RCA Review

March 1973 Volume 34 No. 1
RCARCI 34(1) 1-214 (1973)

RCA Review, published quarterly in March, June, September and December by RCA Research and Engineering, RCA Corporation, Princeton, New Jersey 08540. Entered as second class matter July 3, 1950 under the Act of March 3, 1879. Second-class postage paid at Princeton, New Jersey, and at additional mailing offices. Effective Jan. 1, 1971, subscription rates as follows. United States and Canada: one year $6.00, two years $10.50, three years $13.50; in other countries, one year $6.40, two years $11.30, three years $14.70. Single copies (except for special issues) up to five years old $3.00.

RCA Review

A technical journal published quarterly by RCA Research and Engineering in cooperation with the subsidiaries and divisions of RCA.

Contents

March 1973 Volume 34 Number 1

An Experimental Read-Write Holographic Memory*

W. C. Stewart, R. S. Mezrich, L. S. Cosentino, E. M. Nagle,

F. S. Wendt, and R. D. Lohman

RCA Laboratories, Princeton, N. J.

Abstract—The design, development, and implementation of an experimental, partially populated, 10^6-bit read–write holographic memory with no moving parts and using state-of-the-art components is described. The memory employs an argon ion laser, acousto-optic beam deflectors, a holographic beam splitter, a nematic-liquid-crystal page composer, a photoconductor–thermoplastic erasable storage medium, and a silicon photodiode array. The analysis upon which the design is based is given. Detailed descriptions of the principal components are given, including the performance characteristics of the erasable holographic storage medium.

1. Introduction

This paper describes the design, development, and implementation of an experimental, partially populated, 10^6-bit read–write holographic memory system with no moving parts. Various proposals have been made for realizing read-only memories[1,2] and read–write memories[3,4] for computer applications that utilize high-density hologram storage. Limitations on storage density imposed by component size[5-7] and light scattering[8] in the storage medium have been considered. The purpose

* This work was partially supported by the National Aeronautics and Space Administration under contract NAS8-26808.

of the research reported here is to demonstrate that the concepts underlying the proposed optical systems are compatible and can be realized in practice.

The functions to be performed by the page composer and the erasable storage medium present many formidable problems. Although many materials and devices have been suggested, practical light-valve arrays and storage media that are capable of operating satisfactorily in submillisecond cycle times with existing lasers do not presently exist. Therefore a complete system without moving parts, capable of operating at realistic speeds as a computer memory, is beyond the present state of the art. The alternative approach taken here is to ignore the speed requirement for the moment, and develop an operating breadboard system with presently realizable components. With the exception of laser power, the design of the system optics is essentially independent of the speed requirement. As initially conceived, the prototype memory employed a Q-switched ruby laser, a two-dimensional acousto-optic deflector, a liquid-crystal page composer, a manganese–bismuth film for erasable magnetic holograms, and a silicon-photodiode detector array. Experimental difficulties with the output stability of the ruby laser led to the substitution of an argon ion laser and a photoconductor–thermoplastic storage medium with greater sensitivity and efficiency. The change to a shorter wavelength, along with a decrease in the guard space required around the storage holograms as compared with MnBi, gives a substantial storage density improvement over the original design.

2. Description

The organization specified for the memory system required that there be space in the storage medium for at least 1024 holograms, each of which can store 1024 bits (one page) of information. To provide an extra margin of immunity to image fluctuations from a variety of possible sources, each bit of information in the image of a given page is associated with two diagonally adjacent image spots, A and B. The condition A on and B off represents a binary "0", A off and B on represents a binary "1"; the two remaining combinations where A and B are simultaneously off or on are not used as inputs and represent an error condition if detected.

The memory system is partially populated in a manner that permits a realistic assessment of the performance of a fully populated system. The page composer is fully populated, so that each page written into and read from a hologram contains the full 2048 image spots representing 1024 bits in a page. The deflection system has the full resolu-

tion for addressing any one of 1024 holograms. Only 48 of the possible hologram locations are active—16 in locations near the center of the array and 32 around the periphery. There are five separated groups containing four nearest neighbors, so that full density in local areas at the center and boundaries is included. The photodetector array contains 20 diodes for sensing 10 of the 1024 bits in each image. These are divided among an adjacent bit pair at the center of the image and four bit pairs around the periphery of the image.

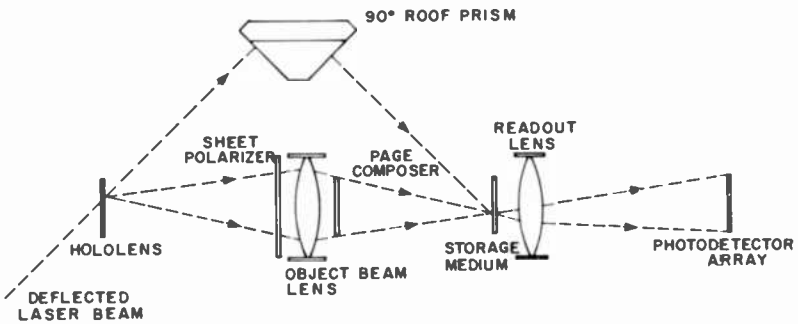


Fig. 1—Holographic optical system.

Other criteria are that the information packing density be reasonably high ($\geq 10^3$ bits/mm^2), that light utilization be efficient, and that the basic design be capable of extension to larger sizes and higher capacities. Fig. 1 is a top-view sketch of the holographic optical system layout selected for the system.

The light input to the system is a small collimated and polarized laser beam that intercepts the hololens at a fixed angle, 45° here. The point of intersection is selected by a deflection system, not shown, that is capable of displacing the beam parallel to itself both horizontally and vertically in fixed increments. Each possible position of the beam will ultimately result in the illumination of a single hologram location on the erasable storage medium. The hololens consists of an area array of permanently prerecorded holograms, one for each page of memory capacity. Light diffracted by the hololens, when polarized parallel to the transmission axis of the sheet polarizer, passes through the object beam lens to illuminate the light valves of the page composer. The object beam lens projects an image of the illuminated hololens area onto the storage medium. Thus, the object beam light from the page composer strikes the storage medium only at the location

where a hologram is to be written. The page composer is a two-dimensional array of light valves. The transmission of each light valve is controlled by electrical signals that compose the page of information to be recorded in a hologram on the storage medium. Since the active area of the page composer containing the light valve array is the aperture stop for the object beam lens, the array has an approximately circular boundary in order to minimize the required lens diameter.

The zero-order component of the collimated input light beam, which is not diffracted by the hololens but is transmitted without deviation. is folded by a 90° roof prism and arrives at the storage medium at a horizontal offset angle of 45°. This beam is used as the reference wave for recording holograms on the storage medium. Two reflections inside the prism produce inversion of the horizontal and vertical positions of the reference beam that match the inverted image of the hololens produced by the object beam lens. Thus, object and reference beams are coincident on the storage medium for any point of incidence of the input beam on the hololens.

Storage medium holograms are read out by rotating the plane of polarization of the input beam by 90°. Diffracted light from the hololens is blocked by the sheet polarizer, so that only the reference wave illuminates the selected hologram on the storage medium. The reconstructed wavefront represents a virtual image of the page composer in the state of transmission that existed during the recording of the hologram. This image is projected at unity magnification onto the photodetector array by the readout lens; it is diffraction limited by the small aperture of the single hologram that is read out at any given time. Since there can be no vignetting of the image rays from any of the holograms arrayed on the storage medium, the total active storage area forms the aperture stop for the readout lens. The boundary of this area is also made approximately circular. The photodetector array contains photosensitive diodes at locations where the image of a page composer light valve can appear. The information output of the system consists of the resulting electrical signals.

In summary, the principal advantages of the present optical configuration are as follows:

[a] Storage medium holograms of high packing density may be recorded in the Fourier transform plane of the page composer.

[b] Use of the hololens provides deflection for the object beam with efficient light utilization for writing.

[c] Locating the page composer and erasable storage medium adjacent to the imaging lenses minimizes the lens diameters for given focal lengths and memory capacity.

[d] Use of the same reference-beam incidence angles for both writing and reading permits alignment of the image on the photodetector array without recording holograms.

The principal disadvantages are:

[a] The object illumination must be blocked during readout.

[b] Different increments of horizontal and vertical displacement of the deflected input beams arriving obliquely at the hololens must be provided.

3. Design Analysis

A specification of parameters such as size or spacing for one component in the optical system usually dictates requirements for other components. In this section, some of the important relationships among the components are analyzed. Simplifying approximations and empirical measurements are used freely. The purpose is to provide a semiquantitative basis for evaluating trade-offs in arriving at a final design.

3.1 Deflector Resolution

The number of resolvable positions N in one dimension that is obtained from an acousto-optic deflector cell is given by

$$N = \Delta\theta/\Delta\phi, \qquad [1]$$

where $\Delta\theta$ is the angular range over which the incident laser beam can be deflected (assuming small deflection angles), and $\Delta\phi$ is the incremental angular spacing between the central rays of two adjacent deflected beams. The far-field diffraction pattern, appearing in the back focal plane of a lens, contains the spatially resolved light spots. The angular range depends on the frequency range of the acoustic waves, the light wavelength λ, and other constants. The allowable angular spacing $\Delta\phi$ depends on the amount of diffraction spreading of the light due to the finite diameter D of the cell's optical aperture, and the amount of overlapping of spots that can be tolerated. Since the diffraction spread is proportional to λ/D, let

$$\Delta\phi = \alpha\lambda/D, \qquad [2]$$

where the constant of proportionality α is chosen according to the desired resolution criterion. For a uniformly illuminated aperture and an effective Rayleigh resolution criterion in which spots are considered resolved when the maximum of one spot falls on the first minimum of

the neighboring spot, $\alpha = 1.22$ and

$$N(\text{Rayleigh}) = \frac{\Delta\theta D}{1.22\,\lambda}. \qquad [3]$$

Since the input laser beam has an inherent Gaussian intensity profile through its cross section, only 37% of the light power can be utilized in forming a perfectly uniform beam. If a sufficiently small-diameter Gaussian beam is used, such that substantially no light is blocked by the cell aperture, the increased diffraction spreading reduces the number of spots resolved according to a given criterion and for a fixed aperture size. A useful compromise for optimizing light throughput and resolution is to expand the laser beam so that the cell aperture truncates the laser beam at its $1/e^2$ intensity radius. This allows 86% of the laser beam through the aperture, and increases the diffraction spreading to approximately 1.4 times the value obtained with no truncation of the same size Gaussian beam. Let $\Delta\beta$ represent the angular spread over which the output beam drops from its central maximum to $1/e$ of this intensity value. It can be shown[9] that

$$\Delta\beta = 0.672\,\frac{\lambda}{D}. \qquad [4]$$

For a resolution criterion that specifies that the separation between centers of neighboring spots be k times $\Delta\beta$,

$$\Delta\phi = 0.672\,k\,\frac{\lambda}{D} \qquad [5]$$

and

$$\alpha(k) = 0.672\,k. \qquad [6]$$

Therefore, for a given $\Delta\theta$ and D, the number of resolvable spots N_k according to the above criterion can be expressed in terms of the number resolved by the Rayleigh criterion:

$$N_k = N(\text{Rayleigh}) \times 1.82/k. \qquad [7]$$

It can be seen that as k increases, the Rayleigh resolution of the deflector must increase in order to obtain a given number of spots resolved by the k criterion.

3.2 Reference-Beam Spreading

A major influence on image contrast arises from the partial illumination of neighboring holograms on the storage medium during readout of a particular addressed hologram. The reference beam (and the object beam) must have sufficient extent to properly expose the storage medium during writing to produce a hologram of the desired diameter. During readout, when the reference beam is centered on the addressed hologram, the total light flux onto the neighboring holograms must be significantly less than on the selected hologram. Illumination of the deflector aperture with a Gaussian beam profile that is truncated at the $1/e^2$ intensity radius results in a deflected beam profile that is very nearly Gaussian in the main lobe. This profile is therefore approximated by the expression

$$I(r) = I_o \exp\left[-(r/r_o)^2\right], \qquad [8]$$

where r is the radial coordinate in the storage medium plane, centered at the addressed hologram, and r_o is the radius at which the intensity falls to $1/e$ times the central intensity I_o.

Assume that the efficiency with which the image of each data spot is reconstructed at the photoconductor is the same for each hologram in the array. The images from neighboring holograms that are partially illuminated are derived from the same coherent wavefronts of the reference beam that produce the desired image. The images from the neighboring holograms arrive at the photodetectors at slightly different angles, however, and produce spatial interference patterns whose periods are smaller than each photodetector. The result is that it is necessary to add the respective intensities of the illumination from neighboring holograms, rather than the amplitudes, because of the spatial integration that occurs over the area of each photodetector. The question of interest is thus reduced to consideration of the light flux on the addressed hologram compared with the total from all the neighbors.

Define F_a as the fraction of the total light flux in the reference beam that falls within the boundary of the addressed hologram. Similarly, F_n is the total light flux on neighboring holograms, expressed as a fraction of the total incident flux. The minimum contrast C_r in the image due to readout beam spreading is

$$C_r = F_a/F_n. \qquad [9]$$

This ratio gives a worst-case estimate for the effect of reference-beam

spreading, since it assumes that bits in the neighboring holograms are binary complements of the corresponding bits in the addressed hologram. For square holograms of width h in a two-dimensional square array with center-to-center spacing s_h, define the one-dimensional packing factor η_h as

$$\eta_h = h/s_h. \qquad [10]$$

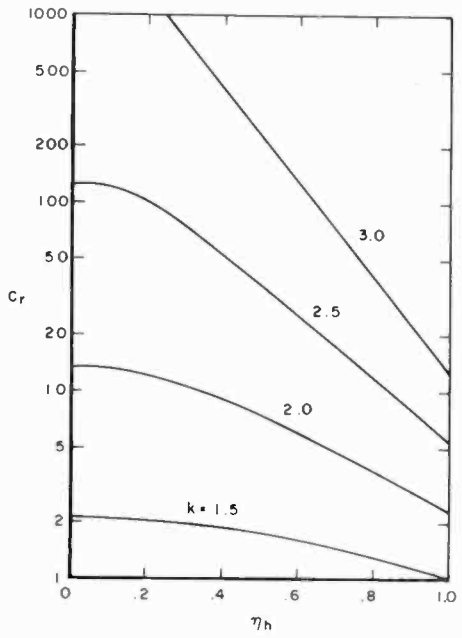


Fig. 2—Worst-case image contrast due to reference-beam spreading as a function of hologram size and spacing.

The resolution factor k is given by

$$k = s_h/r_o. \qquad [11]$$

Fig. 2 is a graph of the results of a numerical computation of the dependence of C_r on η_h for various values of k. Fig. 3 plots the dependence of F_a on these parameters. Whereas F_a increases as both η_h and k increase, C_r increases with decreasing η_h and increasing k. Using circular holograms of diameter h is equivalent to reducing η_h by a small amount in these calculations.

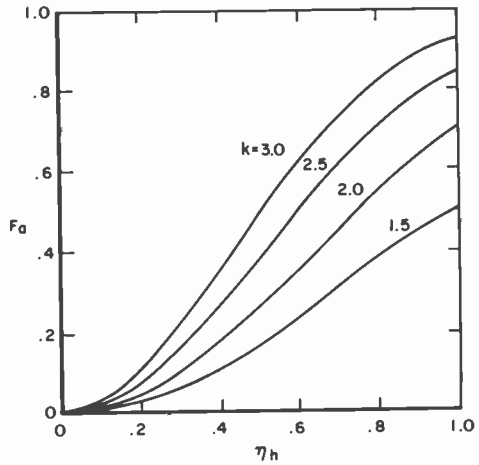


Fig. 3—Fraction of reference-beam light flux on addressed hologram as a function of hologram size and spacing.

3.3 Image Resolution

Consider the diffraction spreading in the image caused by a finite hologram diameter h. The page composer is illuminated by coherent light of wavelength λ that is converging to a point in the Fourier transform plane at an axial distance z from the page composer. Page-composer light valves of diameter p form a periodic array with center-to-center spacings of s_p. Over 80% of the light transmitted by the page composer appears in a circle of diameter d_A in the Fourier plane; this is the size of the central disk in the Airy diffraction pattern of one light valve.

$$d_A = 2.44\,\lambda z/p. \qquad [12]$$

Experimental measurements were used to determine the dependence of diffraction-limited image contrast on the quantities

$$a = h/d_A, \qquad [13]$$

and

$$\eta_p = p/s_p. \qquad [14]$$

The data were obtained using a standard Fourier filtering set-up in which a metal mask with periodic apertures, simulating the page

composer light-valve elements, was placed in the input plane. A circular aperture representing the storage hologram was placed in the Fourier transform plane to transmit only the wavefronts that could be reconstructed from a hologram of the same diameter. The image plane contained a photodetector having an entrance aperture equal to the size of one page-composer element. Measurements were made of the light flux collected by the photodetector in the image of one opaque page-composer element surrounded by all transparent elements, and the light flux in the image of an adjacent transparent page-composer element. The ratio of light in the image of a transparent element to that in the opaque element is called the diffraction-limited contrast C_d. Ratios as high as 4×10^4 were measured without serious error from flare light. The results may be extended to systems with different focal lengths, magnifications, and element sizes by maintaining the same scaling ratio a between hologram aperture diameter and the diameter of the central Airy disk in the Fourier transform (far-field diffraction pattern) of a single page composer element. Some of the experimental data are listed in Table 1.

Table 1—Experimentally Measured Values of Diffraction-Limited Contrast C_d as a Function of a and η_p

	a							
η_p	1.7	1.5	1.4	1.2	1.0	0.8	0.65	0.5
0.40	10,700	3980	1430	1070	15,000	2640	415	111
0.50	2210	3240	872	2440	2600	1730	420	2880
0.67	711	600	814	425	1920	179	260	2.65
0.75	—	—	—	206	1030	100	4.7	—

The fluctuation of C_d as a varies is a real effect caused by the abrupt boundary of the hologram aperture in the experiments. Note that C_d is greater than 10^2 over a $\pm 20\%$ variation in hologram diameter around the value $a = 1$ for all the experimental values of η_p.

An additional insensitivity to the effects of diffraction spreading is built into the geometrical arrangement of image points that constitute an information bit. Diagonally adjacent pairs of detectors represent a single information bit; the pairing of neighboring spots into bits is such that the spillover light from any valid combination of information states in the neighboring spots falls equally on the two detectors. The desired information forms a difference-mode signal, while the noise due to diffraction spreading becomes common mode. The effective contrast from diffraction spreading becomes the product of C_d as measured for single spot detection and the common-mode rejection ratio of the detector electronics. Defects that alter the effective

hologram aperture into a nonsymmetric shape, however, will tend to degrade the contrast in both factors.

3.4 Other Considerations

Intermodulation distortion in Fourier-transform hologram images of a periodic object, such as the page composer, reduces contrast by producing secondary images of an illuminated spot at the primary image locations of other page-composer elements. The effect occurs when the reconstructed wave amplitude is not linearly proportional to the object-wave amplitude used in recording the hologram. Spatial fluctuations of the object beam amplitude that are sufficiently large to exceed the linear dynamic range of the recording medium produce a recorded nonlinearity. Pseudo-random phase plates that uniformly shift the phase of the optical wavefront from each page-composer element greatly reduce the spatial fluctuations.[10,11] An alternative is to use diffuse illumination of the page composer at the expense of introducing speckle noise in the image from a high-density hologram. Nonlinear distortion is also inherent in the reconstruction of linearly recorded phase holograms of high efficiency; however, this effect is essentially negligible for peak diffraction efficiencies less than approximately 10%.

Another aspect of image contrast to be considered is the extinction ratio required of the sheet polarizer used to block the object illumination during readout. The sheet is aligned to transmit the vertically polarized object illumination during writing. The intensity transmittance is m_1 and transmittance of the horizontally polarized object illumination during readout is m_2, where $m_2 \ll m_1$. If the light flux from one page-composer bit is F_o at the hologram during writing, then the flux into the image of a bit due to unextinguished direct object illumination is $F_o \tau m_2/m_1$, where τ is the intensity transmittance of the storage medium. The light flux in the image of a bit reconstructed from the stored hologram is the product of the reference beam flux F_a, and the efficiency E per bit. Define the beam balance ratio K as

$$K = F_a/F_o, \qquad [15]$$

and note that the reference beam flux at the hologram is the same for both reading and writing. The contrast C_p is the ratio of the flux in the readout image of a bit to the flux from unextinguished direct object illumination.

$$C_p = \frac{KEm_1}{\tau m_2}. \qquad [16]$$

3.5 Storage Capacity

This portion of the analysis defines the influence of the design parameters on storage capacity. Since the page-composer elements and the storage-medium holograms each form an array whose unit cell is a square,

$$A_p = N_p(s_p)^2 = N_p(p/\eta_p)^2, \qquad [17]$$

$$A_s = N_s(s_h)^2 = N_s(h/\eta_h)^2. \qquad [18]$$

where p — diameter of one page-composer element

h — diameter of one hologram in the storage medium

s_p — center-to-center spacing of page-composer elements

s_h — center-to-center spacing of holograms

η_p — spatial one-dimensional duty factor of the page composer $\equiv p/s_p$

η_h — spatial one-dimensional duty factor of the hologram array $\equiv h/s_h$

The hologram diameter is a times the diameter of the Airy disk of a page-composer element,

$$h = 2.44a\lambda z/p, \qquad [19]$$

where λ — wavelength of light used for writing and reading

z — axial distance from page composer to storage medium

a — ratio of hologram diameter h to the diameter of the central Airy disk in the Fourier transform of one page-composer element

Combining Eqs. [17], [18], and [19] gives

$$A_pA_s = N_pN_s\left(2.44\frac{a\lambda z}{\eta_p\eta_h}\right)^2. \qquad [20]$$

Choosing the boundaries around the page-composer elements and the storage-medium hologram array to be octagonal rather than square gives

$$A_p = P^2/\sqrt{2}, \qquad [21]$$

and

$$A_s = S^2/\sqrt{2}. \qquad [22]$$

Combining Eqs. [21] and [22] with [20] gives

$$P^2S^2 = 2N_pN_s\left(2.44\frac{a\lambda z}{\eta_p\eta_h}\right)^2. \qquad [23]$$

3.6 Specifications

To fix the design of the optical system requires numerical specification of eight quantities: hologram diameter and spacing, light valve diameter and spacing, wavelength, axial separation between page composer and storage medium, resolution criterion for deflector, and photodetector diode diameter. To maximize the readout light, the diameter of the photodetector is chosen to equal the diameter of a page-composer light valve. In order to obtain identical specifications for focal length and aperture of the object-beam lens and readout lens, the page-composer diameter P is set equal to the storage-medium diameter S. This choice also minimizes the range of angles over which the object beam is incident on the storage medium. The 1024 holograms in the storage medium are arranged as a 36×36 array with the corners truncated to give an octagonal boundary, requiring a deflector resolution of 36 positions in each dimension instead of 32. The remaining parameter values were chosen to give reasonable trade-offs between image contrast on the one hand, and practical size and cost on the other hand.

Table 2 lists the resulting specifications. The initial design was for a ruby-laser system incorporating an MnBi storage medium. The specific Curie-point writing characteristics of the MnBi limit the information recording density and are an important consideration in this design. A calculated worst-case image contrast of approximately 5:1 results in equal parts from reference-beam spreading and from the 100:1 polarization extinction ratio in the object beam.

The final design resulted from a decision to change to an argon laser and a photoconductor–thermoplastic storage medium. Parameter values were chosen, for practical reasons, to minimize the changes in components that were already specified in the initial design; the deflector resolution, page-composer dimensions, and lens focal lengths

were left unchanged. The reduction in required hologram size results from the decrease in wavelength. Resistance heaters one millimeter square, which provide *in situ* development of the thermoplastic storage holograms, establish the maximum allowable size for a hologram. Since the storage-medium diameter is smaller than the readout-lens aperture, and since h is smaller than the size permitted by the heater geometry, the final design is not optimum. Reducing p and increasing N_p to make $h = 1$ mm would give an eight-fold increase in memory capacity with the present optical system. This would result in a reduction of C_r from 12 to 6.

Table 2—Design and Performance Specifications

Parameter	Initial Design	Final Design
λ (nm)	694.3	488.0
a	1	1
k	2	2
z (mm)	356	356
z/P†	5	5
h (mm)	0.62	0.44
s_h (mm)	1.9	1.6
η_h	0.33	0.27
S (mm)	71	60
p (mm)	0.97	0.97
s_p (mm)	1.3	1.3
η_p	0.75	0.75
P (mm)	71	71
N (Rayleigh)	40	40
EN_p	10^{-4}	10^{-2}
K	10^2	10
τ	0.1	.3
m_1/m_2	100	100
C_d	10^3	10^3
C_r	10	12
C_p	10	33

† Effective f/number of the imaging lens system.

4. Laser-Beam Deflection

4.1 Deflector Cells

The holographic optical system design requires that the deflector cells resolve 36 positions with the center of an adjacent spot falling at twice the e^{-1} intensity radius of an addressed spot. The cells are illuminated with an expanded and collimated laser beam that is truncated at the e^{-2} intensity radius by the cell aperture. These requirements are

equivalent to specifying a cell capable of resolving 40 positions according to the usual Rayleigh criterion for uniform incident illumination. Two lead molybdate deflector cells having a specified maximum Rayleigh resolution of 42 positions, and associated drivers, were used. Other characteristics are (a) bandwidth of 29 MHz centered at 91 MHz, (b) angular deflection range across bandwidth of 4 milliradians, (c) diffraction efficiency per cell across bandwidth of 60%, (d) access time of 1.75 μsec, (e) optical aperture of 6.3 mm, and (f) spot position accurate to $\frac{1}{8}$-spot separation at 20% operating duty cycle.

4.2 Deflector Telescope

The laser beam, which may be deflected over an angular range $\Delta\theta$ of 4×10^{-3} radian by the deflector cell, is to be converted by a lens system into small collimated beams, parallel to the optical axis, that can span the 55-mm width of the active areas of the hololens and storage medium. Placement of the deflector cell in the front focal plane of a lens system with an effective focal length f' provides, in the back focal plane a span y given by

$$y = f'\Delta\theta \qquad [24]$$

for small values of angular range. This requires f' to be 13.7 m. With the given deflector cell aperture and this focal length, the diameter of the output waist changes less than 5% over an axial distance of 2 m centered on the output plane. Thus, the deflected beams are parallel and effectively collimated at the hololens plane and the storage medium plane, which are separated by an axial distance of 1 m.

