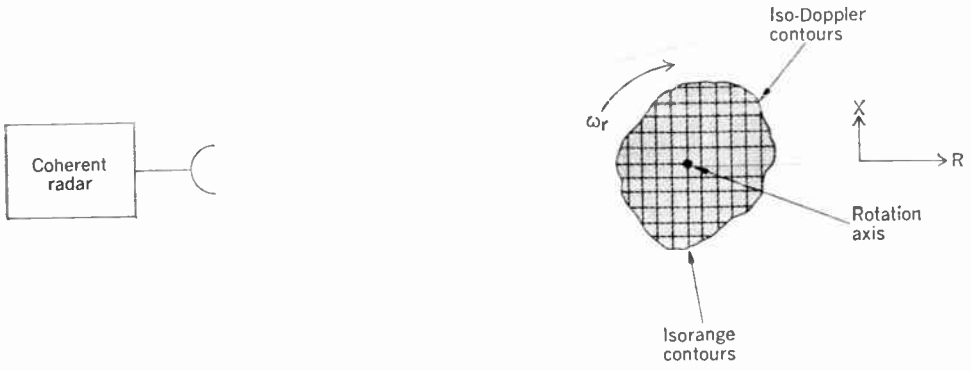



**FIGURE 5. Using a range-Doppler technique, it is possible to get a two-dimensional image of the radar reflectivity from a rotating body.**

Assume that the angular velocity of the body is $\omega_r$ rad/s about an axis perpendicular to the line of sight, with the radar far from the body. Then, the radial velocity of a point at "cross-range" position $x$ on the body is $\omega_r x$, and the Doppler shift of radar returns from this point is

$$\nu_D = \frac{2\omega_r x}{\lambda} \tag{10}$$

If the radar returns are sorted according to Doppler frequency, the scatterers are resolved in the cross-range dimension. If the radar also provides resolution in range, a two-dimensional (radar) image of the body can be obtained. And, if appropriate coherent processing is applied to the signals received over a time $T$, the obtainable Doppler resolution is about $T^{-1}$ Hz. Since the Doppler shift corresponding to the radial velocity $\omega_r x$ is $\nu_D = 2\omega_r x\lambda^{-1}$, a Doppler resolution of $\Delta\nu_D = T^{-1}$ implies a cross-range resolution of

$$\rho_x \approx \frac{\lambda}{2\omega_r T} = \frac{\lambda}{2\Delta\theta} \tag{11}$$

where $\Delta\theta$ is now the change in total aspect angle during the coherent processing time. The rotating-target-field version of the synthetic-aperture concept has been utilized by radar astronomers to obtain radar images of planets.[2,3]

For planetary radar, the contours of constant range are annuli about the point on the planetary surface closest to the earth (subradar point); the contours of constant Doppler are the intersections of the planetary surface with a family of planes—parallel to the rotation axis and to the line joining this axis and the radar. (An ambiguity normally exists between the upper and lower hemispheres of the planet, but this can often be resolved by auxillary methods.) Figure 6 shows, first, a radar image of a lunar crater generated at a wavelength of 3.8 cm by the Haystack radar and an electronic data processor and, second, an optical image of the same region.

Reconsider now the side-looking, terrain-imaging, synthetic-aperture radar situation. Consider a point target at an angle $\theta$ from broadside. The Doppler frequency associated with such a target is $\nu_D = \lambda^{-1}\, 2v \sin\theta$. A $\theta$-interval $\beta$ radians wide is well illuminated, and the corresponding Doppler bandwidth is $B_d \approx \lambda^{-1}\, 4v \sin(\beta/2)$; data with such a bandwidth can be time-resolved to about $\rho_t \approx B_d^{-1}$. In the along-track dimension, $x = vt$, so the corresponding along-track resolution is $\rho_x \approx vB_d^{-1}$ or

$$\rho_x \approx \frac{\lambda}{4 \sin\left(\frac{\Delta\theta}{2}\right)} \tag{12}$$

where $\Delta\theta = \beta$ is the change in aspect angle over which a target is observed by the side-looking radar. If in (12) we let $\Delta\theta = \lambda/D$ and $\sin\Delta\theta/2 \approx \Delta\theta/2$, resolution for the side-looking case is $D/2$. Actually, (12) is the most appropriate cross-range resolution formula for both the side-looking and rotating-target-field cases. In addition, consider a physical aperture and recall that the diffraction-limited angular beam width of approximately $\lambda/D$ yields an azimuthal linear resolution at range $R$ of about $R\lambda/D = \lambda/\Delta\theta$. Here* $\Delta\theta = D/R$ is again the aspect-angle interval over which the physical aperture views the target field. Observe the general property that a system collecting data over an aspect-angle interval $\Delta\theta$ can yield cross-range resolution as in (12) that agrees with (11) for small $\Delta\theta$.

## Converting theory to practice

The remarkable property—that one can enjoy along-track resolution

$$\rho_x \approx D \tag{13}$$

independent of range and of wavelength, that can be improved by *reducing* the dimension $D$—was first reported by Cutrona *et al.*[1] The key is the use of an appropriate data-processing scheme. In the long-antenna context, requirements are: to *store* the reflected data, to *amplitude-weight and phase-shift* the returns from sequential locations of the physical antenna, and to *coherently sum* these weighted returns as for a physical array. In fact, as $L$ becomes large, one finds it necessary to *focus* the synthetic aperture—i.e., to insert a phase shift, quadratic in distance along the aperture, to compensate for the curvature of the reflected wavefront—in order to realize

* The $\lambda/\Delta\theta$ differs from (11) by a factor of two because two-way propagation is involved in both synthetic-aperture situations discussed. Two-way propagation doubles phase shifts and halves resolutions. The same factor of two was ignored earlier when $D$ was derived as the along-track resolution for the side-looking radar.

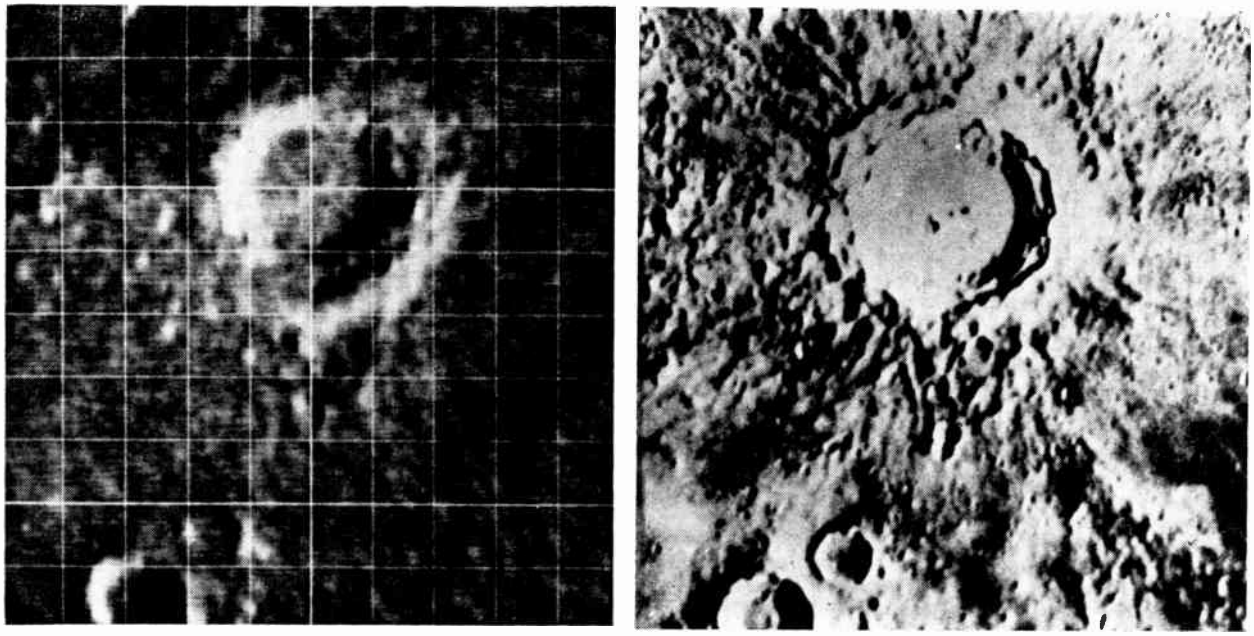


**FIGURE 6. Radar image, left, compares favorably with optical image of a moon crater. (Rectified Lunar Atlas, University of Arizona Press, 1963; courtesy G. H. Pettengill.)**

the ultimate along-track resolution. This focusing operation, which is range dependent, complicates the data processing to some degree, and makes it somewhat more difficult than a simple, range-dependent, Doppler-spectrum analysis. Systems failing to provide this focusing operation are known as *unfocused synthetic-aperture radars*, and have an along-track resolution capability of approximately

$$\rho_x \approx \sqrt{\lambda R_0} \tag{14}$$

Systems that provide focused processing are termed *focused synthetic-aperture radars* and have the resolution potential

$$\rho_x \approx D \tag{15}$$

Various forms of signal processors were attempted by early workers in the field. Goodyear placed initial emphasis on the development of a range-gated filter bank. But it turned out that it was impractical, at that time, to incorporate the focusing operation into this device. The Illinois group pursued three distinct processing approaches, again all limited to the unfocused case: (1) a recirculating delay line, developed at the Philco Corp.; (2) integration on photographic film; and (3) an electronic storage tube integrator. The Michigan group suggested an optical system signal processor for synthetic-aperture generation. (As we shall show, an optical system, illuminated with coherent light, can serve as a versatile analog computer for performing signal-processing operations.) The potential of a coherent optical system as a general-purpose signal processor was soon recognized and a coherent optical data processor that permitted generation of a synthetic aperture properly focused at each slant range was subsequently devised by Cutrona *et al.*[5,6] Finally, in 1957, a synthetic-aperture radar using a coherent optical data processor was successfully demonstrated.

Optical processors were subsequently applied to pulse compression and to range-Doppler radars such as those used in planetary mapping, and to the problem of automatic shape recognition. Some of the optical techniques devised in the context of radar data processing were appropriately modified by E. N. Leith for application to wavefront reconstruction—initially suggested by Gabor[7]—and formed the basis of modern holography.

## System configuration

To employ optical processing, it is customary to generate a photographic transparency on which the unprocessed data are stored. In the case of a synthetic-aperture radar, this transparency has the properties of an array of one-dimensional off-axis holograms. As a result, the optical data processor—the heart of a synthetic-aperture radar—can be described as a viewer in which a one-dimensional wavefront reconstruction process takes place, without losing the range resolution of pulsed radar. The simplified block diagram of Fig. 7 shows the interrelation of a radar, a processor, and an associated storage device, which collectively constitute a synthetic-aperture radar. A stable oscillator serves as a common reference source for both the transmitter and receiver; i.e., the radar is coherent. Therefore, the phase delay associated with a round-trip propagation signal can be determined by synchronously demodulating the received signal with respect to the transmitter output. This permits the physical antenna to sequence observations rather than to make a full set of spatially diverse observations of the reflected wave at one instant, as with a physical array system. Pulsing the radar need not upset the fundamental coherence provided by the stable source.

## Data processing

The synthetic-aperture, terrain-imaging, data-processing concept might be straightforward in principle. But

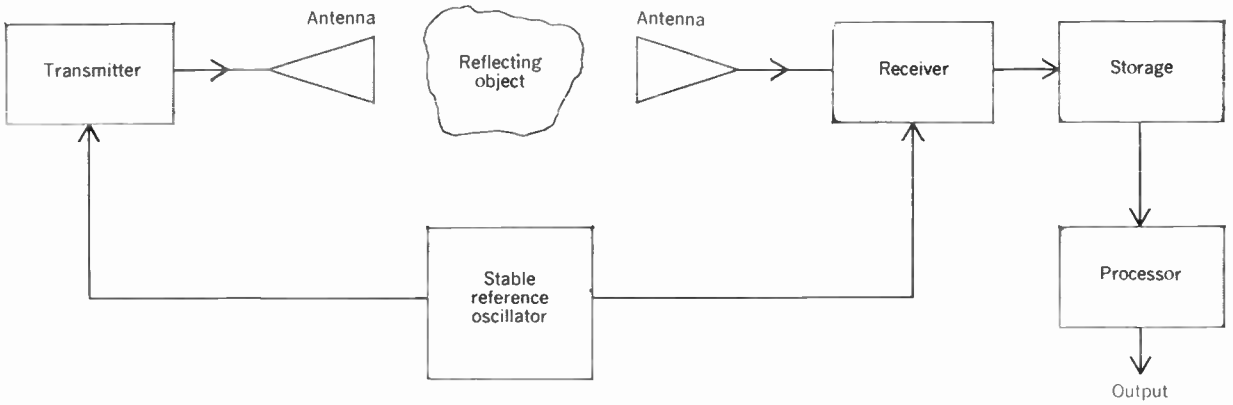



**FIGURE 7. These pieces collectively constitute a synthetic-aperture radar.**

one critical, early question was whether the necessary processor could be implemented in a form neither prohibitively cumbersome nor expensive for large-scale tactical use.

The output of the radar receiver typically has a bandwidth of many megahertz, and the storage device associated with the processor must accept data at a commensurate rate. It was recognized quite early that photographic film is (but for the nuisance of chemical development) well suited to the task of storing the data. The stored data could be read out optically as a viewable image. It made sense, therefore, to consider processing using optical elements alone. (See box.)

In a simple, optical, holographic system, an object is illuminated with coherent light as shown in Fig. 11 (A); each scattering center on the object generates a spherical wave that impinges on a plane $P$. A collimated, reference wave also passes through $P_1$ and a film, in plane $P$, records the intensity distribution of the sum of the reference and scattered waves. This recorded interference pattern, or hologram, has the appearance of a section of a Fresnel zone plate* of focal length $f'$—the distance from the scattering center to $P$. The portion of this recorded zone plate pattern is a function of the size and location of the aperture at $P$ and the offset angle $\alpha$ between the reference beam and the scattered wave's propagation vector.

A recorded portion of the zone plate, when it is removed from the origin is, inferentially, called an "off-axis" zone plate. Inserting a "spatial offset frequency" into the pattern is equivalent to translating the origin of the pattern; this offset frequency is controlled by the angle $\alpha$. If the photographic transparency is developed and then reilluminated with a collimated, coherent light beam, the recorded zone plate gives rise to a pair of spherical waves—one diverging and one converging—each of focal length $f'$, as illustrated in Fig. 11(B). The two waves correspond, respectively, to a virtual and a real image of the scattering center. The propagation directions of the waves are offset by the angle $\alpha$ with respect to the hologram. A third wave associated with the average gray level, or optical bias of the hologram, continues along the optic axis.

A complex scattering object, consisting of a spatial superposition of scattering centers, gives rise to a corresponding superposition of reconstructed waves. The virtual image from the hologram produced by the complex scattering centers appears to be an accurate three-dimensional facsimile of the object itself. One might, therefore, anticipate that optical elements can be used to display the two-dimensional spectrum of a light distribution, appropriately phase-shift and/or amplitude-weight portions of the spectrum by changing the gray scale (through suitable filtering), and then examine the (modified) light distribution. Alternatively, it should be possible to display a one-dimensional spectrum in each of many, independent channels; to modify the spectrum in each channel; and to examine the modified light distribution.

In order to apply these optical filtering properties to radar data processing, the radar data must be stored in an appropriate format.

In its simplest form, the radar emits short pulses of microwave energy, of duration $\tau$ second, periodically at at rate of $\nu_s$ pulses per second. The reflected signal has the form of Fig. 12; $R_1$ is the shortest slant range illuminated by the radar and $R_2$ is the maximum slant range illuminated. The signal emerging from the synchronous demodulator in the receiver is bipolar with respect to some arbitrary bias level; the sum of signal plus bias, a monopolar waveform, is used to intensity-modulate a cathode-ray tube whose beam is swept repetitively in a straight line. To map the full slant-range interval illuminated, the $n$th sweep is started at time $T_0 + (n - 1)\frac{2R_1}{c}$ and is continued for a duration $\frac{2}{c}(R_2 - R_1)$. The intensity-modulated sweep is then photographed onto the radar data film that moves in a direction perpendicular to the sweep direction, as shown in Fig. 13. The variable $y_1$ is then a measure of slant range and $x_1$ corresponds (after scaling) to the along-track position of the air-

* A conventional Fresnel zone plate is a film transparency with a radially symmetric intensity transmission $I(x,y)$ given by

$$I(x,y) = |A_0 + A_1 \cos[\mu(x^2 + y^2)]|^2$$

or a nonlinear function of $I$. If $I(x,y)$ is precisely as given in the equation, the zone plate has a focal length

$$f' = \pm \frac{\pi}{\mu\lambda}$$

where $\lambda$ is the wavelength used to illuminate the zone plate. The ($\pm$) signs denote conjugate foci—i.e., an incident collimated beam gives rise to a real and a virtual point image. If the plate is a nonlinear function of $I$, then higher-order conjugate foci, corresponding to harmonics generated by the nonlinearity, also appear.

craft—hence the position of the physical antenna element. Ideally, the intensity modulation of the CRT and the development of the film are both controlled so that, after development, the light-amplitude transmissivity of the film is a linear function of the bipolar video signal amplitude.

If the data film is illuminated with a normally incident, coherent light beam, the light emerging from the film is uniform in phase (assuming the film itself does not corrupt the phase) and is amplitude modulated as a function of $x_1$ and $y_1$. For any $y_1$ = constant, the $x_1$-dimension displays the elemental signal received at each point along the synthetic aperture for the range corresponding to $y_1$.

### The optical analog to electronic methodology

In a physical radar array, elemental returns would be amplitude-weighted (to control sidelobe levels), phase-shifted (to provide beam steering and/or to focus the array), and then coherently summed. In an optical system, amplitude-weighting can be accomplished by inserting a shaded transparency with uniform phase thickness adjacent to the data film. Phase-shifting is attained by inserting appropriately shaped pieces of glass adjacent to the data film. If for any given $y_1$ corresponding to a slant range $R$ the array is to be focused at $R$, a phase shift quadratic in $x_1$ (with coefficients a function of $R$) must be inserted across the aperture to compensate for the quadratic phase delay that occurs when a spherical wave, emanating from the reflecting object, is observed along the synthetic aperture. Such a quadratic phase correction is provided by a simple plano-convex lens placed in proximity to the data film at the appropriate $y_1$. At each $y_1$, the lens must have an $x_1$-dimension focal length proportional to $y_1$. This is shown in Fig. 14(A). A lens that continuously satisfies this requirement for

### The Fourier property

The most significant property of a coherently illuminated optical system is the Fourier transform property. If coherent illumination of wavelength $\lambda_L$ is incident on a spherical lens, then complex light amplitude distributions $\hat{A}_1(x_1,y_1)$ and $\hat{A}_2(x_2,y_2)$ in the front and back focal planes of the lens are quite nicely related by a Fourier transformation. This is illustrated in Fig. 8. For simplicity, assume the illumination to be linearly polarized; by "complex light amplitude," we mean the magnitude and phase (relative to some arbitrary reference phase) of the corresponding linearly polarized electric field intensity. Employing the coordinate system of Fig. 8, this can be expressed as

$$\hat{A}_2(x_2,y_2) \sim \int\int \hat{A}_1(x_1,y_1) \exp\left[-j\alpha(x_1x_2 + y_1y_2)\right]dx_1\,dy_1 \quad (16)$$

where

$$\alpha = -\frac{2\pi}{\lambda_L f_1} \quad (17)$$

and $f_1$ is the focal length of the lens.

In other words

$$\hat{A}_2(x_2,y_2) \sim \mathfrak{F}_{x,y}\{\hat{A}_1(x_1,y_1)\} \quad (18)$$

where $\mathfrak{F}_{x,y}\{\cdot\}$ denotes a Fourier transform with respect to both $x$ and $y$. This interesting relation follows directly from the Kirchhoff diffraction integral and simply says that the amplitude spectrum of a two-dimensional function displayed at plane $P_1$ can be observed at plane $P_2$. If two such stages are cascaded with the lenses separated by $2f_1$, as illustrated in Fig. 9, then

$$\hat{A}_3(x_3,y_3) \sim \mathfrak{F}_{x,y}\{\mathfrak{F}_{x,y}[(\hat{A}_1x_1,y_1)]\} \quad (19)$$

which renders $\hat{A}_3$ proportional to $\hat{A}_1$ in reversed coordinates:

$$\hat{A}_3(x_1,y_1) \sim \hat{A}_1(-x_1,-y_1) \quad (20)$$

It should be apparent, therefore, that a function $f(x,y)$ can be inserted at $P_1$; the spectrum at $P_2$ can then be modified by inserting slits, stops, shaded transparencies, and/or optical phase shifters, and finally observed at $P_3$.

A companion relationship exists for cylindrical lenses having curvature in one dimension only. In this case, a Fourier transform is effected in only one dimension; the illumination simply diverges in the other dimension. Using a cylindrical lens and a spherical lens of equal focal lengths in the combination of Fig. 10: in the $x$-dimension, a Fourier transform relation exists between $\hat{A}_1$ and $\hat{A}_2$, whereas in the $y$-dimension, $\hat{A}_2$ is an image of $\hat{A}_1$ with the usual inversion of coordinates.

Therefore,

$$\hat{A}_2(x_2,y_2) \sim \int \hat{A}_1(x_1,y_1)\Big|_{y_1=-y_2} \exp\left[-j\alpha x_1x_2\right]dx_1 \quad (21)$$

This represents a multichannel one-dimensional spectrum analyzer, in which the channels are stacked in the $y$-dimension.

all values of $y_1$ is conical, as illustrated in Fig. 14(B). Since the synthetic-aperture length $L$ is also proportional to $R$, and hence to $y_1$, tapering the lens to a point at $R = 0$ is consistent with the required processing.

The coherent light emerging from the film–conical lens combination now has the property that the wave emanating from each $y_1$ (or range channel) is collimated in the $x_1$-dimension—being either uniform in phase or having a phase linear in $x_1$. But it has no focal properties in the $y_1$-dimension. Integration with respect to $x_1$ can be handled via a Fourier transform (see box below), evaluated at $\omega_x = 0$

$$\left(\text{i.e., } \mathfrak{F}_1\{f(x)\}\Big|_{\omega=0} = \int_{-\infty}^{\infty} f(x)e^{-j\omega x}dx\Big|_{\omega=0} = \int_{-\infty}^{\infty} f(x)\,dx\right)$$

To image in the $y$-dimension for the range resolution inherent in the data film calls for a sphere–cylinder lens combination, as shown in Fig. 15. A view at $P_2$ through a vertical slit shows the output of the synthetic aperture for all displayed ranges for one value of the along-track dimension. Note that the slit is displaced from the origin because a bias is needed in order to record a bipolar function and therefore record the radar data in a spatial, bandpass form purposefully to keep its spectrum away from dc. (There can be no negative intensities available from the CRT trace.) A small $x$-dimension carrier frequency is employed and the $x$-dimension at $P_2$ is then evaluated at the corresponding off-axis position.

## Input–output correspondence

It should be apparent, then, that a radar data film made with the optical conversion recording scheme of Fig. 13 is an analog of the all-optical holographic technique.