Shifting the deflector cell axially away from the front focal plane of the lens affects the parallelism of the deflected beams arriving at the back focal plane; the deviation from normal incidence of the arriving beams is greatest at the extremes of the angular range, and is zero at the center of the angular range. It can be shown for an axial translation d of the deflector cell from the front focal plane, that the beam leaving the deflector cell at an angle ϕ measured from the center of the angular deflection range will arrive at an angle δ

$$\delta = d\phi/f' \qquad [25]$$

in the back focal plane, but at the same position in the plane as when d is zero. Our requirement is that a deflected beam arriving at a point on the hololens arrives at the equivalent point obtained by parallel projection onto the storage medium with an error of no more than

one-fourth the diameter of one hologram. Using the design values of 1-mm hologram diameter and 1-m axial distance between hololens and storage medium, the maximum allowable angular deviation δ_{max} from parallelism is

$$\delta_{max} = 0.25 \times 10^{-3} \text{ radian.} \quad [26]$$

Substituting this value along with $\phi = 2 \times 10^{-3}$ radian and the value of f' gives the maximum allowable axial displacement d_{max} of the deflector cell from the front focal plane of the lens:

$$d_{max} = \delta_{max} f'/\phi = 1.7 \text{ m.} \quad [27]$$

This comfortably large tolerance results primarily from the small range of deflection angles involved.

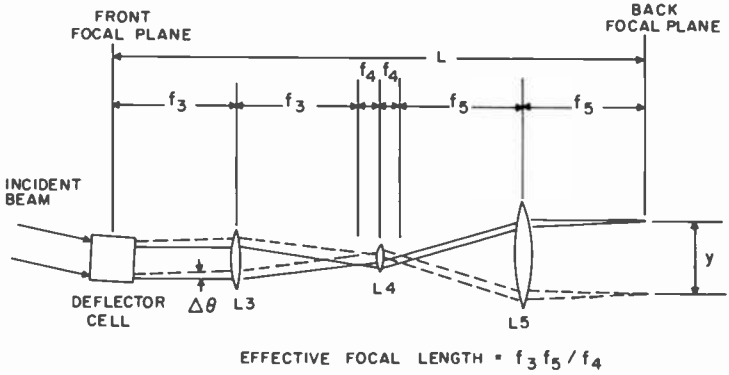


Fig. 4—Compact lens system.

The system of three lenses spaced as shown in Fig. 4 is equivalent to a single lens having an effective focal length of $f_3 f_5/f_4$; the distance between front and back focal planes is $2(f_1 + f_2 + f_3)$. For a given effective focal length, this distance is minimized by choosing $f_3 = f_5$, and making f_4 as small as possible. In the back focal plane of lens 3 is the far-field diffraction pattern of the deflector cell as previously discussed. The size of the pattern is too small to be used directly, since f_3 alone is much smaller than the required effective focal length. The back focal plane of lens 4 contains an image of the deflector cell with the transverse dimensions magnified by the ratio f_4/f_3 (much less than unity), but with all beam angles magnified by f_3/f_4. Lens 5

produces in its back focal plane the far-field diffraction patterns of the image located in its front focal plane. Equivalently, the pattern at the back focal plane of lens 5 is an image of the diffraction pattern in the back focal plane of lens 3, magnified in transverse dimensions by the ratio f_5/f_4 (greater than unity), and with beam angles magnified by f_4/f_5. This system has the same tolerance to axial displacements from the input and output planes as calculated for the single lens to which it is equivalent.

The memory system design requires that the hololens and storage medium planes be inclined at an angle of 45° in the horizontal plane to the parallel beams arriving from the deflection system. The total lateral span, x, for the x-axis deflector, and the effective focal length of the optical system in the horizontal dimension must be smaller than the values for the vertical dimension by the factor cos 45°. The resolution of the x-deflector must be the same as that of the y-deflector. however. The effective focal length f_h of the lens system for horizontal deflection from Eq. [24] is

$$x = y \cos (45°) = f_h \Delta\theta, \qquad [28]$$

or

$$f_h = 9.7 \text{ m}. \qquad [29]$$

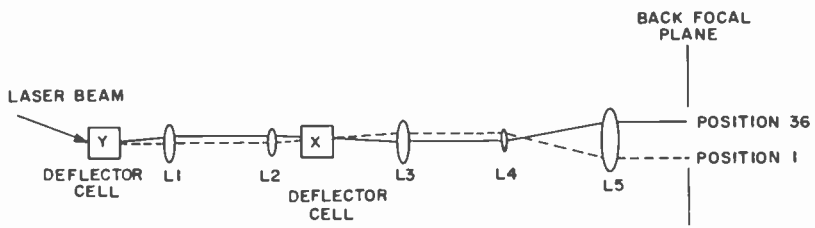


Fig. 5—Deflector lens system.

A lens configuration that satisfies the requirements is sketched in Fig. 5. The expanded laser beam enters the y-deflector cell, which produces vertical deflection with a 4×10^{-3} radian angular range. Lens 1 and lens 2 are cylindrical lenses with refractive power only in the vertical dimension. The cylindrical-lens telescope images the aperture of the y-deflector cell onto the x-deflector cell with a transverse magnification of f_2/f_1, and an angular range magnification of $f_1/f_2 = 1.41$. The diffraction spreading in the horizontal direction between the two cells is negligible. The deflected beams exiting from

the x-deflector cell have a horizontal angular range of 4×10^{-3} radian; the effective center of deflection for both dimensions lies at the x-deflector cell. A three-lens telescope with an effective focal length of 9.7 m thus provides the proper rectangular format of deflected, collimated beams near the back focal plane of the system.

Design values chosen for the deflector optics are

f_1 (cylinder) — 141.4 mm
f_2 (cylinder) — 100 mm
f_3 — 296 mm
f_4 — 12.7 mm
f_5 — 396 mm

Lenses L3 and L5 have a relative aperture of $f/5$ and are corrected for spherical aberration. Lenses L1, L2, L4 are of plano-convex simple form, stopped down to relative aperture values from $f/10$ to $f/20$. The overall axial separation from the first deflector cell to the output plane is 1.9 m.

4.3 Deflector Performance

The overall efficiency of the deflector telescope, including the initial beam expander and truncation aperture, is 70%. The efficiency of the cascaded deflector cells, when operated at nominal drive power, is 20%. Striations of inhomogeneous refractive index in the lead molybdate crystals produce the most serious limitation to deflector performance. These imperfections produce wavefront aberrations of greater than one-half wave and distort the far-field diffraction pattern of a deflected spot.

Fig. 6(a) shows an enlarged photograph of the light distribution in the output plane of the deflector telescope when the deflector cells are removed. This distribution, for which the central disk is overexposed here, is the diffraction pattern resulting from the truncation aperture. Fig. 6(b) is a photograph of the output plane, with the same exposure as (a), after the deflector cells are inserted. The central lobe is distorted, the side lobes are much stronger, and the striations produce additional preferential scattering in the horizontal and vertical directions. Measurements show that the aberrations alone spread the light distribution to approximately 1.5 times the intended extent. Thus, there is relatively less light flux at the addressed hologram and more spillover light at the neighboring holograms. Fig. 6(c) is a photograph of the output plane, without overexposure, showing the deflected spots resulting from scanning all combinations of positions

1, 2, 13, 18, 19, 24, 35, and 36 for x- and y-deflection. With the exception of 16 spots, 4 in each corner, this pattern represents the spatial locations of the 48 active holograms in the storage medium.

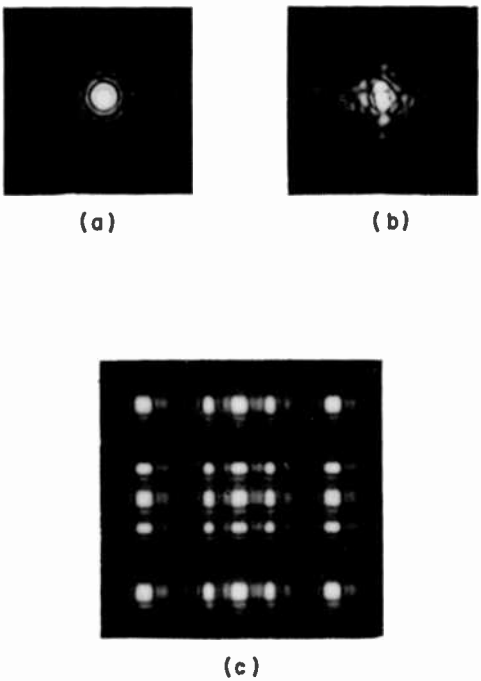


Fig. 6—Output plane of deflector telescope: (a) without deflector cells (central spot is overexposed), (b) with deflector cells (same exposure as (a)) ,and (c) scanning all positions sequentially (normal exposure).

5. Optical Components

5.1 Hololens

The hololens is an array of permanent phase holograms that serves as the beam splitter for recording holograms in the erasable storage medium. Diffracted light from the hololens is used to illuminate all the light valves of the page composer. The hololens holograms represent blank pages of information; these holograms are copied onto the erasable storage medium with the page composer providing the desired information pattern.

Simultaneous recording of the entire hololens array is accomplished with the geometry sketched in Fig. 7. The page composer mask is a metal sheet with clear openings etched in the locations of the page-composer light valves. A random phase plate is also included with the mask to uniformly shift the phase of each aperture according to a computer-generated pseudo-random sequence. Randomization of the phases ensures a sufficiently uniform spatial distribution of object illumination at the holograms. The optical element denoted as the scatterer is imaged by the lens onto the sensitized dichromated gelatin plate, which will become the hololens array. Uniform illumination of the page composer mask must be provided by each discrete scatterer point that is imaged onto a hologram location in the hololens array. One such point is depicted in the figure. A uniform collimated beam incident at 45° to the sensitized plate provides the reference wave for simultaneously recording the entire hololens array.

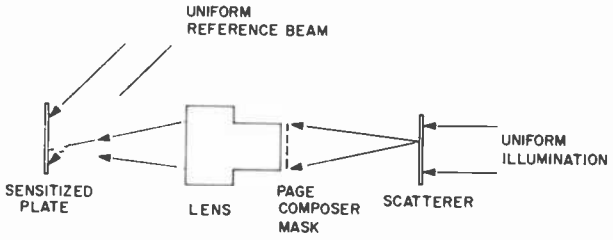


Fig. 7—Hololens recording geometry.

To provide uniform, spatially coherent illumination of the page composer, the scatterer may consist of a multiple-lens array. Focused spots of light, one for each lenslet of the scatterer, are simultaneously imaged onto the sensitized plate at each point where a hologram is to be recorded. Molded plastic lens arrays would appear to be ideally suited for this purpose. Our observations have shown, however, that the surface finish of such arrays is generally too poor to give a uniform cone of illumination with coherent light. The results shown here are obtained with individually selected and mounted simple glass lenses. Fig. 8(a) is a photographic image of the page-composer mask taken through the 1-mm aperture of one hologram in the hololens plane. The opaque aperture near the center is deliberately introduced. The illuminating lenslet in this case has a slight cosmetic defect that projects a faint shadow onto the page-composer mask. Fig. 8(b) shows the object-beam illumination in the hololens plane, to a different scale, that is mixed with the reference wave to form the hololens hologram.

During operation of the memory system, a pattern of identical geometrical shape and size is projected by the object-beam lens onto the erasable storage medium to record the page-composer information.

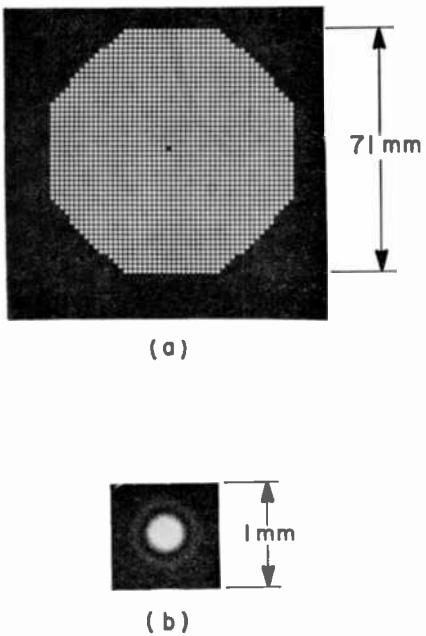


Fig. 8—(a) Image of page composer mask, illuminated by one lenslet, as seen through 1-mm hologram aperture. (b) Object illumination of one hologram in hololens array.

It can be seen in Fig. 8(b) that the ratio of the heater width of 1 mm to the diameter of the central disk of the object beam pattern is 2.3, as specified in the final design of the optical system. Since the hologram size is much larger than is required to resolve the image of the page composer, the extra area can be used to provide additional redundancy. This is accomplished by imaging more than one point source onto each hologram area. Additional point source images are easily obtained by interposing crossed sinusoidal phase gratings between the multiple lens array and the page composer mask in Fig. 7. The result, shown in Figs. 9(a) and (b), is the addition of uniform structure in the image of each element of the page composer mask, and the averaging out of cosmetic defects in the illumination from the lenslets. There remain in Fig. 9(a) some low-contrast interference bands caused by overlapping higher diffraction orders from separate lenslets. This effect is easily eliminated by slightly rotating the crossed

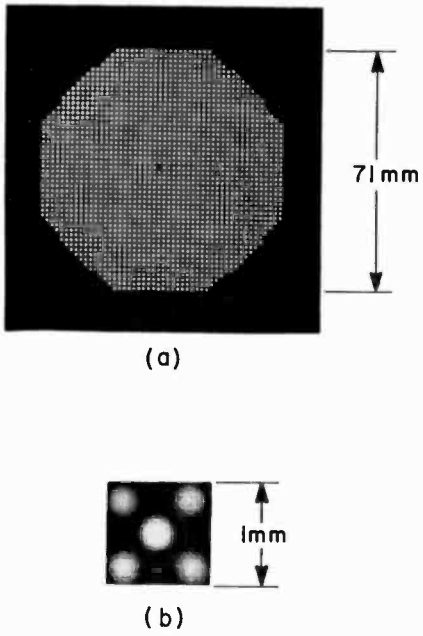


Fig. 9—(a) Image of page-composer mask, with redundancy added by a grating. (b) Corresponding object illumination of one hologram in hololens array.

gratings in their own plane to remove the overlapping orders.

The large area available for recording each hologram on the hololens, and on the erasable storage medium, also permits the use of diffuse illumination. The scatterer of Fig. 7 can be a plate of frosted glass in this case. Fig. 10 is an enlarged image of the portion of the page composer mask surrounding the central opaque element, taken through the 1-mm hologram aperture with diffuse illumination. It can be seen that there are seven or eight resolvable points in the

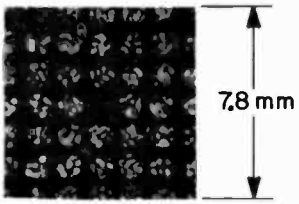


Fig. 10—Image of page-composer mask, diffusely illuminated, as seen through 1-mm hologram aperture.

image of each page-composer element. Although there are large fluctuations in intensity among the resolvable points, the light flux averaged over each page-composer element is much more uniform. When the hologram aperture approaches d_A, the number of resolvable points for each page-composer element approaches unity. The image intensity fluctuations would then be so severe as to rule out the use of diffuse illumination.

The dichromated gelatin holograms were prepared[12] from commercial photographic plates by first removing the light-sensitive silver salts to leave a layer of gelatin on glass, then sensitizing the gelatin in a dichromate solution, followed by development in water and alcohol after the holographic exposure. Typical exposures of 100 mJ/cm^2, with a reference/object intensity ratio of 20:1, required 5 minutes to record over the 6-cm-diameter hololens area. Reproducibility is only fair in these circumstances. Hololens arrays were obtained in which the zero- and first-order outputs were 65% and 8%, respectively, of the incident beam intensity when aligned at the appropriate playout angle. This gives a hololens with an 8:1 beam-splitting ratio and an optical system efficiency of 73%. The remaining light is lost by surface reflections, scattering, and absorption. A sealed cover-glass plate over the gelatin is necessary to prevent gradual loss of diffraction efficiency in environments of high relative humidity ($\geq 70\%$).

5.2 Object and Readout Lenses

Two compound imaging lenses are required for the holographic memory system. The object-beam lens images the hololens array onto the storage medium. The page composer is placed as close as possible to this lens and serves as the effective aperture stop of the lens. The readout lens projects the virtual image of the page composer, as reconstructed by wavefronts from the erasable holograms, onto the photodetector array. The storage medium is located as close as possible to this lens and also forms the effective aperture stop. Both lenses are required by the memory system design to operate at unity magnification with identical focal lengths. Although a somewhat smaller aperture is required for the readout lens than for the object-beam lens in the final design, identical design specifications were chosen.

Specifications that were chosen for the imaging lenses are listed below. The first six items follow directly from the final design specifications for the memory system. The last two items are tolerances on image quality.

[1] Wavelength—488.0 nm, monochromatic.

[2] Magnification—1:1.

[3] Focal length—nominally 175 mm. Axial distance between image plane and aperture stop is 356 mm.

[4] Object diameter—71 mm.

[5] Image diameter—71 mm.

[6] External aperture stop:

(a) Diameter—71 mm.

(b) Location—At least 6 mm outside the end of the lens barrel on the image side. It is required that this aperture be fully and uniformly illuminated by any point in the object plane. This specification guarantees that the page composer can be fully illuminated by every hologram in the hololens array. Also, all readout image rays will reach the photodetector array from every hologram in the erasable storage medium.

[7] Resolution—the image blur from any object point will be less than 0.2 mm in diameter. This specification for the object-beam lens requires that all rays that would ideally arrive at one point on the storage medium must arrive within one-half the diameter of the main lobe of the diffraction pattern from one page-composer element. This ensures that a substantial fraction of the light from all page-composer elements illuminates the hologram area rather than spreading over a larger area. For the readout lens, the specification requires that the image shift from hologram to hologram be less than one-fourth the diameter of one page-composer element.

[8] Distortion—for any object point, the center of the image circle will deviate no more than 0.3 mm from the corresponding undistorted image point. For the object beam lens, this specification ensures reasonably accurate overlap between the object beam and the reference beam for writing holograms anywhere on the erasable storage medium. For the readout lens, distortion shift of the image is less than one-third the diameter of a page-composer element.

The off-axis aberrations of two-element cemented lenses are too severe to fulfill these requirements. A commercially available 8-element 174-mm, $f/1.4$ flying-spot-scanner lens gives adequate image quality, but cannot provide illumination of the entire external aperture as required. The lenses used in the system were custom designed and fabricated. The barrel is 120 mm in diameter by 167 mm long; it contains six elements having an effective focal length of 230 mm, giving $f/3.1$ at infinity. Image blur diameter is less than 0.06 mm and distortion is less than 0.025 mm.

6. Liquid-Crystal Page Composer

The transmissive liquid-crystal page composer (LCPC) used in this system contains 2,048 circular elements arranged in a quasi-octagonal pattern. The elements are 0.976 mm in diameter and are spaced on 1.30-mm centers. The two elements of a bit operate in complementary fashion with one element energized with an ac voltage and the other unenergized. With no voltage across it, the nematic-liquid-crystal layer is transparent and light is transmitted unimpeded. A voltage across the liquid-crystal material causes it to become turbulent so that light passing through it is scattered strongly, with little light reaching the selected storing page to be recorded. In the differential mode of operation, the "1" and "0" are determined by which one of the two elements of a bit is energized.

The LCPC contains 10 bits whose information state is electrically alterable. The remaining 1014 bits are fixed in an unchanging pattern, with half the elements connected to common energizing electrodes, and the other half unconnected. The page composer is fully populated optically; there are 2048 discrete elements of which 1024 are transparent for recording the storage holograms. The 10 alterable bits are located around the periphery of the array and at its center so that results are representative for any bit location. Nearest neighboring bits have been included so that interaction effects would be realistically included.

A cross-sectional view of the LCPC is shown in Fig. 11. Two glass plates with transparent electrically conductive coatings sandwich the nematic-liquid-crystal material between them at a separation of 0.0127 mm as determined by Mylar spacers. The conductive coating on glass 2 is etched to form the pattern shown in Fig. 12 to which voltage can be applied to energize half of the fixed elements and the 10 selected alterable elements. The pads for wire connections to the drive electronics are also visible. The transparent conductive layer on glass 1 is not etched but is left intact to serve as a ground plane or counterelectrode for the electrodes of glass 2. In addition, an opaque layer of aluminum is deposited on the transparent conductive layer of glass 1 to form the pattern shown in Fig. 13, which defines all of the 2,048 elements of the LCPC by masking the regions between and outside of the circular elements. This pattern is then registered precisely with the pattern of Fig. 12 on glass 2 when the sandwich is put together. The task of aligning the LCPC in the optical memory system is significantly eased and simplified because of this built-in mask (see Fig. 14). When the LCPC was electronically activated and tested in a simulated optical memory system using helium-neon laser

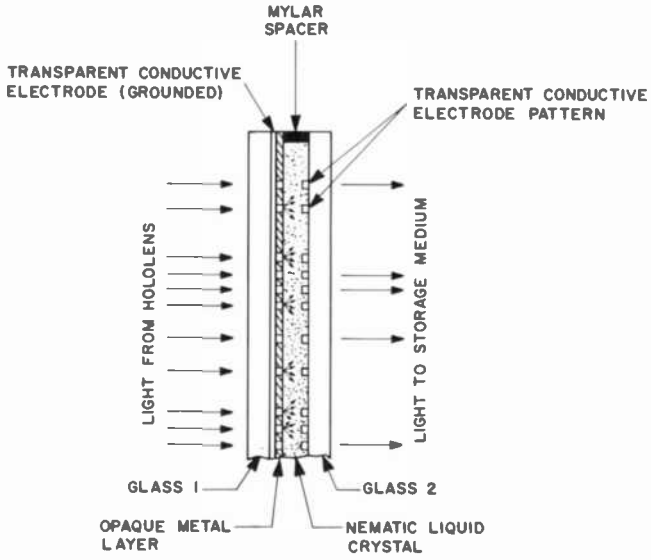


Fig. 11—A cross-sectional view of the liquid-crystal page composer.

Fig. 12—The transparent conductive electrode pattern on glass 2.

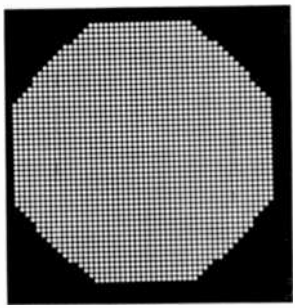


Fig. 13—The opaque aluminum pattern formed on glass 1.

light, contrast of better than 50:1 was measured at a detector for the clear state versus the scattering state of a liquid-crystal element. Birefringent effects in the liquid-crystal material were also measured and were found to reduce the usable light at the corners of the LCPC by about 25%.

The lifetime of liquid-crystal cells is strongly dependent on contaminants that are introduced during the filling and sealing process.

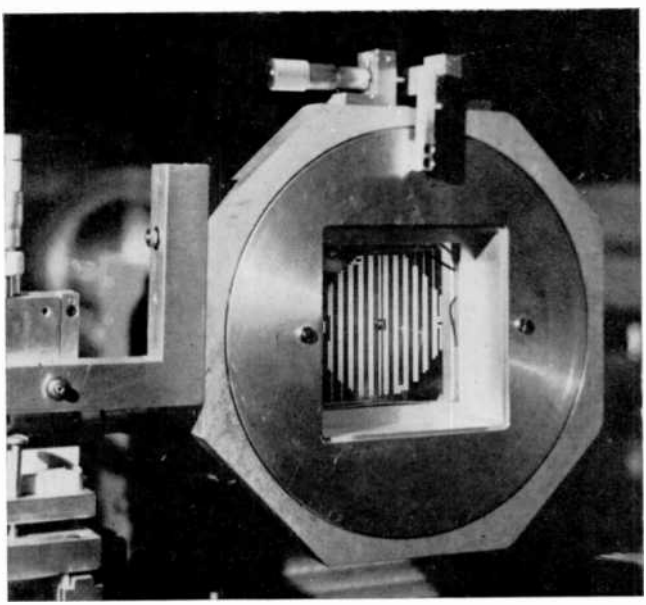


Fig. 14—The 1024-bit liquid-crystal page composer activated to produce a typical pattern of scattering.

Using a precured silicone rubber epoxy seal, we have obtained operation for one year before symptoms of contamination appeared. Lifetimes of several years are obtained in the commercial production of large-area cells with glass frit seals and improved filling methods.

7. Erasable Storage Medium

The requirements for the storage plane in the prototype read–write holographic memory were

Resolution in excess of 1400 lines/mm
Sensitivity to 4880 Å radiation
Transmissive readout
Total size greater than 7.5 cm
Page size—1 mm
Write energy—1 mJ or less
Read efficiency—1% or greater

To satisfy these requirements it was decided that a thermoplastic–photoconductor sandwich would be used for the storage plane. The thermoplastic material used was Staybelite ester 10 (a derivative of natural tree resin), and the photoconductor used was poly-n-vinyl carbazole doped with 2,4,7 trinitro fleurenone. In the following, we will describe material preparation, storage plane fabrication, and the operating parameters of the storage plane. Some interesting techniques that allow operation on a large storage plane will also be covered.

7.1 Materials

a. Photoconductor—The photoconductor used in all experiments was poly-n-vinyl carbazole (PVK) doped with 2,4,7 trinitro 9-fleurenone (TNF). These were mixed in the ratio 5 gm of PVK to 1 gm TNF and then diluted in 100 ml of 1,1,2 trichlorethane. In early experiments the solvent used was an equal mixture of p-dioxane and dichloromethane, but this had the disadvantage of too rapid evaporation of the solvent, which led to problems of repeatability and to problems of film formation in humid atmospheres. The sensitivity of this photoconductor will be discussed later.

b. Thermoplastic—The thermoplastic used in most of the experiments was Stabelite ester 10. Various solvents were used, but the preferred solvent was hexane. The Staybelite ester 10 was diluted in the hexane in the ratio 20 gm of Staybelite to 100 ml hexane.

c. Substrate—The substrate used was InO-coated glass with electrical conductivity in the range 50 to 100 ohms per square. The InO coating was etched into the desired pattern using a dilute hydrochloric solution. The active areas of the partially populated storage medium were 1 mm squares distributed over a 6-cm diameter area. Evaporated gold lines were used to bring power to the active area.

7.2 Sample Fabrication

The technique used for sample fabrication was dip-coating, with the photoconductive layer applied first and the thermoplastic layer applied next. The technique of dip-coating involves the careful withdrawal of a substrate from a solution of the material to be coated. To obtain smooth films, care must be taken to avoid changes in withdrawal speed and to avoid mechanical vibrations. The thickness of the deposited film can be controlled either by varying the speed of withdrawal or by changing the viscosity of the solution; in most experiments the first method was used. The plate was allowed to dry for several minutes after removal from the photoconductor solution, and was then submerged in the thermoplastic solution. After the thermoplastic layer was deposited, the plate was placed in an oven at 60°C for one hour. The uniformity of the deposited film was fair; variations in thickness on the order of several hundred angstroms were observed.

7.3 Recording on Thermoplastic Media

The basic principles of storage have been discussed in the literature.[3,13,14] The operation involves the sequential processes of charging, exposing, recharging, and heating. Erasure is accomplished by reheating the film.

To ensure repeatable results and to avoid catastrophic electrical breakdown of the thin thermoplastic film, the voltage on the film (due to the corona charging) must be precisely controlled. This control is complicated by the need for random selection of storage locations over a large storage plane. First, it is hard to uniformly corona-charge a large area. Secondly, because of random selection, one location may be repeatedly selected (and therefore charged and heated), but another location on the same plane may be only charged. If not controlled, the voltage on the unselected location will rise to the breakdown voltage of the film and cause damage. A third reason is that corona-charging is affected by atmospheric conditions such as humidity and temperature. The variation in these conditions normally encountered will lead to changes in the voltage on the thermoplastic.

The first of these problems, corona-charging a large area, was solved by using fine (50 to 75 μm diameter) wires held in a "serpentine" pattern by a plastic frame over the storage plane. The distance from the wires to the plane and between each wire is approximately 1 cm. The wires are fine enough, and far enough away from the storage plane, so they do not intercept an appreciable part of the light. This arrangement does charge the entire storage plane. It is nonuniform, however, since the edge or external wires emit more charge than the inner wires. (The cause can be attributed to wire-to-wire capacitance effects.)

To solve the problem of nonuniform charging, and the more important problem of precise voltage control, a grid (apertured metal plate) was placed over the storage plane. It contains 1 mm square apertures at the active hologram locations of the storage medium. For a fully populated storage plane the grid would take the form of a mesh. In use, the apertured plate is placed against the storage plane (separated only by a 25-μm-thick insulating film to prevent short circuits), and voltage equal to the desired voltage on the thermoplastic film is applied to it. This prevents the voltage on the thermoplastic caused by corona-charging from rising above the voltage on the apertured plate. The reason is that if the voltage at the thermoplastic were to rise above the grid voltage, the ions of the corona carrying the charge would be deflected to the grid. We found that, with the use of the grid, the maximum voltage on the thermoplastic was substantially independent of charging time and of voltage on the corona applicator. The use of the grid allowed highly repeatable results. The storage plane, with all elements in place, is shown in Fig. 15.

Two of the more important characteristics of the thermoplastic storage medium are its sensitivity (that is, efficiency as a function of exposure) and its spatial frequency characteristics (that is, efficiency as a function of the angle between object and reference beams). Fig. 16 is a typical plot of the efficiency of the reconstruction as a function of the exposure. In this particular experiment the photoconductor thickness was 1.5 μm, the thermoplastic thickness was 0.5 μm, and the grid voltage was 200 V. The beam ratio used in this experiment was 1:1. Note that the thermoplastic medium is quite sensitive (of the order of 649F photographic film) and that the efficiency remains constant even for relatively high exposures. This latter feature is a consequence of using a 1:1 beam ratio in the experiment.

The reconstruction efficiency is not, in general, a monotonic function of the recording exposure. In the following analysis we derive

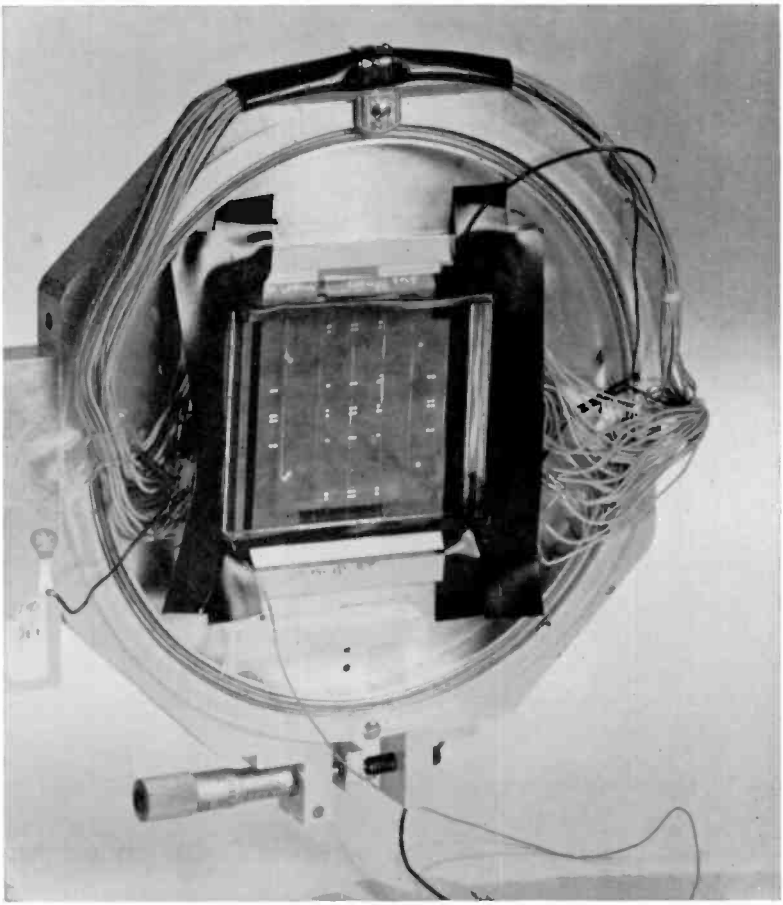


Fig. 15—Assembled and mounted storage plane.

the variation of efficiency versus exposure. We use a lumped-circuit model consisting of a capacitor shunted by a resistor to describe the photoconductor, and we make the assumption that the effect of light on the photoconductor is to vary the conductance of the shunt resistor. Although this model is admittedly somewhat naïve, its virtue is that it accounts for the shape of the experimental curves. In this analysis we confine our attention to the voltage across the photoconductor; this voltage determines the ultimate forces acting on the thermoplastic.

Using the lumped-circuit model, the relationship for the voltage across the photoconductor is

$$C\frac{dV}{dt} + GV = 0, \qquad [30]$$

where C is the capacitance per unit area and G is the conductance per unit area. Assume that the conductivity of the photoconductor varies

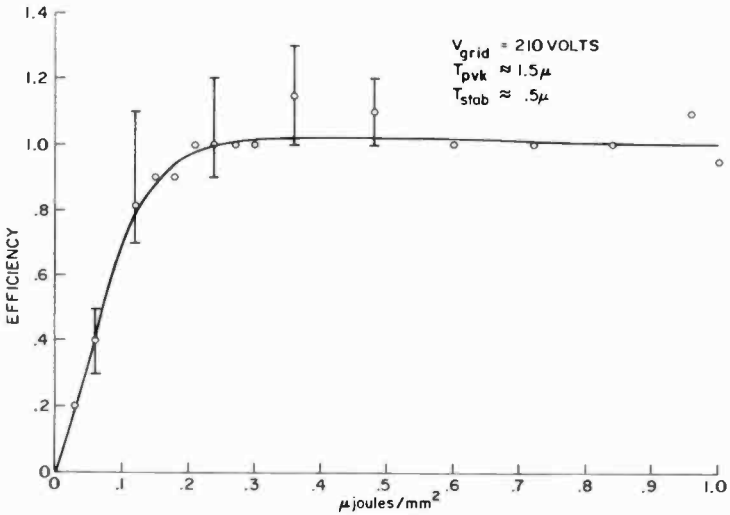


Fig. 16—Measured hologram efficiency versus exposure (unity beam ratio).

with the intensity of the absorbed light, and further assume that the intensity is written as

$$I = I_o\left(1 + m\cos\frac{2\pi x}{d}\right), \qquad [31]$$

where x is the coordinate on the surface of the photoconductor, d is the spatial wavelength, and m is the modulation ratio. The relation between m and the reference to object beam intensity ratio K is

$$m = 2(K)^{1/2}/(1 + K). \qquad [32]$$

Then

$$C\frac{dV}{dt} + DI_o\left(1 + m\cos\frac{2\pi x}{d}\right)V = 0, \qquad [33]$$

where D is a proportionality factor (the sensitivity of the photoconductor). The solution to this equation is

$$V(x,t) = V_o \exp\left[-DI_o\left(1 + m\cos\frac{2\pi x}{d}\right)t\right]. \quad [34]$$

The thermoplastic behaves as an incompressible fluid, which implies that its motion is not dependent on the absolute forces applied, but rather on the spatial variation, or difference, of the applied force. From the previous relationship, we note that the maximum voltage (related to the maximum force) is

$$V = V_o \exp\left[-DI_o\,(1+m)t\right], \quad [35]$$

and the minimum voltage is

$$V = V_o \exp -DI_o\,(1-m)\,t. \quad [36]$$

The difference is

$$\Delta V = V_o \exp\{-DI_o\} \sinh DI_o mt. \quad [37]$$

If we now let $I_o t$ be the total exposure E_o, we may write

$$\Delta V = V_o \exp\{-DE_o\} \sinh DE_o m. \quad [38]$$

In Fig. 17, we plot ΔV versus E_o, with beam ratio as a parameter. Since the ultimate behavior of the thermoplastic is dependent on ΔV, it is evident that the efficiency of the reconstructed image will not be a monotonic function of the exposure.

In the prototype system, beam ratios of the order of 10:1 are used. A plot of the efficiency versus exposure for a hologram made with a 10:1 beam ratio is shown in Fig. 18.

One of the troublesome features of thermoplastic storage is the band-pass nature of the spatial frequency response. The maximum response is centered at a spatial frequency given by $1/(2h)$, where h is the thermoplastic thickness, and the response falls off rather quickly toward high and low spatial frequencies. Fig. 19 shows the response for a film 490 nm thick, which should have had a center frequency of 1000 lines/mm (corresponding to an angle of 30° between reference and object beam), while Fig. 20 shows the response for a film 340 nm thick. The dotted line in both graphs shows the angular extent of the

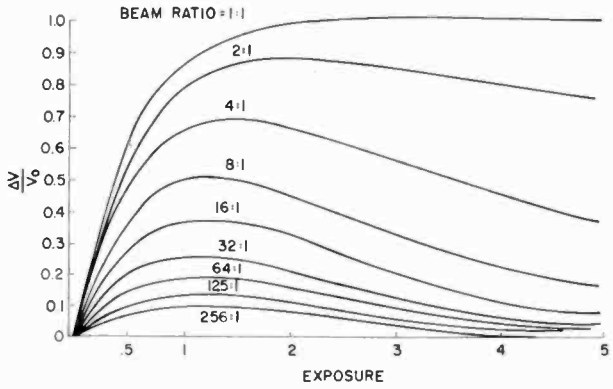


Fig. 17—Photoconductor voltage as a function of beam ratio.

object used in the prototype system, and we can see that the band-pass response leads to shading in the reconstruction. While this is not a disastrous result, it does lead to difficulties in readout. The more important result is that the actual center of the response curve did not occur at the spatial frequency predicted by the thickness measurements made on the film.

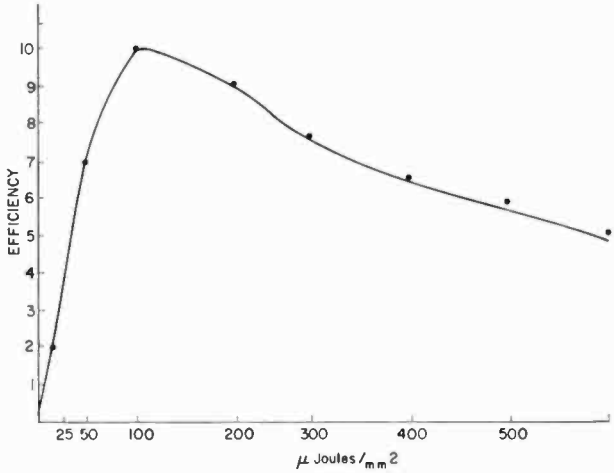


Fig. 18—Measured hologram efficiency at 10:1 beam ratio.

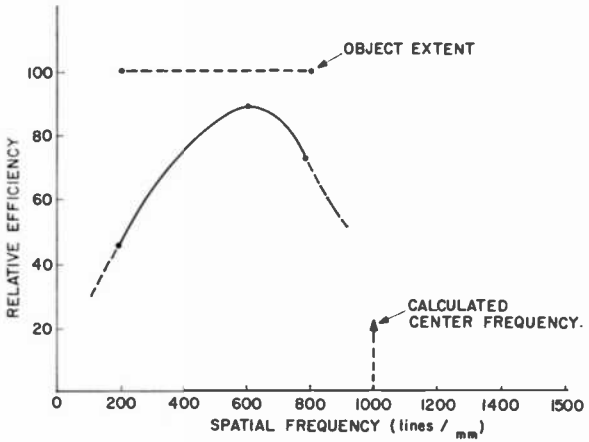


Fig. 19—Hologram efficiency versus spatial frequency for a film nominally 490 nm thick.

These thickness measurements were made by interferometry, but at a location on the storage plane somewhat remote from the storage locations used for the experiment. The reasons for the discrepancy between the predicted and actual center frequency are either an error in thickness measurement, variations in the thermoplastic thickness over the surface of the storage plate, or a failure in the theory that predicts the center frequency. Regardless of the cause, the results point out that it is difficult to fabricate a storage plate with predetermined performance characteristics.

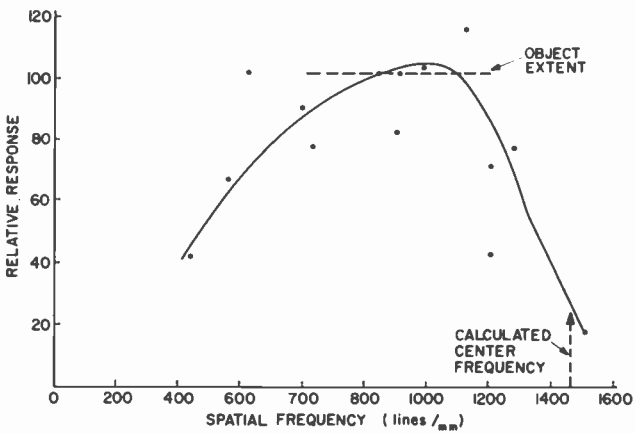


Fig. 20—Hologram efficiency versus spatial frequency for a film nominally 340 nm thick.

In a further experiment, holograms were made with the angle between object and reference beam at 45° (corresponding to 1440 lines/mm) on a storage plane with a thermoplastic thickness of 490 nm (which should result in a center frequency of about 1000 lines/mm). The thickness of the photoconductor was 1.5 μm and the voltage on the grid was 250 V. The object was the page composer used in the prototype system. The key step in the experiment was to sequentially store holograms at ever-increasing exposure and to monitor the reconstruction. That is, the complete process of erase–charge–expose–

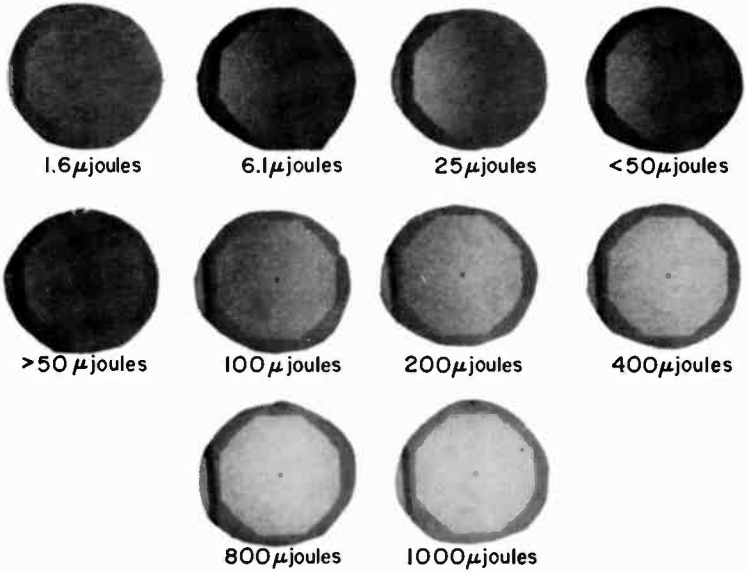


Fig. 21—Spatial frequency response versus exposure, with center frequency of 1440 lines/mm.

charge–read was repeated a number of times, with increasing exposures for each trial. The results of the experiment are shown in Fig. 21. The first reconstruction is expected—only the low-frequency portion of the image is visible. Then, as the exposure is increased we first see the low-frequency portion of the image become brighter (all photographs made with the same exposure) and then—the first surprising result—the high-frequency portion of the image becomes visible, and a dark band separates the two positions. As the exposure is increased further the dark band is observed to sweep across the image, toward lower frequencies. Finally, at an exposure of about 800 μJ, the dark band has left the image. The effect of increasing the

exposure even further appears only to brighten the image.

This experiment was repeated with storage planes of thermoplastic thicknesses that varied from 490 to 200 nm with substantially the same results. The net effect of large exposures is that the exact thickness of the thermoplastic is of no consequence—a result that enormously reduces the complexity of sample fabrication. The reason for this phenomenon is not completely understood; however, it appears that at the high exposures used the photoconductor becomes saturated (or effectively perfectly conductive) while the thermoplastic itself becomes photoconductive. The disadvantage of this technique is that it requires much higher exposures than normally used; fortunately, the prototype system had been designed for just the exposure levels needed for this method.

The final important characteristic of the thermoplastic storage medium is its fatigue behavior. While there are some thermoplastics that exhibit little fatigue, Staybelite is not among them. The upper limit in the number of cycles that could be used before the signal-to-noise ratio of the reconstructed image became intolerable was of the order of 500. The failure was marked by the appearance of many pits and other permanent surface deformations. This was possibly caused by the lack of adequate cleanliness in the materials used, with the resulting impurities initiating the observed defects. A further cause of failure was an apparent increase in the softening temperature following prolonged operation, which was undoubtedly caused by oxidation of the thermoplastic.

Since the purpose of these experiments was the demonstration of the feasibility of holographic storage, and since it is apparent that thermoplastic materials with little or no fatigue are being developed, this aspect of the storage characteristics of Staybelite was of little practical concern. However, it was found that if a pattern was recorded and then retained for long periods—longer than 24 hours—it was difficult to erase. Furthermore, when this pattern was apparently erased and a second pattern was recorded at the same location, the first pattern reappeared. The first pattern was dimmer than the second, but still presented a potentially troublesome signal. It was found that this could be prevented either by not retaining a pattern for longer than 12 hours or by erasing the pattern and not using that location for at least 24 hours. Neither method is ultimately acceptable. The reason for this effect appears to be either cold flow or charge-trapping in the plastic. This is a potentially very serious problem and should be studied further if thermoplastic media are seriously considered for holographic storage.

8. Photodetector Array

An array of 20 p-i-n photodiodes is used to detect the ten active bits in the partially populated optical memory. Two diodes are used for each bit to indicate "1" or "0". Since both diodes feed a differential-input preamplifier, common-mode light and electrical noise are attenuated. The y-axis acousto-optic deflector cell is amplitude modulated with a 100-kHz square wave during readout. This modulation provides a temporal carrier for the detector signal, which, in conjunction with the differential photodetector circuit, provides adequate discrimination against ambient room light. A block diagram of one detector channel is shown in Fig. 22. The active bits are all sensed in parallel; panel lamps indicate the detected information pattern.

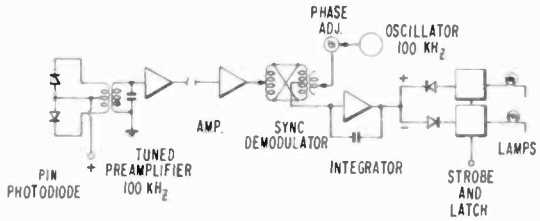


Fig. 22—Detection system diagram.

The equivalent input noise current to each of the ten preamplifiers is about 10^{-11} A rms. This is about 5×10^{-11} W average of light power striking each photodiode (conversion ratio of 0.2 A/W) While the frequency bandwidth of each tuned preamplifier is 2 kHz, the overall bandwidth of the sense system, including synchronous detection and strobing, is about 120 Hz. A redesign of the detection system could reduce the equivalent input noise current from 10^{-11} A rms to probably 2×10^{-12} A rms. The present photodetectors and the associated electronics can detect an optical signal level of approximately 10^{-10} W in 30 msec.

The ten active bits in the partially populated memory were arranged in five locations of two bits each with two photodetectors assigned to each bit. The five locations are separated about 32 mm from each other, which provides sufficient space for mounting the five clusters of four diodes each. The four p-i-n photodiodes of each cluster

are a commercial four-element photodiode chip. Each element is 1.3 mm square with a 0.13 mm separation. The chip is sealed in a TO-5 can with five leads extending from the base. The diameter of the light beam striking each diode is 0.97 mm, and the spacing between each of the four is 1.3 mm. Fig. 23 shows the assembled detector array.

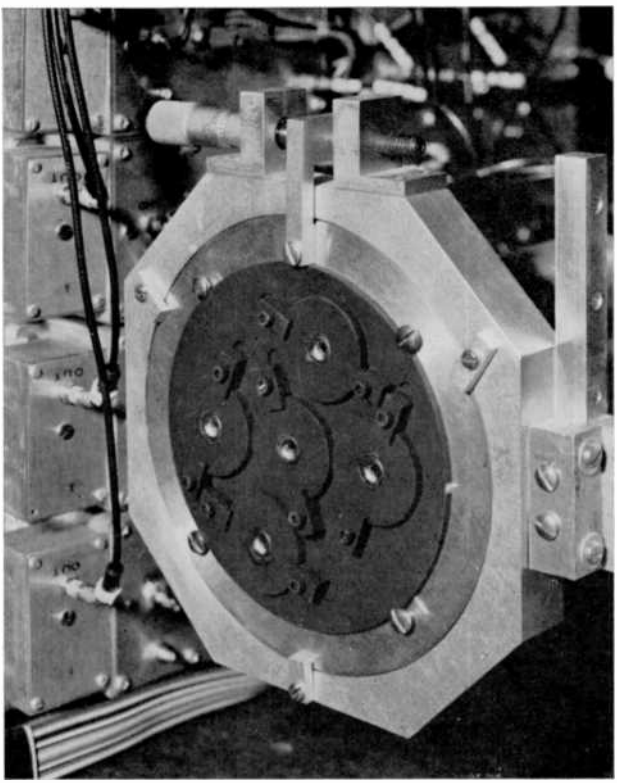


Fig. 23—Assembled and mounted photodetector array.

9. Memory Operation

The automatic sequencer used to operate the memory controls five functions:

Page composition—bits stored temporarily in page composer for later transfer to storage medium.

Page Selection—deflection of laser beam to desired location on storage medium.

Write—transfer of page composer bits to storage medium.

Read—transfer of stored bits from page to photodetectors and light display.

Erase—removal from storage medium of previously written information in any given page or pages.

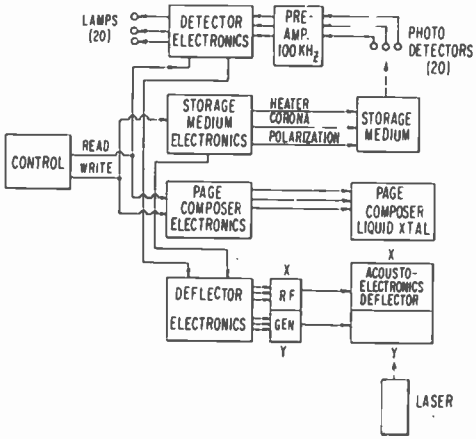


Fig. 24—Block diagram of controller electronics.

An overall block diagram is shown in Fig. 24. The control units for the laser deflector, the page composer, the storage medium, and the photodetector array are each housed on separate chasses with separate front panels. Each unit can function under internal control, independently of the other three, for adjustments and tests of the memory system components. For operation of the entire memory system, the four units function under external control signals from a control console, which initiates the appropriate sequence of operations.

The memory system is assembled on the surface of a 1.3 by 4 m table that is vibration-isolated. Fig. 25 shows the layout of the overall system with the approximate location of the components. The majority of the components are mounted on stages having precision adjustments, since tolerances are fractions of a milliradian for angles, and fractions of a millimeter for linear position. Vibration isolation supports are used between the table top and the argon laser head, which carries circulating cooling water. The strongest noise signals in the readout of an addressed hologram are contributed by spillover illumination from the reference beam hitting neighboring holograms, as expected from considerations discussed in the final design.

Before the final assembly of the memory system, each of the major components was thoroughly tested alone and then in conjunction with others. The assembly procedure consisted of a step-by-step addition of each operating component. Holograms were recorded and read out without error from all 48 active hologram locations, with no mechanical adjustments. The operating cycle with 300-mW laser output was: 2 sec corona charging time, 200 msec effective writing exposure duration, and 30 msec effective reading duration. Effective durations are listed because the temporal duty ratios of the laser beam are 0.1 for writing and 0.05 for reading.

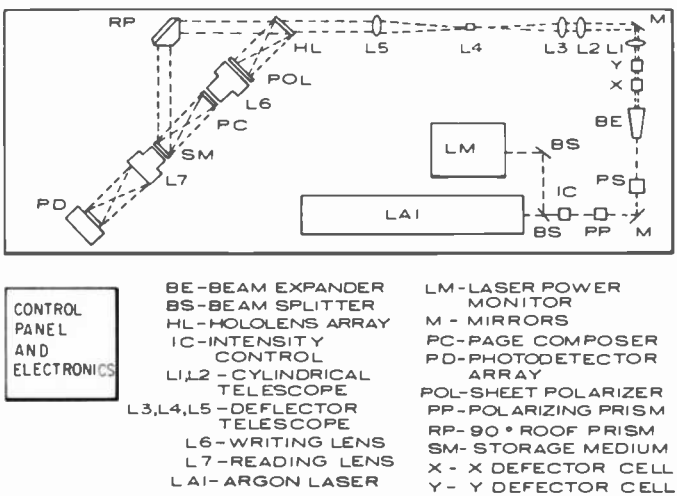


Fig. 25—Approximate layout of optical memory system.