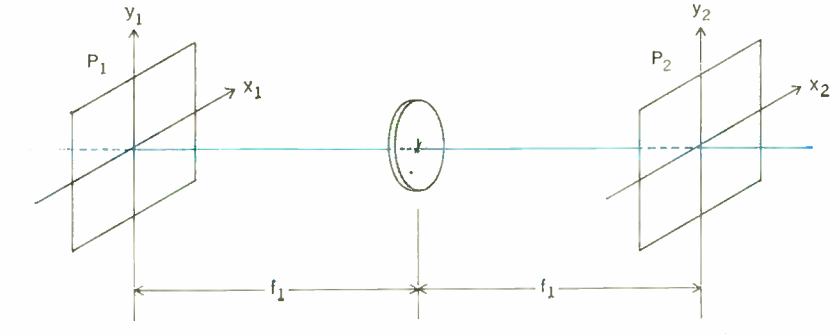



**FIGURE 8. Light distributions at the front and back planes of the lens are related by a Fourier transform.**

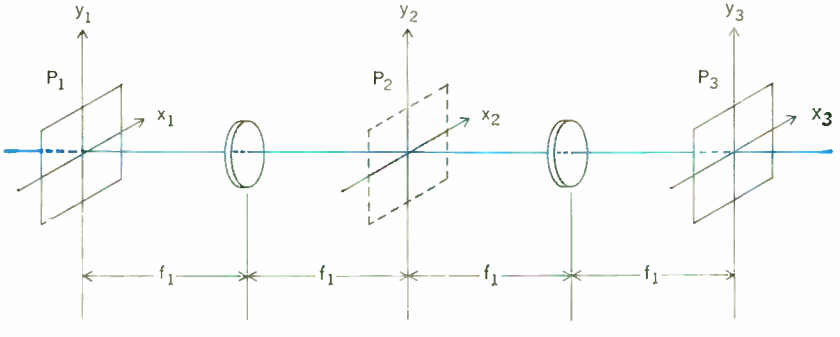



**FIGURE 9. A succession of two transformations, achieved with this double lens arrangement, reverses the coordinates of the original amplitude distribution.**

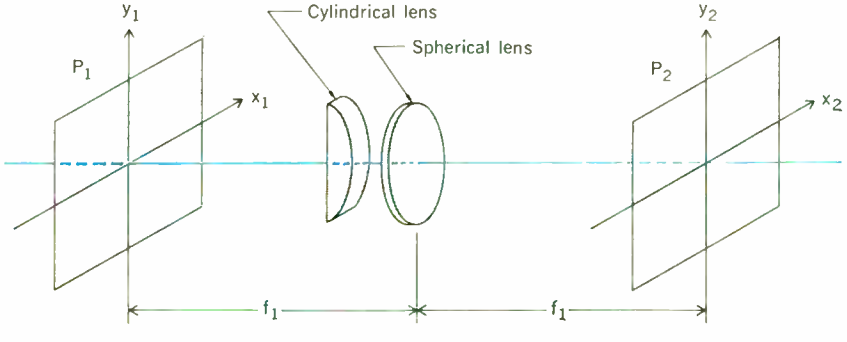



**FIGURE 10. Information from each of the coordinates can be processed differently. Here, $A_2$ is a Fourier transform of $A_1$ in the x-direction and simply inverted in the y-direction.**

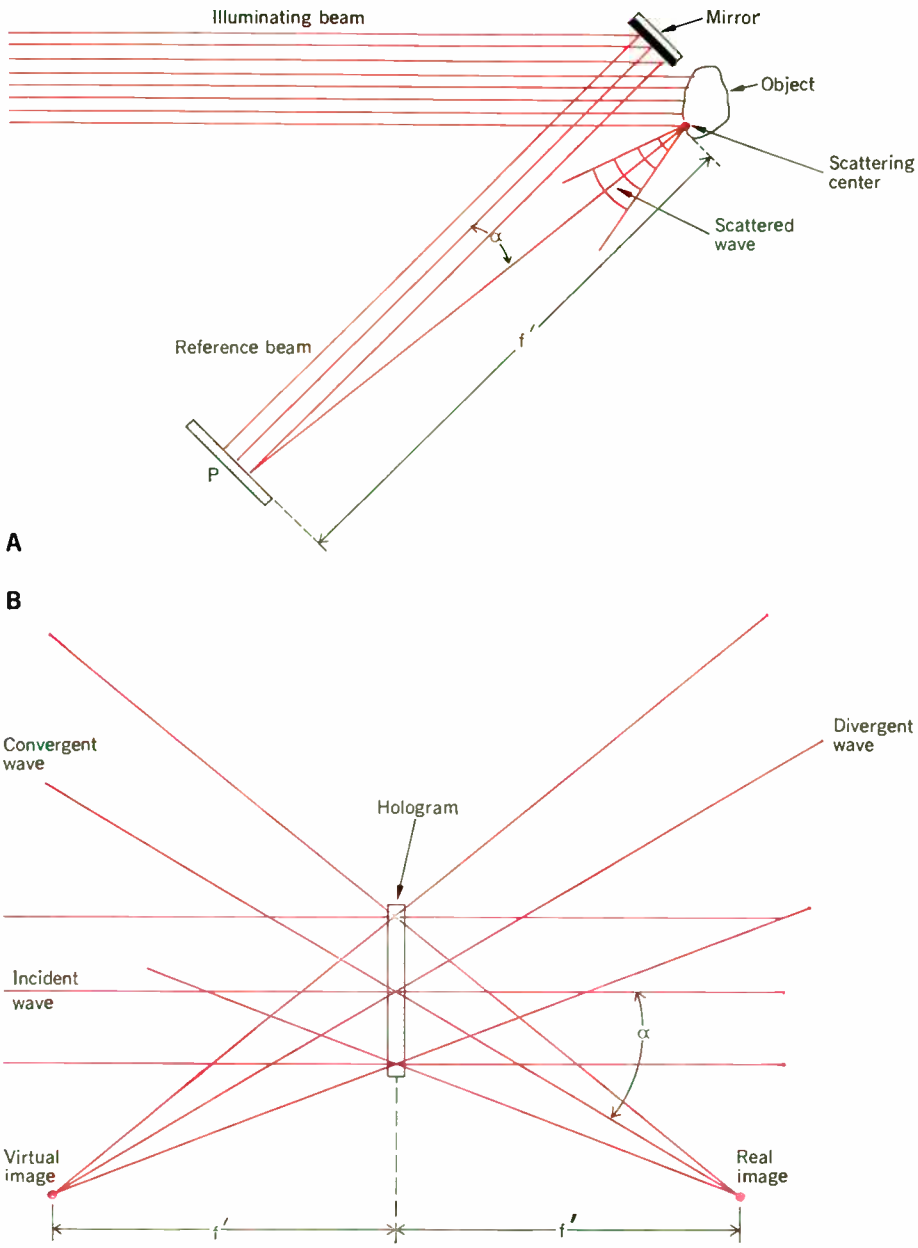



**FIGURE 11. Top diagram depicts conventional optical setup for producing a hologram. Hologram (bottom) can, in turn, reproduce both a virtual and real three-dimensional image.**

Each scattering object on the ground produces a one-dimensional Fresnel zone plate at the value of $y_1$ corresponding to the range of the object, and with a focal length proportional to this range. Reflectors, at the same range but with different along-track coordinates, appear as zone plates with centers displaced in the $x$-direction; their radar cross sections are manifested in the modulation intensity of the zone plate.

When this collection of one-dimensional holograms is illuminated by a coherent, collimated beam of light, a one-dimensional wavefront reconstruction takes place: The spatial pattern associated with each reflector generates a real and virtual image, each having an $x$-dimension focal length proportional to range. This is illustrated in Fig. 16. An image of the target field focused in the $x$-dimension only would be observed by appropriately examining one of these two planes. However, the target responses observed in either of these planes are not focused in the range dimension; furthermore, the tilt angle of the plane may make such observation inconvenient. A special viewer could serve to erect the tilted plane, and to bring the plane of $x$-focus and that of $y$-focus into coincidence. The optical system, Fig. 15, does exactly this: the conical lens erects the tilted plane of $x$-focus, and moves it to infinity; the sphere then moves the plane of $x$-focus to $P_2$; and the sphere–cylinder combination images the $y$-focused structure of plane $P_1$ at $P_2$. However, the one-dimensional holograms are of an off-axis reference beam variety (a consequence of using an offset frequency in the recording process), and the energy associated with the virtual image does not spatially overlap that associated with the bias term or with the real image. The desired image may, therefore, be viewed through a slit in $P_2$.

The data processing required for a rotating-target radar (or, for that matter, for any range-Doppler radar) is some-

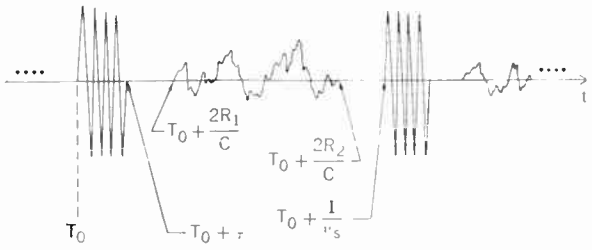



**FIGURE 12. Delay relationships exist between outgoing and incoming side-looking radar pulses. $R_1$ and $R_2$ represent the nearest and farthest slant range probed.**

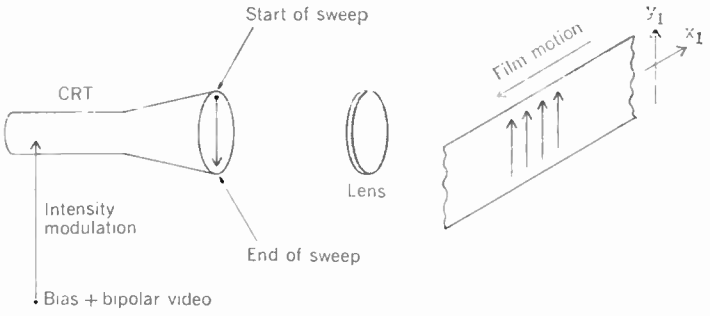



**FIGURE 13. Radar signals are sent through a biased CRT, and then recorded on film for subsequent optical processing.**

what simpler than just described. In such cases, a conventional image is obtained for range and a Fourier transform is formed for the orthogonal dimension. Optical compression of the range pulse may easily be incorporated in either the range-Doppler or the airborne terrain-imaging case. This is done by optically synthesizing the "matched filter" associated with the range-pulse waveform, and inserting this into the optical channel. The processor becomes more complicated (for both the range-Doppler case and the airborne terrain-imaging case) when, over the integration time used by Doppler analysis, the range to points in the target field changes by a distance comparable to or greater than the range resolution; i.e., when the "range-walk" problem is significant. These problems can be solved, but the solutions are beyond the scope of this article.

## Ambiguity constraint

Finally, a remark on the ambiguity problem associated with synthetic-aperture side-looking radar—and one that does put a constraint on its usefulness.

If the vertical cross section of the physical antenna pattern illuminates a slant-range interval of $\Delta R$ and if a periodic pulse train is transmitted, range ambiguities can be avoided, provided that the interpulse period satisfies the relation

$$T \geq \frac{2\Delta R}{c} \tag{22}$$

However, the Doppler bandwidth of the received signals imposes an upper bound on the interpulse period. The highest well-illuminated Doppler frequency is $(\beta/2)2v\lambda^{-1}$, and sampling theorem dictates a minimum sampling rate of twice this highest frequency:

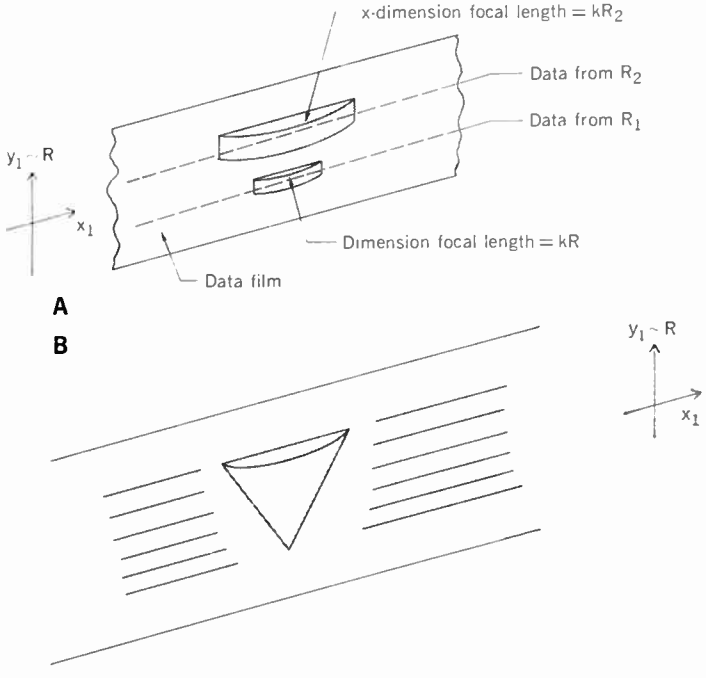
$$T^{-1} \geq \frac{2\beta v}{\lambda}$$

Equivalently, since $\beta \approx \lambda/D$, the minimum rate required for proper sampling of the Doppler frequencies is

$$T^{-1} \geq \frac{2v}{D} \tag{23}$$

The constraint of Eq. (23) indicates that at least one pulse must be transmitted whenever the aircraft moves a distance of $D/2$, this being approximately the azimuthal resolution limit of the system; i.e., as one might expect, the spacing of the samples in the along-trace dimension must be as fine as the achievable resolution.

**FIGURE 14. A quadratic phase delay, introduced by a plano-convex lens, (top) is used to focus abscissa information for a particular slant range R. To bring entire area between two slant ranges into focus, therefore, requires a conical lens (bottom).**




Various additional elementary conclusions can be drawn. With operation at the ambiguity limit [i.e., equality in both (22) and (23)], an image is acquired at a rate of $W$ (two-dimensional) resolution cells per second, where $W$ is the RF bandwidth of the radar in hertz. Also, let $\rho_x = D/2$ be the azimuthal resolution of the system; then (22) and (23) give

$$\Delta R = \frac{c}{2v}\rho_x \tag{24}$$

as the unambiguous obtainable swath width. Since $c$ is so great, (24) usually permits generous coverage for side-looking radar. However, if these same ideas are considered for sonar, $c$ becomes the speed of sound in water. In turn, the resulting slant-range interval of (24) is disturbingly small; several orders of magnitude in available slant-range interval are typically lost when one goes over to the sonar case.

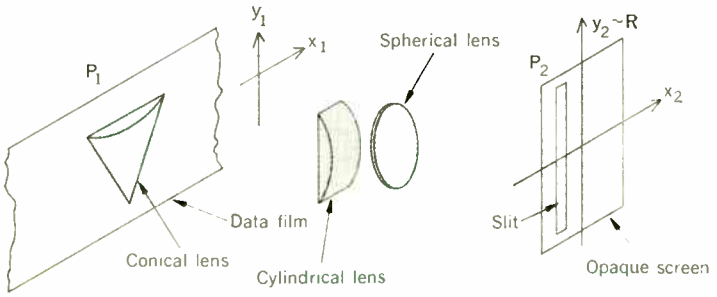
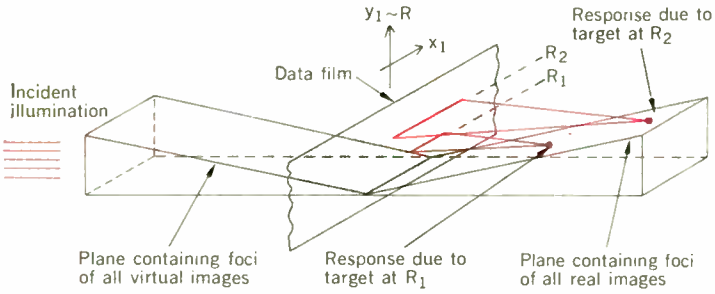



**FIGURE 15. With this optical arrangement, a view through the vertical slit at $P_2$ displays output of the synthetic aperture for all ranges with the same along-track dimension.**

**FIGURE 16. A one-dimensional wavefront reconstruction generates a real and virtual image of the slant-range reflections so that they fall in a plane.**





REFERENCES

1. Sherwin, C. W., Ruina, J. P., and Rawcliffe, R. D., "Some early developments in synthetic aperture radar systems," *IRE Trans. Military Electronics*, vol. MIL-6, pp. 111–115, Apr. 1962.

2. Shapiro, I. I., "Planetary radar astronomy," *IEEE Spectrum*, vol. 5, pp. 70–79, Mar. 1968.

3. Evans, J. V., and Hagfors, T., *Radar Astronomy*. New York: McGraw-Hill, 1968.

4. Cutrona, L. J., Vivian, W. E., Leith, E. N., and Hall, G. O., "A high-resolution radar combat-surveillance system," *IRE Trans. Military Electronics*, vol. MIL-5, pp. 127–131, Jan. 1961.

5. Cutrona, L. J., Leith, E. N., Palermo, C. J., and Porcello, L. J., "Optical data processing and filtering systems," *IRE Trans. Information Theory*, vol. IT-6, pp. 386–400, June 1960.

6. Cutrona, L. J., Leith, E. N., Porcello, L. J., and Vivian, W. E., "On the application of coherent optical processing techniques to synthetic-aperture radar," *Proc. IEEE*, vol. 54, pp. 1026–1032, July 1966.

7. Gabor, D., "A new microscopic principle," *Nature*, vol. 161, p. 777, May 1948.

BIBLIOGRAPHY

Brown, W. M., "Synthetic aperture radar," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-3, p. 217, 1967.

Brown, W. M., and Fredricks, R., "Range-Doppler imaging with motion through resolution cells," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-5, p. 236, Jan. 1969.

Brown, W. M., and Palermo, C. J., "Effects of phase errors on resoluton," *IEEE Trans. Military Electronics*, vol. MIL-9, pp. 4–9, Jan. 1965.

Brown, W. M., and Palermo, C. J., *Random Processes, Communications, and Radar*. New York: McGraw-Hill, 1969.

Cook, C. E., and Bernfeld, M., *Radar Signals: An Introduction to Theory and Application*. New York: Academic Press, 1967.

Cutrona, L. J., and Hall, G. O., "A comparison of techniques for achieving fine azimuth resolution," *IRE Trans. Military Electronics*, vol. MIL-6, pp. 119–121, Apr. 1962.

Develet, J. A., "Performance of a synthetic-aperture mapping radar system," *IEEE Trans. Aerospace and Navigational Electronics*, vol. ANE-11, pp. 173–179, Sept. 1964.

Greene, C. A., and Moller, R. T., "The effect of normally distributed random phase errors on synthetic array gain patterns," *IRE Trans. Military Electronics*, vol. MIL-6, pp. 130–139, Apr. 1962.

Harger, R. O., "An optimum design of ambiguity function, antenna pattern, and signal for side-looking radars," *IEEE Trans. Military Electronics*, vol. MIL-9, pp. 264–278, July/Oct. 1965.

Harger, R. O., and Crimmins, T. R., "The effect of phase errors on weighted spectra," *IEEE Trans. Military Electronics*, vol. MIL-9, pp. 298–299, July/Oct. 1965.

Heimiller, R. C., "Theory and evaluation of gain patterns of synthetic arrays," *IRE Trans. Military Electronics*, vol. MIL-6, pp. 122–129, Apr. 1962.

Klauder, J. R., Price, A. C., Darlington, S., and Albersheim, W. J., "The theory and design of chirp radars," *Bell. Sys. Tech. J.*, vol. 39, p. 745, 1960.

Kozma, A., and Kelly, D. L., "Spatial filtering for detection of signals submerged in noise," *Appl. Optics*, vol. 4, p. 387, Apr. 1965.

Leith, E. N., "Photographic film as an element of a coherent optical system," *SPSE*, vol. 6, p. 75, Mar.–Apr., 1962.

Leith, E. N., Upatnieks, J., and Haines, K. A., "Microscopy by wavefront reconstruction," *J. Opt. Soc. Am.*, vol. 55, p. 981, 1965.

Leith, E. N., and Upatnieks, J., "Recent advances in holography" *Progress in Optics*, vol. 6, E. Wolf, ed. Amsterdam: North Holland Pub. Co., 1967.

Leith, E. N., "Optical processing techniques for simultaneous pulse compression and beam sharpening," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-4, p. 879, 1968.

Leith, E. N., and Ingalls, A. L., "Synthetic antenna data processing by wavefront reconstruction," *Appl. Optics*, vol. 7, p. 539, 1968.

McCord, H. L., "The equivalence among three approaches to deriving synthetic array patterns and analyzing processing techniques," *IEEE Trans. Military Electronics*, vol. MIL-6, pp. 116–119, Apr. 1962.

Turin, G. L., "An introduction to matched filters," *IRE Trans. Information Theory*, vol. IT-6, pp. 311–329, June 1960.

Tyler, G. L., "The bistatic, continuous-wave radar method for the study of planetary surfaces," *J. Geophys. Res.*, vol. 71, p. 1559, 1966.

Vander Lugt, A., "Operational notation for the analysis and synthesis of optical data-processing systems," *Proc. IEEE*, vol. 54, pp. 1055–1063, Aug. 1966.


**William M. Brown** (SM) received the B.S.E.E. degree in 1952 from the University of West Virginia. He transferred to Johns Hopkins University and earned the M.S.E. (1955) and D. Eng. (1957) while working at the Air Arm Division of Westinghouse (1952–1955) and at the Johns Hopkins Radiation Laboratory (1954–1957). Except for last year—when he was a visiting professor at the Imperial College of London—and 1958—while with the Institute of Defense Analysis—he has been with the University of Michigan: head of the Radar and Optics Laboratory (1960–1968) and since an advisor there and now professor of electrical engineering. He's also vice president of the Chain Lakes Research Corporation.

**Leonard J. Porcello** (SM) came to the University of Michigan in 1955 after receiving the B.A. degree from Cornell University. He has since been affiliated with the school's Radar and Optics Laboratory of the University's Institute of Science and Technology. He received the M.S., M.S.E., and Ph.D. degrees in 1957, 1959, and 1963, respectively. He has headed the Radar and Optics Laboratory since 1968 and is currently also a member of the Electrical Engineering department.

# Human experience in artificial intelligence

***Thinking machines?***
***Some people are skeptics, others blind believers; a true appraisal lies somewhere in between. With man to guide the way, computers can display a form of rationale***

**Carl V. Page** *Michigan State University*

*When computers were yet in their infancy, some experts fancied them to be—much as they fancied humans were—structurable to develop intelligence on their own. It has since been learned that the mind is not developed with the spontaneity originally conjectured and that computers are not wont to "bootstrap" their way to higher levels of intelligence. The computer, if it is to adapt to new situations, must be given a "helping human hand". There are various ways of programming a computer to "acclimate": One is based on a logic of syntax; another uses semantics; still a third, dwelled upon here at length, is based on repeated human intervention and computer interplay.*

A preoccupation in early research into the field of artificial intelligence included attempts to construct systems that could somehow "bootstrap" their way from total ignorance to a creditable performance of some difficult task. Thus, researchers held conferences on such subjects as "self-organizing" systems and "learning systems." The philosophy behind some of this research was to provide a minimal initial structure to be modified as required by training. However, systems designed with this minimalist philosophy have been mediocre performers despite a constantly improving technology of implementation. The minimalist viewpoint is still sometimes expounded in the popular press, but artificial intelligence now has many distinctly human parts: Most artificial intelligence systems have substantial a priori human-supplied information.

Regardless of the theories researchers professed, there has always been some kind of "instinct" or unlearned information, relevant to the problem environment, built into successful artificial intelligence projects. At first, workers were apologetic, and buried the instinctive portions deep in their computer programs. One reason for this, undoubtedly, was the fear of being considered charlatans.

A possible origin of this fear dates back to the late 18th and early 19th centuries. A Hungarian named Wolfgang von Kempelen astonished large audiences with a small chess-playing machine known as the Maelzel Chess Automaton.[1] Many ostensible explanations of its operation were presented—one by Edgar Allan Poe—but the explanations were not adequate to explain the high quality of chess played by the machine. Of course, although the machine appeared too small to contain even a child, a small man, made smaller by amputations received while serving in a European war, was inside.

Early workers in artificial intelligence were clearly aware that building the "little man" directly into their programs, through storage of human responses to all environmental situations, was as undesirable as it was impractical. Yet attempts to cast out the "little man" entirely, as the minimalist philosophy would have them do, were not successful.

The minimalist attitude derived from the old view of the human nervous system as a randomly connected network that organized itself to carry out complex information processing in a surprisingly short time. But recent biological data suggest that maturation of the organization and structure of nervous systems continues for several

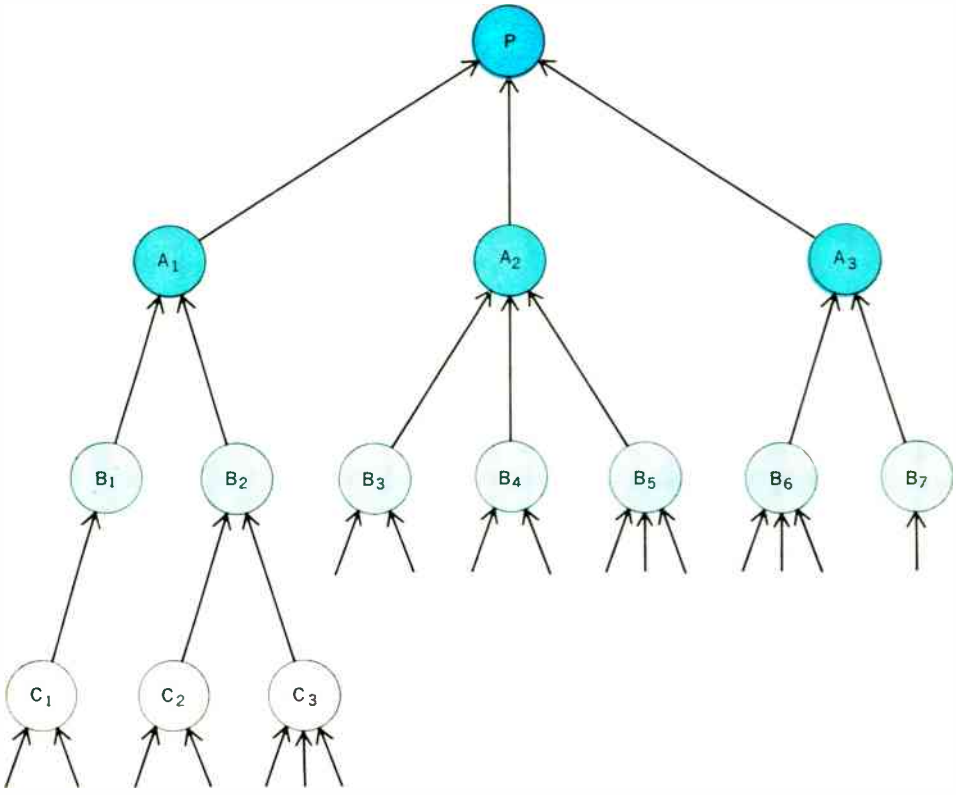



**FIGURE 1. There are usually many routes available to prove a proposition P. And, regressing backward, the routes multiply through each of the subaltern steps. To prove P, $A_1$ or $A_2$ or $A_3$ must be proved; and to prove any of the A's, two B's must be proved. The tree can therefore grow too large for even a computer to tackle.**

years after birth[2] so that natural nervous systems are not "self-organizing" in the minimalist sense, but possess a sequence of mechanisms—called *instinct* when they occur in their grosser manifestations.

Many kinds of functions are "built-in" animal response, and there is considerable preprocessing of sensory inputs[3] that greatly affects the data-manipulation ability of nervous systems. Consider, for example, the classical study of the frog's visual system done by Lettvin *et al.*[4] The frog eye is a beautifully evolved system. The optic nerve is optimally stimulated by insect-sized objects moving within range of the frog. What a great help to internal processing!

Consider a few classic artificial intelligence projects, looking for those traces of the "little man"—human experience in artificial intelligence.

## Analyzing the means from the end

There are a number of programs to find the "means" to obtain certain "ends," as, for instance, finding a proof (the means) for a theorem (the end) in propositional calculus.[5] Such systems also frequently work in reverse. Usually, in interesting situations, the "tree" of alternatives—as a practical matter, "uncomputably" large—is implicitly defined by the problem and methods. Figure 1 shows such a tree. Enumerative methods are of no value in such cases.

A program—the General Problem Solver (GPS)—for such a problem environment was developed by Newell and Simon.[6]

The GPS, like some people, is a "symbolizer" in its approach to problem solving. That is, GPS manipulates the symbols, according to formal connection rules or syntax, independently of the feasible meanings of interpretations possibly attached to them. The only information GPS has about the effectiveness of its methods of proof is in the ordering of the syntactic differences provided by humans, or, in principle, by self-adaption, perhaps in a manner based on one of Samuel's programs (discussed later). Nevertheless, man still has a critical part in adjusting the adaptive mechanism.

Interestingly, GPS-like programs select their own subgoals when attempting a proof. This is both more efficient and more esthetic than being "man-picked." However, the subgoal selection mechanism in part depends on an ordering of the differences that GPS detects between the node connections (strings) shown in Fig. 1. For instance, GPS won't attempt to prove a subgoal if it is more difficult than the goal containing it. Thus, if the problem were contained in a symbolic logic environment, GPS would not, as a subgoal, try to change the main connective of a string of symbols in order to change the grouping of the string, because changing connectives would be harder than changing the grouping. Of course, the ordering of difficulty of subgoals is supplied a priori by researchers—in effect, being the instinct of GPS concerning the environment of problems and methods.

Heuristics, or rules of thumb for discovery that often work but are not guaranteed to do so, also appear in GPS. The unreliability of heuristics together with their stated intent to copy human behavior make them obvious

to consider in a search for traces of human experience. Nevertheless, although humans may use certain heuristics to solve problems, their performance is usually much better than attributable to the mechanistic application of the heuristics they espouse. Hence, although heuristic methods are often unabashed attempts to embed human xperience into artificial intelligence, they are frequently unsuccessful. And, if they are not the way people really do it, they are not a human part of artificial intelligence.

### A semantic approach

In contrast to GPS, a geometry theorem-proving pro-ram by H. Gelernter *et al.*[7,8] works at the semantic level as well as syntactic level. The result is an ability to prove geometry theorems about as well as a good high school sophomore can. The semantic "understanding" of theorems in geometry is provided to the program by means of analytic geometry subroutines used to "examine" a diagram included along with the statement of the theorem to e proved. Like GPS, the geometry machine works backwards from a theorem, using axioms and previously proved theorems, together with rules of inference, to generate various proofs for the theorem. The number of proof pathway possibilities is very large, resulting in a tree of alternatives with about 1000 branches at each node. To find an eight-step proof randomly requires searching on the average about $^1/_2 \times 10^{24}$ branches. By using semantics and checking an intermediate result such as, say, "line *AB* equals line *BC*" in the diagram, obviously false subgoals can be screened out. If the length of line *AB* and the length of line *BC* are equal within the precision of calculation, the program may try to prove them equal. If the opposite case holds and the lines are clearly unequal in the diagram, this portion of the tree is pruned, or more realistically, "nipped in the bud."

This semantic approach reduces the effective number of alternatives at each node from 1000 to just five. This means that a trial-and-error approach on the remaining branches requires only about $2 \times 10^6$ trials to find an eight-step proof. Of course, exponential law is still valid, and long proofs are impossible to find by methods of exhaustion. Consequently, other types of human experience, called "specific geometry heuristics," are put in the program.

The specific geometry heuristics include an *ad hoc* function for the distance between a subgoal to be proved and the axioms. This distance is used to help select the next subgoal from the remaining paths—clearly the same type f human experience stored for GPS in its difficulty-of-difference list. The importance of such a syntactic tool was demonstrated in Gelernter's experiments: Specific geometry heuristics with their built-in human knowledge of the problem environment produced a fivefold speedup in the program.

The geometry program exposes another kind of human ssistance also present in GPS but hidden by its symbolic nature. In particular, for both programs, relevant features are extracted from the environment, without also extracting extraneous signals (noise), for insertion into the program. That is, an attempt is made to feed the computer as much preprocessed, germane information as possible. low much? Well, how much geometry does a human ave to learn before realizing the importance of triangles? The geometry program, in effect, should have preprocessing every bit as good as the preprocessing that the frog's eye provides the frog's brain. In GPS, this preprocessing service is trivial because the problem environment is expressed by symbols. But in the geometry program, the human programmer initially supplies lists of such relevant features as lines, angles, and triangles.

The geometry program, being a "visualizer," has an important advantage over a symbolizer like GPS. When all else fails—if the preprocessed information is insufficient—the geometry program can define a line to exist between two points, thereby creating new angles and triangles to work with. This construction, in one fell swoop, changes the description of the environment and introduces new syntactic objects. Manipulation of the new syntactic objects may provide a proof where none was available before. Of course, there is always the risk that the program may construct too many new syntactic objects and be overwhelmed. This could also happen if the program were given a needlessly complicated diagram. Hence the presentation of a fairly good diagram to the machine by the human, as embodied in the geometry program, is an important part of artificial intelligence. Setting up a machine for a simple diagram or a simple semantic interpretation describing a theorem, it seems to me, is an interesting problem in its own right.

### Game-playing research

Games are convenient vehicles for study of artificial intelligence. They possess clearly specified rules as contrasted with the complicated problems of everyday life. More important, the simple rules of a game lead to a large amount of complexity for a minimum investment in programming. There is also a background of published information concerning major games that makes it easier to evaluate the artificial intelligence system. Finally, there is an opportunity for a person who doubts that any learning has taken place to match his intelligence against the intelligence of the machine. Indeed, a very vocal critic of artificial intelligence research was recently silenced after being beaten at chess by the Greenblatt program.[9]

Samuel's programs[10,11] now play challenging games against master checker players although their performance is not superior to humans—yet. One important aspect of his research has been Samuel's persistent interaction with his programs. During the life of the Samuel project—extending more than 12 years over several machines—Samuel stood by, ready to intervene if the program played a bad series of games. When necessary he made small changes in certain learning rules until the immediate difficulty was solved. This type of dialogue between Samuel and his program is now typical of some modern interactive pattern recognition programs.[12,13]

What is the nature of the dialogue between Samuel and the program? It certainly cannot be expert checker-play advice since Samuel was beaten by a very early version of his own program. The learning techniques used by Samuel were originally developed on faulty machines (free time being obtained on new machines being tested in the factory). However, the learning techniques apparently converged despite an occasional error, much like certain algorithms for differential equation solutions that converge despite small occasional errors.

### Bestowing assistance

The most obvious aid provided by Samuel was an answer to the problem of subgoal selection. The goal of ulti-

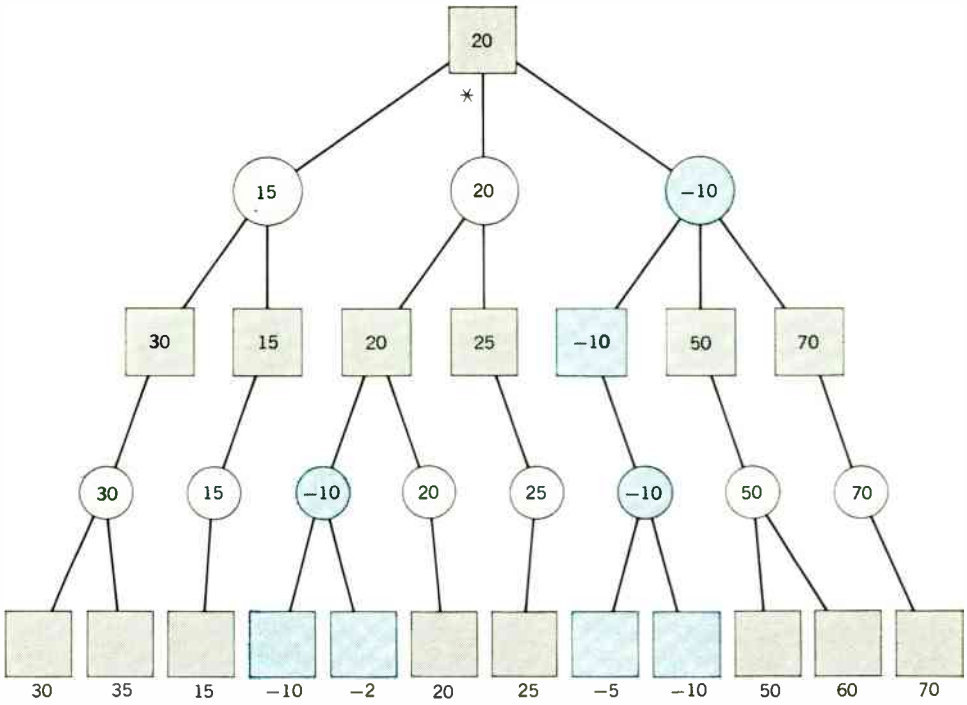



**FIGURE 2. A computer's life may be just a game of checkers. Here is a simplified model of a segment of a computer program designed to make the computer victorious over its human opponent. The computer's alternatives are indicated by boxes and the man's by circles. Enclosed numbers are subscores. The computer looks ahead and chooses the best route for the highest ultimate score—assuming that its opponent also knows the tree of scores.**

mately winning the game has the subgoal of making the best possible move each time. But, the best move depends on what the opponent will do, and so on. If all moves and their replies were considered, a tree of all possible board moves, similar to the alternatives or trial-and-error tree of GPS and the geometry program, would be obtained. The initial board tree has about $10^{40}$ possible terminal paths, some resulting in wins and some in losses. Of course, this tree is too large to enumerate.

Samuel's program analyzes or computes different lines of play to different depths. The end to an analysis of play is controlled both by the amount of storage available and the board configuration. For instance, the program attempts to follow chains of jumps to their conclusion. Given the look-ahead tree, the positional score-function is computed for each terminal node. Assuming that a high score is good for the program and a low score is good for its opponent, and that both do the best they can with respect to the score, the best move for the program, based on the information available at the nodes of the tree, can be computed. The process by which this is done is the minimax process of game theory. (For those not familiar with the minimax concept, examine Fig. 2: A move is determined by the terminal nodes' scores of each segment of the move-tree available from the look-ahead process.)

Currently, a computational device called alpha–beta technique—that also appears in dynamic programming—provides an efficient method for computing the minimax score without evaluating all the terminal nodes. (That is, each player does *not* look at continuations of moves that are worse for him than some other move he has already evaluated.) The alpha–beta technique then uses the evaluation to govern the look-ahead tree to some extent. In order for it to be efficiently employed, the technique must "look" at the strongest moves first. Samuel now uses a "plausibility" analysis prior to look ahead that approximately ranks the moves—making the alpha–beta technique effective.

In the best case, when the moves are ranked correctly, the alpha–beta technique allows *doubling* of the depth of look-ahead with no increase in computing time. So, the process automatically gives an estimation of responses from the environment via the board evaluation function provided by the human.

Unlike the geometry program or GPS, the checker program does not have complete control of the path to be negotiated through the tree because the opponent picks every other branch when he makes his move, and some method must be used to estimate the opponent's replies to various moves. Samuel decided to use a linear numerical function of certain features of the board configuration as a kind of "goodness of position" score. If the linear function were accurate, then the best move for each side could be computed by a look-ahead process.

The linear-positional score-function has terms that are functions of certain features of the board, including some rather obviously important properties of a checker position such as the piece advantage, mobility, etc. In addition, there are terms like second moment of the pieces around the diagonal through the double corners (dispersion), and some binary-valued terms that are defined to be 1 over certain regions representing intersections of inequalities with other numerical features. Samuel com-

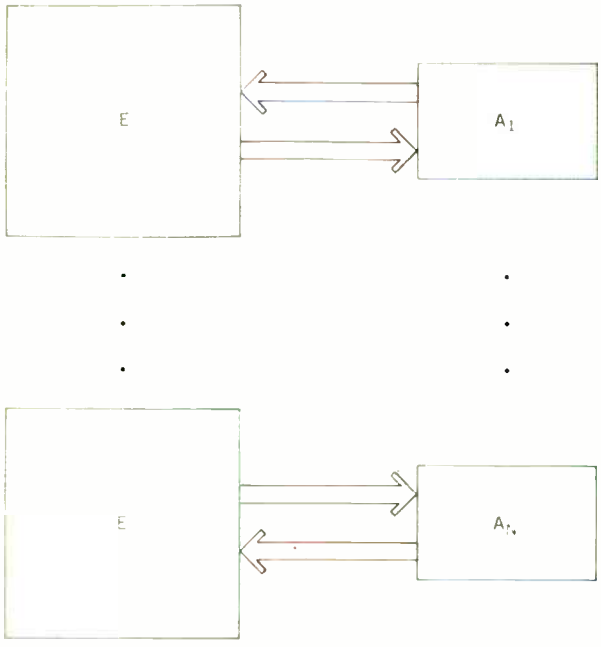



**FIGURE 3. Schematic representation of a machine–environment interaction.**

bined as many as 16 such parameters of board position with linear weights to form the linear-positional evaluation function.

All features of the function are supplied by humans, but there is one great difference between how this is done by Samuel and how it is done in GPS and the geometry program. Indeed, this is the subject of much of the interactive-like dialogue between Samuel and his program—an adaptive process monitored by Samuel to determine features and their numerical weights. The 16 features are used at any time, selected from a set of, say, 40, with the rest kept in reserve. The coefficients of the polynomial function are adjusted to how well the features work.

Samuel controls an adaptive process somewhat akin to hill climbing—used to find a good set of features and weights. Hence, there is here a different and, I believe, general kind of human part of artificial intelligence, the regulation of an adaptive process by a human. Modifying J. Holland slightly,[14] *an adaptive system* consists of a "family" of four interrelated items:

$$A, \epsilon, T, \chi$$

where

- $A$ = set of possible devices (sometimes a program with different parameters)
- $\epsilon$ = set of admissible environments (outputs of $E$ in $\epsilon$ are called "payoffs")
- $E$ = environment
- $T$ = adaptive strategy—a method of specifying a new family of devices from $A$ to be tested against $E$ using the results of test of the last family
- $\chi$ = fitness criterion—maximization of accumulated payoff, escape from extinction, etc.

Imagine that the environment can act on all members of the family at once, separately.

Figure 3 describes a family of machines interacting in parallel with different copies of the environment.

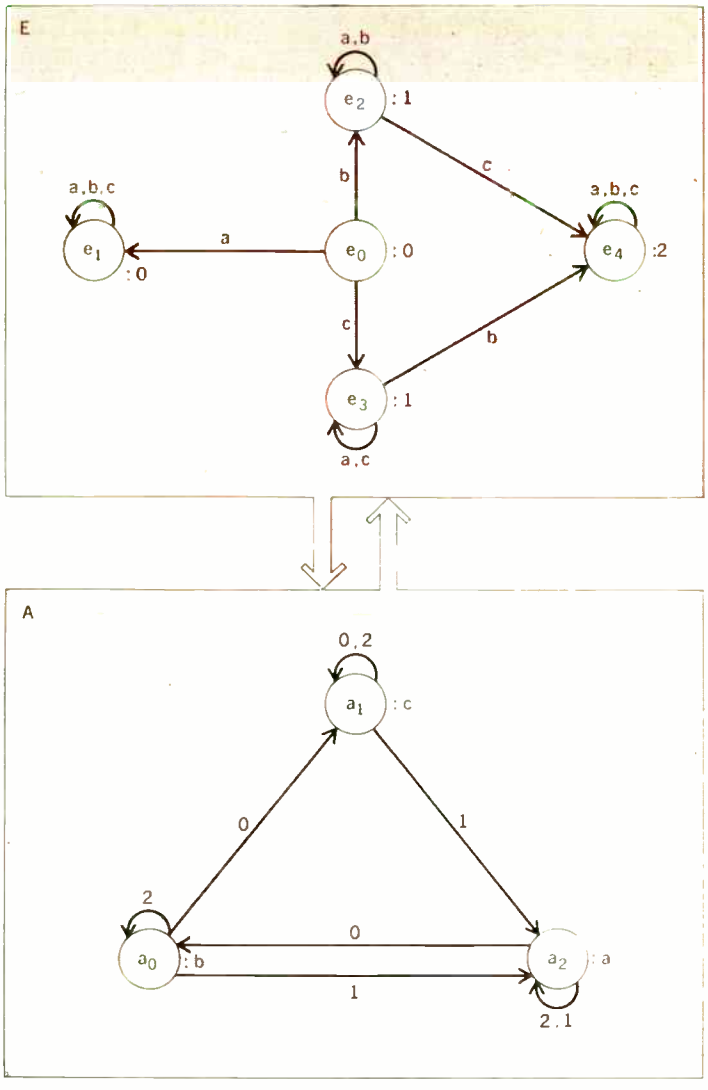
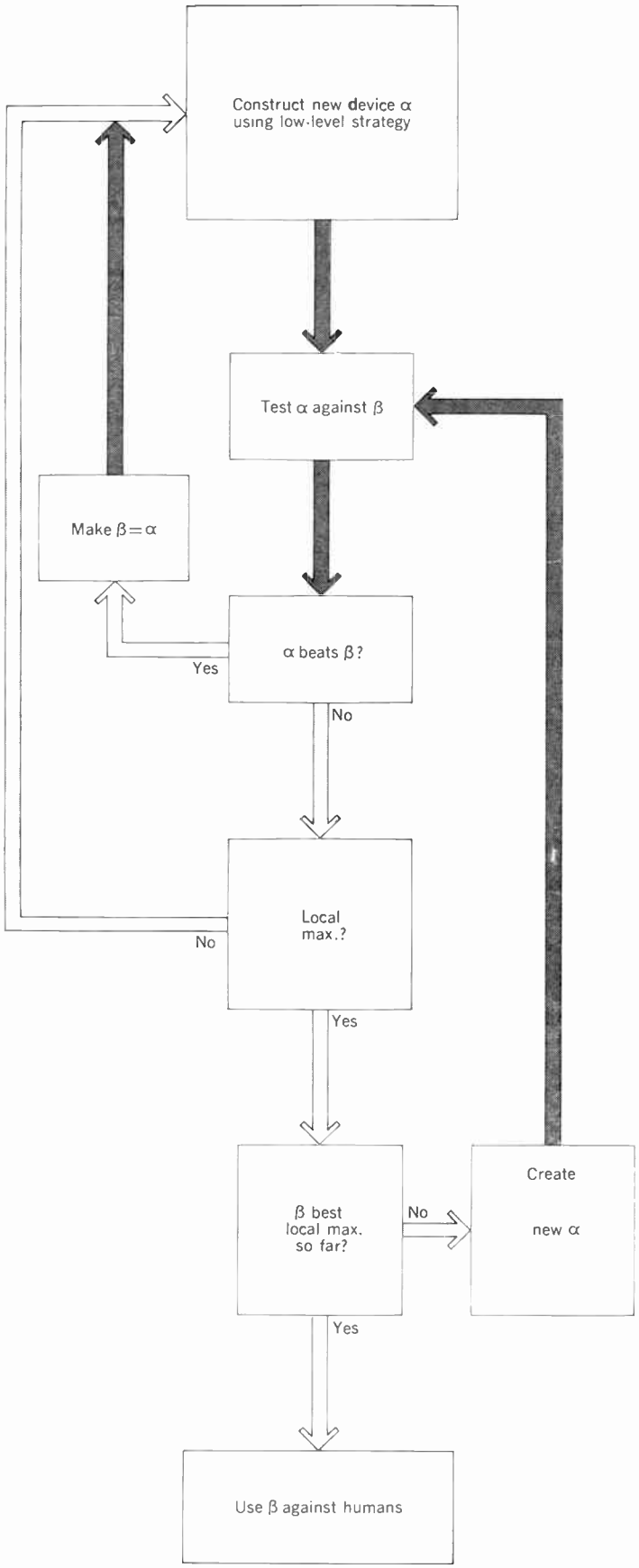



**FIGURE 4. As with animal species, new generations of computers can be developed along a "survival-of-the-fittest" doctrine. Nomenclature prefaced by a colon are outputs; those not prefaced are inputs. Inputs alter existing states as indicated by the transition lines (sometimes curved). Thus, an input of "a" or "b" to environmental state "$e_2$" will leave "$e_2$" unchanged, whereas input "c" will switch state "$e_2$" to "$e_4$." Obviously, since inputs "a," "b," or "c" leave states "$e_1$" and "$e_4$" unchanged, the states are dead-ended.**

Figure 4 represents an $N = 3$ case where three machines play against the environment. All three machines and the environment are finite-state devices. All machines happen to have the same transition function and output, differing only in initial state. The adaptive strategy is to sample the environment on the next generation with a distribution over machines proportional to the expected average payoff of the machines of that type in the last generation.