The predominant source of optical crosstalk in the detected images comes from reference beam spreading to neighboring storage holograms, as expected. When there is no hologram in the addressed location, partial readout of the information from the neighbors is detected. When a hologram is present at the addressed location, however, adequate rejection of the noise signals is obtained with the differential detection system.

10. Discussion

We may extrapolate from the performance of the partially populated memory to conclude that a fully populated 10^6-bit system of this design would operate successfully. Replacement of only the aberrated deflector cells with ones displaying the full 40-spot Rayleigh resolution would be required. The calculations presented in the design analysis

suggest that an eight-fold increase in capacity represents the upper limit of the specific optical components used here.

As might be expected in a prototype system of this complexity, a number of engineering difficulties were encountered. Standard lens designs are not particularly well suited for the unusual imaging functions required in the memory. Aberrations in the deflector cells, and lifetime problems with the laser, the dichromated gelatin hololens, the liquid-crystal page composer, and the thermoplastic storage medium were all present. These factors, as well as improved high-speed page composers and erasable storage media, represent the areas to which increased attention is required in future efforts to develop larger systems.

Acknowledgment

The authors wish to express their appreciation to Jan A. Rajchman for his encouragement and support during this work. The invaluable contributions of C. W. Robbins, G. W. Leck, G. E. Bodeep, P. Mitnaul, J. L. O'Neill, D. Jose, J. N. Hewitt, J. A. Castellano, and R. N. Friel in the fabrication of the devices and components described here are gratefully acknowledged.

References:

[1] F. M. Smits and L. E. Gallaher, "Design Considerations for a Semi-permanent Optical Memory," **Bell Syst. Tech. J.**, Vol. 46, p. 1267 (1967).

[2] L. K. Anderson, "Holographic Optical Memory for Bulk Data Storage," **Bell Lab. Record**, Vol. 46, p. 319 (1968).

[3] L. N. Lin and H. L. Beauchamp, "Read-Write-Erase In Situ Optical Memory using Thermoplastic Holograms," **Appl. Opt.**, Vol. 9, p. 2088 (1970).

[4] J. A. Rajchman, "An Optical Read-Write Mass Memory," **Appl. Opt.**, Vol. 9, p. 2269 (1970).

[5] R. M. Langdon, "Design of a Large Read-Only Holographic Memory," **Radio and Electronic Engineer**, Vol. 38, p. 369 (1969).

[6] W. C. Stewart and L. S. Cosentino, "Optics for a Read-Write Holographic Memory," **Appl. Opt.**, Vol. 9, p. 2271, (1970).

[7] B. Hill, "Some Aspects of a Large Capacity Holographic Memory," **Appl. Opt.**, Vol. 11, p. 182 (1972).

[8] W. Lee, "Effect of Film Grain Noise on the Performance of Holographic Memory," **J. Opt. Soc. Am.**, Vol. 62, p. 797 (1972).

[9] L. D. Dickson, "Characteristics of a Propagating Gaussian Beam," **Appl. Opt.**, Vol. 9, p. 1854 (1970).

[10] C. B. Burkhardt, "Use of a Random Phase Mask for the Recording of Fourier Transform Holograms of Data Masks," **Appl. Opt.**, Vol. 9, p. 695 (1970).

[11] W. C. Stewart, A. H. Firester, and E. C. Fox, "Random Phase Data Masks," **Appl. Opt.**, Vol. 11, p. 604 (1972).

[12] L. H. Lin, "Hologram Formation in Hardened Dichromated Gelatin Films," **Appl. Opt.**, Vol. 8, p. 963 (1969).

[13] J. C. Urbach and R. W. Meier, "Thermoplastic Xerographic Holography," **Appl. Opt.**, Vol. 5, p. 666 (1966).

[14] T. L. Credelle and F. W. Spong, "Thermoplastic Media for Holographic Recording," **RCA Review**, Vol. 33, p. 206 (1972).

A Membrane Page Composer*

L. S. Cosentino and W. C. Stewart

RCA Laboratories, Princeton, New Jersey 08540

Abstract—The feasibility of producing a page composer for optical memory systems using thin, deformable, membrane-mirror elements as light valves was investigated. The electromechanical and optical performances of such elements were determined both analytically and experimentally. It was found that fast switching (10 μsec), high-contrast (10:1 or greater), fatigue-free operation over millions of cycles, and efficient utilization of input light could be obtained with membrane light valves. Several arrays of 64 elements were made on substrates with feedthroughs allowing access to individual elements from the backside of the substrate. Single light valves on such arrays were successfully operated with transistors designed for selection and storage at a bit location. This simulated the operation of a prototype page composer with semiconductor chips beam-lead bonded to the back of the substrate.

1. Introduction

Research on and development of optical memory systems[1] have resulted in a need for an array of light valves, which we call a page composer, to control the light pattern recorded at the storage medium. Ideally

* The research discussed in this paper was jointly sponsored by NASA George C. Marshall Space Flight Center under Contract No. NAS8-26808, and RCA Laboratories, Princeton, N. J.

this device should be fast, produce a high contrast ratio at the detector, accept light over a large cone of angles, operate without fatigue or degradation over billions of cycles, and be efficient in utilization of the light impinging on it. In addition it must be available in large enough sizes to accommodate big arrays. Virtually every candidate investigated to date suffers from one or more shortcomings in satisfying these requirements. For example, liquid crystals are slow, requiring several milliseconds, at best, to relax. Most electro-optic crystals have very limited angles of acceptance and large perfect pieces are difficult to produce. PZT ferroelectric ceramics are quite fast but fatigue effects have been found to limit operational lifetime.[2]

At present, we are studying the feasibility of producing a reflective page composer using thin electrostatically deformable metallic mirror elements as light valves. Array operation requires coincident pulse selection and storage at each bit location that can be provided by semiconductor circuitry on the back side of the substrate. Electrical connections to the elements are made by means of conductive feedthroughs extending from the back surface of the substrate to the front one.

Earlier efforts with membrane light valves included the membrane light modulator (MLM) of Preston[3], which was utilized in an optical computer, the photo-activated MLM of Reizman[4], which was used as an image converter, and the schlieren light valve (SLV) of van-Raalte[5], which was used in a television projection system. The present work extends the results in utilizing membrane light valves as elements of a page composer for an optical memory system.

2. Membrane-Light-Valve Geometries and Fabrication

All of the membrane light valves consist of a thin, metal film suspended over a support structure that defines the basic shape of the elements. A counter-electrode must be provided beneath the element, to which voltage is applied to activate the light valve with the membrane grounded. The electrostatic attraction between the membrane and the drive electrode causes the metal film to deform from its initially flat state. In the process, the reflection of light from the film surface changes in angular extent. An aperture can be appropriately placed in the optical system, such that the light reflected from the flat membrane passes through it while most of the light reflected from the deformed film does not get through the aperture. Thus, optical contrast can be achieved at a detector located after the aperture.

A line element is shown in Fig. 1(a), which is grossly exaggerated

in the vertical dimension for clarity. This type of light valve is similar to those used by J. A. van Raalte and his group at RCA Laboratories for a television projection system.[5] We have adopted, with some modifications, the techniques developed in the course of that project. For our purposes, a typical line element might have a separation between membrane and electrode of a few micrometers, an element width between supports of 0.25 mm or larger, a membrane thickness of about 0.5 μm, and support structure widths that are 5-10% of the element width. The line element has a line electrode beneath it running the entire length of the element.

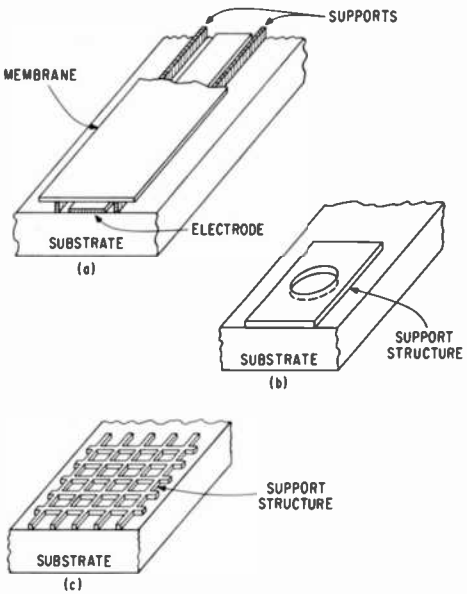


Fig. 1—Membrane light-valve geometries: (a) line element, (b) circular cylindrical element (membrane and electrode not shown), and (c) square element (membrane and electrode not shown).

Another membrane light-valve geometry is shown in Fig. 1(b). Here, cylindrical holes in a support structure define circular elements in the membrane that is eventually suspended on the supports. The electrode (not shown) beneath the element may also be circular, or it may be of another shape if a portion of it is opposite the active area of the membrane.

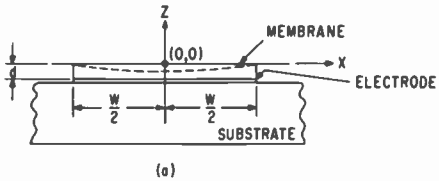
Other geometries are also possible for membrane light valves. Fig. 1(c) shows a support structure that defines square elements when a membrane is suspended on it. An individual electrode (not shown) can be located beneath each square to activate one element, or one

large electrode can be located beneath several elements and simultaneously activate the group of elements. For instance, an electrode could be located to activate a 3×3 group of elements simultaneously as a single light valve. With such a multiple-element light valve, redundancy is obtained so that perfection in the membrane is not required. Another advantage is that a large light valve could be built up from smaller elements, which might make fabrication simpler in some cases.

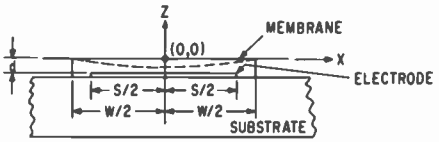
The fabrication of membrane light valves includes standard photolithographic techniques to make the electrode pattern and the support structure. The forming of the membrane, however, involves special processing. A layer of photoresist thick enough to be higher than the support structure is spread on the sample. After drying, the resist is polished until all the support structure surfaces are visible and level with the resist elsewhere. The remaining material serves as a temporary substrate on which the membrane may be deposited. Before that, however, another step is necessary in order to be able to remove the photoresist after the membrane is completed. Basically the technique is to incorporate tiny micrometer-sized holes into the membrane so that the resist solvent can go through the membrane and carry out the resist with it. A gelatinous solution is first poured over the substrate, and then rinsed off with deionized water leaving a thin film of gel. Then, with the substrate still wet, a dispersion of micrometer-sized particles in water is poured over the substrate and allowed to settle. The substrate is gently rinsed with deionized water and blown dry, leaving a uniform distribution of particles, anchored in gelatin, over the resist and support structure. The substrate is placed in a vacuum system where the membrane material is then evaporated over the powder. Since the membrane is about 0.5 μm thick, it does not quite cover the powder. Therefore when the sample is gently wiped with wet tissue, the tiny particles are dislodged and are then removed by a rinse with deionized water leaving holes in the metal where the particles had been. The tiny voids in the membrane do not appreciably affect the surface reflectance. Essentially the same technique can be used to electroplate a film, except that only a thin layer of "seeding" material needs to be evaporated on the photoresist, after which additional metal can be plated until the desired thickness is reached. The final step is the removal of the photoresist beneath the membrane by immersing the sample in solvent, which gets to the resist through the tiny holes in the membrane and diffuses out of the region. Flat films of good quality have been successfully produced with this technique.

3. Electromechanical Operation

When voltage is applied between an electrode and the membrane of a line element, such as the one shown schematically in Fig. 2(a), electrostatic attractive forces act to move the membrane towards the electrode until elastic restoring forces in the film are sufficient to prevent further deformation. The nature of the deformation can be



(a)



(b)

Fig. 2—Line element with symbols used in calculation of deflection: (a) electrode extends over the full width of the element and (b) electrode is smaller than the element width.

determined with the aid of a free-body diagram of a section of the deformed membrane, with the sum of the forces equated to zero since the body is in equilibrium. When this is done, the resulting equation is

$$F = T_o \frac{d^2z}{dx^2}, \qquad [1]$$

where F is the electrostatic force per unit area, and T_o is the film tension at the center of the element. Assuming an ideal parallel-plate capacitor, such as the one formed by the electrode and membrane if the deflection is negligible compared to the separation of the conductors, then

$$F = \frac{\epsilon V^2}{2d^2}, \qquad [2]$$

where ϵ is the permittivity of the medium, V is the applied voltage, and d is the membrane-to-electrode separation. Therefore

$$T_o \frac{d^2z}{dx^2} = \frac{\epsilon V^2}{2d^2}. \qquad [3]$$

Solution of this equation with the appropriate boundary conditions gives the form of the deformation as

$$z = \frac{\epsilon V^2 W^2}{16 T_o d^2}\left(\frac{4x^2}{W^2} - 1\right), \qquad [4]$$

which is seen to be parabolic. The maximum deflection, which occurs at $x = 0$, is given by

$$z_o = -\frac{\epsilon V^2 W^2}{16 T_o d^2}. \qquad [5]$$

The foregoing assumes that the electrode extended beneath the entire width of the element. When this is not the case, as in Fig. 2(b), a similar analysis results in

$$z = \frac{\epsilon V^2}{4 T_o d^2}\left(x^2 + \frac{S^2}{4} - \frac{SW}{2}\right) \qquad \text{for } \frac{S}{2} \geq x \geq 0, \qquad [6]$$

and

$$z = \frac{\epsilon V^2}{4 T_o d^2}\left(\frac{S}{2} - x\right) \qquad \text{for } \frac{W}{2} \geq x \geq \frac{S}{2}, \qquad [7]$$

which describe a parabolic deformation region where the membrane is over the electrode and straight line segments connecting the ends of that region to the supports. The deformation at the center is somewhat reduced with narrow electrodes. If an electrode covered only half of the element width ($S = W/2$), Z_o would be three-quarters of the central deformation with a full width electrode ($S = W$).

The above analysis is valid whether the film tension (T_o) at the center of the deformed element is built into the film at the time of deposition, whether it results from elongation of the film, or even if it is a combination of these two tensions. For the case in which T_o is a constant not substantially affected by the stretching of the

film, the derived equations are complete. If on the other hand, T_o arises only because of elongation, with the tension in the undeformed state being equal to zero, the above equations are still valid, but T_o is determined by other factors, such as the mechanical properties of the film and the amount of elongation of the film.

Young's modulus (E) is defined as the ratio of stress (σ) to strain (e). Since in the thin film $\sigma = T_o/t$, where t is the film thickness, and $e = \Delta W/W$, then

$$T_o = \left(\frac{\Delta W}{W}\right) Et. \qquad [8]$$

For a curved line, such as the deformed film of Fig. 2(a),

$$\Delta W = \int_{-W/2}^{W/2} \left(\frac{ds}{dx}\right) dx - W = 2\int_{0}^{W/2} \sqrt{\left(\frac{dz}{dx}\right)^2 + 1}\, dx - W,$$

where ds is the differential distance along the curved line. Since $dz/dx \ll 1$ for all practical cases,

$$\sqrt{\left(\frac{dz}{dx}\right)^2 + 1} \approx 1 + \tfrac{1}{2}\left(\frac{dz}{dx}\right)^2,$$

and

$$\Delta W \approx 2\int_{0}^{W/2} \left[1 + \tfrac{1}{2}\left(\frac{dz}{dx}\right)^2\right] dx - W = \int_{0}^{W/2} \left(\frac{dz}{dx}\right)^2 dx.$$

From Eq. [4],

$$z = \frac{\epsilon V^2 W^2}{16 T_o d^2}\left(\frac{4x^2}{W^2} - 1\right),$$

so that

$$\left(\frac{dz}{dx}\right)^2 = \frac{\epsilon^2 V^4}{4 T_o^2 d^4} x^2.$$

After substitution and integration,

$$\Delta W = \frac{\epsilon^2 V^4 W^3}{96 T_o^2 d^4},$$

and

$$\frac{\Delta W}{W} = \frac{\epsilon^2 V^4 W^2}{96 T_o^2 d^4}. \qquad [9]$$

Substitution in Eq. [8] yields

$$T_o = \left(\frac{\epsilon^2 V^4 W^2 E t}{96 d^4} \right)^{1/3}, \qquad [10]$$

and since from Eq. [5]

$$z_o = -\frac{\epsilon V^2 W^2}{16 T_o d^2},$$

substitution of the value of T_o of Eq. [10] gives

$$z_o = -0.287 \left(\frac{\epsilon V^2 W^4}{d^2 E t} \right)^{1/3}, \qquad [11]$$

with

$$z = z_o \left(1 - \frac{4x^2}{W^2} \right)$$

as before. Note that for a given sample, Eq. [11] shows that

$z_o \propto V^{2/3}$ for the case of tension caused by elongation only,

while Eq. (5) shows

$z_o \propto V^2$ for the case of a constant built-in tension.

The change in tension along the membrane has been assumed to be negligible in all of the previous analysis so that T_o was not a function of x. This can be justified by a simple calculation to deter-

mine the tension at one of the supports. Since

$$T(x) = T_o \frac{ds}{dx} = T_o \sqrt{1 + \left(\frac{dz}{dx}\right)^2},$$

and from Eqs. (4) and (5)

$$z = -z_o \left(\frac{4x^2}{W^2} - 1\right),$$

then

$$\left(\frac{dz}{dx}\right)^2 = 64 \frac{z_o^2 x^2}{W^4},$$

and

$$T\left(\frac{W}{2}\right) = T_o \sqrt{1 + \frac{16 z_o^2}{W^2}}.$$

Since z_o is typically about 1 μm and W about 250 μm,

$$T\left(\frac{W}{2}\right) \approx T_o \left[1 + \tfrac{1}{2} \frac{16}{(250)^2}\right] = T_o \left(1 + \frac{8}{62{,}500}\right),$$

and

$$T\left(\frac{W}{2}\right) \approx T_o.$$

Therefore the tension is essentially constant and equal to T_o all along the membrane.

It is also of interest to calculate the incremental strain of the membrane for a typical deformation of 1 μm. From Eq. [9],

$$\frac{\Delta W}{W} = \frac{\epsilon^2 V^4 W^2}{96 T_o^2 d^4} = 2.67 \left(\frac{z_o}{W}\right)^2$$

Again letting $W = 250$ μm,

$$\frac{\Delta W}{W} = 0.43 \times 10^{-4} = 0.0043\%.$$

This strain is quite small and well below the limit of elastic deformation which, for most metals, is a strain of about 1%.

The dynamic response of a membrane line element is obtained by using Newton's law, $F = ma = m\ d^2z/dt^2$, and equating this force to the elastic restoring force equal to $T_o d^2z/dx^2$. The resulting wave equation, with appropriate boundary conditions applied, has a solution that gives a vibration frequency of

$$f = \frac{n}{2W}\sqrt{\frac{T_o}{\delta t}}, \qquad n = 1,2,3,\ldots \qquad [12]$$

and

$$f_1 = \frac{1}{2W}\sqrt{\frac{T_o}{\delta t}}, \qquad [13]$$

is the fundamental mode. In these formulas, δ is the density of the membrane material, t is the thickness of the membrane, and T_o is assumed to be constant and unaffected by the motion. Also the formulas apply only to vacuum. In air there is mechanical damping caused by the gas below the membrane that resists deformation that tries to compress the volume of air. More is said of this later.

Similar analyses can be performed for the other geometries of membrane elements. For the case of the circular element, the corresponding equations of static deflection are

$$z = \frac{\epsilon V^2}{8T_o d^2}\left(r^2 - \frac{D^2}{4}\right), \qquad [14]$$

and

$$z_o = -\frac{\epsilon V^2 D^2}{32T_o d^2}, \qquad [15]$$

where D is the diameter of the element. Note that the central deflection is about one-half that of a line element with width D for the equivalent conditions of tension, voltage, separation, etc.

The resonant frequency equation for the circular membrane is

$$f = \frac{p}{\pi D}\sqrt{\frac{T_o}{\delta t}}, \qquad [16]$$

where p equals the values at which the zero order Bessel function of the first kind $J_o(p)$ equals zero ($p_1 = 2.40$, $p_2 = 5.53$, $p_3 = 8.65$ etc.).

The fundamental mode is therefore given by

$$f_1 = \frac{0.764}{D}\sqrt{\frac{T_o}{\delta t}}. \qquad [17]$$

For the case of a square element, the corresponding static deflection equation is[6]

$$z = -\frac{2W^2\epsilon V^2}{\pi^3 T_o d^2} \sum_{n\,=\,1,3,5,\ldots}^{\infty} \frac{1}{n^3}(-1)^{\frac{n-1}{2}} \left[1 - \frac{\cosh\dfrac{n\pi y}{W}}{\cosh\dfrac{n\pi}{2}}\right] \cos\frac{n\pi x}{W}, \qquad [18]$$

where W is the length of one side of the square. Similarly

$$z_o \approx -(0.037)\frac{\epsilon V^2 W^2}{T_o d^2}. \qquad [19]$$

The central deflection is not very different from that of the circular element with a diameter of W and is therefore also about one-half that of a line element of width W.

The resonant frequency equation for a square element is

$$f = \frac{1}{2W}\sqrt{\frac{T_o}{\delta t}}\sqrt{m^2 + n^2}, \qquad \begin{matrix} m = 1,2,3,\ldots \\ n = 1,2,3,\ldots \end{matrix} \qquad [20]$$

and

$$f_{11} = \frac{0.707}{W}\sqrt{\frac{T_o}{\delta t}}. \qquad [21]$$

4. Optical Considerations

In high-density read–write holographic memory systems, the holograms are recorded at a location on the storage medium where the

light from each undeformed page-composer element has been focused to a diffraction-limited spot. The size of this spot is inversely proportional to the width of the membrane element. When an element is deformed by applying voltage, the light spreads over an area much larger than the hologram, and little light is recorded. In the image reconstructed from the recorded hologram, a substantial contrast is obtained between the deformed and undeformed states of the membrane. An optical system that simulates this aspect of holographic memory operation is shown schematically in Fig. 3. This configuration

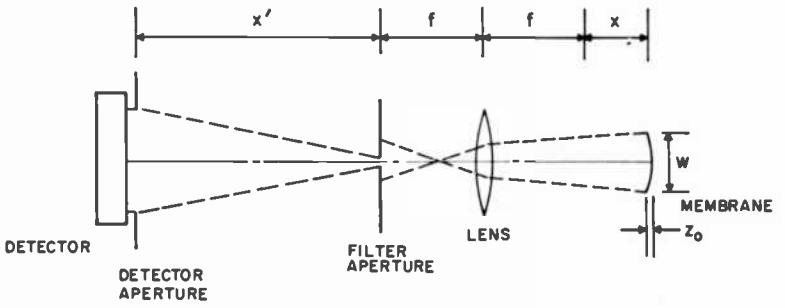


Fig. 3—Simulated optical memory system.

is a Fourier-transform filtering system having a low-pass spatial frequency response that is determined by the width of the filter aperture. With the condition $X'X = f^2$, the lens projects an image of the membrane element onto the detector aperture with a magnification $(M) = (X' + f)/(X + f)$. The detector aperture width is chosen to be MW, the size of the magnified image of the membrane element.

For this one-dimensional analysis of a line element, the light incident on the membrane is assumed to be collimated. The membrane is assumed to deform as a parabola, with a central deflection of z_o. When the membrane is undeformed, the collimated light is reflected through the lens, which focuses it to a spot small enough to pass through the small filter aperture. As the membrane deforms, the reflected light spreads over a large area in the plane of the filter aperture, and less light is transmitted through the aperture.

First, consider the geometric-optics approximation, which ignores diffraction effects. The parabolic mirror formed by the membrane has a focal length (f_m) of

$$f_m \sim W^2/16z_o \qquad [22]$$

for small deformations. The deviation of the originally collimated rays by the membrane produces a region of uniform illumination of width d in the plane of the filter aperture, where

$$d = fW/f_m. \qquad [23]$$

For a filter aperture of width

$$d_a = 2a\,\lambda f/W, \qquad [24]$$

where a is a constant to be chosen, and λ is the wavelength of the light; the light flux I transmitted by the aperture is

$$\begin{aligned} I &= I_o d_a/d, & d &> d_a, \\ I &= I_o, & d &\leqslant d_a. \end{aligned} \qquad [25]$$

This light forms the geometric image of the membrane element. I_o is the transmitted light flux for no membrane deflection. Combining the equations gives

$$\begin{aligned} I/I_o &= a\lambda/(8z_o), & z_o &> a\lambda/8, \\ I/I_o &= 1, & z_o &\leqslant a\lambda/8. \end{aligned} \qquad [26]$$

With $a = 1$, a contrast of 8:1 is predicted for a central deflection of one wavelength by this simple analysis.

A more rigorous analysis includes the effects of diffraction caused by the finite filter aperture. The constant a is recognized as the ratio of the filter aperture width to the width of the main lobe of the diffraction pattern caused by an isolated, undeformed membrane element. Over 90% of the light from an isolated, undeformed membrane would be transmitted by the aperture for $a = 1$. The image intensity distribution is given exactly by the squared modulus of the convolution of the complex amplitude of the light reflected from the membrane element, with the point-spread function of the imaging system as set by the filter plane aperture size. This distribution is integrated over the detector aperture to give the detector output I. Figs. 4, 5, and 6 show representative numerical results of the exact analysis for $a = 2.0$, 1.0, and 0.5, respectively. Also shown for comparison is the geometric-optics approximation represented by Eq. [26]. The approximation obviously becomes more accurate as the size of the filter aperture increases and the effects of diffraction are relatively

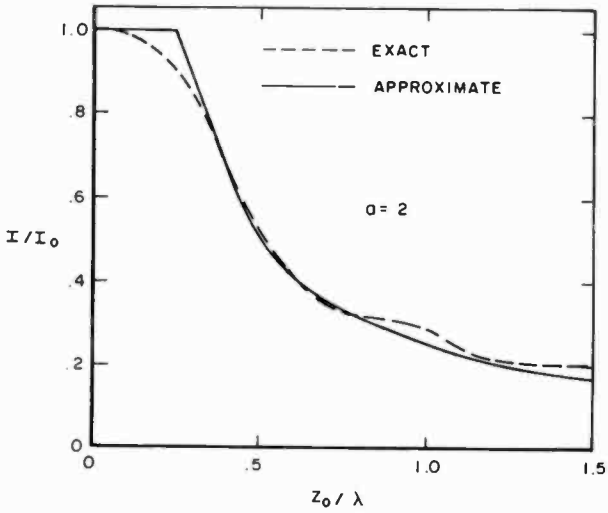


Fig. 4—Detector output versus central deflection computed for large filter aperture, $a = 2$.