At the start of $A$–$E$ interplay, Fig. 4, $E$ is always taken to be the initial state $e_0$. Each machine sends inputs to the other simultaneously. Call $A$ with initial state $a_0$ machine $A_0$; it sends $b$ to $E$ at the same time that $E$ sends a 0 to $A_0$. This causes $E$ to be switched to $e_2$ while $A_0$ goes to $a_1$. Being in state $e_2$ causes $E$ to send $a_1$ to $A_0$ while being in $a_1$ causes $A_0$ to send $c$ to $E$. Finally, $A_0$ ends up in the $a_2$

**FIGURE 5. To make the computer's human checker adversary play honestly (the human can fool the computer by early bad play) a program has been devised whereby the computer also plays against itself. This is called the alpha–beta strategy, as diagrammed.**




mode and $E$ is in $e_4$. At this point $A_0$ remains in state $a_2$ sending $a$'s to $E$; and $E$ remains in state $e_4$ sending 2's to $A$. Since the outputs of $E$ are the payoffs of the adaptive system, it is clear that $A_0$ obtains as much payoff per input as any machine possibly can. But examine the behavior of machines $A_1$ ($A$ with initial state $a_1$) and $A_2$ ($A$ with initial state $a_2$) in order to see how the adaptive strategy converges on $A_0$. $A_2$ starts off on the wrong foot by sending an $a$ to $E$ and drives it to the "dead-end" state $e_1$, that has output and hence payoff of 0. Consequently, $A_2$ receives the lowest payoff possible from the environment. Finally, consider $A_1$, which is between these two extremes. Starting in $a_1$, $A_1$ drives $E$ from $e_0$ to $e_3$ while being kept in $a_1$ by the 0 sent by $E$. Next $E$ sends 1 forcing $A_1$ to a $a_2$ while $A_1$ keeps $E$ in $e_3$ with $c$. At last a stable situation occurs in which $A_1$ in state $a_2$ sends $a$'s to $E$ keeping it in $e_3$ and causing 1's to be send to $A_1$, locking it in $a_2$. Therefore, the payoff obtained by $A_1$ is about one per input step.

The simple reproductive adaptive strategy used here introduces a new generation of machines whose composition is based on the payoff success of machines from the previous generation. Since $A_0$ averaged about two units of payoff per input, the next generation contains two copies of $A_0$. Likewise, $A_1$ received about one unit of payoff so one copy of $A_1$ is contained in the next generation. $A_2$ becomes extinct since it obtained 0 units of payoff; i.e., no copies of $A_2$ are included in the next generation to be tested against the environment. After ten generations, there will be 1024 copies of $A_0$ and just one copy of $A_1$. So while $A_1$ does not become extinct, it does get "birthed out" by the superior machine $A_0$. In this sense, the adaptive system converges on the best of all machines for dealing with $E$.

Each new generation of machines should be more like the ideal than the last as long as the environment is fixed. A truly successful adaptive strategy must, perforce, do well over a range of environments. Thus, a more realistic and self-serving adaptive strategy might introduce a mutation of $A_0$ or $A_1$ from time to time to increase the variety of the population as a hedge against a change in $E$.

In the Samuel case, each linear positional evaluation function, together with the checker program, corresponds to a device. (A special-purpose sequential machine could be built to simulate it.) The environment $\epsilon$ consists of all possible checker strategies; during a game the machine is confronted by a particular strategy. The adaptive strategy $T$ only creates a family of one device corresponding to the linear-positional evaluation function. The evaluation function created is based on results from the last evaluation function tested.

## Alpha vs. beta

An adaptive strategy that generates a better evaluation function can be viewed as two adaptive strategies in one. A low-level strategy, alpha, satisfying its own fitness criterion, is used by a higher-level strategy, beta, to get trial evaluation functions to test against a higher-level fitness criterion.

The low-level system, alpha, exploits the fact that the greater the projection, the less accurate need be the evaluation function—as long as it contains the piece-advantage term. Were it possible to look to the end of the game, knowing the piece-advantage term would be enough to determine the winner.

The idea of the strategy is to force the positional-evaluation score of the existing board to be the same as the score of the final board actually arrived at in play. A suitable fitness function $\chi$ for the lower-level system is implicitly defined by comparing the evaluation score two board moves (ply) earlier with the minimax score of the look-ahead tree beginning at the existing position. Weights of the feature functions are adjusted by the adaptive strategy to reduce any difference between the comparative quantities. Features consistently "out of step" with this scheme get "black marks" from the adaptive strategy and are eventually dropped to the end of the reserve features list. Any feature, other than piece advantage, may be dropped and picked up again several times, but the best features begin to stabilize and acquire stable coefficients.

Incidentally, the lower-level strategy was independently tried for a short time in actual play against human opponents, but instabilities soon forced its retirement except as used by the high-level strategy. The low-level strategy can be fooled by bad initial play on the part of the computer's human partner. Remember that the fitness function required the program to try to predict the score two ply away, including a move by the opponent. If the opponent consciously picks bad moves early in the game, the adaptive strategy will give black marks to good terms and they will soon be dropped; a bad evaluation function evolves that plays almost randomly.

To avoid this situation, Samuel embedded the low-level system in the high-level system so that it would have a different environment than the uncooperative human player. By playing alpha against beta a certain number of times, if alpha beats beta a majority of times, then alpha appears to be superior to beta. Figure 5 indicates how the lower-level strategy is embedded in a higher-level strategy.

Of course, alpha might not be better than beta in beating human opponents.

Therefore, a higher-level adaptive strategy, a kind of self-improvement, was instituted by Samuel using this ordering on evaluation functions. If alpha beat beta a majority of times, then beta was given alpha's evaluation function and alpha was required to do better! Hence, there is a kind of hill climbing through the space of evaluation functions to find a superior evaluation function. It is not true hill climbing because "beating" is not necessarily a transitive relation (i.e., $a > b, b > c \therefore a > c$). However, if it is transitive, with a probability greater than ½, it should drift toward a local maximum.

But what guarantee is there that an evaluation function occupying a local maximum arrived at in this manner will cause the program to play checkers better against humans than, say, one of its near neighbors? Samuel kept the process honest by testing various evaluation functions generated in this way against book games and puzzles to be sure that the evolution was not proceeding in the wrong direction. Hence Samuel ranked evaluation functions in the environment of human checker players using a rather subjective fitness function $\chi$. In addition, the program was tested against human players of known ability and the performance noted. When the low-level process had clearly gotten stuck on a local maximum, Samuel made some arbitrary, severe change in alpha's evaluation function, in effect adjusting the adaptive strategy.

Returning to our search for distinct human parts of artificial intelligence, note that both the adaptive and the fitness function were supplied by the human. Both were skillfully adjusted by Samuel in what could have been an interactive dialogue. Regulation of an adaptive system seems to be a very general kind of human aid for an artificial intelligence system.

## Other adaptive mechanisms

Another example of the evolutionary approach to artificial intelligence is found in the recent book by Fogel, Owens, and Walsh.[15] Here, at first glance, appears to be a retreat to the minimal a priori information philosophy. A finite-state machine is evolved to solve a particular task—typically, prediction of the next symbol of a sequence given the previous elements of the sequence. A set of inputs is given, the fitness function $\chi$ is specified, and various adaptive strategies are applied to the initial family of machines—i.e., transitions, initial states, and outputs are changed; states are added and deleted. Sometimes three machines are "mated" by running them simultaneously with majority rule determining the outputs. The frequency with which states are added and deleted is critical to the adaptive strategy.

For example, generate a sequence of 1's and 0's in the following way: the $i$th bit is 1 if the $i$th natural number is prime. There is no finite-state machine that generates or recognizes this sequence. Yet after a few hundred bits are processed, it is possible for a machine to evolve that recognizes that multiples of 2, 3, and 5 are not prime. This is rather impressive even considering how much human direction and interpretation is going on, especially in deciding when to stop the evolution in order to unbiasedly evaluate a program's merits. For instance, a person might ascribe unwarranted perspicacity to a very young child by watching what he draws and, at the right time, stopping him. Thus, by preventing the child from adding a bushy coat of fur and four legs, he might obtain what appears to be a picture of a spaceship. More seriously, the problem of experimenting with adaptive processes without prejudicing the results seems hard to solve.

There has been human-provided feature extraction in all the projects discussed so far. Uhr and Vossler[16] attempted to make feature extraction automatic to the extent that it was controlled by an adaptive process. But again, this merely replaces one kind of human skill—the selection of important features—with another—the specification of an adaptive system that evolves them.

Let me point out that human direction of such evolutionary processes seems to be essential for their success in any reasonable amount of time. In realistic problems of modeling it is not straightforward to decide what the inputs to the machine should be. Much harder is the question of what is a good adaptive strategy for a particular problem. The fitness function $\chi$ is, of course, critical to what evolves. Indeed, anything that evolves is implicitly defined by the information from the environment and from the human insight that specifies the adaptive system.

## Future trends

By this time I hope you are convinced that there are several types of human parts of artificial intelligence. But what direction will the future take? The answer lies in extrapolating upon the robot projects at Stanford, Stanford Research Institute, and at M.I.T.[17] Early workers built special machines to test artificial intelligence

theories but simulation on general-purpose digital computers soon proved to be more economical. However, the reduction in size and cost of digital equipment together with the increasing cost of computer software is reversing the trend.

Simulation of the environment has been the most difficult and unsatisfactory part of research.[18] Hence, rather than trying to simulate the real-world environment, the well-funded projects use specially built devices to interact with it. Progress has been made in the "hand–eye" projects combining a television camera and range-finding mechanism with a mechanical hand to do such tasks as stacking blocks. At Stanford Research Institute, a very strong mechanical hand is encased in a wire cage to prevent it from menacing programmers and innocent bystanders when there are bugs in its programs.

What will be the practical applications of such systems? Many, but an obvious one will be the exploration of outer space.

Let me discuss a problem brought to my attention by Bill Kilmer from Michigan State. Professor Kilmer is involved in a research project with Sutro of M.I.T. to build a robot to explore Mars.[19] Interesting constraints are placed on the human aid that can be given the robot because it takes at least three minutes to transmit information between the Earth and Mars. The robot needs to incorporate as much human experience as possible to survive and carry out its mission. Suppose the robot falls into a hole. Can you imagine trying to direct —from Earth—its climb out, taking into account the 6-minute-long time delay between command and verification of response? The robot must be able to make on-the-spot decisions as to how to commit its resources when something interrupts its mission.

Kilmer and Sutro are taking a bionics approach to the problem. The reticular formation found near the base of the brain integrates input data to commit an animal to various modes of behavior such as fight, flight, sleep, etc. Kilmer and Sutro propose to build an artificial reticular formation in the Mars robot to handle interruptions that can't wait for help via the 6-minute feedback loop to earth.

While on the subject, I might speculate on other kinds of human aid a robot, built to explore Mars, might find useful. Certainly it will need on-line training as to what is interesting to its cameras. It will need adaptive features just as many of our earthbound computers have. Only a human can decide what are interesting features of the Mars landscape, and he can't decide beforehand; a robot on a mission to dig for water might discard shovelsful of nuts and bolts as uninteresting—not bothering to transmit their presence back to earth. So, there is a premium on adaptability of behavior.

Behavior can be adjusted as done by Samuel, but from long range. The Mars' probe will need to construct a symbolic map of its environment. The real environment provides us with Gelernter's semantic interpretation. The robot must be able to generate its own subgoals to follow long-term missions—perhaps in the manner of GPS and Gelernter—using human-stored measures of difficulty of certain potential actions.

Many important research projects[20] have been ignored in this survey for lack of space, but their inclusion would have only added support for the thesis that human experience needs no apology when it is built into an artificial intelligence system. Good research in artificial intelligence is not to be rated in terms of how little human instinct it possesses but by how intelligent it is.

REFERENCES


1. Shannon, C. E., "A chess playing machine," *The World of Mathematics*, vol. 4, Newman, J. R., ed. New York: Simon and Schuster, 1956, pp. 21–25.

2. Lectures by Leon Harmon.

3. Manning, A., *An Introduction to Animal Behavior*. Reading, Mass.: Addison Wesley, 1967.

4. Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, H. W., "What the frog's eye tells the frog's brain," *Proc. IRE*, vol. 47, pp. 1940–1951, Nov. 1959.

5. Newell, A., Shaw, J. C., and Simon, H. A., "Empirical explorations with the logic theory machine: A case study in heuristics," *Computers and Thought*, Feigenbaum and Feldman, eds. New York: McGraw-Hill, 1963, pp. 109–133.

6. Newell, A., and Simon, H. A., "GPS, a program that simulates human thought," *Computers and Thought*, Feigenbaum and Feldman, eds. New York: McGraw-Hill, 1963, pp. 279–293.

7. Gelernter, H., "Realization of a geometry-theorem proving machine," *Computers and Thought*, Feigenbaum and Feldman, eds. New York: McGraw-Hill, 1963, pp. 134–152.

8. Gelernter, H., Hansen, J. R., and Loveland, D. W., "Empirical explorations of the geometry-theorem proving machine," *Computers and Thought*, Feigenbaum and Feldman, eds. New York: McGraw-Hill, 1963, pp. 153–163.

9. Greenblatt, G., Eastlake, D., and Crocker, S., "The Greenblatt chess program," *Proc. Fall Joint Computer Conf.*, 1967.

10. Samuel, A. L., "Some studies in machine learning using the game of checkers," *IBM J.*, 1959; also *Computers and Thought*. New York: McGraw-Hill, 1963, pp. 71–105.

11. Samuel, A. L., "Some studies in machine learning using the game of checkers—II—Recent progress," *IBM J. Res. Dev.*, vol. 6, pp. 601–617, Nov. 1967.

12. Sammon, J. W., "On-line pattern analysis and recognition system (OLPAS)," Rome Air Development Center Technical Report NO RAOC—TR–68–283, Aug. 1968.

13. "SARF signature analysis research facility," Report S68–11A, AC Electronics, General Motors Corp., Goleta, Calif., Oct. 1968.

14. Holland, J., *A Theory of Adaptive Systems* (in preparation).

15. Fogel, L., Owens, A., and Walsh, M., *Artificial Intelligence Through Simulated Evolution*. New York: Wiley, 1966.

16. Uhr, L., and Vossler, C., "A pattern-recognition program that generates, evaluates, and adjusts its own operators," *Computers and Thought*. New York: McGraw-Hill, 1963, pp. 251–268.

17. Feigenbaum, E. A., "Artificial intelligence themes in the second decade," Stanford Artificial Intelligence Project, Memo No. 67, Aug. 13, 1968.

18. Nilsson, N. J., and Raphael, B., "Preliminary design of an intelligent robot," *Computer and Information Sciences—II*, J. Tou, ed. New York: Academic Press, 1967, pp. 235–259.

19. Kilmer, W. L., and Sutro, L. L., "Assembly of computers to command and control a robot," *Proc. Spring Joint Computer Conf.*, 1969.

20. Minsky, M., *Semantic Information Processing*. Cambridge, Mass.: M.I.T. Press, 1968.



**Carl V. Page** (M) earned all three science degrees—B.S., engineering sciences; and M.S. and Ph.D., communication sciences—from the University of Michigan. While there, he also was a member of the Logic of Computers Group and lectured at the Dearborn campus. After receiving the Ph.D in 1965, he joined the University of North Carolina as a member of the Computer and Information Science Department. Dr. Page, having since returned to more familiar territory, is now an assistant professor of computer sciences at Michigan State University. His professional associations include membership in the Association for Computing Machinery, the Pattern Recognition Society, Tau Beta Pi, and Sigma Xi.


# The CRT in phototypesetting systems

*The printed word can be electronically painted on a cathode-ray tube for rapid and versatile photo-offset reproduction if certain CRT operating criteria are met*

*R. J. Klensch, E. D. Simshauser* R C A Corporation

*Rapidity and flexibility in photo-offset printing are synonymous with computer-driven CRTs. Yet, designing such a system has its problems. One big consideration is beam deflection linearity. Another is resolution. Reproduction accuracy resolved, design alternatives remain: Should the CRT be step-scanned or scanned continuously (as for television)? How much of the tube should be used? What should the absolute motion of the film be with respect to the generated motion of the characters? Naturally, there are tradeoffs.*

The first recorded invention of a phototypesetter occurred in France at the turn of the century. This optomechanical device was a major breakthrough in the printing industry since phototypesetters can handle a greater variety of type fonts and operate at higher speeds than metal typecasting machines.

Metal-casting machines, because of their bulk and the difficulty of moving relatively heavy casting forms (materials), can only accommodate a maximum of eight different type fonts of a limited size range. In contrast, recent phototypesetters can store up to 24 type fonts in master photographic matrices on glass plates or drums. Up to eight different type sizes are available from each matrix by optical enlargement. Therefore, a typical modern phototypesetting machine theoretically has 192 different type fonts available on short recall. In one such machine, where each font consists of 88 characters, up to 16 896 characters are thus available, representing a significant advantage over metal linecasting machines usually having only 480 characters available in four styles within a narrow size range.

Besides their great versatility in type storage, modern phototypesetting machines have established a breakthrough in composition speed. A typical phototypesetting machine can generate 25 to 30 mixed characters a second, which is approximately ten times the speed of modern, tape-driven, metal linecasting machines.

Cathode-ray-tube (CRT) phototypesetters represent the latest advance in phototypesetting systems, increasing performance capabilities by many orders of magnitude. At least one CRT phototypesetter on the market today is capable of practically unlimited font storage, and each font can be recalled in much faster time than the recall time in a conventional optical phototypesetting machine. Furthermore, it is capable of typesetting up to 6000 characters a second. Reasons for this significant stepup in performance level are the digital storage of type fonts and the computer-driven CRT.

The CRT phototypesetter forms characters on the face of a cathode-ray tube by stroking the beam in television style. Tubes with internal character masks are not used in high print quality machines because the font style cannot be changed while in use, and resolution is limited. Characters are positioned on film by a variety of means; the most common locates the character horizontally by writing in the proper horizontal position on the tube face, and vertically by moving the film. Another method uses both horizontal and vertical positioning on the tube face. Vertical film motion generally achieves better quality.

## Performance tradeoffs

The cathode-ray tube to be used in an electronic phototypesetter should provide a very small spot diameter, very high surface brightness, small values of deflection-induced distortion, and rapid deflection speeds. Unfortunately, all of these characteristics can not be optimized at the same time:

1. *Brightness and spot size.* Up to a limit, the maximum surface brightness increases with phosphor thickness. But, because the light diffuses in all directions through the phosphor, the effective spot diameter also increases with thickness: *High brightness and small spot size tend to be opposing requirements.*

Previously, high-brightness, high-resolution screens were impractical. Small (for best resolution) phosphor particles settled slowly and erratically out of the slurry to form a thick screen. Now cataphoretic deposition electrostatically propels slurry-suspended phosphor particles between a charged grid and conducting, transparent face-plate coating for fast, uniform results.

2. *Spot size and deflection angles.* As the electron beam focusing coil is moved closer to the face of the tube, the diameter of the electron spot striking the face of the tube

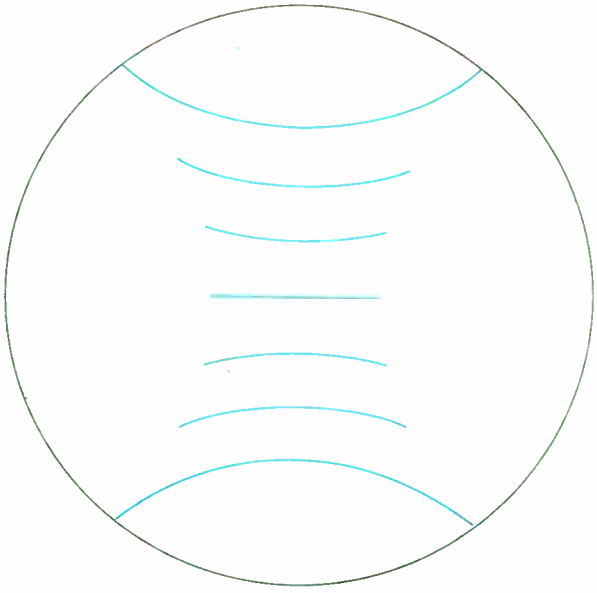


**FIGURE 1. With pincushion distortion, straight lines take on more and more curvature toward the periphery of the cathode-ray tube.**

decreases. This is analogous to the case of optical lenses where shorter focal length produces a smaller image for a given object size and distance. Unfortunately, because the CRT deflection coil must be between the focusing coil and the face of the tube, a short focus distance necessitates a large deflection angle for a given scan size. This severely increases the linearity, astigmatism, and defocusing problems to be discussed in this article: *Small spot size and low deflection angles tend to be opposing requirements.*

3. *Deflection speed and spot size.* In order to obtain minimum deflection and flyback time it is necessary to use low-inductance deflection yokes with relatively few windings. But few turns make magnetic field shaping difficult, resulting in increased spot astigmatism: High deflection speeds and low distortion are conflicting requirements.

## Choosing the CRT

The three major categories of CRT are (1) the electrostatically focused and deflected, (2) the electrostatically focused and magnetically deflected, and (3) the magnetically focused and deflected. The all-magnetic kind are used in most CRT phototypesetting machines because of their generally superior resolution. Furthermore, because they

## The math-physics of electron ballistics for magnetic deflection

Figure 2 shows the geometry for an electron of charge $-e$ and velocity $V_0$ orthogonal to magnetic field $B$. The magnetic field is directed into, and also orthogonal to, the plane of the paper and exists over the distance $l_m$. Therefore, $\tan \theta_1 = D/L$ and by construction $\theta_2 = \theta_1$, which leads to $\sin \theta = l_m/r$.

The radius of curvature of the electron path $r$ is found from two equations that describe the forces on an electron moving through a magnetic field. The field force: $f = -eB \times \bar{v}$ acts perpendicularly to the electron trajectory along the radius $r$ and directed toward $A$ as shown in Fig. 2. The centrifugal force on the electron is $mv^2/r$. When set equal to the field force of magnitude of $eBv$, $r$ equates to $mv/eB$. Inserting this $r$ into the equation $\sin \theta = l_m/r$, produces $\sin \theta = l_m eB/mv$ or alternatively, $\tan \theta = l_m eB/\sqrt{(mv)^2 - (l_m eB)^2}$ (see Fig. 3). Since $\tan \theta = D/L$, $D = L \tan \theta = L\,(l_m eB/\sqrt{(mv)^2 - (l_m eB)^2}$. Dividing each term by $mv$:

$$D = \frac{Ll_m eB/mv}{\sqrt{1 - (l_m eB/mv)^2}}$$

Assume for simplicity that $l_m e/mv = K$, a constant — a valid assumption for a given operating condition. Then $D = LKB/\sqrt{1 - K^2B^2}$.

The deflection $D$ obviously is not linear with $B$, the deflecting field, but has the shape shown in Fig. 4. This is the source of pincushion distortion.

Figure 5 represents a front view of a flat-face CRT with variously deflected spot positions.

The radius vector $\sqrt{2}d$ represents the desired or correct location of a spot deflected when the $Y_d$ and $X_d$ fields are simultaneously applied; $d'$ is the actual spot location. $Y_d' - Y_d = X_d' - X_d = \Delta$ (spot location error) for separately considered deflections, but because of the nonlinear nature of the error, $d' - \sqrt{2}d$, as measured along the 45° line, is greater than $\sqrt{2}\,\Delta$. Stated simply, the percent error increases with deflection.

To determine the required "deemphasis" of the deflection field to produce a linear spot deflection, equation $D = LKB/\sqrt{1 - K^2B^2}$ is solved for $B$:

$$B = \frac{D/LK}{\sqrt{1 + (D/L)^2}} \qquad (1)$$

Equation 1, shown graphically in Fig. 6, when decomposed into its $X$ and $Y$ components, yields the following two equations that give the corrected $B$ fields in the $X$ and $Y$ directions for essentially error-free deflection:

$$B_x = \frac{X}{LK\sqrt{1 + (X^2 + Y^2)/L^2}} \qquad (2)$$

$$B_y = \frac{Y}{LK\sqrt{1 + (X^2 + Y^2)/L^2}} \qquad (3)$$

These equations show their dependence on both $X$ and $Y$. $B_x$ depends on not only $X$, but also $Y$, and $B_y$ is dependent on $Y$ and also $X$. Therefore, the dashed lines in Fig. 7 indicate the required cross-coupling.

need no deflection or dynamic-focus amplifiers operating at 400 volts or more, much freer use of transistors is possible. Certain electrostatically focused tubes appear to equal or exceed the resolution of magnetic types, but high-voltage, dynamic-focus systems are required.

However, electromagnetic deflection has some problems, too. The deflection across the faceplate is not a linear function of the deflection-yoke drive current in electromagnetic tubes. Although a uniform deflection field across the deflection region—a necessary condition for minimizing spot size—produces a deflection angle that is linear with current, deflection of the spot on the flat faceplate is not linear with yoke current.

The kind of nonlinearity produced by magnetic deflection on a flat faceplate is called, from its appearance, pincushion distortion. Figure 1 indicates the presence of pincushion distortion on a series of equally spaced scan lines of equal length.

This problem can be resolved by analyzing electron ballistics for magnetic deflection (see box). Not only are the corners pulled out but there are positional errors over the entire faceplate—except at the nondeflected center position. A brief look at Fig. 8 suggests one technique for pincushion correction; the family of curves gives the required $X$-axis $B$-field versus $X$-axis distance for various values of $Y$-axis deflection.

A typical method of generating a nonlinear current function (to negate what would otherwise be nonlinear deflection) is to construct it from a number of linear approximations leading to the system illustrated in Fig. 9. The low-impedance voltage sources $e_1, e_2, \ldots, e_n$, used to back-bias diodes $d_1, d_2, \ldots d_n$, determine the level at which the input signal $e_{in}$ is shunted by resistors $R_1, R_2, \ldots R_n$, respectively. The slope control resistors $R_1, R_2, \ldots R_n$ are adjusted (as are voltages $e_1, e_2, \ldots e_n$) to produce the best approximation to the desired curve. The number of $R$, $d$, and $e$ combinations determines the final accuracy. By readjusting the diode break-control voltages $e_1, e_2, \ldots e_n$, it is possible to accurately approximate any one of the curves in Fig. 8. This readjustment can be made by some functional relationship proportional to the amount of off-axis, $Y$ deflection. If this functional relationship is properly chosen, then for any value of $Y$, the values for $e_1, e_2, \ldots e_n$ will be adjusted to produce the particular curve in Fig. 8 for optimum correction in the $X$ direction. Correction in the $Y$ direction is similarly accomplished.

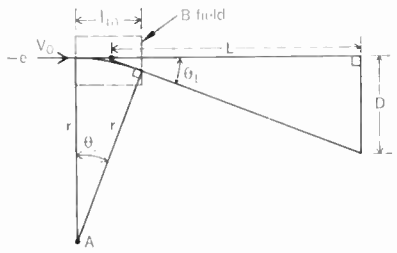



**FIGURE 2. The force acting on an electron moving in a magnetic field is perpendicular to both its direction of motion and the direction of the magnetic field. In the diagram, L is the perpendicular distance from the center of deflection to the flat faceplate of the CRT, D is the deflection distance, and $\theta$ is the deflection angle. Box text explains remaining symbols.**

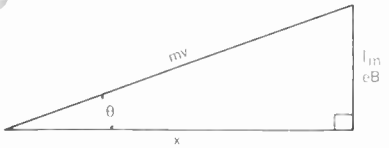



**FIGURE 3. The trigonometric relations created by electron deflection in a magnetic field are better visualized in this section expanded from Fig. 2.**

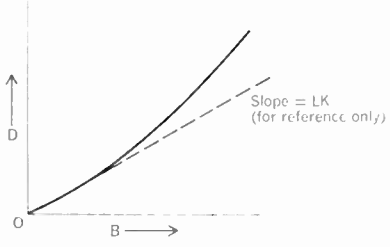



**FIGURE 4. The deflection of electron beam across a flat faceplate of a CRT as a function of field strength becomes less linear as the field amplitude increases.**

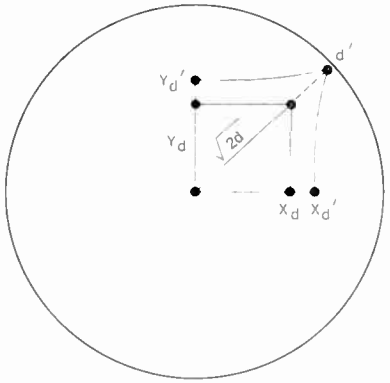



**FIGURE 5. This is a schematic front-view of a flat-face CRT with variously deflected spot positions: $Y_d$ is the desired or correct location of a vertically deflected spot; $Y_d'$ is its actual position owing to a nonlinear sweep. $X_d$ and $X_d'$ are the horizontal counterparts of the Y's.**

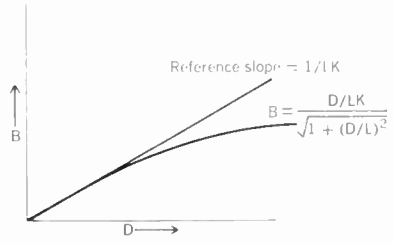



**FIGURE 6. The magnetic field may be compensated to provide linearity over the entire CRT faceplate.**

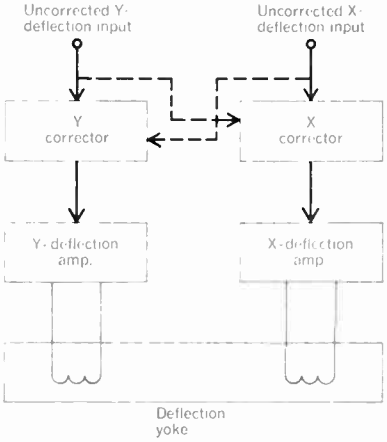



**FIGURE 7. Because correction for the X and Y magnetic deflection fields are interrelated, coupling is needed between the two—indicated by the dashed cross-coupling lines.**

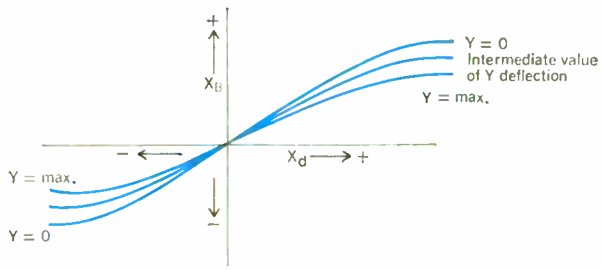



**FIGURE 8. A family of curves is shown for the corrected magnetic deflection field B in the X direction as a function of the distance X for values of Y deflection.**

**FIGURE 9. A ladder of biased rectifiers and shunt resistors is one device for creating nonlinear output, field-generating currents to correct for pincushion distortion.**

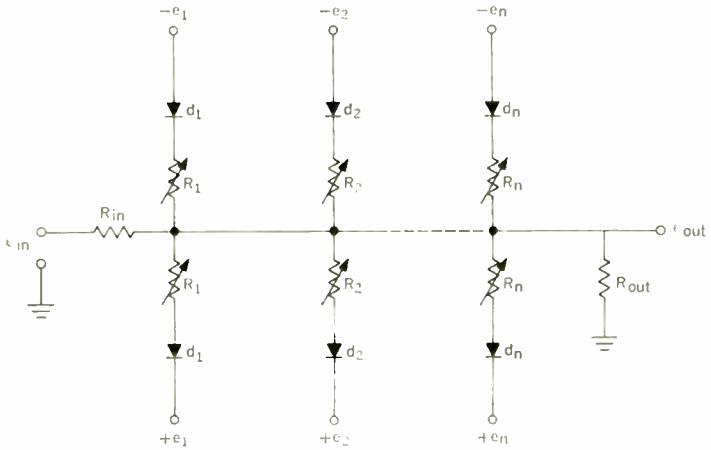



**FIGURE 10. Strength of the focusing field must necessarily be less as beam deflection increases. This block diagram shows, among other parts, the focusing coils that take care of static and dynamic errors.**

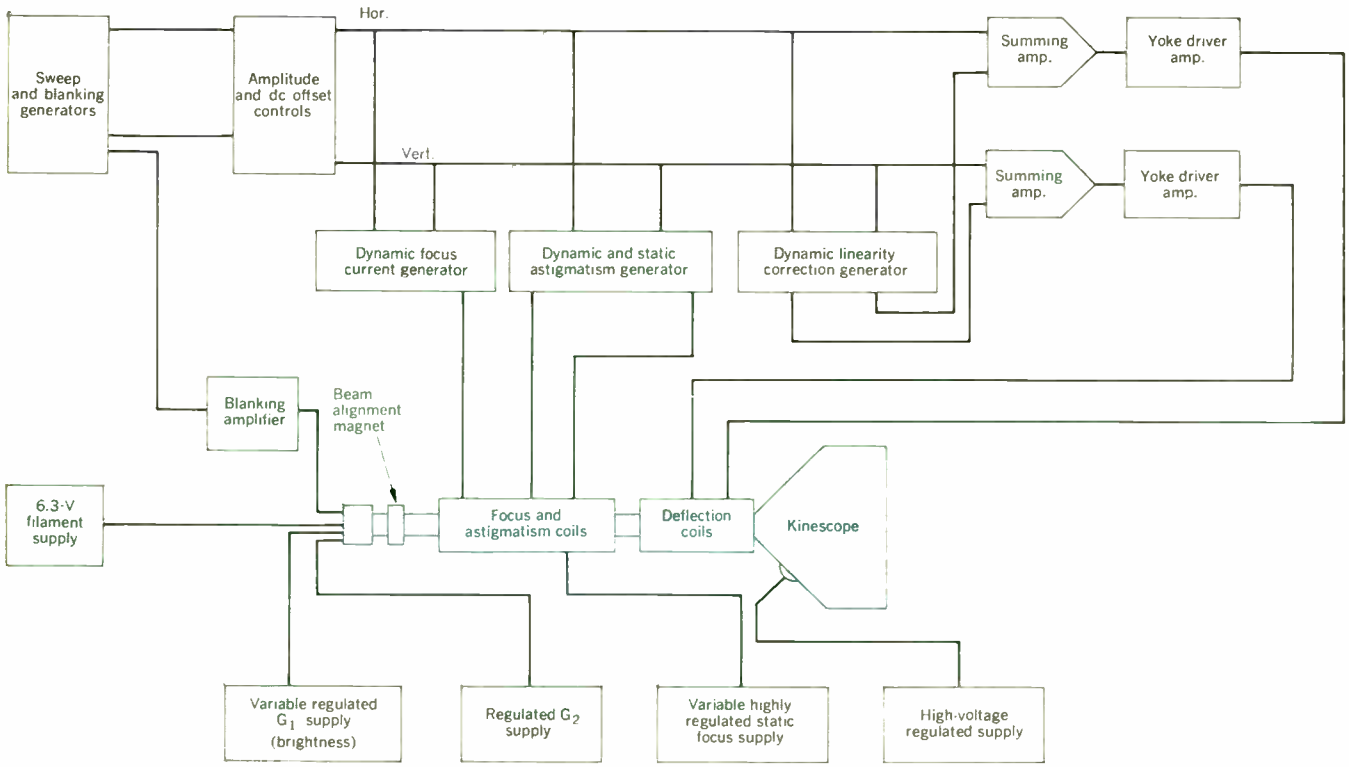
Alternately, a compensating approach is to produce a binomial expansion of Eqs. (2) and (3) [see box]. The accuracy of correction, of course, depends on the number of terms used in the expansion. Multipliers, squarers, and adders are the basic building blocks needed to simulate the expansion.

To determine the degree of distortion in a given magnetic deflection system, consider the deflection equation $D = L \tan \theta$ (see Fig. 2). This equation can be rewritten as $D = L \sin \theta/\cos \theta$, and since $\sin \theta = l_m eB/mv = KB$, $D = LKB/\cos \theta$. Solving for $B$: $B = (D \cos \theta)/LK$.

Therefore, the correction factor is $\cos \theta$, and for a $\pm 20°$ deflection system, the maximum correction factor is approximately 0.94, roughly representing a 6 percent error at full deflection. Reductions of this figure by an order of magnitude have been accomplished, with indications that $\pm 0.1$ percent residual error over the entire faceplate is possible. For that matter, using only the first three terms in the binomial expansion of $[1 + (X^2 + Y^2)/L^2]^{-1/2}$ to synthesize the actual function, it is theoretically possible to approximate the function to within $\pm 0.065$ percent for a maximum deflection of about $\pm 20°$. At this level of accuracy, practical considerations, such as stability and accuracy of electronic components, mechanical stability of the CRT/yoke combination, and extraneous deflecting fields play an important role.

## Pinpoint focus

Once the beam strikes the phosphor accurately positioned, the spot on the face of the CRT must be made as small as possible. For best focus, the magnitude of the longitudinal focusing field, $B$, must be altered as a function of deflection if the radius of curvature of the CRT faceplate is not equal to the distance from the faceplate to the center of deflection. Usually the radius of curvature is



greater—being infinite for flat-face CRTs. In such cases, therefore, the electrons travel a greater distance as the deflection increases and, consequently, require less focusing current than for the undeflected condition.

The system block diagram, Fig. 10, shows the location of the focusing coil, and its input signal requirements. The focusing function is broken into two parts: static focus and dynamic focus. For modest deflection angles, the required change in focusing field strength is under 10 percent. Quite frequently, a separate coil of low inductance is used for the dynamic focusing signal. This is done primarily to allow high-speed correction currents to be applied through the low coil impedance when rapid changes in deflection occur. The dynamic coil is generally an integral part of the much larger static focusing assembly that requires a well-regulated, adjustable current source. Adjustment allows accommodation for different CRTs and operating potentials. The proper waveform for dynamic-focus correction is approximately $i_c = ad^2$, where $d$ is the radial deflection. Since $d^2 = X^2 + Y^2$, this equation leads directly to the manner in which the correction waveform is generated. The $X$ and the $Y$ deflection waveforms are squared, summed, and fed through a level-control potentiometer to the dynamic focusing coil driver transistor. At this point all required corrections, except for a spot distortion, called "astigmatism," are accomplished.

### Astigmatic distortion

Lack of roundness of a deflected spot causes loss of resolution. Consequently, astigmatism as a function of deflection must be prevented. In addition to the purely geometrical cause of astigmatism shown in Fig. 11 any stray deflection fields that alter the focusing field can also cause spot-shape distortion. To compensate for astigmatic distortion, a correction field is added to the focusing field. This causes the axial component of the focusing field to vary in magnitude across the region where deflection occurs—increasing in the direction the beam is deflected—and shapes the total axial focusing field to compensate for geometric and deflection-derived distortion. In other words, as the deflection increases, so does the degree of correction.

To achieve correction, current is introduced into sets of astigmatic coils mounted adjacent to the focusing coils. One coil distorts the axial field along the $X$- and $Y$-axes. The other coil operates on a set of axes usually rotated 45° from the original $X$ and $Y$. Theoretically, the required current waveform is $i_1 = K_1 (Y^2 - X^2)$ for the first set of coils, and $i_2 = K_2 (XY)$ for the second set.[1] These two functions can be generated using squarers and adders operating on the $X$ and $Y$ deflecting waveforms. The circuit details are omitted, but the general location within the system can be seen in Fig. 10. In practice, the required correction currents may deviate from their theoretical values because of nonuniform deflection and focusing fields. Modification of the ideal dynamic-astigmatism correction current may be needed for a particular CRT.

The stability of the various correction circuits must be commensurate with their accuracy requirements. That is, if 0.1 percent deflection accuracy is needed, then the stability of the deflection system must be at least that good. The stability of the high-voltage power supply may be approximately 0.2 percent. Fractional-percent accuracies in focus are also necessary for maintaining the high

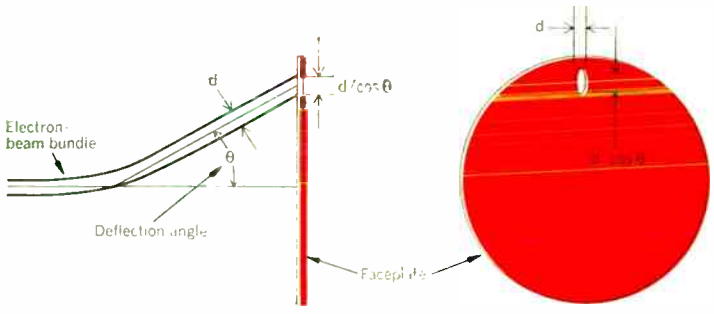



**FIGURE 11. Widening of the electron beam at focal plane is an astigmatic effect. One cause, shown here, is purely geometric.**

resolution, but astigmatism correction is less critical.

Linearity, focus, and astigmatism correction accounted for, certain other considerations still remain.

### Other matters to consider

Hum is a problem, and the usual precautions of ground-loop removal and magnetic shielding around the cathode-ray tube must be taken. A detailed analysis of the hum problem is given in another article.[2]

Because character type must be completely uniform over the CRT face, deflection-amplifier axis-crossing distortion must be virtually nonexistent. Since such axis-crossing distortion occurs in zero-bias class B transistor amplifiers, amplifiers should be operated toward class A, thereby trading efficiency for linearity.

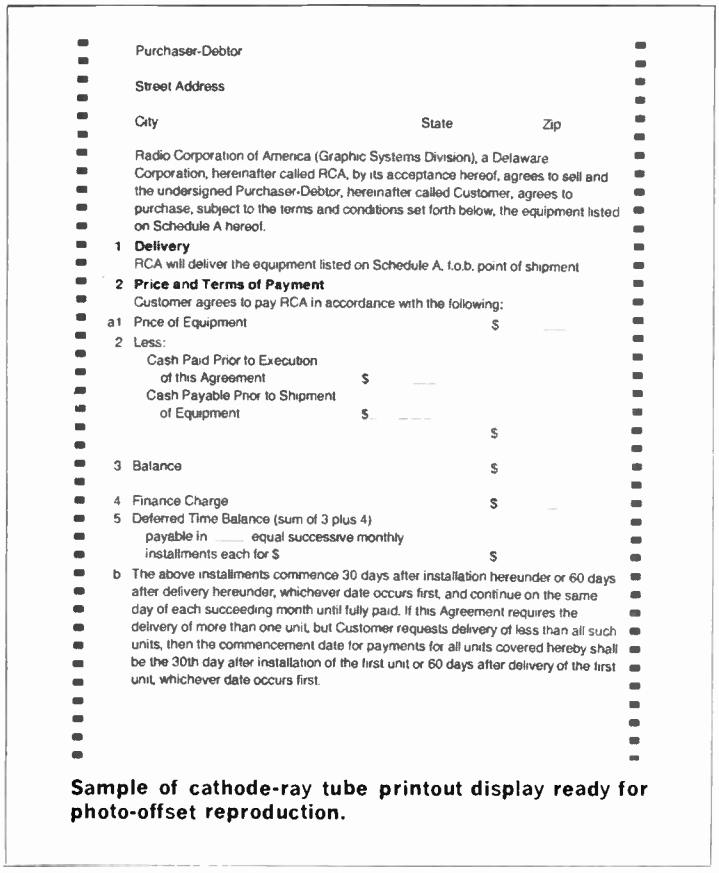
### Comparing continuous and step scan

Two general modes of scanning are possible on a cathode-ray tube: (1) the continuous, television type whereby the entire tube face is scanned with continuous strokes and the beam turned on where desired; and (2) step scan, whereby the beam is moved directly to location where it is to be on. Step scan appears to save time wasted by continuously scanning—including blank spaces. A problem with the step-scan method is that the beam cannot be deflected instantaneously. A finite settling time ranging from 2 to 20 μs or more is associated with each step. Thus, if each picture element is written in a step-settle-write sequence to produce a true dot picture, the cumulative settling time of the deflection process may be equal to, or greater than, the time spent scanning blank areas by the continuous method. However, a combination of the two methods may be used to increase response of the television method used alone by a factor of 3 or 4 to 1, depending on circumstances.

In this combination method, the beam is step-scanned to the general vicinity of the character to be written. Then, the actual character is written using a small, continuous scan limited to the area of the character. Additionally, this method relieves the computer of the time- and memory-consuming task of storing a full page of "text" and then synthesizing television-style sweeps from this information. The true step scan, nevertheless, has been found useful in the generation of certain types of halftone photographs.[2]

### Full-face vs. single-line scan

In the context of the printed word, resolution can be defined as the number of characters distinguishable on a



Purchaser-Debtor

Street Address

City State Zip

Radio Corporation of America (Graphic Systems Division), a Delaware Corporation, hereinafter called RCA, by its acceptance hereof, agrees to sell and the undersigned Purchaser-Debtor, hereinafter called Customer, agrees to purchase, subject to the terms and conditions set forth below, the equipment listed on Schedule A hereof.