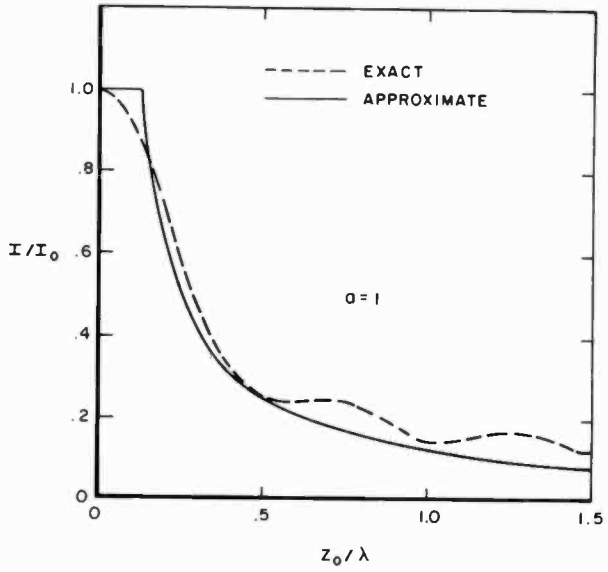


Fig. 5—Detector output versus central deflection computed for medium filter aperture, $a = 1$.

decreased. For a given membrane displacement (z_o), the trend is generally toward better contrast for smaller filter aperture width.

The exact curves shown in the figures are for the case in which there are wide, rigid supports on either side of the deformable membrane that are also illuminated and reflect the incident collimated light. Since diffraction by the filter aperture tends to blur the image of the sharp boundaries between deformed and undeformed regions

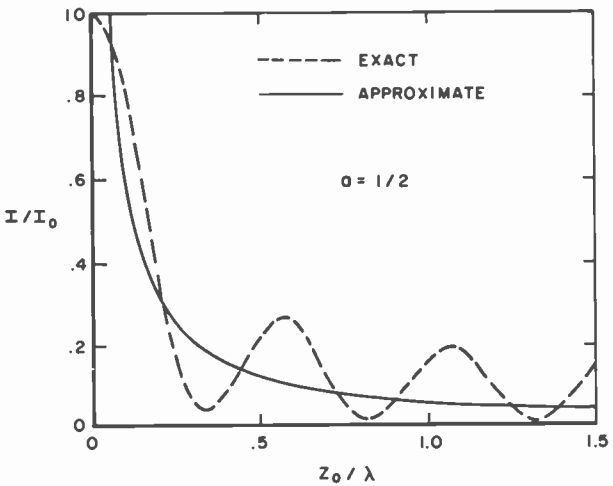


Fig. 6—Detector output versus central deflection computed for small filter aperture, $a = ½$.

of the reflecting surface, some of the illumination from the supports spills over into the detector aperture. This effect limits the asymptotic values of I/I_o for large deformations to 0.03 for $a = 2.0$ and to 0.06 for $a = 1.0$ and $a = 0.5$.

The oscillations in the detector output as z_o varies are caused by interference in the image between the spillover light from the supports and the phase-shifted light reflected by the deformed membrane. For the case where only the deformed region is illuminated, and no light is incident on the supports, the exact calculations show that the oscillations diminish in magnitude, contrast increases, and the curves more closely resemble the geometric-optics result. Also, as in the approximate analysis, localized illumination gives an asymptotic value of zero for I/I_o as z_o becomes very large.

The exact numerical calculations have, in addition, been made to include the cases of (a) initial flat displacement of the membrane from the level of the supports, (b) different reflectivities of the deformable

membrane and the support regions, (c) lateral displacement of the filter aperture, (d) a triangular-membrane deformation profile, and (e) detector aperture different from the size of the geometric image of the membrane element. Although the detailed variation of I/I_o with z_o is somewhat sensitive to these additional factors, the qualitative behavior is quite similar to that in Figs. 4, 5, and 6. Of particular interest is the case of a smaller detector aperture, for which the magnitude of the oscillations increases somewhat, but the contrast for large deformations is improved by the exclusion of spillover light when the supports are illuminated.

In summary, the geometric-optics analysis provides a reasonably close first approximation to the calculated diffraction-limited performance of an idealized membrane page composer line element in the intended optical system. For small deformations, the contrast improves as the filter aperture width is reduced. However, diffraction blurring reduces the maximum available contrast as the filter aperture is narrowed when the membrane and its supports are uniformly illuminated. This degradation is reduced by narrowing the detector aperture, and is eliminated by restricting the incident illumination to the deformable region of the element only.

5. A Membrane Page Composer

The foregoing involved membrane light valves without concern for how array operation can be attained. Since a light-valve element does not have true threshold or memory properties, these must be provided at each bit location by other means. We have chosen semiconductors for this purpose. Static storage is obtained with flip-flops at each bit, while dynamic storage is possible if MOS transistors are used as gates at each bit to control the charge and discharge of the membrane-electrode capacitance. Still remaining is the problem of connecting each element electrode to the semiconductor circuitry. Since, in a large array, it is impossible to bring out leads from each element, we make use of feedthrough connections in the substrate itself leading from electrodes on one side to the semiconductor drivers on the opposite side of the substrate. Interconnections and pads for external wires to the control and selection electronics can also be included on the backside of the substrate, and semiconductor chips can be beam-lead bonded at the appropriate locations.

The three major parts of a prototype 8×8 page composer can now be described. First is the membrane light-valve array. This consists of circular elements, 0.75 mm in diameter, spaced on 1.5-mm centers.

The substrate with feedthroughs is the second major part of the device. After studying several alternatives, we chose Fotoform or Fotoceram materials. This glassy material can be exposed with ultraviolet light passing through a mask and can be subsequently developed and preferentially etched to close tolerances. Holes etched through the thickness of the substrate can then be filled in with conductive material by various techniques. We have successfully used several kinds of silver epoxy for this purpose. The advantage of this material is that regular patterns can be etched that register precisely with other patterns needed for fabrication of the device. The glassy substrate can also be converted into a tough ceramic-like form by a final heat-treatment step. The substrate thickness is about 1.25 mm.

The third major part of the device includes the semiconductor chips, the metallization patterns, and their interconnections on the backside of the substrate. One method of obtaining dynamic storage utilizes a single MOSFET gate at each element. This simplifies the circuitry on the semiconductor chip and the back-plane wiring considerably over what would be needed if flip-flops were used. In addition, by designing the chip to include "carryovers" that bridge interconnection lines on the substrate on either side of the chip, only a single layer of metallization is needed on the backside of the substrate. Each chip includes four transistors, each of which controls one light valve so that only 16 chips are needed for the 8×8 array.

As an intermediate step, devices were designed to have light valves on a substrate as described, but with a metallization pattern on the backside that brought out each feedthrough to a connection pad, to which an external wire could be connected. This device would allow testing and evaluation of the light-valve–substrate combination to proceed without the semiconductor circuitry, which could be included on later samples.

A clearer picture of the prototype device can be obtained by considering the details of a 4×4 array for illustrative purposes. Fig. 7 shows the development of such an array. In the final fabrication sequence, step (e) precedes step (d) because the membrane is the most delicate component of the device and is therefore put on last. Fig. 8(a) shows a pictorial representation of section AA of Fig. 7(c). The backside of the substrate is shown in Fig. 8(b) with the chips in place. Fig. 8(c) shows a schematic layout of the transistors and the wiring on a single chip. Two other views of the chip on the substrate can be seen in Fig. 9. Note that the chip is suspended above the surface of the substrate by the beam lead connections. Also note that the X and GND leads run beneath the chip to continue across the substrate,

while the Y leads are continued across the substrate by means of a bridge or tunnel on the chip itself. The actual chip size is 1 mm square, while the conducting lines are 0.125 mm wide.

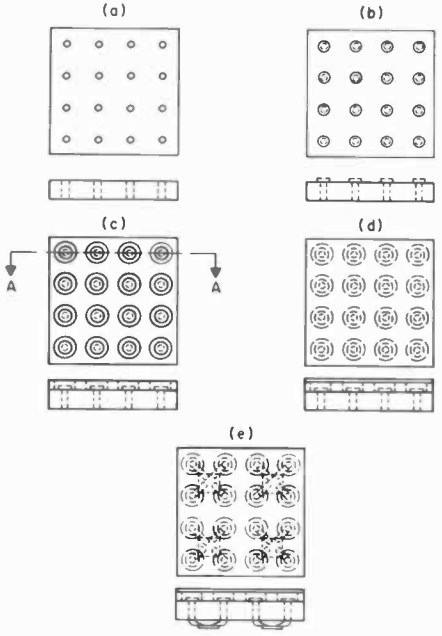


Fig. 7—Sequential steps in the development of a 4 × 4 array of membrane light values: (a) Fotoceram substrate with holes, (b) after filling holes with conductive material and forming electrode caps, (c) after forming a support structure, (d) after adding membrane, (e) after bonding semiconductor chips to backside of substrate.

As can be seen in Fig. 8(c), an X line is connected to one row of MOS gates while a Y line is connected to one column of drains. The source of each MOS is connected to a respective feedthrough. In operation, an X, or word, line is activated causing the MOS channel to be conductive. The desired voltage pulse ("1" or "0")on each Y line then appears at the corresponding feedthrough electrode, and acts on the grounded membrane of that element accordingly. With the gate disabled before the data pulse ends, the charge on the capacitance element formed by the electrode–membrane combination will leak off slowly in a time determined by the leakage current of the transistor. This current can easily be made small enough for storage times of many milliseconds. As long as the cycle time of the page composer is short compared to the leakage time, or if refreshing of the array is allowed, effective memory is obtained with this technique.

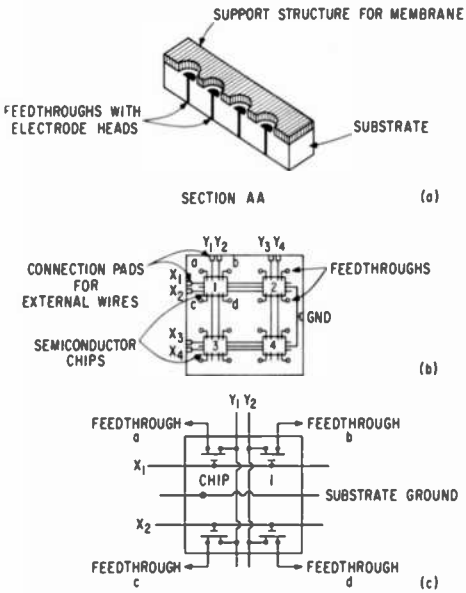


Fig. 8—(a) A pictorial representation of Section AA of Fig. 7(c). (b) A schematic representation of the backside of the substrate showing the semiconductor chip in place and its associated wiring. (c) A schematic representation of a semiconductor chip showing 4 MOS transistors and their associated wiring.

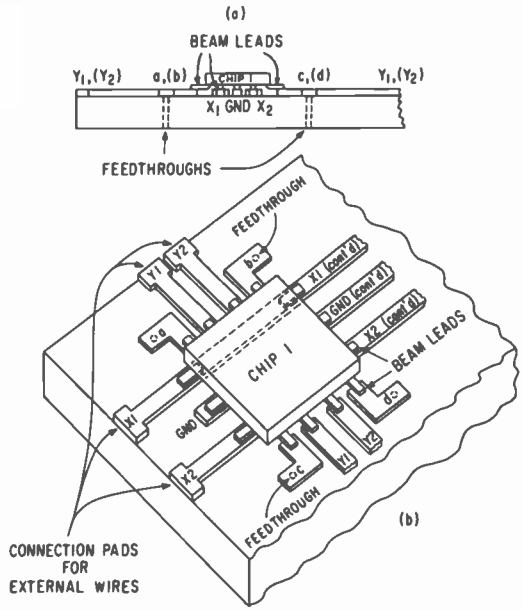


Fig. 9—(a) A detailed view of a semiconductor chip and its associated leads on the backside of the substrate. (b) A pictorial view of the chip on the subtrate.

6. Experimental Results

6.1 Line Elements

The first experiments were made with membrane line elements that were readily obtainable because of previous work that had been done.[5] Samples were tested to determine the separation between membrane and electrode and the metal membrane thickness needed to obtain deflections sufficient to produce a contrast ratio of about 10:1 at the detector with an applied voltage of 50 V or less. The contrast number was felt to be the minimum that might be tolerated in an optical memory system, while the voltage was the minimum breakdown voltage specified for the drive transistors of the prototype device. The vehicle used for this purpose was a sample on a 5×5 cm glass substrate with nine line elements of the type shown in Fig. 1(a), except that wide supports (~ 3 mm) were used between elements. An aluminum support structure was used. Each cell was about 0.25-mm wide with electrodes half as wide. It was found to be advantageous to deposit silicon monoxide over the aluminum line electrodes to prevent a local defect from shorting out the entire line element. Dozens of such samples with numerous variations were made. Electrical and optical properties of the line elements were determined and recorded. Different materials, either evaporated or plated, were tried for the membrane, but because of early success and better control with plated nickel films, most of the samples were made with this type of membrane.

After a visual inspection of the sample, checks were made of electrode continuity and for shorts between membrane and electrodes. Next, data were taken in the simulated optical memory system of Fig. 3, with a helium–neon laser light source and a photomultiplier tube detector. Usually, the detector aperture was about the size of the geometric image and the filter aperture was about equal to $2\lambda f/W$. The light intensity (I) at the detector was measured as a function of the applied voltage, and photos such as that of Fig. 10 were taken, from which contrast could be determined. For all such photos, the zero light level is at the top line of the graticule and light intensity increases in the direction shown; the scope sweep speed is 5 msec/div., unless otherwise noted. Data from interferometer measurements were used to plot the central deflection (z_o) as a function of applied voltage as seen in Fig. 11. The same data have been replotted on log-log scales as seen in Fig. 12. The plated-nickel membrane was typically quite flat, but was located below the support structure surface by about 0.1 μm or less, as indicated on Fig. 11. This probably occurred

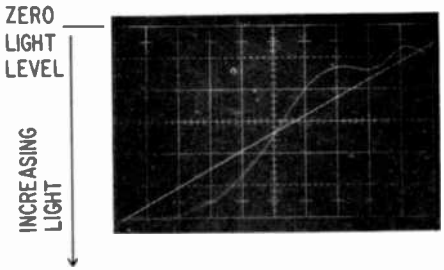


Fig. 10—Detector output versus applied voltage. The straight line shows the sawtooth voltage, while the curved waveform shows the light variations. The voltage scale is 20 V/div. The data are for line 3 of sample R-161, which had a membrane 0.36 μm thick and a support structure 4.2 μm high.

when the photoresist was polished before the membrane was put on. Since the resist is softer than the support structure it would polish faster, resulting in an undercut of the type seen. Another contributing factor might be a slight sagging or shrinkage of the resist after polishing and before the membrane is deposited. Because the membrane surface is quite flat, however, the small offset from the surface of the support structure does not introduce significant problems. By combining the two previous sets of data, I versus z_o could then be determined as shown in Fig. 13.

Test samples were fabricated with support heights ranging from less than 2 to more than 6 μm. Membrane thickness ranged from 0.18 to over 0.5 μm. The best results obtained from these elements were

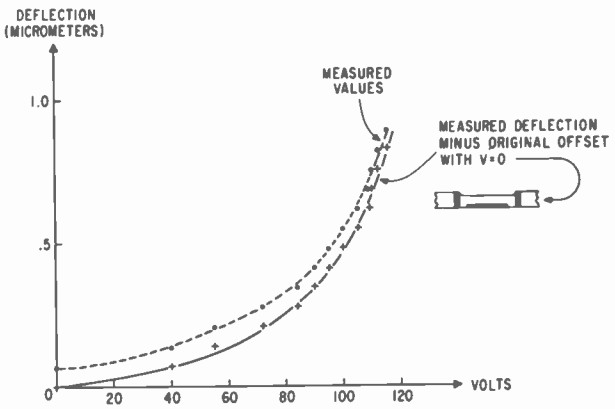


Fig. 11—Deflection versus voltage for line 3 of R-161.

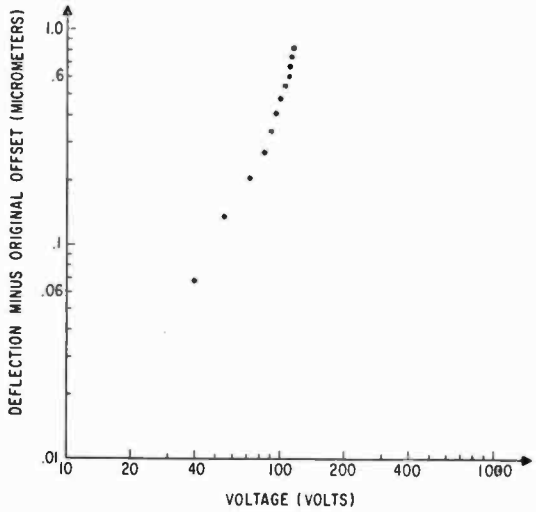


Fig. 12—Deflection versus voltage for line 3 of R-161 replotted on log-log scales.

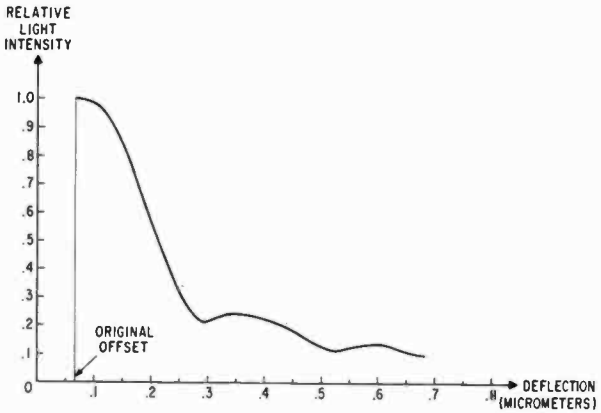


Fig. 13—Detector output versus deflection for line 3 of R-161. This plot was made by combining the data of Fig. 10 with the curve of Fig. 11 drawn through the deflection values as measured from the support structure height.

well within our requirements. For example, one sample produced a 15:1 contrast with 40 V applied. Several others showed a contrast of about 10 with applied voltages ranging from 25 to 50 V. The favorable performances indicated that a line element with a plated-nickel membrane about 0.4 μm thick, and a membrane-to-electrode spacing of about 2 μm would meet our goals, while maintaining a reasonable yield from the fabrication cycle.

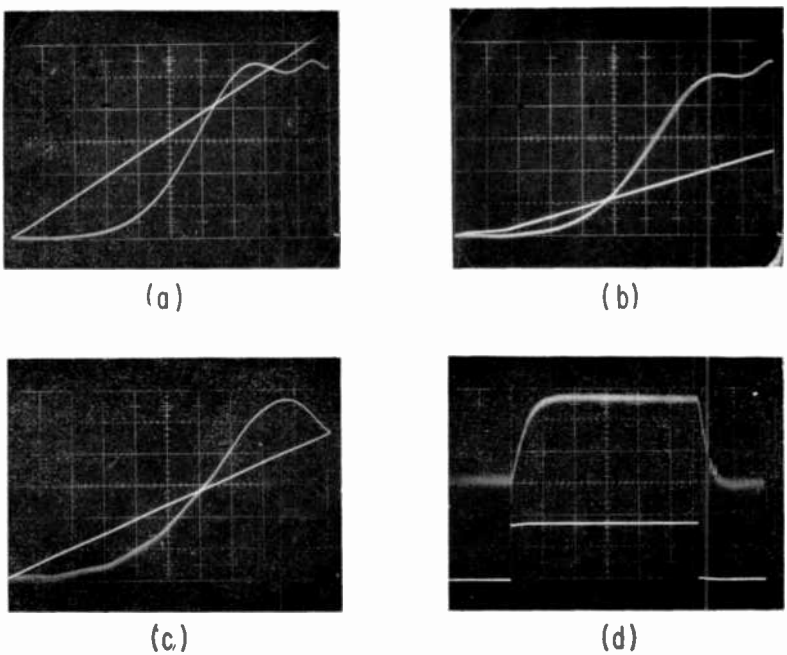


Fig. 14—Detector output versus applied voltage. The voltage scale is 20 V/div. for all photos except (c), where it is 10 V/div. (a) line 5 of R-172, which had a membrane thickness (t), of 0.4 μm and a support structure height (d), of 4.0 μm; (b) line 3 of R-178 for which $t = 0.22$ μm and $d = 3.3$ μm; (c) line 1 of R-239 for which $t = 0.5$ μm and $d = 3.3$ μm; (d) pulse response of line 4 of R-238 for which $t = 0.4$ μm and $d = 3.15$ μm. The top waveform shows the detector output with the bottom voltage pulse applied. The horizontal time scale is 20 μsec/div. for this photo.

Some other data obtained from line elements are shown in Fig. 14. Light intensity versus voltage is seen for three samples in (a), (b), and (c) of that figure. Pulse experiments were also run to determine switching speed and fatigue behavior of the elements. The pulse response in air of one element is shown in Fig. 14(d). Another element of this sample was cycled 189×10^6 times (5 hours and 15 minutes of pulse excitation at 10 kHz). No change was discernible either

in the light reflected from the element, or in the appearance of the element itself after excitation, indicating that fatigue effects, if any, are small.

Switching of a line element was done both in air and in a pumped-out enclosure. Operation is slower in air than in vacuum because the gas in the space beneath the membrane resists compression as the film deforms. An early sample (MC7) with an evaporated aluminum–zinc membrane 0.60 μm thick and 6.5 μm high support structure was pulse excited in vacuum and caused light oscillations at 330 kHz. By using a slower rise time excitation pulse (e.g., 10 μsec), the amplitude of the oscillations could be reduced to very small values. In air, the same line element was seen to have a fast region of response followed by a slower tail as the air was forced out of the element space. More results in vacuum were obtained with other types of elements and are described later.

6.2 Square Elements

Membrane elements 0.25-mm square were fabricated on a two-dimensional grid structure made on a 5×5 cm glass substrate. Line electrodes were formed that were 0.15 mm wide, and ran in one dimension only under one row of square elements. A silicon monoxide blanket about 1 μm thick was deposited over the substrate to insulate the electrodes from the aluminum support structure, which was about 1.25 μm thick. Except for a few shorts, there were no problems with this technique and adhesion of the supports to the SiO was good. The membrane was plated nickel about 0.4 μm thick. Three samples of this type were extensively tested. Data were taken for 0.25-mm-square elements activated by a single line, and also for simulated large 0.75-mm-square elements composed of a 3×3 group of the 0.25-mm-square elements, and activated by energizing the three line electrodes beneath that group simultaneously. Several elements of each kind were tested on every sample. In the simulated optical memory system, the average contrast measured for the single elements tested was 10.4 at an average applied voltage of 33.1 V, with an aperture in the focal plane of the imaging lens of about 1 mm diameter. This aperture is large enough to let through the first-order light of the support grid structure. With the aperture reduced to 0.8 mm diameter to exclude the first-order light from the image, contrast increased substantially—as much as a factor of 2 in some cases. Contrast as high as 20:1 was obtained for a single element for this case. Simulated multiple-element light valves were also found to operate with a contrast of 10:1 or greater at voltages below 50 V, with various aperture sizes at the focal plane of the

imaging lens, thereby meeting our performance requirements. Another test was run on a single-element valve using a fixed focal-plane aperture of 1-mm diameter and various aperture sizes at the detector. The results are shown in Fig. 15. It can be seen that in the range measured, the contrast improves as the detector aperture is decreased. The highest contrast of 30 was obtained with an aperture about 0.68 times the nominal image size. Of course, a reduced detector aperture also means reduced light at the detector, but a trade-off might be desirable in some cases.

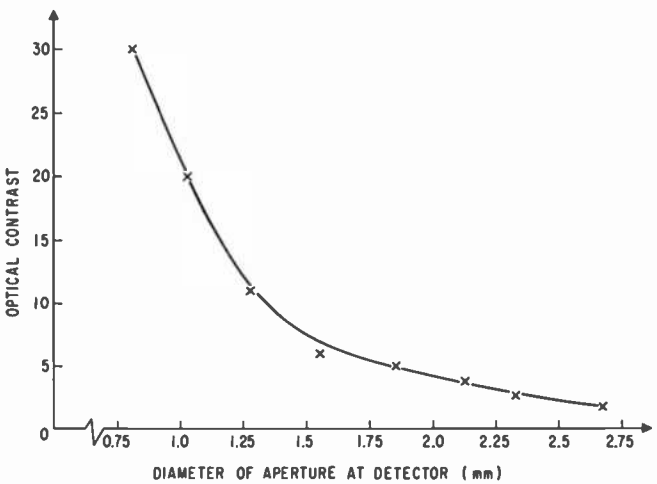


Fig. 15—Contrast versus detector aperture size for one square element on line S1 of sample 2D-5.

A curve of light intensity versus voltage for one square element is shown in Fig. 16(a), while the pulse response of an element is shown in Fig. 16(b). Corresponding pulse responses of elements from two other samples are shown in Fig. 16(c) and (d). Note that in air, the rise time for the three elements varies from less than 10 μsec for one element, to 200 μsec for another. The response in vacuum of three elements is seen in Fig. 17, which shows the oscillations in the reflected light from the membranes as they mechanically vibrate. Note that the resonant frequencies range from 530 to 610 kHz, which is a small variation compared to that of the responses in air. The reason for this appears to be that the response in air is dependent on the number of micrometer-size holes in the membrane of an element, which

are a by-product of the fabrication process. The greater the number of holes, the faster is the response, presumably because of less damping in air. A microscope inspection of the samples showed that 2D-5 had the greatest number of holes per unit area, 2D-3 had fewer holes per unit area, and 2D-4 had the fewest holes per unit area, supporting the above hypothesis.