1 **Delivery**
RCA will deliver the equipment listed on Schedule A, f.o.b. point of shipment

2 **Price and Terms of Payment**
Customer agrees to pay RCA in accordance with the following:

a1 Price of Equipment $ ___

2 Less:
Cash Paid Prior to Execution of this Agreement $ ___
Cash Payable Prior to Shipment of Equipment $ ___

$ ___

3 Balance $ ___

4 Finance Charge $ ___

5 Deferred Time Balance (sum of 3 plus 4) payable in ___ equal successive monthly installments each for $ $ ___

b The above installments commence 30 days after installation hereunder or 60 days after delivery hereunder, whichever date occurs first, and continue on the same day of each succeeding month until fully paid. If this Agreement requires the delivery of more than one unit, but Customer requests delivery of less than all such units, then the commencement date for payments for all units covered hereby shall be the 30th day after installation of the first unit or 60 days after delivery of the first unit, whichever date occurs first.

**Sample of cathode-ray tube printout display ready for photo-offset reproduction.**

single line. Naturally, the smaller the line width, the fewer the characters that fit. It follows that maximum resolution for a CRT is across its diameter (diagonal). If the entire face of the CRT is used, then resolution will be cut in proportion to the percentage difference that each line deviates from the diagonal of the tube. Therefore, exposing a full page with about a 4/3–height/width ratio (diagonal 5) wastes about 40 percent [$(5 - 3)/5 \times 100$] of the potential resolution across the tube diameter. Sweeping the writing beam along *the* horizontal diameter while moving the film vertically is one solution to this resolution problem. However, this imposes severe film-motion restrictions and necessitates generation of a television type of scan by the electronic system; this, as mentioned previously, imposes more computer load.

Considering various limitations, particularly in early phototypesetters, a compromise, was adopted for text writing. A writing "window" the height of the largest character to be written is used. Since such a slot can be nearly the full-tube diameter in width, this arrangement neatly overcomes the problems of wasted resolution and television-style scans. It also greatly reduces pincushion distortion and astigmatism. For a given tube, it achieves the best resolution.

Recently, however, demands for increased flexibility have led to development of full-face machines despite their more limited resolution. Utilizing the full face of the tube to write a full page without film motion has many advantages, chiefly high writing speed and simplification of film-handling mechanisms. It also contributes to better edge matching between the first and the last written parts, after moving the film.

## Mechanical requirements

Positioning of critical components such as the focusing yoke, kinescope, lens, and film plane must be held to typical tolerance of $\pm 0.0025$ cm or closer over distances of one meter or less. Transverse vibration must be held to less than 0.001 cm on these components. In addition, these tolerances must be held over temperature variations of perhaps 30°C and under conditions of vibration by film-handling mechanisms, changing of film cassettes, vacuum pumps, and blower motors. The system must be mechanically rigid to withstand shipment and installation processes with minimum realignment of the optics.

## An overview

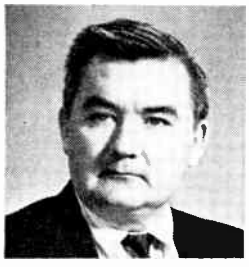
When used with precision deflection and correction components, the high-precision CRT is an excellent device for generating graphics material and is currently in widespread use. Further improvement, particularly in resolution and brightness (about 2 to 1 in each area) would be very useful. But while some is in the offing, conventional CRTs are approaching the limits of performance and so the rate of improvement is slow.

The authors are grateful to G. O. Walter, chief engineer of the RCA Graphic Systems Division, who supplied much of the introduction to this article.

REFERENCES

1. Schlesinger, K., and Wagner, R. A., "Correction of deflection-aberrations by analog computer," *IEEE Trans. Electron Devices*, vol. ED-12, pp. 478–484, Aug. 1965.

2. Hallows, R. L., and Klensch, R. J., "Electronic halftones," *IEEE Spectrum*, vol. 5, pp. 64–72, Oct. 1968.



**Richard J. Klensch** joined the RCA Laboratories after receiving the B.S.E.E. degree from the University of Illinois in 1952. During his early days with the lab, he worked on high-resolution radar detection devices. Starting in 1954, he spent the next two years as a U.S. Army radar instructor. Back again with RCA, Mr. Klensch researched such areas as microwave scanning antennas, time-division multiplex systems, and color television. While at RCA, he also did graduate work at the Electrical Engineering Department of Princeton University during 1952–58. More recently, Mr. Klensch has been investigating new electronic halftone generation techniques and CRT display systems.

**Elvin D. Simshauser** came to RCA with the bachelor degree in physics in 1951. For the next three years—until inducted into the Army—he did development work on sound-powered telephones, aircraft and submarine intercom systems, and miscellaneous transducers. His army duties took him to White Sands Proving Grounds, and here he worked with computers. Returning to RCA in 1955, Mr. Simshauser developed headsets and microphones for the Air Force, and then worked on several classified communications programs. After a stint at the RCA Tucson plant from 1963 to 1966 he transferred to the company's Graphic Systems Division in 1966; since then he has worked on kinescope photocomposing systems, including dynamic astigmatism generators, and line-drawing and halftone reproduction systems.

A special staff report

# A look at Apollo electronics

***Without electronics, manned spaceflight would be impossible. But space electronics requirements have forced the development of new components, systems, and techniques that aid man on earth as well as place him high among the stars***

**W. J. Evanzia** *Associate Editor*

Rocco Petrone* denied it, but everyone thought the rocket seemed mightier, the flames brighter, and the smoke more dense. It had all happened before. No! This was different; everything had built up to this moment. The drama, tension, and anxieties of the Mercury, Gemini, and previous Apollo flights were buried in the past. This is the beginning.

Truly, it is the beginning of a new age of scientific exploration; of compiling knowledge, confirming old laws, and presenting new theories. For science, a golden era has begun.

But the "man in the street" knows little of what made the moon shot possible. He might have heard that the cost of the Apollo 11 mission is estimated at $355 million but he knows little of the technical backup provided the astronauts and the space program.

Just about all scientific and engineering disciplines contributed; and it is difficult to say that one technology is, or was, more important than the other. But perhaps it is fair to assume that the computer technology crossed the boundary lines of more disciplines than any other. For example, an RCA 110A general-purpose digital computer was used to check out the first stage of the Saturn V launch vehicle and a separate system—consisting of a Control Data 924A general-purpose computer and a number of test stations—was used to check out the second stage.[1] Computers were used to guide the command and service module (CSM) to the moon, to set the lunar module (LM) down on the surface, and for communications control. In fact more than 600 computers performing in excess of 50 million calculations per minute were required to get Apollo 11 to the moon and back.

## Guidance and navigation

Perhaps the most complex and sensitive component in both the CSM and the LM is the guidance and navigation (G&N) system. Basically, it is composed of three subsystems: an inertial guidance subsystem—inertial measurement unit (IMU) and associated equipment—to gauge changes in spacecraft attitude and in velocity due to thrust, and to assist in generating steering signals; an optical navigation subsystem—space sextant and scanning telescope—to determine spacecraft position and velocity and to align the IMU; and the command module computer (CMC), which calculates the steering signals and engine discretes necessary to keep the spacecraft on the required trajectory, positions the stable member in the IMU to a coordinate system defined by precise optical measurements, points the optical unit at celestial objects, conducts limited malfunction isolation of the G&N system by monitoring the level and rate of system signals, and supplies spacecraft condition information to the display and control panel.

Like the CSM's computer the LM's guidance computer (LGC) has four major functions: it calculates steering signals and engine discretes; positions the stable member in the IMU; conducts limited fault tests in the Primary

**FISH-EYE VIEW of real-time computer complex at NASA's Manned Spacecraft Center. The IBM System/360 Md. 75 computers monitor spacecraft operations, and analyze tracking information for display to mission controllers.**


* Launch Director, Kennedy Space Center

Guidance and Navigation Control Subsystem (PGNCS); and supplies LM condition data to the display and control panels. During the moon landing phase of the Apollo 11 mission, the LM computer overloaded, causing alarms to flash and a near mission-abort. But more on that later.

As everyone knows by now, Apollo 11 took off from Cape Kennedy and went into a near circular parking orbit around the earth. At that time, it was moving at about 8000 m/s. Then the engines fired again, adding about 3000 m/s to the spacecraft's velocity (translunar injection), and the vehicle was sent winging its way to the moon.

To get to the moon, the G&N system had to answer three questions: Where am I and what am I doing? Where do I want to go? What should I do? The answer to the first question comes from the on-board inertial subsystem. It has accelerometers and gyros that give velocity and direction data to the on-board digital computer. The computer uses this information to arrive at a best estimate of the state vector—seven quantities: time, three coordinates of position, and three values of velocity. Periodically, the state vector is updated by ground information fed into the digital computer via the data link.

The Apollo's navigation system also includes the ground radar, which keeps track of the vehicle in space; the ground computer, which processes radar data; and the ground communication system. This is, of course, in addition to the on-board inertial and celestial navigation subsystems. Information from these subsystems plus data concerning the vehicle's origin are fed to the computer to derive the best estimate state vector.

Planning data preprogrammed into the computer memory plus information sent from the ground as the flight plan changes is used to answer the question, "Where do I want to go?"

Finally, there is the question, "What should I do?" Here, "I" is actually the spacecraft's pilot and/or the ground control system. By utilizing changes in spacecraft velocity (engine burns) and attitude, either of them, or both, guide the vehicle to some point in space. In other words, the pilot uses the engines to provide thrust over which he has a degree of directional control.

The point in space at which the astronauts aimed was the Sea of Tranquility, and all went well until Eagle separated from Columbia.

At about 12 000 meters the LM's landing radar pointed at the moon's surface and began feeding altitude and rate-of-descent data to the computer. At the same time, the LM's crew began asking the computer for readings. Immediately, alarms began to flash and ground control started to worry.

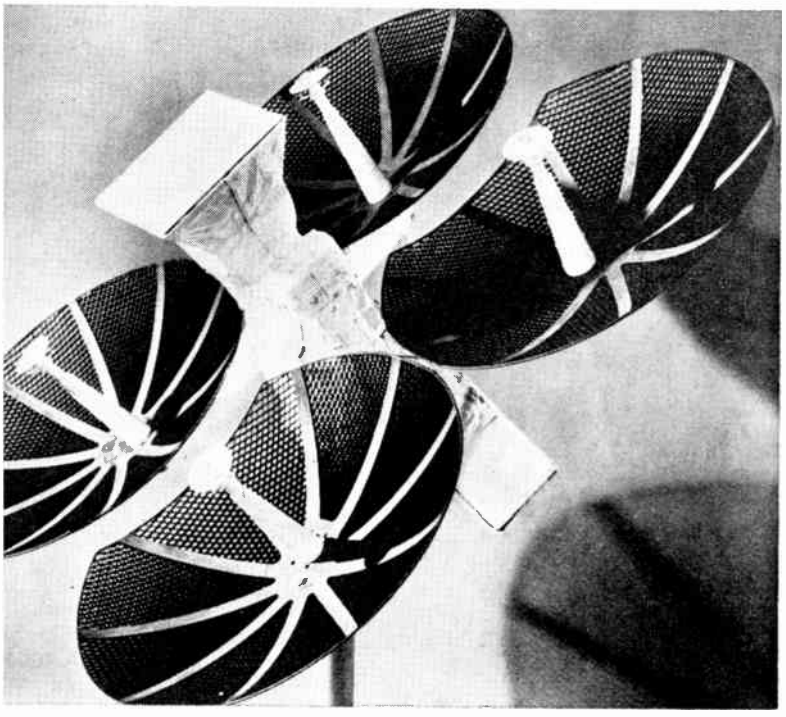
Apparently the computer overloaded. It couldn't do its job properly. No one knew why because it had successfully been flown on previous Apollo missions. However, the situation was so bad that the mission was in danger of being aborted.

According to Christopher Kraft, director of flight operations at the Manned Spacecraft Center, the rendezvous radar on LM was tracking the CSM and triggering a transponder that sent back distance, angle, and rate-of-closure information to the LM. He also said that the radar's mode switch was in the "auto track" position instead of the "LGC (LM guidance computer)" position; this meant the radar was using the spacecraft's power supply instead of the computer's power supply, and, as a result, the rendezvous radar's signals were out of synch, creating a noise or dither that ate up much of the computer's capacity.

Supposedly, the computer has enough capacity to take care of almost any emergency—36 864-word fixed (rope) memory and 2048-word erasable (ferrite-core) memory. However, up to 90 percent of the computer's capacity may be used during a normal landing. But in this case, the rendezvous radar, which has the capability of inputing 6400 bits per second into the computer, supposedly required 15 percent of the capacity. See the problem?

All of the companies concerned—AC Electronics, who made the guidance and navigation subsystem's inertial platform; the Raytheon Company, who made the computer; and RCA, who developed the rendezvous radar—say each of their systems worked well, and as required. Perhaps the problem lies in the computer program, or with the mission planners at MSC, but whatever the problem is or wherever it lies, it will have to be dealt with and cleared up before Apollo 12 blasts skyward.

**CLOSE-UP of a typical deep-space, high-gain, S-band antenna for Apollo communications.**



## Communications made Apollo 'go'

Getting to the moon is only a part of the Apollo story; the mission would not have been a success if it were not for the huge communications backup.

In a real sense, mission success or failure depended upon the instantaneous real-time processing of communications to and from the spacecraft. Incoming data were fed into a command computer and, in seconds, evaluated and compared against the mission profile. This enabled controllers to determine immediately if the mission was proceeding as planned.

Nascom, NASA's multimillion dollar communications network, had the primary responsibility of transmitting, receiving, conveying, and routing the flood of data and messages that flowed between Apollo 11 and the Mission

Control Center in Houston, Tex. The computerized data and communications processing system known as the Automatic Data Switching Systems (ADSS) is part of this network.

The hub of ADSS is located at the Goddard Space Flight Center, Greenbelt, Md., where three Univac 494 communications processors act as an electronic switchboard for Teletype messages. The second part of the network is the Communications Command and Telemetry Systems (CCATS) located at Houston's Manned Spacecraft Center. The CCATS was designed to handle the traffic of up to 30 two-way 100-word-per-minute telemetry links and up to 20 high-speed data lines.

The third segment of the Nascom network is the Remote Site Data Processing (RSDP) systems. These are located at 14 global ground tracking sites and on four Apollo instrumentation ships. The RSDP systems job is to accept, record, and transmit data originating from the spacecraft ("down" data) and to compute and issue commands to the spacecraft ("up" data).

"Up" information is transmitted over an ultra high-frequency radio (Apollo Unified S-Band) link at a 2400 bit-per-second data rate. Communication between the ground tracking sites and Houston, via high-speed links, occurs at the same rate.

In the case of "down" data, sensors built into the spacecraft continuously sample the pressure and temperature inside the capsule, its attitude and position in space, and such physical factors as the astronauts' respiration, heart beat, and temperature. These data are transmitted to ground stations at a rate of 51 200 bits per second.

The Apollo command module's telemetry system, which transmitted the aforementioned sensor data, was developed by the Electronics Group of the Harris-Intertype Corporation. It was a complex package containing 21 400 electronic components, weighing 22 kg, and measuring less than 0.028 cubic meter.

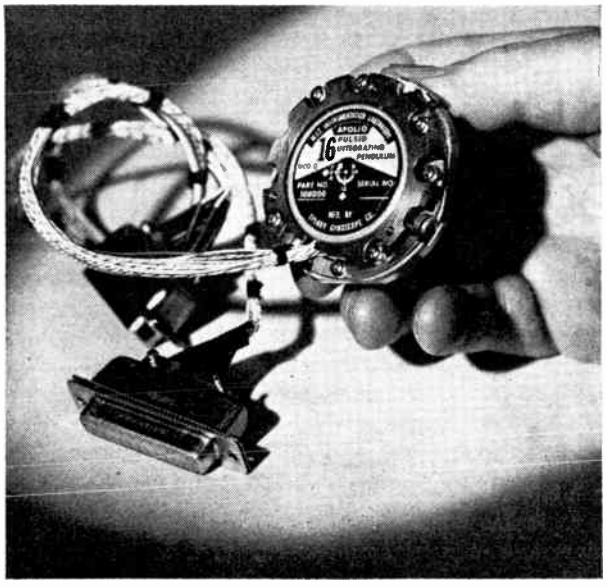
The LM's telemetry unit had 9100 components packed into a box 50.8 cm long by 12.7 cm wide by 15.2 cm high. It weighed 10.4 kg.

During the lunar landing phase, this system served as the primary communications link between the LM and the orbiting CSM, processing and relaying vital data on all aspects of the craft and its crew. Information such as vehicle attitude; temperature readings from the outside skin, cabin, and engine; and data on the descent/ascent propulsion system were processed.

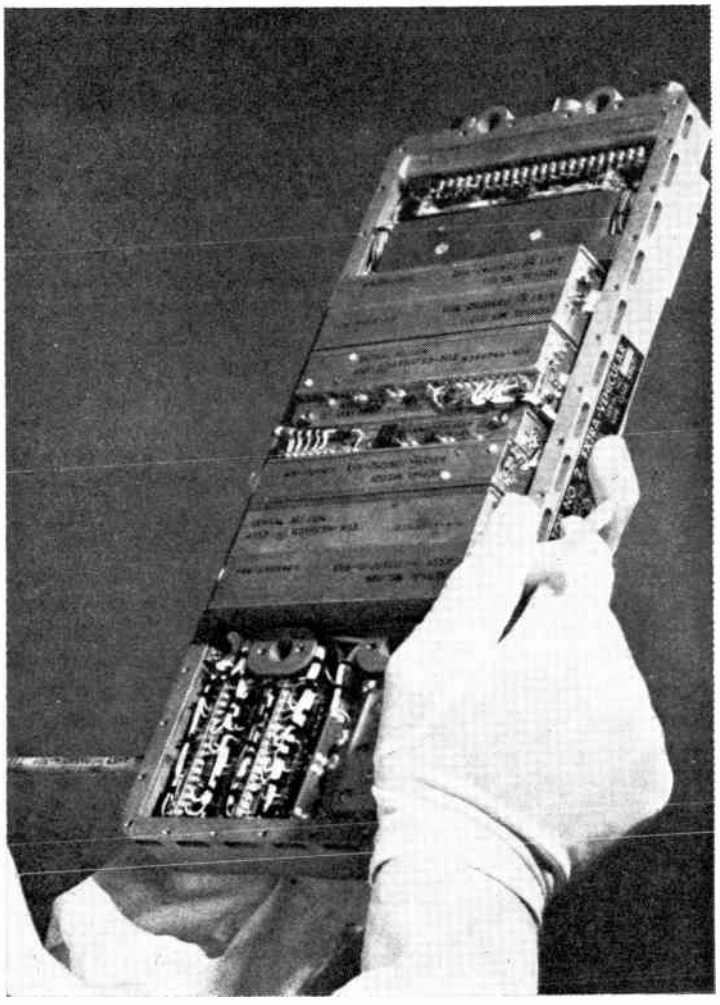
Keeping Capcom talking with the astronauts was a big job for the Apollo instrumentation ships and range instrumentation aircraft. Some 450 000 kg of sensitive electronics worth $100 million are packed into each of the huge 182-meter-long vessels, operated by the U.S. Air Force and Military Sea Transport Service. During a mission, 122 technical personnel are on board to maintain and operate the systems. Each aircraft has 13 600 kg of instrumentation crammed into it; and, in addition to a flight crew of four, carries a mission controller and six electronics operators.

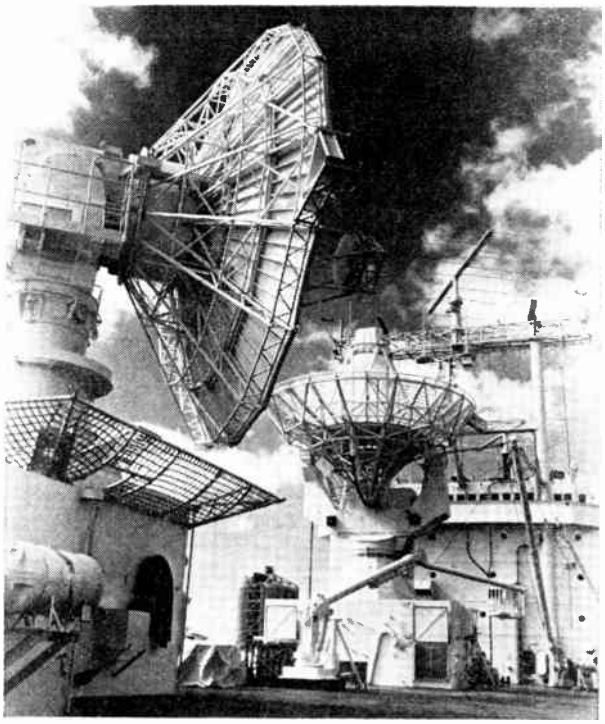
Here's a quick rundown on some of the antenna systems on board the ships: a 9.1-meter parabolic dish for satellite communications; a C-band precision tracking

**TINY ACCELEROMETERS like this sensed the amount and direction of change in acceleration of the Apollo 11 spacecraft and fed the data to the guidance system.**



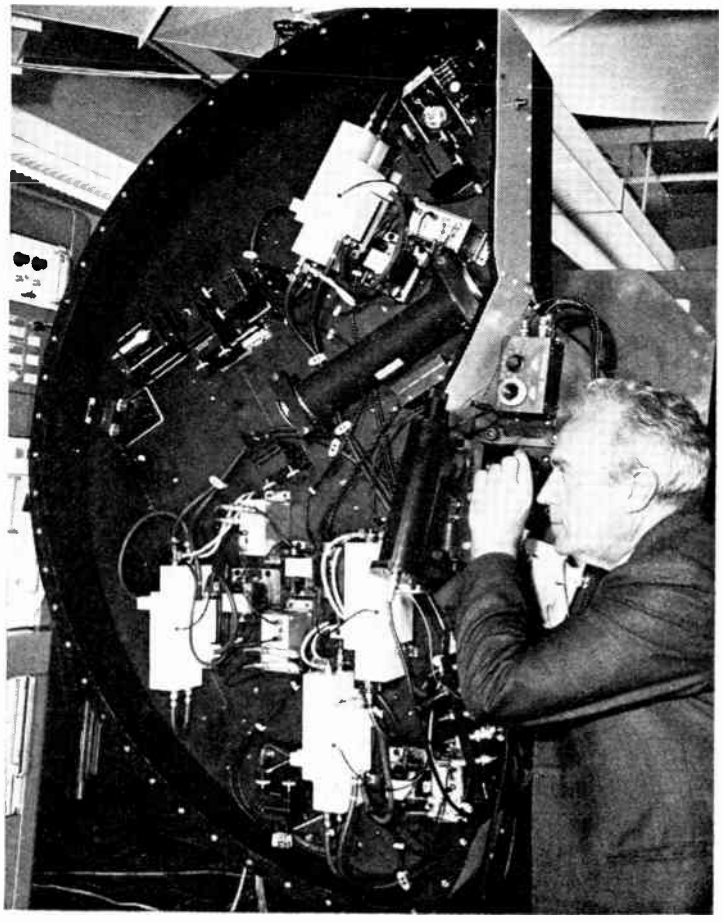

**TECHNICIAN checks an Apollo Extra-Vehicular Communications System. The set contains five transmitters and receivers for voice and telemetry to radio data transmissions between astronauts, the CSM, and the earth.**





**UNIFIED S-band antenna (front) on board the USNS Vanguard helped send Apollo 11 to the moon. The large pedestal-mounted antenna in the rear is for receiving telemetered data from the CSM.**

**SCIENTIST Dr. Renne Julian sights through the alignment device of a laser rangefinder that will be used to compute the precise distance between the earth and moon.**



radar with two 9.1-meter antennas, one dish to receive television and biomedical data from the spacecraft and the other to relay data to and from the Mission Control Center in Houston; and a dual quad-helix array for command and control as well as voice communication between ship and spacecraft.

The airplanes carry 2.13-meter parabolic dishes (world's largest airborne steerable antenna) in their noses for telemetry and communications reception. A probe antenna is on each wing tip for high-frequency work.

In addition to communications systems, the planes also carry Airborne Lightweight Optical Tracking Systems (ALOTS), which provide optical coverage of the missile launch, including liftoff, staging, and reentry.

### The world watched on television

To the mass of men, women, and children hunched in front of a million television sets, eyes glued to screens and ears straining to hear every word, the black-and-white television camera was perhaps the best known of the electronic systems. Actually, two Westinghouse cameras were involved in the space show: A color camera was used in the CSM to photograph the astronauts and the earth, and the black-and-white camera was used exclusively on the moon's surface.

Though the television coverage provided entertainment and the means by which millions of people the world over participated in the walk on the moon, its primary purpose was to furnish scientists with a supplemental real-time data source. That is, scientists used television as an aid in determining the LM's exact location on the lunar surface; to evaluate the extravehicular mobility unit (EMU); to evaluate man's capabilities in the lunar environment; and as an aid in documenting sample collections. By means of television, scientists were also able to correlate the crew's activity with telemetered data, voice comments, and other photographic coverage.

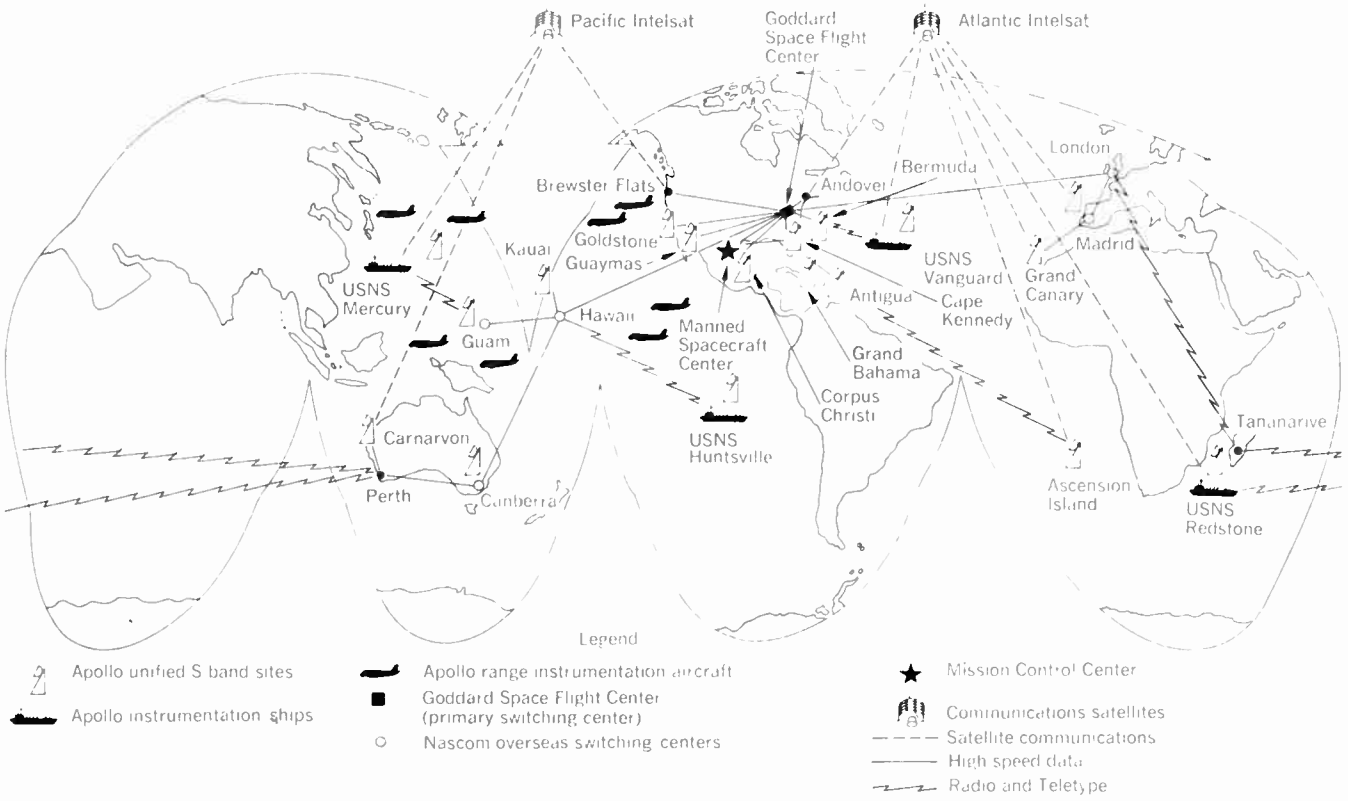
The 3.3-kg Westinghouse black-and-white camera needed only 6.5 watts of power—less than a night light—to photograph astronaut Armstrong as he climbed from the LM. A special secondary electron conduction (SEC) imaging tube that operated in light ranging from 0.007 to 12 600 footlamberts, or from near total darkness to glaring brightness, gave it the capability of observing Armstrong and Aldrin moving about on the moon.

Unlike the black-and-white lunar camera, the 5.9-kg color camera was not designed to work in a space vacuum, but it was equipped with a more sophisticated lens system. It had a variable-focus zoom lens with a focal length ranging from 12.5 to 75 mm that provided a diagonal field of view from 54 to 9 degrees.

It also used only one imaging tube (standard broadcast cameras are built with three tubes, each with a separate color photographic function). To take color pictures, the camera used a method called the field sequential technique. This technique was developed by CBS Laboratories 28 years ago but never used because it was not compatible with black-and-white television broadcast standards.

The chief component in the color camera is 7.62-cm-diameter wheel divided into six sections containing red, blue, and green filters.

The wheel spins at 600 revolutions per minute so that the sequence of color filters passing in front of the imaging tube during one revolution is red, blue, and green, red, blue, and green.




**SATELLITE, ships, and planes are links in NASA's worldwide communications network for Project Apollo.**

As each color filter passes in front of the imaging tube, it collects all of the information on red colors and hues, then blue colors, then green, and then repeats the sequence. In effect, the camera is taking many one-color pictures at an extremely rapid rate.

Back on earth, a video recorder developed by Data Disc Inc. converted the special color signal from the spacecraft to a form compatible with commercial television.

Basically, the recorder consisted of a fixed-head parallel video disk rotating at 3600 r/min, a modulator/demodulator unit, a servo drive, and a video delay unit, plus logic and switching circuits.

### Getting to know the moon

As the astronauts climbed into Eagle and prepared to head home, data from one of the experimental packages left on the moon's surface were already being transmitted to scientists on earth.

The EASEP (Early Apollo Experiments Package) was put together by the Bendix Aerospace Systems Division of the Bendix Corporation; that is, Bendix was the program manager, and as such, responsible for the design, integration, and test of the package. A number of industrial companies contributed their expertise by supplying the components that made up EASEP. For example, Dynatronics, Ling Temco-Vought, Lockheed, Philco, and Spectro Labs, as well as Teledyne, furnished parts for the Passive Seismic Equipment Package (PSEP). Components for the Laser Ranging Retro-Reflector (LRRR) were supplied by Arthur D. Little, Inc., and Perkin Elmer.

The objective of the seismic experiment was simply to measure moonquakes (analogous to earthquakes) and meteoroid impacts. In this way, says Dr. Gary Latham of the Lamont Geological Observatory, scientists hope to obtain some idea of the internal structure of the moon.

How does the PSEP work? Basically, the relative motion of a suspended weight (which tends to remain immobile as the experiment package moves with motions of the moon) causes an electrical change, which then becomes a reading of the amount and frequency of motion. Four such units in PSEP report long- and short-period vibrations along the horizontal and vertical axes.

PSEP measurements are sent to earth by a transmitter that shares a helical ribbon antenna with a command receiver. An earth command is received as a phase-modulated digital signal, which, when decoded, is directly to the experiment as a discrete command. Scientific data from the experiment are first sent to a data-processing unit and then combined with other PSEP information in a special format and transmitted in digital form to the ground.

Solar panels that convert the energy of sunlight to electricity and have an output of 33 to 43 watts provide the power for PSEP. Two isotope heaters (each producing 15 watts of heat from Pu-238 fuel) help PSEP survive the $-300°$ F temperature of the lunar night.

The PSEP's four seismic sensors—three long-period (LP) and one short-period (SP) sensor—and associated electronics are contained in a 10.2-kg 27.9-cm-diameter by 38.1-cm-high cylindrical beryllium container. The LP sensors measure frequencies from approximately one to 0.004 Hz and contain 0.73-kg masses mounted on the ends of three booms that allow two masses to swing horizontally and one mass to swing vertically. The masses are part of a capacitance circuit, so they produce an output proportional to their displacement.

The SP sensor is a single-axis device that detects vertical motions of approximately 20 to 0.05 Hz. It consists of a magnet suspended in a coil. When the magnet moves its cuts the coil's magnetic field, inducing a voltage that is proportional to the velocity of the relative motion.

The electronics amplify and filter the sensor's outputs and convert them to 10-bit digital words. These are stored until the PSE receives a signal indicating that the data system is ready to accept the information for transmission to earth. The digital data are formatted and transmitted as eight distinct measurements: six signals for the three LP seismic outputs, the SP output, and the sensor temperature. More than 15 functions may be ordered by earth commands.

The other section of EASEP, the LRRR experiment, is entirely passive and contains no electronics. The reflector unit, an array of 100 cyclindrical cavities, is a corner cut from a perfect cube of synthetic fused silica.

Presumably, scientists will be able to use laser beams to measure earth–moon distance with an uncertainty of 15 cm. Thus, using the moon as a reference point, they may be able to study the wobbling of the earth on its axis, or track continental drift.

As was previously mentioned, the PSEP began to send back data before the astronauts left the moon's surface, and although the package was supposedly damaged by the LM's blast-off, it has continued to transmit a steady stream of information. A number of seismic events have been recorded and these are the subject of intense study by scientists. At this writing, the laser experimenters haven't had much luck. On July 25, some moon laser flashes were believed to have been recorded, but they were so weak that scientists couldn't identify them with certainty. However, as movies and other data of the moon landing are studied, scientists say they will be able to pinpoint the location of the LM's descent stage exactly and thus aim their laser beams more accurately.

## In the future

ALSEP (Apollo Lunar Surface Experiments Package) will be carried by future moon ships. It will be more extensive than EASEP and will attempt to answer more questions. It will contain an active as well as a passive seismometer for detecting moonquakes and other lunar activity; a lunar surface magnetometer to measure the magnitude and direction of the surface magnetic field; an extensive solar wind experiment; and a suprathermal ion detector to measure flux, number density, velocity, and energy per unit charge of positive ions in the vicinity of the lunar surface. There is also a cold-cathode gauge experiment to provide data pertaining to the density of the lunar ambient atmosphere; a heat-flow experiment that will measure the lunar temperature profile at depths up to 3 meters; and a charged particle lunar environmental experiment that will study the energy distribution of proton and electron fluxes.

All of these experiments will be tied together by a sophisticated data-processing system; and as with EASEP, data and commands will be transmitted back and forth in digital form. However, ALSEP will not depend upon the sun for power. A SNAP 27 Radioisotope Thermoelectric Generator using plutonium-238 for fuel will produce 1500 watts of thermal energy, enough to run ALSEP for a year.

It's impossible to discuss all the electronic systems or components involved in the Apollo mission. They were all important: the ground radar systems that tracked the Saturn V vehicle as it lifted off the pad; the lunar landing radar built by Ryan and first used successfully in the Surveyor program; the rendezvous radar built by RCA Aerospace Division; the indicating meters supplied by Weston Instruments; the command service module communications developed by the Collins Radio Company; the MOS integrated circuits supplied by Philco-Ford Microelectronics division; the fuel gauges made by Simmonds Precision; all contributed to the mission's success. It is evident that electronics sophistication was shown by the products these companies and others supplied the Apollo program, but continued electronics advancement is necessary if the industry is to keep pace with fast-moving space requirements.

## The citizen's benefits

"If we can put a man on the moon, we certainly can. . ." has become a popular catch-all phrase. It doesn't matter if we're talking about fixing up commuter railroads, fighting poverty, abolishing ghettos, curing cancer, or whatever; almost everything is being compared to the space program. But perhaps this is more right than wrong. According to Dr. Wernher von Braun, Director of the Marshall Space Flight Center at Huntsville, Ala., "The real payoff (of the space program) does not lie in mining the moon, but in enriching our economy and our science in new methods, new procedures, new knowledge and advanced technology in general."

Every American was with Armstrong, Aldrin, and Collins when they lifted off from launch complex 39 on July 16; and every American was, justifiably so, proud when on July 20, Neil Armstrong made man's first step on the moon. But at the same time, it has been difficult for many Americans to see how the space program could improve their personal life.

The lists of space program "spinoffs" or "fallouts" goes on and on, but a few of the more important are:

Management techniques and systems, and teams of highly trained managers and scientists able to cope with the problems of urban development, mass transportation, etc., have resulted.

Computer technology took on new vigorous growth. Virtually every aspect of human endeavor has been enhanced by the commercial application of the digital computer. New computers, developed as a result of space programs needs, work at traffic control and industrial process control, run automated hospitals, and perform sophisticated medical diagnosis.

Law enforcement departments utilizing aerospace industry management techniques could produce in the near future an "instant cop"—a sophisticated, computerized system of information, communication, storage, and retrieval to speed up investigative processes and improve the quality of justice.

Indeed, the citizen is entitled to a return on his investment. With a little thought and a careful look around, he can spot some of the more obvious dividends. For others, he'll have to look harder, but the returns are there.

REFERENCE

1. Evanzia, W. J., "Automatic test equipment: a million dollar 'screwdriver,'" *Electronics*, Aug. 23, 1965.

# New product applications

## Scratch pad memory features fast switching speeds

The Intel 3101 is a 64-bit Random Access Memory. Its high speed makes it ideal in scratch pad applications. The use of Schottky barrier diode clamped transistors to obtain fast switching speeds results in higher performance than obtainable with equivalent devices made with a gold diffusion process.

The 3101 is packaged in a hermetically sealed 16-pin dual in-line package and is organized as a 16-word by 4-bit array. The storage flip-flops are addressed through an on chip 1 of 16 binary decoder using four input address leads. In addition to the four input address leads and the inhibit lead, there is a write input that allows data presented at the data leads to be entered at the addressed storage cells.

Since the 3101 is DTL and TTL logic compatible, wiring precautions are no more stringent than those for a DTL or TTL system. The accompanying circuits are examples of the 3101 used as a building block in constructing memory systems of different sizes. In each example, four 3101s are used although a similar approach may be used to expand the memory to other sizes.

Figure 1 shows a 16-word by 16-bit memory. All chip-select, write-enable, and address inputs are connected in parallel. A full 16-bit word is made up of four bits in each of the four packages.

A 64-word by 4-bit memory is shown in Fig. 2. It is made up of four 3101 memory devices. The word expansion is made possible by using the chip-select input as an additional address line. A 1 out of 4 decoder is used to drive the chip-select inputs. The outputs of each 3101 are OR-tied.

In a 32-word by 8-bit memory, shown in Fig. 3, two 3101 devices are used to increase the number of bits per word to eight. This is accomplished by connecting the chip-enable and address inputs in parallel. To increase the number of words to 32, two such parallel combinations are used.

More information on the 3101 is available from Intel Corp., 365 Middlefield Rd., Mountain View, Calif. 94040.

Circle No. 85 on Reader Service Card

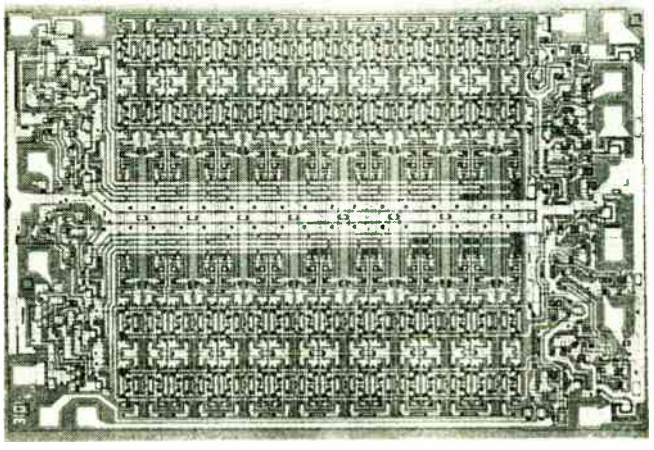


FIGURE 1. A 16-word by 16-bit memory.

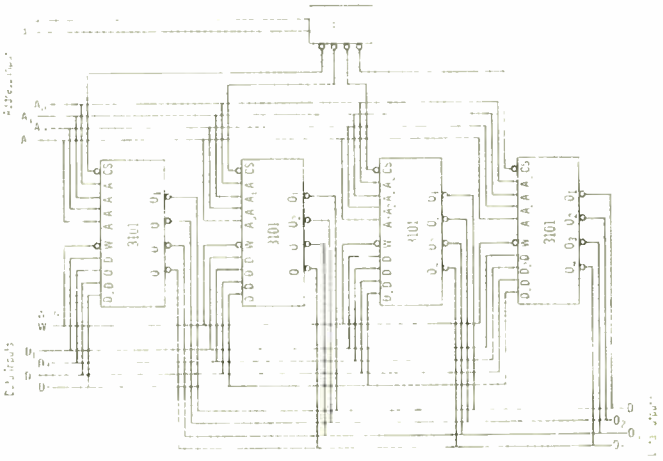


FIGURE 2. A 64-word by 4-bit memory.

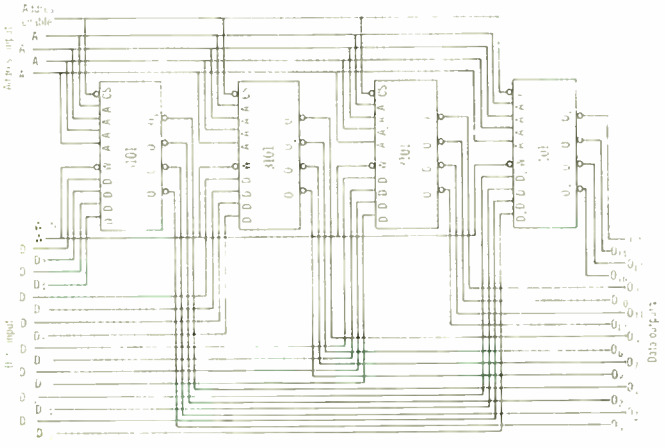
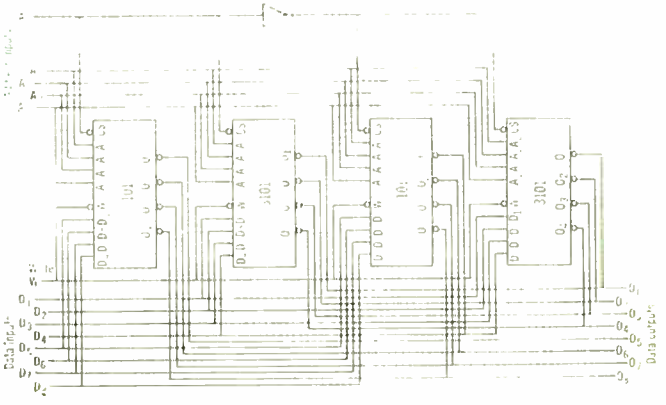


FIGURE 3. A 32-word by 8-bit memory.



# New product applications

## Digital acquisition reporting technique enables lowest possible drilling cost per foot

A new system — the Digital Acquisition Reporting Technique — when used in conjunction with computer time-sharing facilities, permits almost constant adjustment of drilling variables to a predetermined optimum combination that results in the lowest possible drilling cost per foot.

In the past, computer programs used to solve drilling programs have had two disadvantages. First, there has been lack of timely information to plug into the programs. Second, the scheduling of time to run analyses and the ability to get quick results has been difficult.

The new Martin-Decker system eliminates the need to use chart information to solve drilling problems. The punch-tape concept gives the driller almost immediate, useful information for maximum drilling efficiency — including the capability of correction when drilling conditions change.

The system components are shown in the accompanying illustrations.

Two features of the system are the recording of rotary table revolutions and cumulative depth information. Penetration per revolution is the best way to analyze drilling with the calculating speed of a computer.

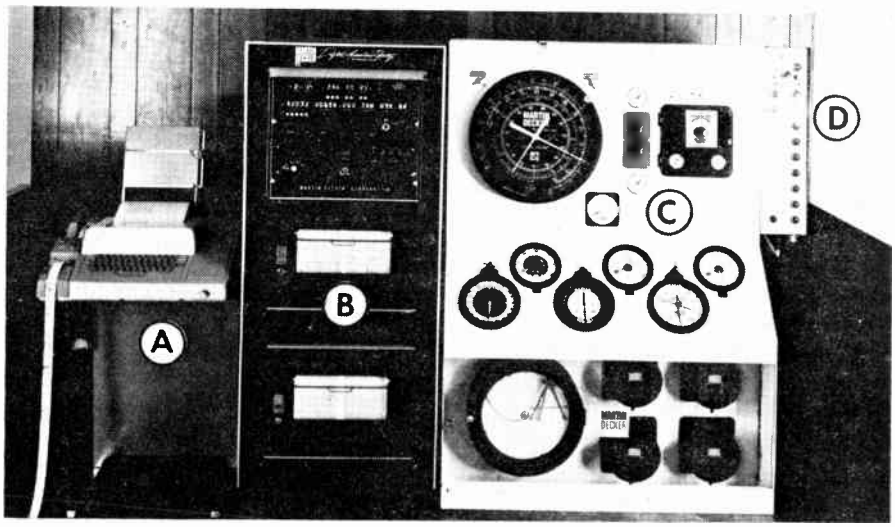
Other drilling variables presented in digital form include: weight on bit, pump pressure, rotary speed, and pump speed. On the panel, thumb wheels are used to set well number, bit-run number, mud weight, mud viscosity, and yield point. Another thumb wheel tells the recording system at what depth interval to record data.

The main objective of the new system is to optimize drilling variables to get the lowest cost per foot. A drilling optimization program uses the appropriate model well and the data contained in the punch-tape library to propose a schedule of bit weights and speed programs within predetermined arbitrary limits.

Other programs that can be used in connection with the system include: rig hydraulics optimization, drilling fluid analysis, rig-cost accumulation, directional survey analysis, and wireline cutoff determination.