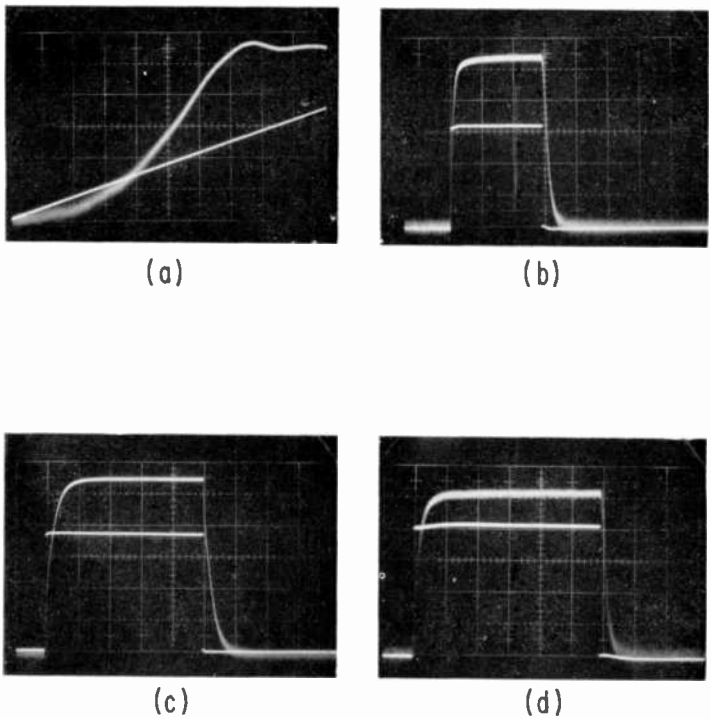


Fig. 16—Detector output versus applied voltage. The voltage scale is 10 V/div. for all photos. (a) An element of line S1 on sample 2D-3; (b) the pulse response of an element of line S1 on 2D-3 (the time scale is 50 μsec/div., the applied voltage pulse is shown on the same base line as the light output); (c) an element of line S1 on 2D-4 (the time scale is 200 μsec/div); (d) an element of line S1 on 2D-5 (the time scale is 10 μsec/div).

6.3 Circular Elements

Membrane elements 0.75 mm in diameter were made only on Fotoform–Fotoceram substrates with feedthroughs. They were accessed by individual wires connected to the metallization pattern on the backside of the substrate, which brought out a conductor from each feedthrough to a pad. In this respect, the circular elements differ from those of the line and square elements, which were made on ordinary glass with

evaporated electrodes and connection pads on the same side of the substrate as the membrane. The use of feedthroughs introduced complications that did not exist for the elements previously described, and that affected operation of the circular elements. For example, the silver epoxy used as the feedthrough material is quite grainy and soft relative to the substrate. Consequently after the 0.25-mm-diameter

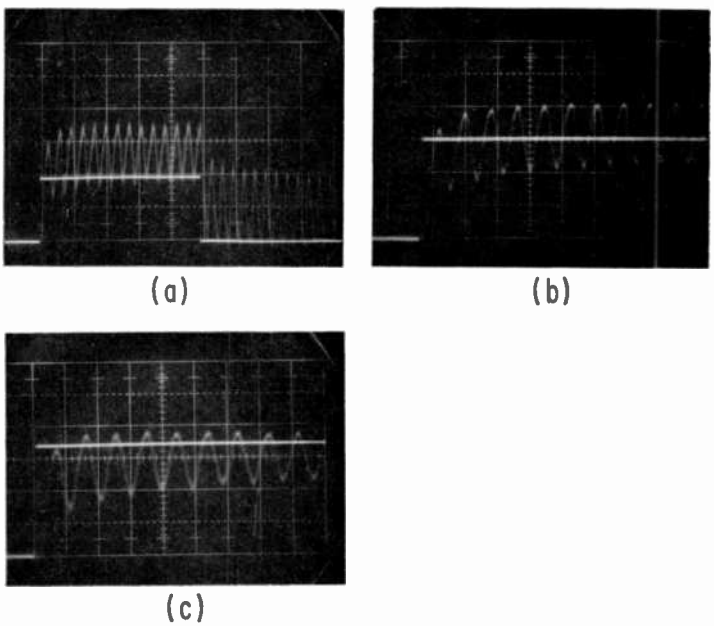


Fig. 17—Pulse response in vacuum. The voltage scale is 10 V/div. for all the photos. (a) An element of line S1 on 2D-3 (the time scale is 5 μsec/div., $f_{11} = 580$ KHz); (b) an element of line S1 on 2D-4 (the time scale is 2 μsec/div., $f_{11} = 610$ KHz); (c) an element of line S1 on 2D-5 (the time scale is 2 μsec/div., $f_{11} = 530$ KHz).

holes were filled with the epoxy and the substrate surface was ground and polished, the silver feedthrough surface was typically undercut and irregular. Therefore, if the feedthrough itself were used as the drive electrode, as it was in the earliest samples, it would be farther from the membrane than the support structure height, and nonuniform over its area as well as from element to element. Some improvement was obtained by evaporating 0.5-mm-diameter aluminum caps over and around the silver feedthroughs, and then grinding them back to get smoother, flatter electrode surfaces. This helped somewhat but

the grinding operation again produced nonuniformities in electrode height that caused corresponding variations in the voltage needed to deflect membrane elements a given amount. The next modification involved offsetting the feedthrough element from the optical element area, and insulating the feedthrough–electrode combination from the

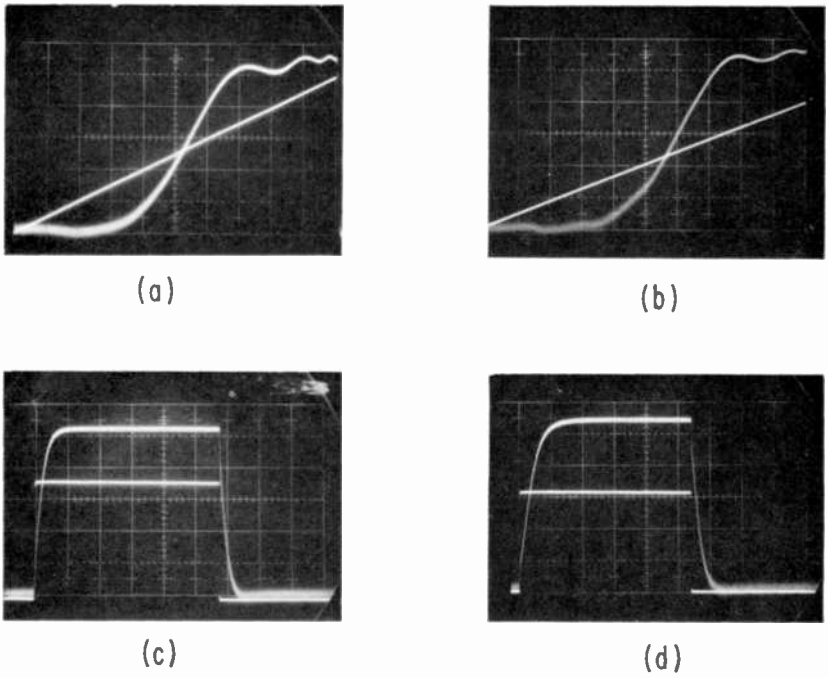


Fig. 18—Detector output versus applied voltage. The voltage scale is 10 V/div. for all photos. (a) Element (5,5) of sample RX1; (b) element (4,5) of RC6; (c) pulse response of element (5,5) of RX1 (the time scale is 0.5 msec/div); (d) pulse response of element (4,5) of RC6 (the time scale is 1.0 msec/div).

metal support structure with an SiO blanket. Thus, the drive electrode in the effective element area would be a deposited aluminum layer similar to that used for elements made on ordinary glass. Samples with plated-nickel membranes were made and tested with each of these types of electrodes used to drive the circular elements.

Circular elements were tested in the usual optical system to determine contrast and pulse response. Results for two samples are shown in Fig. 18. Maximum contrast is seen to be about 15:1 in the light intensity versus voltage curves. The pulse response of the two

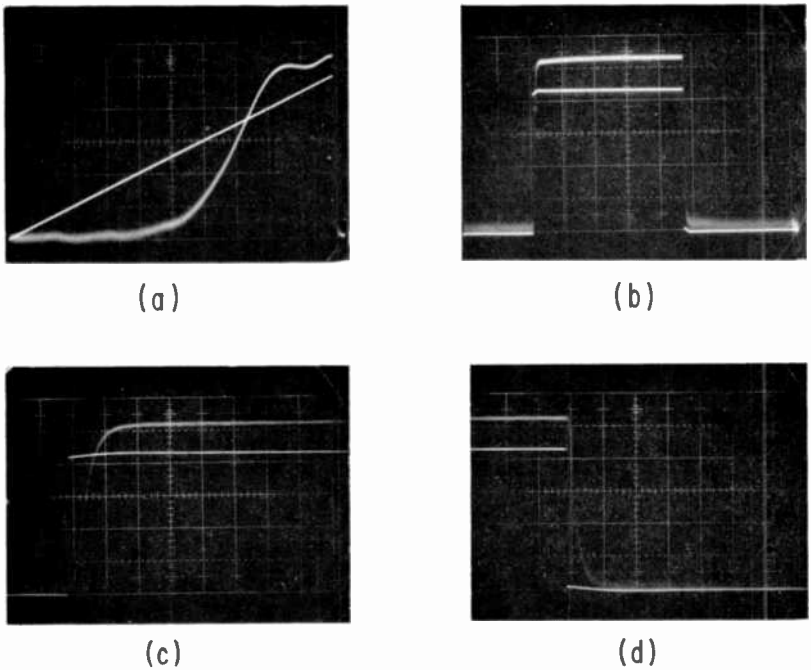


Fig. 19—Detector output versus applied voltage for element (1,1) of sample RC2. The voltage scale is 10 V/div. for all photos. (a) Response to sawtooth voltage; (b) pulse response (the time scale is 100 μsec/div); (c) rise time of response (the time scale is 10 μsec/div); (d) fall time of response (the time scale is 10 μsec/div).

elements shown is seen to be rather slow. Again, it was found that switching time would vary markedly, with the fastest time associated with elements having many holes in the membrane. A fast response can be seen in Fig. 19, which shows light intensity versus voltage for the element as well as rise and fall times. One circular element of sample RX1 with a 0.6-μm-thick membrane was also tested in vacuum and found to resonate at 125 kHz when pulse-excited.

Experience with the 0.75-mm-diameter circular elements indicated that a device with a 0.8-μm-thick plated-nickel film for the membrane and a separation of about 2 μm between electrode and membrane would meet our performance requirements.

6.4 8 × 8 Devices

Completed 8 × 8 arrays with individual access to each element were mounted on a holder and electrically connected to an electronics test

panel that could select and drive a single element or scan an array in a sequence determined by an external trigger. By viewing the image of the array on a screen or letting it fall on a photodetector in the simulated optical memory system, operation of devices could be evaluated. More than half a dozen samples were tested this way including some with multiple-element light valves.

Some of the data obtained from light valves on 8×8 arrays have already been described in the section on circular elements. Other data involve uniformity over the array and a variation of properties with changes in dimensions or the geometry of the element. Using the test electronics an 8×8 array would be scanned at a low voltage, and the corresponding contrast of elements would be noted at the image of the array on a screen. As the voltage was increased in steps, further changes would be observed until either the entire array was operating with good contrast, or the reasons why it was not operating with good contrast were determined. One fabrication problem involved collapsed or partially collapsed elements on a completed sample. The first arrays were made using the feedthroughs directly as electrodes, which resulted in many problems as previously described. The samples with 0.5-mm-diameter aluminum caps performed much better but were still nonuniform. The grinding step would leave the caps higher in the center of an electrode than at the periphery, and produce considerable variation in the height of an electrode from element to element on the same sample. One sample with cap electrodes, when completed had all 64 elements intact, flat, and quite specular to the eye. However, two shorts between membrane and electrodes were found during electrical checkout. While other elements worked, they did not operate at the same applied voltage. Some elements had high contrast at 30-35 V, while others did not have good contrast until perhaps 45 V had been applied. Later samples were made with the feedthroughs offset from the optical element area and the drive electrode beneath the entire 0.75-mm-diameter element to achieve better uniformity. A blanket layer of silicon monoxide was used to insulate the feedthroughs from the aluminum support structure. Several shorts between the feedthroughs and supports were typically found on such samples. We then decided to try silicon monoxide support structures, thereby eliminating the need for aluminum supports and also easing the short situation. Early results with this technique have been quite encouraging and we are pursuing it further.

Other tests made on 8×8 devices involved simulations of the prototype device by connecting a MOSFET to an electrode of a sample, and evaluating the performance of the combination. Several semicon-

ductor chips with four transistors on each, as designed and made by RCA Somerville, were cut, packaged in standard 14-lead DIP cases, and delivered to us. Tests showed that all of the original specifications had been met. The breakdown voltage between any two terminals of a MOSFET exceeded 50 V and in some cases was 70 V or more. In addition, pulse tests with simulated loads verified that the transistors could drive an element to 50 V in 1 μsec or less, with appropriate pulses on the gate (X) and data (Y) lines. Finally, the transistors on a chip were used to drive membrane elements on an 8×8 device by making the proper connections to the individual access pads. Some of the results are shown in Fig. 20. The top waveform is the data pulse (Y), the middle waveform is the gate pulse (X), and the bottom waveform is the light output from the detector. In (a), the gate pulse overlaps the data pulse completely, so that the element charges and discharges again when the data pulse occurs. In (b), the pulses are narrowed so that the gate width is 10 μsec. The storage effect is seen when the gate pulse terminates while the data pulse is still present. In (c), the pulses are narrowed further so that the gate width is about 1 μsec. Note that the contrast has decreased because the element did not charge fully in 1 μsec. In (d), the gate and data pulse amplitudes were increased to 40 from 37 V. Note that full contrast is now obtained since the element does charge sufficiently in 1 μsec. In (e), the same waveforms are seen on an expanded time scale. These results are seen to verify proper operation of the transistor-light valve combination.

7. Discussion

It is of interest to consider the experimental data with reference to the analyses previously performed for electromechanical operation. Note that Eq. [5] predicts that the central deflection (z_o) of a line element with a constant built-in tension, is proportional to the square of the applied voltage (V), while Eq. [11] predicts that z_o of a line element with tension resulting only from elongation of the film is proportional to the two-thirds power of V. Fig. 11 shows the experimentally found dependence of z_o on V. Fig. 12 indicates two broad regions of exponential influence. For small deflections, the exponent of V (the slope of the lines through the points) appears to have a value of about 1.7 to 1.9, depending on how one draws a line through the points. For deflections larger than about 0.3 μm, the exponent is about 3.6. For the region of small deflections, we would expect that the built-in tension dominates, since there is little elongation. Therefore, the expon-

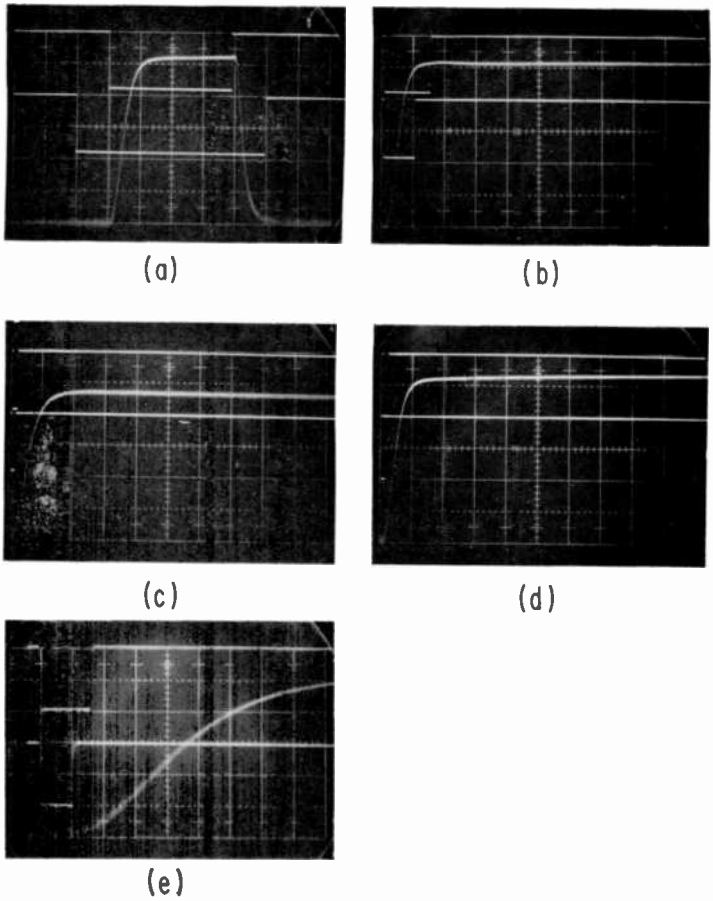


Fig. 20—Waveforms showing simulated operation of a prototype device. The voltage scale is 20 V/div. for all the photos. The zero light level is at the top line of the graticule. The time scale is 10 μsec/div. for (a), (b), (c), and 1 μsec/div. for (d).

ent should be close to 2. As the deflections get larger, the assumption that $z_o \ll d$ no longer applies and one would expect that for fixed T_o, an incremental change of voltage would produce even greater deflection than predicted by Eq. [5], since the center of the element membrane is closer to the electrode than the nominal separation. This effect appears to completely override the effect of increasing tension caused by elongation. In fact, for deflections that are an appreciable fraction of the original separation, a runaway situation can and sometimes does eventually take place with the thin film collapsing on the electrode, or tearing. The above data suggest that the built-in tension

in the deposited film is much larger than that contributed by stretching of the film. The vacuum tests support this view because the vibration frequency of the membrane did not vary appreciably with changes in the amplitude of the exciting pulse. Another point in support of this view can be made by calculating the tension in a film from Eq. [13] repeated here for convenience

$$f_1 = \frac{1}{2W}\sqrt{\frac{T_o}{\delta t}}.$$

For the line element of MC-7 (for which the resonant frequency of 330 kHz was measured and for which $W \approx .25$ mm, $\delta \approx 2.7$ gm/cm^3, and $t \approx 0.6$ μm), it turns out that T_o from the above equation is 44 newtons per meter (N/m). This represents the total tension in the film regardless of the cause. Now we can estimate the component of tension in the film caused by elongation from Eq. [8], $T = (\Delta W/W)Et$. We have also previously calculated the incremental strain, $(\Delta W/W)$, for a deflection of 1 μm of a cell 0.25 mm wide and found it to be 0.43×10^{-4}. Using $E = 7.3 \times 10^{10}$ N/m^2 for aluminum–zinc in the equation, we can solve for the tension to find $T \approx 1.9$ N/m, which is less than 5% of the total tension in the film, even with a deflection of 1 μm. Thus, the tension introduced in the film during fabrication appears to be the dominant one. The total strain in the film would be about 0.1%, which is still small and well within the elastic limit of most metals. The experimental data support this point since no change in the operating characteristics of an element was seen after 189 million cycles of excitation. It should also be noted here that in other work with membrane elements,[3] no fatigue effects were seen even after 10^{12} cycles of operation.

The pulse response of membrane elements both in air and in vacuum showed that fast operation of these light valves is possible. In vacuum, vibrations at resonance must be avoided. This can be accomplished by using drive pulses having rise times that are slow relative to the natural vibration period of the element, but that can still be in the 5-10 μsec range. Another possibility is to bleed in enough air to critically damp the oscillations. In air, it was noted for both the square and the circular membrane elements that the speed of response is strongly affected by the number of micrometer-sized holes in the membrane. Fastest response (10 μsec) was seen for elements with the most holes, which allow for the easy passage of air into and out of the membrane-to-electrode space.

The analyses of optical performance were useful in understanding

operation of the light valves and in predicting experimental responses. The geometric optics approximation was useful in predicting that about 1 μm of deflection would result in an optical contrast of about 10:1 at the detector. This value in turn was used to calculate that 50 V or less might be sufficient to operate a 0.25-mm-wide or larger element. The more complicated computer analysis, which included considerations of diffraction effects, also yielded useful results. One of these was the prediction of the oscillatory portion of the curve of the light intensity at the detector as a function of deflection, as can be seen in the calculated plot of Fig. 5, as well as in the numerous photos showing I versus V. The agreement between the calculated and observed behavior can be seen by comparing Fig. 5 with Fig. 13. The two plots are quite similar, with the only appreciable discrepancy occurring in the location of the peaks and valleys of the oscillatory portion of the curves. The factors that may be contributing to these differences are the nonparabolic deformation of the element in the actual case, the differences in the reflectivity of the membrane and support structure, and the actual placement of components such as the focal-plane aperture in the experiment. The qualitative analytical results concerning variations in illumination and in aperture size at the focal plane and/or at the detector have also been verified by experimental observations. It is clear that system trade-offs are possible to obtain high-contrast operation of a membrane light valve by the appropriate choice of aperture size used at the detector and in the Fourier transform plane.

Experience with the fabrication of light valves on ordinary glass substrates and on Fotoceram substrates with feedthroughs clearly pinpointed the problem areas which need to be worked on. Common to all the valves is the need for a defect-free membrane with no collapsed elements. Although difficult, this has been accomplished and can be done. In addition, the use of multiple-element valves alleviates the problem somewhat. There are also difficulties associated with the feedthroughs. Different materials and improved techniques should be investigated for filling the etched holes and smoothing the feedthroughs.

8. Conclusions

The production of a membrane light-valve page composer for optical memory systems appears feasible. Membrane light valves located over a ½ inch square area were operated at high speed with good contrast, efficient utilization of light, and without fatigue. The electromechanical and optical performance of membrane elements have been analyzed

and agree well enough with experimental results to give confidence in the models. It is now clear that membrane elements make excellent light valves. What remains to be shown is that arrays of such valves can be produced without serious defects and with enough uniformity to operate within practical tolerances. The remaining problems appear to be small ones that should yield to technological effort.

Several arrays of 64 elements on substrates with feedthroughs allowing access to individual electrodes from the backside have been made and tested. Single light valves of the arrays have been successfully operated with transistors designed and produced for selection and storage at a bit location. In this way, simulation of the operation of a prototype page composer with semiconductor chips bonded to the substrate was accomplished.

Acknowledgments

It is a pleasure to acknowledge the numerous contributions to the achievements of this project. Discussions with V. Christiano and J. A. van Raalte in the early phases of the effort were informative and helpful. V. Christiano, J. B. Harrison, and J. N. Hewitt showed much skill and ingenuity in fabricating the light valve devices. E. M. Nagle was responsible for the design and construction of the test electronics packages. Thanks are also due many members of the RCA Solid State Technology Center at Somerville for their aid with the page-composer semiconductor circuitry.

References:

[1] J. A. Rajchman, "Promise of Optical Memories," **J. App. Phys.**, Vol. 41, p. 1376, March 1970.
[2] W. C. Stewart and L. S. Cosentino, "Some Optical and Electrical Switching Characteristics of a Lead Zirconate Titanate Ferroelectric Ceramic," **Ferroelectrics**, Vol. 1, p. 149, July 1970.
[3] K. Preston, Jr., "A Coherent Optical Computer System using the Membrane Light Modulator," **Trans. IEEE Aerospace and Electronic Sys.**, Vol. AES-6, p. 458, July 1970.
[4] F. Reizman, "An Optical Spatial Phase Modulator Array, Activated by Optical Signals," **Proc. 1969 Electro-Optical Systems Design Conference**, p. 225, Sept. 1969.
[5] J. A. van Raalte, "A New Schlieren Light Valve for Television Projection," **Appl. Opt.**, Vol. 9, p. 2225, Oct. 1970.
[6] S. Timoshenko, **Theory of Elasticity**, McGraw-Hill Book Co., New York, (1934), p. 245.

The Insulated-Gate Field-Effect Transistor —A Bipolar Transistor in Disguise

E. O. Johnson

RCA International Licensing, New York, N. Y.

Abstract—Straightforward physical arguments show that when the IGFET specific gate capacitance is increased, such as by decreasing insulator thickness or by increasing the dielectric constant, device performance will change toward the low impedance, high transconductance characteristics uniquely typical of bipolar transistors. At relatively large specific capacitances, the characteristics become remarkably similar to those of bipolar transistors. The analysis shows that there is a fundamental trade-off between impedance level, transconductance, transconductance/current ratio, and power gain in IGFET's. The high input impedance uniquely typical of the IGFET stems from the gate–insulator capacitance and not from any unique behavior of field-effect action.

An explanation is given for the previously unexplained limiting value of **e/KT** experimentally observed for the transconductance/current ratio at low IGFET drain currents. Modification is suggested to correct existing IGFET theory, which incorrectly predicts unlimited increase in the transconductance/current ratio as the specific gate capacitance is increased.

The above results are derived from the behavior of the high/low dynamic junction that connects the source and channel, a junction that has the **e/KT** behavior of the emitter/base junction in the bipolar transistor, and whose presence and behavior seem to have been generally unappreciated. One interesting consequence of this junction is that it results in a dynamic control capacitance identical in form and magnitude to that of the diffusion capacitance in the bipolar transistor.

Introduction

Although both the Insulated-Gate Field-Effect Transistor (IGFET) and the bipolar transistor operate on charge-control principles,[1,2] these devices are widely believed to be intrinsically different in the details of their operation.[3] On the one hand, the bipolar transistor is characterized as a device with a perfect e/kT control action that results in a high transconductance, a transconductance/current ratio that reaches the theoretical maximum value of e/kT, and a low input impedance that is caused in a large measure by the shunting action of a dynamic capacitance known as the diffusion capacitance. On the other hand, the IGFET is widely characterized as a device having a relatively low transconductance that is somewhat akin to that of a vacuum tube,[2] a transconductance/current ratio that is in practice well below the theoretical limit e/kT, and a relatively high imput impedance. Conventional IGFET theory,[3] however, does predict a transconductance/current ratio that approaches infinity as the gate-insulator capacitance is made arbitrarily large as, for example, by making the insulator thickness approach zero in value. This prediction seems intuitively incorrect. At least it is at marked variance with the behavior of other charge-control devices such as the bipolar transistor and vacuum tubes.[2-4]

The following analysis was inspired by three items. The first of these was the interesting experimental observation that field-effect transistors universally exhibit a transconductance/current ratio that closely approaches e/kT in value as their drain currents are decreased to low values.[5] As far as the author can tell, no detailed explanation of this unexpected result was ever given. The second item is the intuitively incorrect result predicted for the transconductance/current ratio by conventional IGFET theory. If it were true that this ratio approached infinity with decreasing gate-insulator thickness, the IGFET would be a most unusual device indeed![4] The third, and potentially most important item, is the extant implication that IGFET devices are restricted to relatively lower transconductances than bipolar transistors and thus suffer a basic speed disadvantage, particularly in digital integrated circuits[6] wherein load and stray capacitance driving capability must be considered in addition to the usual device measures of frequency performance, for example, the current gain cutoff frequency F_T.

IGFET Potential Distribution

Simple physical arguments added to the classical picture of IGFET behavior lead one to believe that the source–drain potential distribu-

tion existing within a few kT/e potential units of the silicon–insulator interface is as shown in Fig. 1. Here, the gate is biased to create an inversion channel and the device is operating in drain saturation. The point of particular interest in this paper is the potential barrier V_s that must exist in the source–channel interface region to balance the large charge-density difference across this interface. The many consequences of this will be described below.