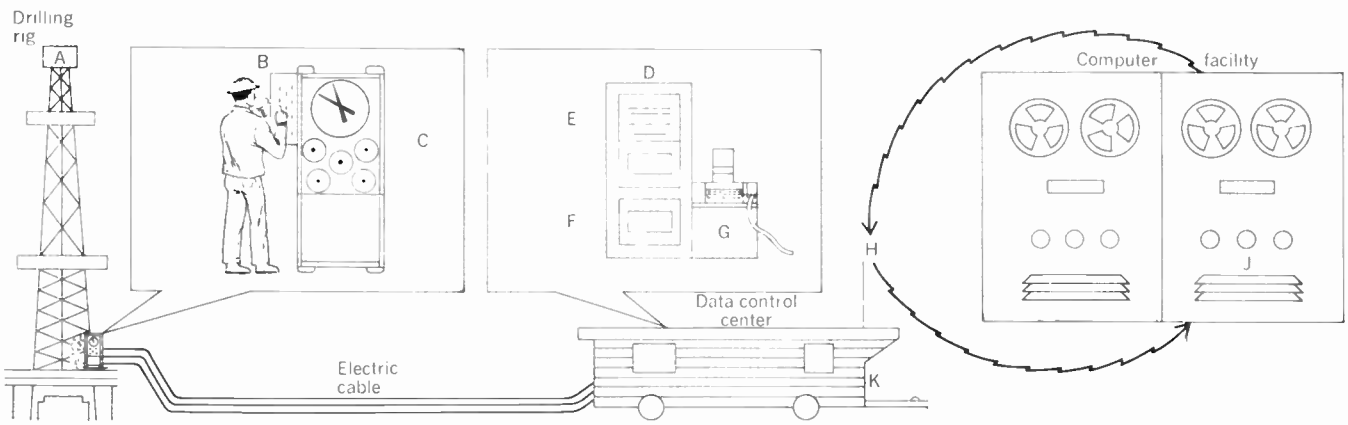
Additional information is available from Martin-Decker Corp., 1928 S. Grand Ave., Santa Ana, Calif. 92705.

Circle No. 86 on Reader Service Card




**COMPONENTS OF THE SYSTEM. A — Teletype with punched tape. B — Digital acquisition reporting technique console with digital logic and presentation component (top) and analog recorders (bottom). C — Driller's instrument panel with complete instrumentation. D — Driller's selection panel for retrieving drilling functions.**

**HOW THE DIGITAL ACQUISITION REPORTING TECHNIQUE WORKS**
**Driller on rig (A) selects drilling functions he wishes to retrieve by using selection panel (B) mounted on drilling control instrument panel (C). Transducers on drilling panel send information to data reporting system (D) where information is converted to digital readout (E). Analog recorders are at (F). Punch tape at (G) is used for Teletype transmittal (H) to a time-sharing computer (J). The best possible program is selected by the computer and transmitted by Teletype back to the data control center (K) for immediate use on the rig (A).**



# New product applications

## High-speed digital switch performs six separate, unrelated functions

Collectron Corporation's new "universal" rotary switch is a versatile device. And it has a sampling rate of up to 10 000 Hz, which exceeds the saturation time required by phototransistors and photodiodes and makes the switch adaptable to high-speed digital response systems.

Up to 50 output poles and 50 counts per pole may be used although the standard low-priced version contains three output poles and eight counts per pole. The switches may be ganged.

When the new switch is used as a sampling or selector switch, Fig. 1, information generated by several inputs is decoded and distributed to a recorder by the wiper pickoff circuit of the switch. The wiper arm is set at any one of the switch addresses by an appropriate stepping motor or hand knob/detent arrangement.

When the application is a telemetry switch, Fig. 2, the arrangement is identical to the sampling switch of Fig. 1 except for a transmitter–receiver

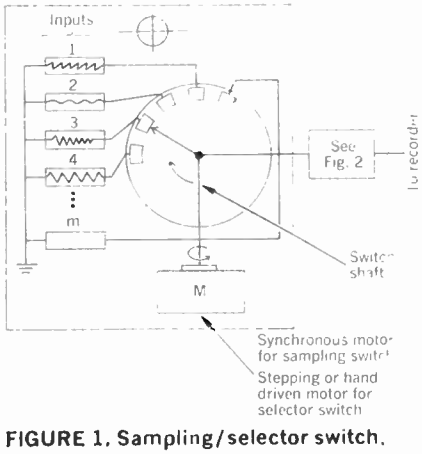
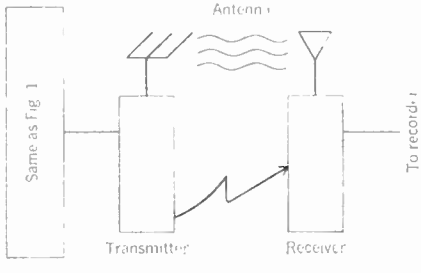



**FIGURE 1. Sampling/selector switch.**

**FIGURE 2. Telemetry switch.**




required to transmit through the dielectric interface. The telemetry switch has extensive use in radar, missile, satellite, airborne guidance, navigation, and fire control systems.

An incremental encoder is designed to deliver a train of pulses as it is rotated. A form of accumulation and storage circuitry is required to determine shaft angle at any given moment. Figure 3 illustrates how the universal switch, accommodated with lead and lag wipers, feeds this pulse train into two flip-flops (required to determine direction sensing) and, subsequently, into the accumulator counter. The output of the counter is read directly as shaft position. It is possible to obtain 3600 counts per revolution with a single gear pass.

Some of the many uses for the switch as an incremental encoder are in automatic weighing devices, military and commercial navigation and control systems, digital feedback on analog tape recorders, and in automatic conveyor systems.

When the switch is used as a tachometer, as shown in Fig. 4, it can measure angular velocity up to 16 000 r/min. The switch monitors rotational velocity directly with the output fed into an interval counter to display with 1/24-r/min accuracy the average speed of the device being measured.

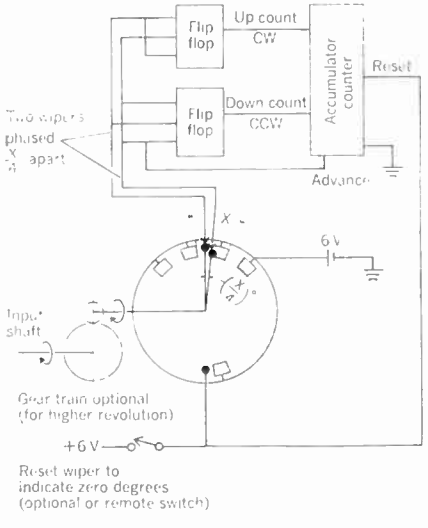
When the switch is used as a program switch, Fig. 5, it must be custom designed. The switch pattern is fabricated to each specific code pattern required. Units with up to 50 poles can be supplied with one or more inputs. Speeds encountered range from one rotation per day to 16 000 r/min.

When used as a pulse generator, Fig. 6, the switch exhibits a true square wave shape. There is a noticeable absence of any on-time or off-time noise perturbations and edge noise.

Additional details for specific applications are available from Collectron Corp., 304 E. 45 St., New York, N.Y. 10017.

Circle No. 87 on Reader Service Card

**FIGURE 3. Incremental encoder.**



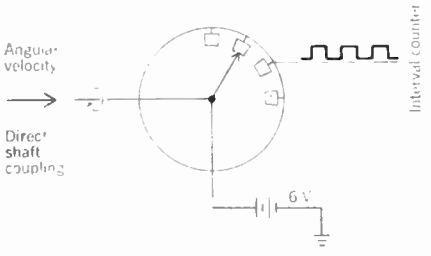




**FIGURE 4. Angular velocity measurement.**

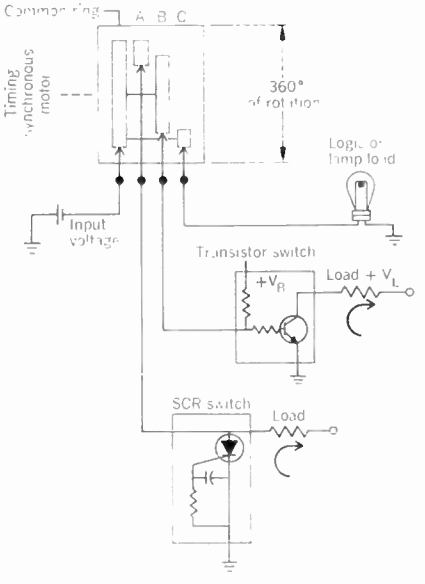
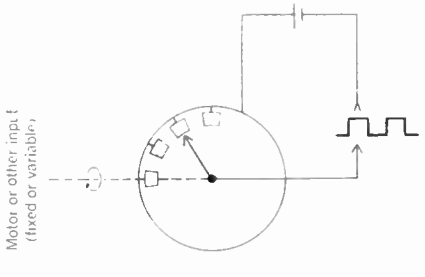



**FIGURE 5. Program switch.**

**FIGURE 6. Square-wave pulse generator.**



# New product applications

## Inexpensive instrument displays logic levels

The Logic Scope is a versatile digital instrument that displays static logic levels and detects and identifies pulses as narrow as 50 ns at repetition rates up to 10 MHz regardless of duty cycle or fall times. It is entirely self-contained and operates from 105 to 125 V, ac, at 50 to 60 Hz.

Quiescent voltage levels, as well as the positive and negative peaks of pulses, are read directly in volts from a calibrated threshold adjust dial on each of four independent channels. The threshold adjustment variable from −10 to +10 volts assures compatability with virtually any family of digital circuits.

The instrument is designed specifically to replace or complement oscilloscopes in field service, production tests, and general troubleshooting. It is well suited for use with digital computers, numerical control equipment, and desk calculators, as well as other digital equipment.

The use of only one control and one indicator per channel assures ease of operation even by unskilled personnel.

For a logic "1" condition, the lamp is ON and flashes OFF for 60 ms when a negative going pulse of 50 ns or greater is sensed. For a logic "0" condition, the lamp is OFF and flashes ON for 60 ms when a positive going pulse occurs. At low repetition rates (to 30 p/s), the lamp flashes ON and OFF in synchronism with the monitored pulses. At higher repetition rates (to 10 MHz), the lamp continues to flash rapidly but visibly.

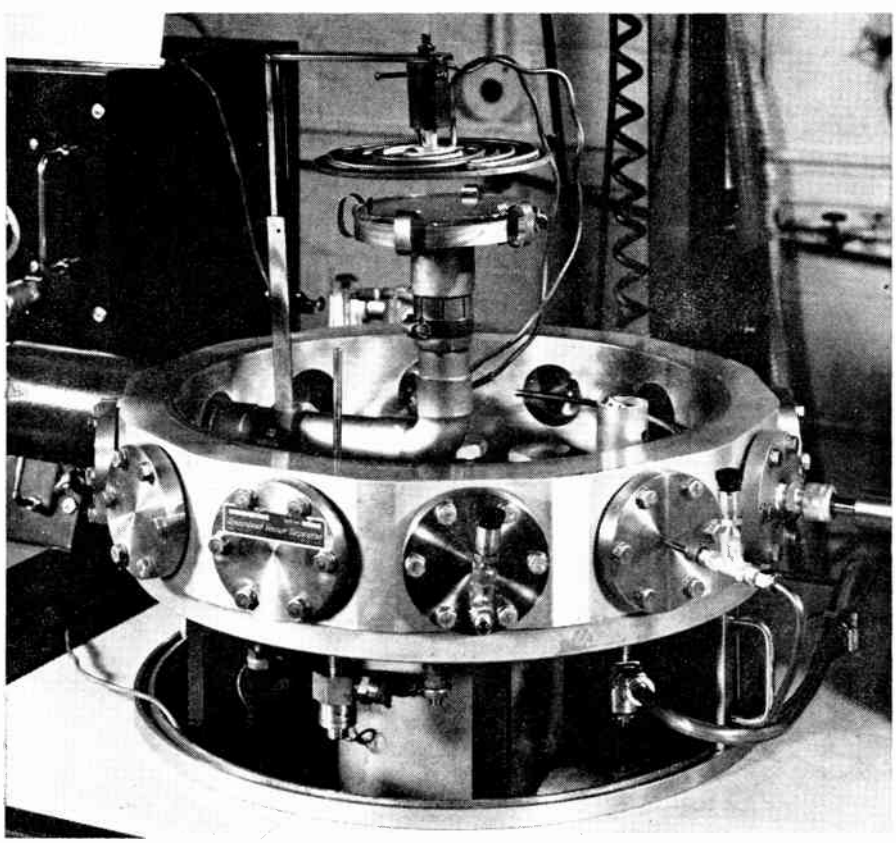
More details are available from Automated Control Technology, Inc., 3452 Kenneth Drive, P.O. Box 10501, Palo Alto, Calif. 94303

**Circle No. 88 on Reader Service Card**

## Low-cost diode sputtering system for limited batch production

A new Diode RF Sputtering System, PlasmaVac 350, features a low positioned target electrode resulting in a sputter-up configuration. The system is relatively inexpensive and highly versatile and can be used in limited batch production or pilot production operations as well as process development and feasibility studies in the laboratory.

One of the more interesting and promising applications for diode RF sputtering is in reactive deposition. This method differs from conventional sputtering in that deposition is done with argon plus some partial pressure of another gas that reacts or combines with the target material as it sputters. Reactive sputtering may be selected because the material that it is desired to deposit is not available as a target.



By using a variegated target with alternate areas of metal and dielectric materials, it is possible to deposit cermet films. Cermets are used for high sheet resistance films in thin-film resistor patterns because of the inherent instability of very thin metal resistors.

Since all material will sputter, the PlasmaVac 350 can be used to advantage for the removal of material in applications such as precise etching. Normally, chemical etchants are used to create circuit patterns defined by the photoresist technique. Several of the difficulties experienced with chemical etching, such as undercutting of the resist image, poor resolution, and the inability to remove inert materials like platinum and cermets, can be overcome by sputter etching.

More information is available from The Bendix Corp., Scientific Instruments & Equipment Div., 1775 Mt. Read Blvd., Rochester, N.Y. 14603.

**Circle No. 89 on Reader Service Card**

# New product applications

## Monolithic differential comparator


The new Model 351 Differential DC Comparator is a monolithic IC device designed for accurate sensing and measuring applications. It is capable of resolution in the fractional millivolt and submicroampere regions and will handle signal sources of $10^5$ ohms. It eliminates the auxilliary preamplifier, output booster, protection circuit, and special power supply regulator frequently required with earlier IC types.

The comparator is based on 5-ohm /cm substrate material and does not use the gold doping process. Most instrumentation applications demand accuracy rather than speed so that the enhanced input impedance, gain, current stability, common mode rejection, and ±15-volt supplies outweigh the penalty of decreased speed.

Typical applications include analog–digital converters, set-point controllers, zero-crossing detectors, and precision integrator resets. In such applications the comparator indicates when one current or voltage reaches within microvolts or microamps of another. The circuit distinguishes between two signal levels instead of providing only a GO/NO-GO indication of the presence or absence of a signal.

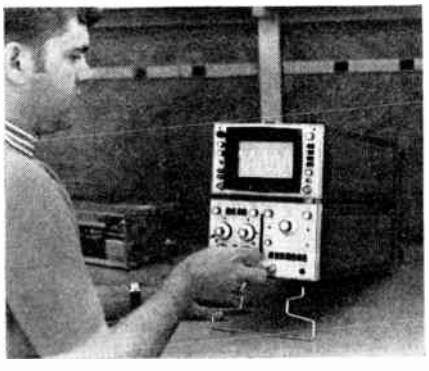
For example, a high-resolution analog–digital converter uses sensitive and stable comparators to discern when an internally generated reference voltage (or current) has been adjusted to within half a least significant bit of the unknown input. Half a least significant bit for a 14-bit converter handling 1-volt signal levels is only about 30 microvolts. But the instrument's accuracy and resolution rest on the comparator's ability to indicate when the reference signal has been trimmed to within 30 microvolts of the input.

Owing to its good common-mode performance, the new comparator can be operated "off the ground" in circuits requiring differential configuration at high levels of common-mode voltage, among which can be included various function generators such as Schmitts, one-shots, square wave, sawtooth, ramp, and other waveform sources.

More information is available from Analog Devices, Inc., 221 5th St., Cambridge, Mass. 02142

**Circle No. 90 on Reader Service Card**

## New oscilloscope displays low-level transients



Very fast pulses of not very great amplitude often elude visualization if they occur only now and then. A source of error-inducing transients in a digital system, for instance, is hard to track down without a scope that has both sensitivity and a real-time wide-band frequency response that is coupled with a fast CRT writing rate to trace transients as they occur.

These requirements also exist in electromagnetic interference evaluation of high-energy and other fast-pulse experiments when the engineer needs to know how much interference is coupled into ancillary equipment.

The demand for a sensitive, ultra-wide-band scope for these applications has been intensified by the growing need for viewing groups of short pulses that are repeated only occasionally, like the digital words in the newer computers using pulses only a few nanoseconds wide. Sampling scopes aren't applicable because the requirement for up to 1000 repetitions of the waveform to complete one scan prevents these instruments from responding quickly to changes in a slow rep-rate waveform. High-frequency real-time response is required.

The new Hewlett-Packard Model 183A Oscilloscope fills these needs. It has a dc to 250-MHz bandwidth and a sensitive 10-mV/cm minimum deflection factor.

The 183A's high-frequency real-time performance is also useful for displaying single short pulses that occur at low repetition rates, like those generated by laser beam detectors. And, in analysis of communications system performance, the 250-MHz response of the scope makes possible predetection display of modulation envelopes on RF carriers.

High-frequency performance, with corresponding fast rise time, is essential for photographing fast, single-shot transients, as in nuclear and high-energy experiments. The 183A has an internal flood gun that illuminates the entire phosphor display surface. This illumination sensitizes the phosphor, thereby increasing the photographed writing speed (to about 4 cm/ns). It also "fogs" photographic film slightly to increase effective film sensitivity. To simplify single-transient photography, the flood gun can be flashed in synchronism with the horizontal sweep, allowing the camera shutter to be left open for the event. The flood gun turns on only with the sweep.

More information is available from Hewlett-Packard Company, 1501 Page Mill Road, Palo Alto, Calif. 94304

**Circle No. 91 on Reader Service Card**

# New product applications

## Miniature magnetic field sensor installs easily

A new magnetic field sensor — a thin-film Hall generator and a hybrid circuit integral amplifier — requires only a single direct voltage input to produce an output at a usable level.

The magnetic field sensor can be installed in a variety of applications by simply connecting only three leads — power, ground, and output. The following applications are typical:

In elevator floor positioning, Fig. 1, the linear use of the sensor in the elevator circuitry provides information on both the floor at which the car is located and the car's relative position to the floor level. The output signal can be used to position the car properly. By sensing car position at the floor, cable stretch, floor differences, and other small variables can be compensated for without delicate adjustments.

In a brushless dc motor application, Fig. 2, two sensors are used linearly for detecting the rotating field of the permanently magnetized rotor of the motor. Their outputs are amplified by power amplifiers to drive *X* and *Y* field coils. The *X* sensor feeds the *Y* coil and the *Y* sensor feeds the *X* coil so that the magnetic field of the coil always leads the magnetic field of the rotor by 90 degrees to produce torque in the desired direction of rotation.

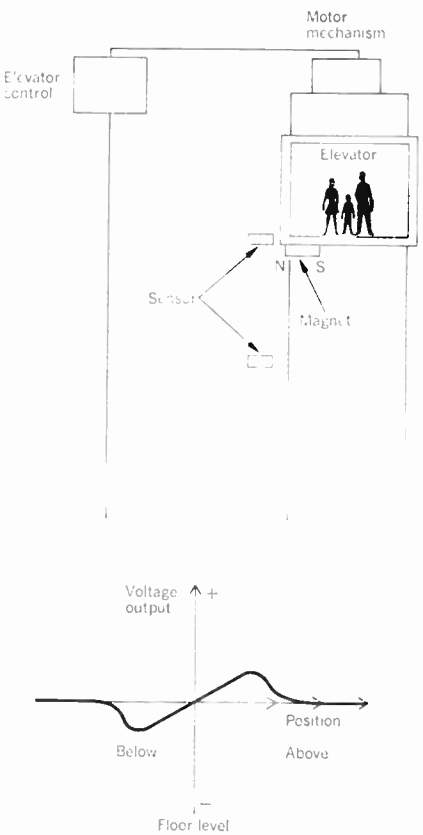
For limit switching applications, Fig. 3, a magnetic field sensor with a wide switching hysteresis centered around a zero magnetic field is used. As the cylinder shaft moves out, the magnet with the north pole moves over the sensor. The sensor output switches reverse the valve to cause the ram to move back. When the south pole magnet is detected, the valve is again triggered to move the ram forward. Stroke adjustment is made by changing magnet placement.

In an ignition trigger application, Fig. 4, a magnetic field sensor with a narrow switching hysteresis is used. Unlike a limit switching application, switching is not around the zero magnetic field. As the distributor shaft rotates, a north pole moves into a position adjacent to a sensor. The field level rises to the switch point, causing the ignition trigger to produce a pulse to the coil. As the shaft continues to turn the field will drop to below the trip point, readying the trigger for another pulse.

More information is available from F. W. Bell, Inc., 1356 Norton Ave., Columbus, Ohio.

**Circle No. 92 on Reader Service Card**

**FIGURE 1. Elevator floor positioning.**



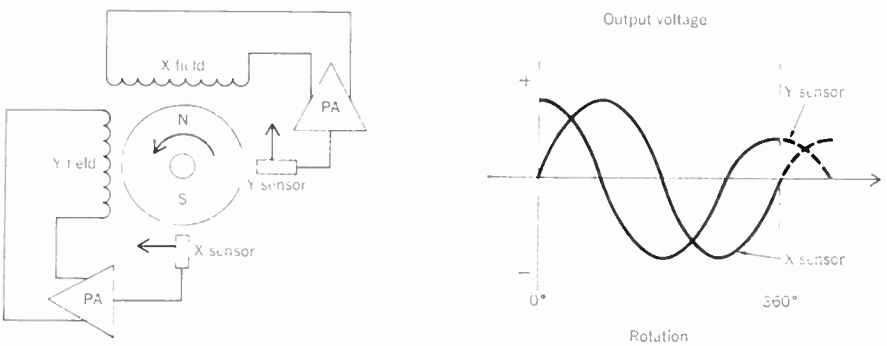




**FIGURE 2. Two sensors used to detect rotating field of dc motor rotor.**

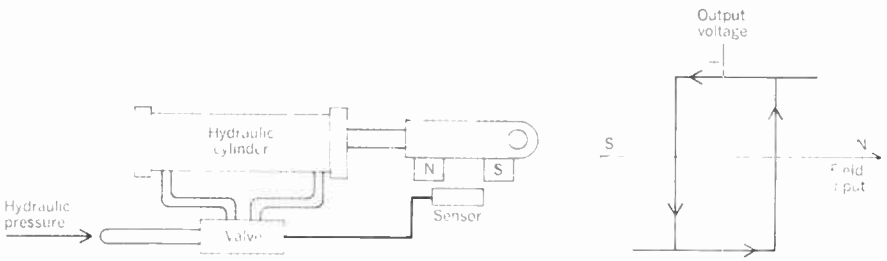



**FIGURE 3. Stroke adjustment of a hydraulic cylinder.**

**FIGURE 4. Control of ignition trigger pulses to an ignition coil.**