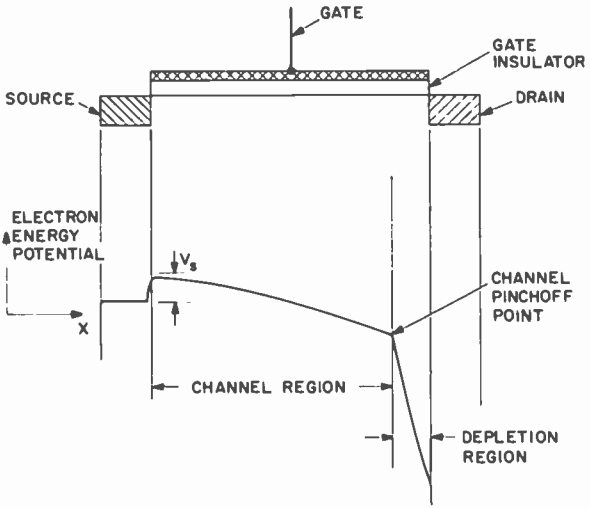


Fig. 1—General form of the source–drain potential distribution in an n-channel IGFET operating in drain saturation.

The potential decrease along the channel shown in Fig. 1 is due to the IR drop caused by the load current and channel resistance. The pinch-off point is reached when the net voltage across the gate insulator falls to zero. A depletion region exists between this point and the drain. The striking general similarity to the emitter–collector potential distribution in the typical bipolar transistor is to be noted. Similarities in the channel–drain region, such as the Early effect, have been noted elsewhere.[7] One presumes that other similarities exist, such as "channel widening", corresponding to bipolar base widening, and also a complex mixture of both field and diffusion flow in the channel,[8] as in the base of the bipolar device.

Let us now consider the small but important potential barrier V_s in more detail.[9] The charge density induced in the conducting channel of an IGFET is at least a few orders of magnitude smaller than the dopant charge density in the usual source electrode. Accordingly, just

as at the emitter–base junction in the bipolar transistor, the carrier densities across the source-channel interface can be connected by a Boltzmann relation:

$$\frac{n_c}{n_s} = \exp\left(\frac{-eV_s}{kT}\right). \qquad [1]$$

For an n-channel device n_c is the electron density on the channel side of the interface, n_s is the electron density on the source side, e is the electron charge, V_s is the previously noted barrier potential of Fig. 1, k is Boltzmann's constant, and T is the absolute temperature of the region in question. An analogous relation holds for p-channel devices. For our purposes Relation [1] will be considered meaningful down to one inversion channel depth, that is, a few kT/e potential units from the semiconductor–insulator surface interface. Since carrier densities in both the source and channel regions generally vary with depth from the surface interface, the potential V_s will in general vary with depth. This detail should not affect the general conclusions of this paper and for simplicity we shall assume a single average value of V_s.

A further detail to be noted is that the source–channel barrier is a high/low junction in contrast to the pn-junction at the emitter–base interface in the bipolar transistor. A high/low junction is more ohmic in behavior than a p-n junction and thus need not produce an offset effect near the origin of the I-V characteristic, as is typical of bipolar transistors but not of field effect devices.

Potential Distribution in the Input Circuit

In IGFET devices of conventional design, virtually all of the input voltage applied between the gate and source electrodes appears across the gate insulator. One electrode of the gate-insulator capacitor is the gate electrode, and if we neglect stray capacitances and edge effects, the other electrode is the conducting channel itself. This channel returns the circuit to the source through the previously cited high/low junction. Under both dc and ac operating conditions the potential V_s across this junction is in series with the voltage across the gate insulator and any IR drops that may appear in the input circuit. These resistive drops act in much the same way as the base and emitter series resistance in the input circuit of the bipolar transistor; the potential V_s acts in almost exactly the same way as the potential across the emitter-base junction. All mobile load-current carrying charges

that traverse the active region of the device must first climb a kT/e-type potential barrier to be "activated" for use in the load path.

If the device input voltage is increased to increase the inversion charge in the channel, the potential V_s must decrease according to Eq. [1] to account for the new charge-density ratio across the junction. A change in V_s of several kT/e potential units can account for a relatively large change in channel inversion charge and so is not normally noticed in the usual IGFET input circuit where the largest fraction of the input voltage is absorbed across the insulator. However, in the limit when the gate insulator capacitance is made large, such as by decreasing the insulator thickness or increasing the dielectric constant, the input voltage drop across this part of the circuit will approach zero and the entire input voltage will approach V_s except for any IR drops that may remain. The situation then becomes almost identical to that in the input circuit of the bipolar transistor.

The IGFET Control Capacitance

In the same manner that changes in base minority-carrier charge in the bipolar transistor can be associated with a change in the emitter-junction voltage, the changes in IGFET channel charge can be associated with changes in the potential V_s. This recognition, combined with the well-known definition of capacitance as a change in charge for a change in associated voltage, leads to what might be defined as a control capacitance C_c. In the following derivation Q_c is the total channel inversion charge:

$$C_c = \frac{dQ_c}{dV_s} = \frac{d}{dV_s}\left(\text{const.} \exp \frac{eV_s}{kT}\right) \qquad [2]$$

$$= Q_c \frac{e}{kT}. \qquad [3]$$

The channel charge Q_c relates to n_c in Eq. [1] for the reasons shown in the Appendix. The capacitance C_c is the same in form and magnitude as the well-known diffusion capacitance that shunts the input of a bipolar transistor. Of the total device input voltage some fraction will appear across the high/low junction kT/e and thus across C_c which is effectively in series with the gate insulator capacitance. The fraction of the input voltage that appears across C_c can be viewed as the useful portion of the input voltage relative to the modulation of channel charge and drain current. The fraction of the input voltage

absorbed across the gate insulator capacitance is lost as far as charge modulation is concerned. The charge on both capacitances is equal since the capacitances are effectively in series.

Transconductance Relations

The drain current I is related to the channel charge Q_c by

$$I = \frac{Q_c}{\tau} \qquad [4]$$

where τ is the charge carrier transit time in the channel. This relation simply states that a charge Q_c is swept out into the drain each time period τ.

In the extreme limiting case of an infinitely large specific gate-insulator capacitance, all of the input voltage appears as V_s; none is lost across the insulator. At this hypothetical limit the transconductance g_m is given by

$$g_m = \frac{dI}{dV_g} = \frac{dI}{dV_s} = \frac{1}{\tau}\frac{dQ_c}{dV_s} = \frac{Q_c}{\tau}\frac{e}{kT}$$

or, by Eq. [4],

$$= I\frac{e}{kT}; \qquad [5]$$

the transconductance/current ratio is given by

$$\frac{g_m}{I} = \frac{e}{kT}. \qquad [6]$$

This new result for the limiting transconductance/current in IGFET's is consistent with that of other charge-control devices, such as the bipolar transistor and the grid-controlled vacuum tube. This new limit is to be compared with that predicted by classical IGFET theory

$$\frac{g_m}{I} = \frac{2}{V_g}, \qquad [7]$$

which is found in standard texts or can be straightforwardly derived from Eq. [24] in Appendix 1. Relation [7] predicts an unlimited range of values, in particular, ones that can be in excess of e/kT if V_g is made sufficiently small. This excursion can be achieved in principle by arbitarily increasing the value of the specific channel capacitance and decreasing V_g correspondingly to keep the channel charge constant, or Eq. [24] in the Appendix satisfied. In IGFET's of conventional design, V_g is of the order of a volt, and the incorrect limit result predicted by Eq. [7] does not come into question.

Generalized Transconductance and Input Impedance

Consider the general situation wherein the total gate input voltage V_g divides across the total gate insulator capacitance C_o ($= CL$ of Appendix 1) and the control capacitance C_c associated with V_s and the channel charge. For simplicity, we shall ignore stray capacitances at the channel ends and those arising from depletion layers between the channel and the underlying semiconductor body. Following the procedure in Eq. [5], the transconductance will then be

$$g_m = \mathrm{I}\,\frac{e}{kT}\,\frac{dV_s}{dV_g}. \qquad [8]$$

The derivative term can be evaluated by using the distribution of voltage across C_o and C_c,

$$V_s = V_g\,\frac{C_o}{C_o + C_c}, \qquad [9]$$

which gives

$$\frac{dV_s}{dV_g} = \frac{C_o}{C_o + C_c} - V_g C_o (C_o + C_c)^{-2}\,\frac{dC_c}{dV_g}. \qquad [10]$$

Substituting Eq. [4] into Eq. [3] and differentiating gives

$$\frac{dC_c}{dV_g} = \tau\,\frac{e}{kT}\,g_m. \qquad [11]$$

Use of Eq. [11] and several other simple substitutions in Eq. [10]

gives the final results:

$$g_m = \frac{e}{kT} I \left[\frac{C_o}{C_o + 2C_c} \right] \qquad [12]$$

$$g_m = \frac{e}{2kT} I \frac{C_o}{C_c} \qquad \text{for } C_o \ll C_c \qquad [13]$$

$$g_m = \frac{e}{kT} I \qquad \text{for } C_o \gg C_c \qquad [14]$$

$$g_m = \frac{e}{3kT} I \qquad \text{for } C_o = C_c. \qquad [15]$$

Relation [13] corresponds to the situation in IGFET's of conventional design. That this is indeed the case can be seen by using substitutions from Eq. [3] and from Eqs. [28] and [29] of the Appendix, which reduce Eq. [13] to

$$g_m = I \frac{3/2}{V_g}. \qquad [16]$$

Aside from the small difference in the numerical constant, this relation is identical to that derived from classical IGFET theory as typified by Eq. [7]. Note, also, that Eq. [13] can be written

$$\frac{1}{C_o} g_m = \frac{e}{2kT} \frac{I}{C_c} = \frac{1}{2} \frac{1}{\tau}, \qquad [17]$$

which shows that there is a basic trade-off in an IGFET between transconductance and input impedance as represented by $1/C_o$. High input impedance or dynamic voltage range capability results in low transconductance and vice versa.[2]

The condition represented in Eq. [14] is identical to that of a bipolar transistor[3] and is approached in an IGFET of conventional design at very low drain currents where C_c, as seen from

$$C_c = \frac{e}{kT} Q_c = \frac{e}{kT} I \tau, \qquad [18]$$

is relatively small. This result seems to be a reasonable explanation for the IGFET experimental observations noted earlier.[5] Condition [14] would be observed at higher drain currents if the specific insulator capacitance is increased. Numerical calculations based upon surface C-V data[10] seem consistent with the results noted above.

With respect to Eq. [15], there is a general physical argument to the effect that the parallel plate spacing must be approximately a Debye length for the gate capacitance to match the control capacitance. A similar type of argument shows that $C_o \approx C_c$ when the gate voltage is equal to kT/e. More generally, $C_c/C_o \cong V_g/(kT/e)$.

Frequency Capability

As the IGFET specific gate capacitance is increased to make the device approach the bipolar transistor earmark characteristics of high transconductance and low input impedance, the IGFET frequency capability will improve because of an improved ability to rapidly charge parasitic device and circuit capacitances.

The above conclusion as well as the effect on power gain can be seen from another viewpoint also. This viewpoint was developed in an earlier analysis based upon charge control considerations as well as upon solid-state material constraints.[12] The analysis showed in a very general manner that there are basic trade-offs between transistor power gain G_p, device input voltage dynamic range V_T, maximum allowable device output voltage V_m, and operating frequency f at which the foregoing parameters were defined. For either a bipolar or field-effect transistor at their ultimate performance limit,

$$(G_p V_T V_m)^{1/2} f = \frac{E v_s}{2\pi}. \qquad [19]$$

Here, the basic device material constraints are E, the dielectric breakdown field of the semiconductor, and v_s, the carrier saturation velocity. For a given value of V_m, G_p and f will be maximum when V_T is a minimum. The minimum attainable value of V_T is the kT/e characteristic of bipolar transistors. The value of V_T ($\cong V_g$) for IGFET's of conventional design is typically several volts. Accordingly, as V_T is decreased toward the kT/e value of bipolars by increasing the specific gate capacitance, IGFET power gain and frequency response characteristics will improve.

Other Effects

In accord with the classical picture of IGFET operation, the drain saturation voltage will tend to decrease with the gate input voltage as the specific input capacitance is increased. This can be seen from Eq. [24] of the Appendix, which shows that

$$V_g\ (=\text{drain saturation voltage}) = \sqrt{\frac{2LI}{\mu C}}. \qquad [20]$$

Accordingly, the drain saturation voltage will vary inversely with the square root of the gate insulation capacitance in the range represented by Eq. [13]. In the range near that described for Eq. [15], the behavior of the drain saturation voltage with gate-insulator capacitance is more complicated. At the hypothetical limit for V_g of kT/e, the drain saturation voltage would be of the order of kT/e. This, combined with the absence of the typical voltage offset of bipolar transistors, would put the current knee at a lower voltage than that seemingly possible with bipolar transistors.

The dynamic power consumption P_d of a digital circuit is given by the well-known relation

$$P_d = fCV^2 = fQV.$$

If we assume a given value of the operating frequency f and a given switched charge Q, then P_d will decrease with V, the operating voltage level. This level is directly related to the drain saturation voltage.

Junction Field-Effect Transistors

I have not considered in detail the reasons for the approach to e/kT limits for JFET's, which was also noted experimentally at low drain currents.[5] Crudely speaking, the reasons for this behavior would seem to be as sketched below.

The equivalent of the IGFET insulator capacitance in JFET's is the gate depletion capacitance. The carrier charge activation barrier is not at the channel-source electrode interface, as in the IGFET, but effectively exists around the edges of the channel where the channel joins to the gate depletion layer. As the gate voltage is decreased, the channel expands its dimensions, and total charge, by the thermal motion of majority carriers which climb a kT/e hill at the channel–depletion-layer interface. In other words, the location of

the channel edge moves outward toward, and is defined by, the potential surface at which the depletion layer retarding potential is of the order of kT/e units higher than that of the underlying channel.

If the channel current is very small, as would be the case for very small total channel charge, then the fraction of total channel charge that feels the kT/e barrier will be relatively large, and we should not be surprised to find g_m/I ratios approaching e/kT. To state it another way, which is analogous to the IGFET case, the channel control capacitance is relatively small compared to the depletion layer capacitance, which is in series with it. At large channel currents, the situation is reversed, and we should not expect to find kT/e results.

To maximize the drain current at which kT/e characteristics are evident, the device should have a very heavily doped, thin, channel. This would maximize the depletion layer capacitance, maximize the transconductance, and minimize the input impedance and dynamic input voltage range. Indeed, high performance JFET devices are fabricated in this manner.[13]

Analysis Limitations

The preceding analysis was intended as a demonstration of basic physical principles and consequent limits, rather than as a detailed analysis of any particular device under specific conditions of operation. For reasons of simplicity, the effects of stray capacitances, resistances, and frequency were neglected. As with bipolar transistors, these effects can almost certainly be expected to deter the attainment of the theoretical e/kT limit in commercially realizable devices. The kT/e limit, itself, should be inviolate.

Summary

The preceding analysis proposes that an IGFET can be viewed as a hypothetical bipolar transistor with a base that couples to the outside world through a series capacitance, the gate-insulator capacitance. The bipolar device is unusual in that it has no dc base current, that is, it has an alpha identically equal to unity, and has no dc net leakage current. If the series capacitance is increased sufficiently, the IGFET performance takes on the distinguishing characteristics of the bipolar device.

To Warner's statement[6] "It is noteworthy that attainable g_m in the MOSFET is tied directly to the state-of-the-art, whereas g_m in the bipolar device is essentially fixed by nature," I would add, "But the MOSFET state-of-the-art leads toward precisely the same natural limit!"

Acknowledgment

It is a pleasure to acknowledge stimulating and helpful discussions with D. O. North, W. Bosenberg, R. Williams, K. H. Zaininger, and D. P. Kennedy.

Appendix 1—Channel Charge Relations

The following derivation closely parallels conventional derivations that take into account the variation of voltage along the channel length of the gate capacitance due to the resistive voltage drop along the channel. Relationships are developed between applied gate voltage, physical gate capacitance, channel charge, potential drops along the channel, and the channel current. Particular attention in the following derivation is given to the relationship between total channel charge and the charge density conditions at the source end of the channel. These need to be related to justify Eqs. [2] and [3] of the text.

The inversion charge $Q(x)$ per unit length of the channel at the point x units from the source is

$$Q(x) = C\,[V_g - V(x)], \qquad [21]$$

where C is the gate-insulator capacitance per unit length of the channel, V_g is the applied gate voltage (which is assumed totally across the gate capacitance), and $V(x)$ is the voltage drop at the point (x) due to the resistive drop along the channel associated with the flow of the load current I. The differential drop dV per differential length dx of the channel at the point x is

$$dV = \frac{I}{\sigma(x)} = \frac{I\,dx}{\mu Q(x)} \qquad [22]$$

where $\sigma(x)$ is the channel conductivity at the point x, and μ is the charge-carrier mobility, which is assumed constant along the channel. This assumption should have no effect on our final conclusions over a rather wide range of behavior of μ.

Substituting Eq. [21] into [22] and integrating the voltage between the limits 0 and V, and the distance over the limits 0 and x, we get

$$V^2 - 2V_g V + \frac{2I}{\mu \mathrm{C}}\,x = 0. \qquad [23]$$

For the pinch-off condition, $V = V_g$ at the drain end of the channel where $x = L$. In consequence Eq. [23] yields:

$$\frac{2I}{\mu C} = \frac{V_g^2}{L}. \qquad [24]$$

This shows several well-known relations that apply at channel pinch-off.

Inserting Eq. [24] into [23] and solving for V we get

$$V = V_g \pm V_g \sqrt{1 - \frac{x}{L}} \qquad [25]$$

The positive sign is spurious and will hereafter be dropped. Eq. [25] gives the potential drop along the channel under pinch-off conditions.

For conditions below pinch-off, Eq. [25] becomes

$$V = V_g - V_g \sqrt{1 - \frac{bx}{L}} \qquad [26]$$

where b expresses the ratio of the actual current I to the pinch-off value at a given value V_g of the applied gate voltage. Eq. [26] states that V never rises to a magnitude V_g unless pinch-off occurs, that is, unless $b = 1$. It is interesting to note that under non-pinch-off conditions, there will be a low/high junction at the channel/drain interface in addition to the high/low junction at the source/channel interface. The junction voltage at the drain end of the channel will be the larger of the two, because the channel charge density is lower at this end.

Inserting Eq. [25] into [21] we obtain the expression for the inversion charge $Q(x)$ per unit length of the channel at any point x along the channel:

$$Q(x) = CV_g \sqrt{1 - \frac{x}{L}} \qquad [27]$$

At the source end of the channel $x = 0$, and

$$Q(x) = Q(0) = CV_g. \qquad [28]$$

The total inversion charge Q_c in the channel is obtained by integrating Eq. [27] over the length of the channel:

$$Q_c = 2/3\, LQ(0). \qquad [29]$$

To relate $Q(0)$ to the channel volume charge density n_c adjacent to the source, and hence to Eq. [1] of the text, we make use of a previously derived semiconductor surface relation[14] that links the inversion charge density n_c with the surface potential ψ, the distance x from the surface into the material's interior, the intrinsic carrier density n_i, the doping density n_A of the background material, the minority carrier charge density n_o in the background material, and the intrinsic Debye length $\mathcal{L}$:

$$n_c = \frac{n_o}{\left[\left(\frac{n_i}{n_A}\right)^{1/2} \frac{x}{2\mathcal{L}} + \exp\left\{\frac{-e\psi_o}{2kT}\right\}\right]^2} \qquad [30]$$

An analogous thermal equilibrium relation applies to p-channel devices.

The potential at the semiconductor surface is ψ_o. Eq. [30] is only valid in the inversion layer. Terms involving the fixed dopant charge and the majority carrier charge were dropped. We now integrate the relation

$$Q(0) = W \int_0^x n_c dx, \qquad [31]$$

wherein W is the lateral gate width. We obtain

$$Q(0) = \frac{2W\mathcal{L}n}{\left(\frac{n_i}{n_A}\right)^{1/2}}, \qquad [32]$$

wherein $n = n_o \exp\{e\psi_o/(2kT)\}$ is the inversion charge density at the semiconductor surface adjacent to the source. The integration was carried out to a depth x into the material sufficient to include the majority of the inversion charge.

Eq. [32] is the desired relation. It shows that the channel volume charge density adjacent to the source is directly proportional to $Q(0)$ and hence to the total channel charge Q_c (Eq. [29]).

Relation [32] is quite general. It is not affected by the inclusion of nonthermal equilibrium conditions into Eq. [30], as for example,

by including the effect of the load current IR voltage drop along the channel. The above relationship between the total charge in the channel and the channel charge density adjacent to the source is analogous to the familiar bipolar transistor relationship between total base charge and the base charge density adjacent to the emitter.

References:

[1] J. J. Sparks and R. Beaufoy, "The Junction Transistor as a Charge Control Device," **Proc. IRE,** Vol. 45, p. 1740, Dec. 1957.
[2] E. O. Johnson and A. Rose, "Simple General Analysis of Amplifier Devices with Emitter, Control, and Collector Functions," **Proc. IRE,** Vol. 47, p. 407, March 1959.
[3] See, e.g., A. S. Grove, **Physics and Technology of Semiconductor Devices,** John Wiley and Sons, New York (1967).
[4] W. G. Dow, **Fundamentals of Engineering Electronics,** Second Ed., pp. 243-244, John Wiley and Sons, N. Y. (1952).
[5] L. Evans and K. A. Pullen, Jr., "Limitations of Properties of Field-Effect Transistors," **Proc. IEEE,** Vol. 54, p. 82, Jan. 1966.
[6] R. Warner, "Comparing MOS and Bipolar Integrated Circuits," **IEEE Spectrum,** p. 50, June 1967.
[7] See, e.g., Reference (3), p. 255.
[8] P. Richman, **Characteristics and Operation of MOS Field Effect Devices,** p. 58, McGraw-Hill Book Co., N.Y. (1967).
[9] Specific references to this in the literature were a brief remark in Sah and Pao, "Fixed Bulk Charge in MOS Transistors," **IEEE Trans. Elec. Devices,** Vol. ED-13, p. 395, April 1966; and a brief note by R. R. Troutman, "Subthreshold Design Considerations for Insulated-Gate Field-Effect Transistors," Paper THAM 9.4, ISSCC, Philla., Pa., Feb. 15, 1973.
[10] K. Zaininger and F. Heiman, "The C-V Technique as an Analytical Tool," **Solid-State Tech.,** Vol. 13:5-6, May-June 1970.
[11] See, e.g., A. Rose, "An Analysis of the Gain-Bandwidth Limitations of Solid-State Triodes," **RCA Review,** Vol. 24, p. 627, Dec. 1963.
[12] E. O. Johnson, "Physical Limitations on Frequency and Power Parameters of Transistors," **RCA Review,** Vol. 26, p. 163, June 1965.
[13] First brought to the author's attention by J. Moll.
[14] See Eq. 3.17, p. 38, of Reference (8).

Practical Use of III-V Compound Electron Emitters*†

A. H. Sommer

RCA Electronic Components, Princeton, N. J.

Abstract—Aspects of the physics and chemistry of III-V electron emitters have been discussed in many publications and have been summarized in several review papers (3, 29). The emphasis of the present paper is on the practical use of these new materials. After a brief recapitulation of the principles on which the new cathode materials are based, the paper discusses the technological problems encountered in the formation of the cathodes and gives an account of the most important practical III-V emitters and the devices in which they are successfully used. Some materials that are still in the developmental stage are described, followed by a few remarks dealing with the stability of III-V electron emitters.

1. Introduction

In 1965, Scheer and Van Laar[20] reported that they had obtained higher white light sensitivity with cesium-activated p-type GaAs photocathodes than had been obtained with any photocathode material up to

* The research reported in this paper has been sponsored by the U.S. Army Electronic Systems Command, Fort Monmouth, N.J., the U.S. Army Night Vision Laboratory, Ft. Belvoir, Va. and the RCA Corporation.

† This paper also appears in the Proc. 4th Int. Symp. on GaAs, Inst. of Phys. Conf. Ser. No. 17, 1973, published by the Institute of Physics, Bristol, England.

that time. This result stimulated research on the electron emission characteristics of cesium- and cesium-oxygen[28] activated III-V compounds in many laboratories, with the goals of obtaining a better understanding of the emission process and of producing new materials for use in photoemissive devices such as photomultipliers, image tubes, and television camera tubes. In both directions great progress has been made during the intervening years.

2. Comparison of Negative-Electron-Affinity and Conventional Electron Emitters

The principle of negative electron affinity (NEA) and its use in III-V compound electron emitters has been discussed in many papers (see, for instance, References (3) and (29)). For the present purpose, it is therefore sufficient to summarize in greatly simplified form the essential points that lead to the practical usefulness of these materials as photoelectric and secondary electron emitters. (The use of NEA III-V compounds for cold cathodes will not be discussed in this paper, because it is still in the developmental stage. Promising results with GaAs cold cathodes have recently been reported by Schade et al[19].)

The advantages of the NEA emitters over previously used (hereafter called "conventional") semiconducting electron emitting materials, such as the alkali antimonide photocathodes[24] and the alkaline earth oxide (MgO, BeO) secondary emitters, are illustrated in Figs. 1 to 3. Figs. 1(a) and 1(b) show the simplified energy-band diagrams for strongly p-type-doped conventional and NEA electron emitters, respectively. In case (a) the work function (difference between Fermi level and vacuum level) is larger than the bandgap energy E_G by the amount E_A, the electron affinity; in case (b) it is smaller, hence the electron affinity is negative. No material is known that corresponds to Fig. 1(b). However, in Cs-activated p-doped materials the bands bend downward near the surface[26] as shown in Fig. 1(c), resulting in "effective" NEA because then the vacuum level lies below the bottom of the conduction band in the bulk material.

The advantage of NEA for electron emitters is based on the following facts. Excited ("hot") electrons produced in the bulk of the material by light (photoemission) or by the impact of electrons (secondary emission) are able to escape into vacuum only if on their path to the vacuum interface they retain enough energy to overcome the surface barrier E_A. Hot electrons lose energy rapidly (about 10^{-12} sec) through interaction with phonons until they have been "thermalized" at the bottom of the conduction band; therefore, in case (a),

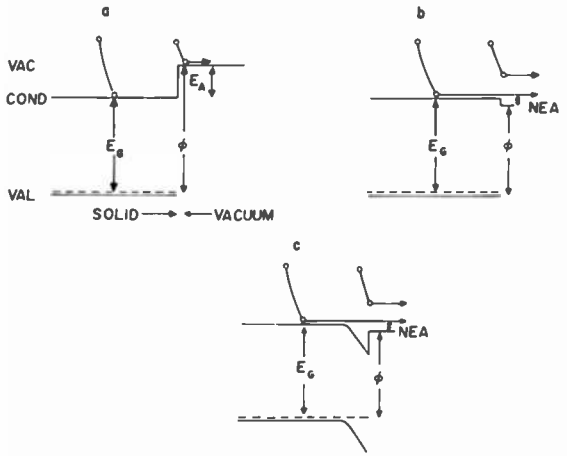


Fig. 1—Energy-band models, Val. = top of the valence band, Cond. = bottom of the conduction band, Vac. = vacuum level, E_G = bandgap energy, ϕ = work function, E_A = electron affinity, NEA = effective negative electron affinity. Three basic models are shown: (a) photoemissive p-doped semiconductor, (b) NEA semiconductor, and (c) NEA semiconductor with band bending.

only those electrons can escape that have been produced so close to the surface (usually within a few hundred angstroms) that their energy still exceeds the value $(E_G + E_A)$. On the other hand, "thermalized" electrons have a much longer life time (about 10^{-9} sec); therefore, in case (b), electrons that have their origin at much greater distance from the vacuum interface (order of microns) can diffuse to the surface and still have sufficient energy to escape. The NEA emitter therefore differs basically from the conventional (positive-electron-affinity) emitter by the much longer escape depth of the excited electrons.

Fig. 2 shows schematically the effect of NEA, i.e., of enhanced escape depth, on the spectral response curves of photoemitters. Curve

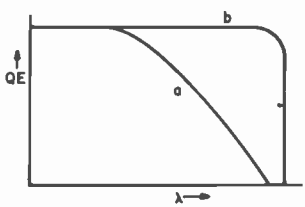


Fig. 2—Schematic plot of quantum efficiency (QE) versus wavelength, (a) for conventional and (b) for NEA photoemitters.

(a) represents a conventional emitter (positive electron affinity) and curve (b) a NEA emitter. Equal E_G values are assumed for both cases to emphasize the basic difference. Comparison of the curves shows the following:

1. At short wavelengths, where the photon energy is much larger than E_G and, hence, the light absorption is high, the quantum efficiency (QE) is substantially the same for both materials. The reason is that most of the light is absorbed within the short escape depth (a few hundred angstroms) associated with positive electron affinity materials.
2. The threshold wavelength of curve (b) is longer than that of (a) because, for equal E_G, the threshold photon energy of curve (a) is greater than that of (b) by the energy corresponding to the electron affinity E_A.
3. The most important difference between curves (a) and (b) is that with increasing wavelength, the QE of the NEA cathode remains high until close to the threshold, and therefore is far superior to that of the conventional cathode in the threshold region. The reason is that the light absorption decreases with increasing wavelength, with the result that an ever increasing fraction of the photoelectrons are produced beyond the escape depth of positive-electron-affinity emitters, whereas practically all the light is still absorbed within the escape depth of the NEA material.

In addition to QE and threshold wavelength, the amount of thermionic emission of a photocathode at room temperature (dark current) also affects the usefulness of the photoemissive device. (Thermionic emission can be eliminated by cooling, but cooling is usually inconvenient.) An important advantage of the NEA III-V compound emitters is that they have lower thermionic emission at room temperature than do conventional cathodes with comparable threshold wavelength. The factors determining the thermionic emission of III-V emitters have been discussed by Bell.[2] It should be emphasized that low thermionic emission cannot be an inherent characteristic of NEA cathodes, because the thermionic emission of NEA Si cathodes[14] is many orders of magnitude higher than that of the III-V materials.

To illustrate the effect of NEA on secondary electron emission, typical gain versus primary-electron-energy (E_p) curves for conventional and NEA materials are shown in Fig. 3. To emphasize the basic difference, it is assumed that both materials have the same "intrinsic" gain factor, i.e., that at low primary energies they require the same E_p value for the production of one secondary electron.

With increasing E_p, the secondaries are produced at an increasing distance from the vacuum interface; hence the number of emitted electrons depends on the escape depth. Because NEA materials have a longer escape depth, curve (b) increases linearly with E_p up to high values of E_p, whereas the gain of conventional materials, curve (a), begins to drop at values where too many electrons are produced beyond the relatively shorter escape depth.

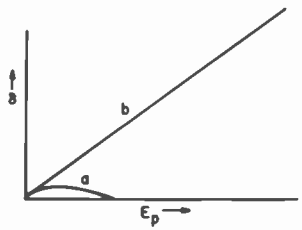


Fig. 3—Schematic plot of gain (δ) versus primary energy E_p for (a) conventional and (b) NEA secondary electron emitter.

It is important to note that, at low primary energies (up to about 300 eV) the NEA materials have no advantage over conventional secondary emitters for the same reason that NEA photocathodes have no advantage at short wavelengths. In both types of emitter, the superiority of the NEA material manifests itself only when the electrons are produced beyond the limited escape depth associated with positive electron affinity.

3. Technological Problems

The outstanding difference between NEA and conventional electron emitters is that NEA emitters require the use of single crystals of great perfection and well-controlled doping. Imperfections of the crystal reduce the diffusion length of the electrons and thus reduce the large escape depth that is the reason for the high efficiency of NEA emitters. (However, as mentioned in Section 4.2, there are some special cases where polycrystalline NEA materials have found practical use.) By contrast, polycrystalline materials are quite adequate for conventional emitters, because the escape depth is small compared with the size of the microcrystals, i.e., the material is "single crystal" for the electrons produced within a few hundred angstroms from the surface. The need for single crystals imposes the following limitations on the formation of NEA cathodes.

1. The desired III-V compound with the desired doping is generally grown either by vapor-[27] or liquid-phase[16] epitaxy, on a single-crystal substrate. These processes are inherently costly and time consuming when compared with the formation of conventional cathodes; for instance, the simple evaporation of Sb in the formation of Cs_3Sb. Attempts to produce single-crystal films of III-V compounds by simpler techniques, such as simultaneous or flash evaporation of the components or sputtering,[9] have at the time of writing not produced crystals of satisfactory quality.

2. Whereas cathodes made by evaporation techniques can be made of any desired size, it is difficult to make single crystals of large area. At present, III-V compounds are used only for emitters having areas in the 1 to 3 cm^2 range.

3. Single crystals have a plane surface and are therefore not useable where the electron optics of the device requires a curved electrode. Evaporated films of conventional materials can, of course, be deposited on a substrate of any desired shape.

4. Conventional cathodes can be formed entirely in high vacuum. For instance, Cs_3Sb is made by evaporating Sb in the vacuum of the final device and then exposing it to Cs vapor. It is virtually impossible to do the Cs-activation of III-V single crystals without exposing the crystal to air between removal from the growth equipment and mounting in the device. The exposure to air contaminates the surface; hence an additional process is required to clean the crystal in the final vacuum. In the early work,[20] this was accomplished by cleaving the crystal in vacuum, but by this method only very small area surfaces can be obtained, and the procedure is not practical in complex tubes. Cleaning is now usually accomplished by the simpler method of heating the crystal to a temperature close to the decomposition temperature of the III-V compound. Several problems are associated with this heat treatment, which is done by radiation, by electron bombardment, or by resistance heating. In all three cases, great care must be taken that the heating be uniform over the whole area because the temperature is critical. Moreover, accurate temperature measurement is not easy. An infrared pyrometer is generally used, but, to obtain reliable results, the emissivity of the material must be known or, at least, must be reproducible from crystal to crystal.

5. Probably the greatest obstacle to the general use of NEA emitters up to the present has been the difficulty of making semitransparent cathodes that can be used with light incident from the substrate

side. This requirement is almost essential for all imaging devices, such as image tubes and television camera tubes, and is easily accomplished with conventional cathodes by simply evaporating the base material (such as Sb for the formation of Cs_3Sb) to the required thickness. The best solution for making semitransparent III-V cathodes would be epitaxial growth on a transparent substrate, such as quartz or sapphire. Attempts in this direction, however, have so far been unsuccessful.[13] At present, best results for long-wavelength response are obtained by growing a thin film of the desired compound epitaxially on a relatively thick substrate consisting of a single-crystal III-V compound with a lattice constant similar to that of the cathode but with a larger bandgap. This substrate then acts as a "window" in that it absorbs only shorter wavelength light, but permits the long wavelength light to be absorbed in the NEA material.

The above problems arising from the requirement for single-crystal materials are not completely solved, but sufficient progress has been made to allow the use of III-V emitters in the practical devices described below.

4. Practical NEA Electron Emitters

4.1 Introduction

It is generally found that the photoemissive sensitivity of cesium-oxygen activated III-V compounds decreases with decreasing bandgap energy E_G, especially near the threshold wavelength. Qualitatively, this agrees with expectation, because, if one assumes that the work function does not vary substantially from compound to compound, the magnitude of the effective negative electron affinity will decrease with decreasing E_G (see Fig. 1). Lower NEA causes lower escape probability of the electrons. The ultimate limit to the reduction of E_G is, of course, set by the fact that E_G must be greater than the work function to make NEA possible.

The greater escape probability at larger E_G values has practical consequences for both photoemission and secondary emission. For photoemission, it means that increased threshold wavelength (lower E_G) is generally obtained at the expense of reduced quantum efficiency. Hence, for any specific application, the III-V compound having the largest bandgap compatible with the desired threshold wavelength should be used. However, because the energy of the exciting photons is below 2 eV for most applications, only III-V compounds with bandgaps well below 2 eV are of practical interest. On the other hand, in

secondary emission, the primary electron can be given any desired energy up to many hundreds of eV; hence, the slightly higher primary energy required to produce a secondary electron from a material with larger bandgap is more than compensated by the greater escape probability associated with the larger bandgap.

4.2 NEA Materials for Secondary Electron Emission

Cesium-activated GaP ($E_G = 2.3$ eV) was the first NEA III-V compound to be used in a practical device[21] because—as discussed above—this material has higher escape probability than the smaller-bandgap materials required for photoemission. For Cs-activated single-crystal GaP, Simon and Williams[23] reported gain factors that increase linearly with primary electron energy in accordance with Fig. 3. The high gain obtainable with GaP is particularly valuable when applied to the first dynode of a photomultiplier. If the gain of the first dynode exceeds approximately 30 (as compared with gain values of approximately 5 for previously used conventional materials), the photomultiplier becomes more useful at extremely low light levels (as encountered, for instance, in scintillation counters for very low energy radiation), because it is able to discriminate single, double, and up to five or more photoelectron events.[12,15] Moreover, the high-gain first dynode improves the signal-to-noise ratio of the tube. The high gain also ensures that every photoelectron emitted by the cathode contributes to the signal output, because the statistical probability that a primary photoelectron produces zero secondary electrons becomes neglible.

For optimum electron-optical performance, the first dynode of a photomultiplier should be curved rather than flat. This requirement would be a considerable handicap if single-crystal GaP had to be used, but experience has shown that polycrystalline GaP, deposited from the vapor phase onto a suitably shaped metal substrate (tungsten or molybdenum), is quite satisfactory for use on dynodes. The reason can be seen in Fig. 4. As expected, the escape depth of polycrystalline GaP is smaller than that of single crystals; therefore, the gain-versus-E_p curve falls below the straight line as E_p approaches 100 eV. However, the overall performance of the photomultiplier does not improve substantially for first-stage gains above 40 to 50, and Fig. 4 shows that, at the E_p values corresponding to these gains, the curve for polycrystalline GaP is only slightly below that for single-crystal GaP.

Additional improvement in the performance of the photomultiplier is obtained if all the conventional dynodes are replaced by GaP dynodes. Thus, a five-stage GaP tube is superior to conventional 10 to 14 stage tubes in speed of response, not only because of the smaller

number of stages but also because the higher inter-stage voltage at a given total applied voltage reduces the time of flight between successive stages.[11,12]

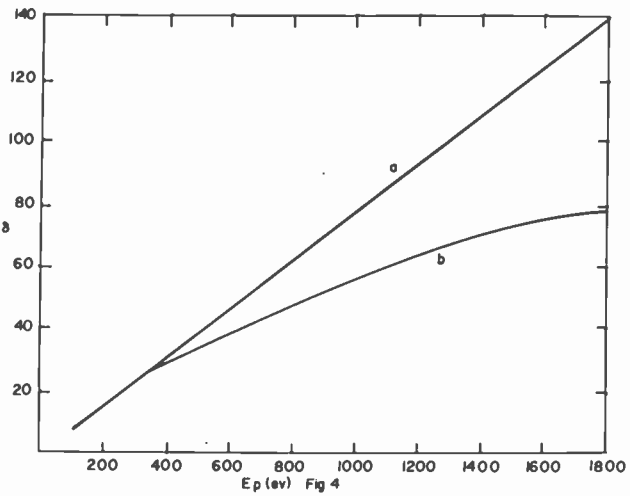


Fig. 4—δ versus E_p plot (a) for single-crystal and (b) for polycrystalline GaP.

GaP dynodes can be combined with different photocathodes, depending on the application. To give some examples, bialkali cathodes are used for scintillation and Cerenkov counters.[16] S-20 cathodes have been used for such projects as laser ranging to the moon.[17] Bialkali cathodes in combination with five GaP dynodes are of great value in high-energy physics.

4.3 NEA Materials for Photoemission

In addition to their superiority over conventional photocathodes in performance, NEA III-V cathodes have the advantage that, within wide limits, any desired bandgap energy, and hence threshold wavelength, can be produced by using the appropriate composition of ternary materials, such as (Ga,In)As, In(As,P), Ga(As,P), etc. As was pointed out in Section 4.1, the photoelectric efficiency decreases with decreasing bandgap. As a result, Ga(As,P) ($E_G > 1.4$ eV) is easier to activate than GaAs ($E_G \approx 1.4$ eV), and GaAs, in turn, is easier to activate than (Ga,In)As and In(As,P) ($E_G < 1.4$ eV). In the following, practical cathodes will be discussed in the sequence of decreasing E_G values.

(a) $Ga(As_{.75}P_{.25})$

In the case of conventional cathodes, the Cs_3Sb cathode is inferior in every respect to some of the more recently developed complex alkali antimonides, but, because of its lower cost, it is still widely used in applications where optimum performance is not essential. Similarly, there are applications where the $Ga(As_{.75}P_{.25})$ cathode ($E_G = 1.7$ eV) is used because it has economic advantages over the GaAs cathode, although it has lower long-wavelength response.

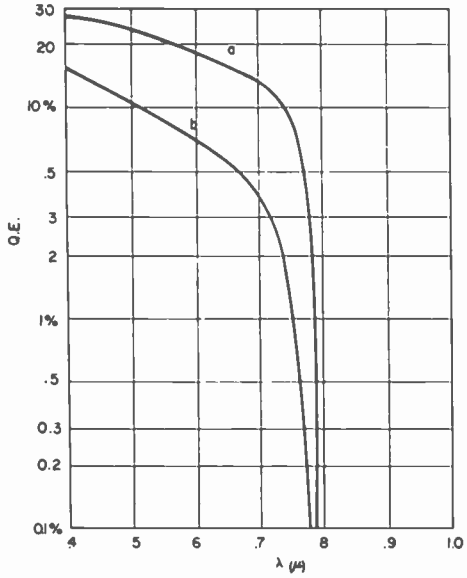


Fig. 5—Quantum efficiency versus wavelength for (a) single-crystal and (b) polycrystalline $Ga(As_{.75}P_{.25})$.

The spectral response curve of a single-crystal $Ga(As_{.75}P_{.25})$ cathode is shown in curve (a) of Fig. 5. In the reflecting (as distinct from the semitransparent) mode, this cathode is superior to any conventional cathode covering the same range. Curve (b) shows the response of the same material in polycrystalline form. As was discussed in connection with the polycrystalline GaP dynode, polycrystallinity reduces the escape depth and therefore reduces the response to the less efficiently absorbed long wavelengths. In practice, the polycrystalline cathode has found more widespread use than the single-crystal material because it is less costly and can more readily be formed on large areas and on electrodes of any desired shape.

Photomultipliers incorporating $Ga(As_{.75}P_{.25})$ cathodes[22] are characterized by high quantum efficiency over the whole visible spectrum and extremely low thermionic emission at room temperature ($<10^{16}$ A/cm^2). The main application is in the field of spectroscopy for instruments where infrared response is not required.

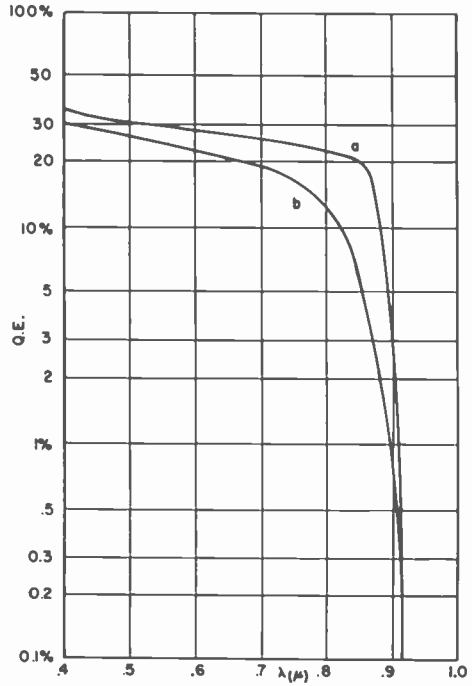


Fig. 6—Quantum efficiency versus wavelength for GaAs (a) under optimum conditions and (b) as typically obtained in photomultipliers.

(b) GaAs

At the present time, single-crystal GaAs is the most widely used of all the NEA III-V photocathodes because it has higher response above 7000 Å and a flatter response throughout the spectrum than any other cathode. Fig. 6 shows typical response curves. Curve (a) represents the best performance that has been obtained in simple experimental tubes (about 1500 μA/lm). In photomultipliers, the sensitivity is usually lower, but sensitivities above 800 μA/lm, corresponding to curve (b), and up to 1200 μA/lm are now frequently achieved.

The superiority of the GaAs cathode over previously used materials

may be illustrated by a few examples. In many spectroscopic instruments, response over the whole visible and near-infrared regions is required. In the past, two photomultipliers had to be used, one for the blue and one for the red and infrared region. The use of a single GaAs tube greatly simplifies operation. Additional advantages in this application are the flat response throughout the spectral range and the very low dark current, which is not obtainable with the previously used infrared-sensitive cathodes. The high infrared response of GaAs, combined, again, with the low dark current, has been found to be particularly useful for Raman spectroscopy.[7] Astronomy has also benefited from the use of GaAs tubes, especially in the study of weak stars.[4]

For the above-mentioned applications, photomultipliers with cathodes operating in the reflection mode are quite adequate. The processing of semitransparent GaAs cathodes is only just emerging from the developmental stage, but an image tube with a thin GaAs cathode[8] has given promising results. The GaAs film is grown on a GaP-Ga(As,P) substrate that absorbs the incident light of wavelengths below approximately 6500 Å, i.e., the cathode is only useful for the red and infrared regions. Even at long wavelengths, the sensitivity is still well below that obtained with GaAs cathodes in photomultipliers, but it begins to outperform the conventionally used S-20 cathode.

(c) (Ga,In)As

As was mentioned before, (Ga,In)As compounds, because of their lower E_G, are more difficult to activate than GaAs and these difficulties increase with increasing indium content (decreasing E_G). However, there are many applications where the longer threshold wavelengths of the ternary compounds are essential. Two specific examples are (1) the detection of the wavelengths emitted by the GaAs laser, which are just beyond the threshold of the GaAs cathode, and (2) the detection of the Nd-YAG laser radiation at 1.06 μm.

For the former, an indium content of only 6% ($E_G = 1.28$ eV) is adequate, and photomultipliers with $Ga_{.94}In_{.06}As$ cathodes have been activated successfully. For useful response at 1.06 μm, an indium content of about 18% ($E_G = 1.15$ eV) is required. Under optimum experimental conditions, quantum efficiencies up to 3% at 1.06 μm have been obtained[8] with $Ga_{.82}In_{.18}As$, but, up to the present, sensitivities in photomultipliers tend to be lower. Thus, this cathode, although successfully used for special applications, represents a borderline case between "practical" and "developmental" materials. It should be noted, however, that even a quantum efficiency of only 0.1% at 1.06 μm is

higher than that of the only available alternative, the Ag-O-Cs (S-1) cathode, and that the dark current of the (Ga,In)As compound is much lower at room temperature.

In Fig. 7, typical curves are shown for compounds with 6% and 18% indium content, respectively. It is evident that, as expected, the smaller bandgap increases the threshold wavelength, but at the same time the associated lower escape probability, for the reasons discussed in Section 4.1, reduces the efficiency, especially in the region adjacent to the threshold.

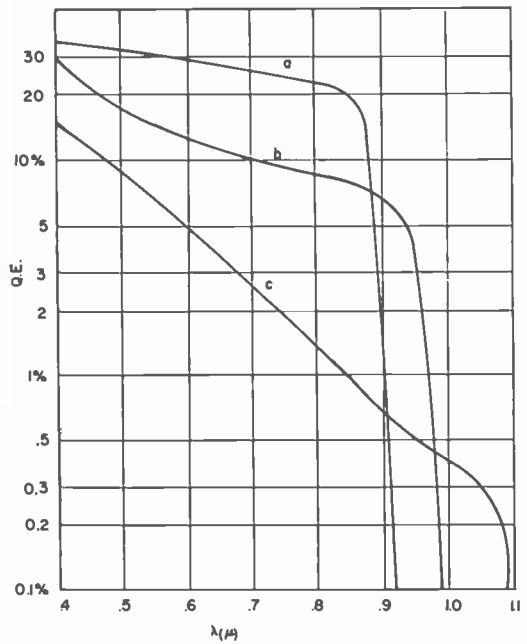


Fig. 7—Quantum efficiency versus wavelength under optimum conditions in sealed-off tubes for (a) GaAs, (b) $(Ga_{.94}In_{.06})As$, and (c) $(Ga_{.82}In_{.18})As$.

(d) Developmental Cathodes

The preceding discussion shows that there are two main requirements for extending the use of NEA III-V emitters. (1) In order to obtain higher infrared response, crystal growth and processing of ternary compounds with $E_G < 1.4$ eV must be further improved. (2) Better methods must be developed for growing and activating semitransparent cathodes of GaAs as well as of ternary compounds.

A detailed discussion of the experimental work on these problems is not within the scope of this paper, but a brief summary of the materials that are under study in various laboratories, together with relevant references, may be useful.

With regard to crystal growth, good results for both GaAs and ternary III-V compounds of varying qualitative and quantitative composition have been obtained by using epitaxial growth from either the vapor[27] or the liquid phase.[16] Most of the work on (Ga,In)As has been done with crystals grown from the vapor phase,[5] whereas In(As,P) has been made predominantly by liquid-phase growth.[10] A more recent development is the use of the quaternary compounds InGaAsP, also grown from the liquid phase.[1] One of the major problems with all these materials is to adapt activation methods that work well under optimized experimental conditions in ultra-high vacuum to the processing used in the manufacture of photomultipliers and other devices.

In the formation of semitransparent cathodes, additional difficulties are encountered because crystallographically matching substrates are required. Moreover, the thickness of both substrate and cathode film must be carefully controlled to obtain optimum sensitivity and good uniformity. This problem does not arise with reflecting cathodes because, above a minimum thickness, sensitivity and uniformity are independent of thickness.

(e) Stability of III-V Electron Emitters

The stability of an electron emitter is determined by a complex combination of variables.[25] To begin with, one must distinguish between "shelf life" and "operational" stability. An essential requirement for good shelf-life stability is, of course, that the vacuum and activation techniques employed eliminate the possibility of chemical contamination after the device is removed from the pump system. In Cs-activated devices, it is also important that a Cs equilibrium be established such that the finite Cs vapor pressure of the emitter does not lead to a gradual loss of Cs from the cathode and, hence, a change in emission characteristics. During operation of the device, additional causes of instability arise. Probably the most important is the bombardment of the emitter by positive ions when the device is operated at high voltage. The ions have their origin either in unavoidable traces of residual gas and/or in gas released by electron bombardment from electrodes that are at high potential with respect to the electron emitter. In this case, the effect on the emitter will obviously increase with the amount of current taken from the device.

Of the materials described in this paper, the GaP dynode material

and the $Ga(As_{.75}P_{.25})$ photocathode, when operated in photomultipliers, are at least as stable in performance as conventional cathode and dynode materials. The shelf-life stability of GaAs and (Ga,In)As is also satisfactory but operational life in photomultipliers is lower. For example, at present the maximum permissible output current over long periods of time is 0.1 μA as compared with 10 to 100 μA for conventional materials. In practice this limitation is, fortunately, relatively minor, because in most applications the currents involved are much lower and are drawn only for short periods. The operational stability of the GaAs cathode has been substantially improved during the short period of 2 years since it was first used, and there is reason to believe that, with further refinements in the formation process, it will eventually approach that of conventional cathodes.

In general, it appears that the stability decreases in the sequence GaP-GaAs-(Ga,In)As, i.e., with decreasing bandgap energy. The reason for this decrease may be that the Cs vapor pressure of the materials (which has not been measured so far) increases with decreasing bandgap. Furthermore, minor changes in the surface characteristics are likely to be much more critical at the smaller NEA values associated with lower E_G.

5. Summary

Photoemissive and secondary electron emitting III-V compounds are superior to conventional electron emitter materials in most characteristics. Their use in practical applications has been held back by the more difficult and hence more costly technology for making the cathode materials. These difficulties include the necessity of using single crystals, the unavoidable exposure to air before cesium activation, the special problems encountered in making semitransparent cathodes, and some problems of operational stability. Despite these difficulties, III-V electron emitters of many different compositions are now in practical use and progress is being made in transferring new materials from the experimental to the practical stage. The most important applications are the following.

GaP is used as the secondary electron emitter for photomultiplier dynodes. The high gain factor in the first stage has enhanced the resolution at very low light levels, and the use of GaP in all stages has increased the speed of response and the time resolution.

GaAsP (approx. 75% As content) is used as a very stable photocathode with high response throughout the visible spectrum and extremely low thermionic emission at room temperature.

GaAs is used as a photocathode with uniformly high response from the UV up to approximately 9000 Å. Photomultipliers with this cathode have found widespread use in spectroscopy and similar applications because they are superior to previously available cathodes in at least three respects: (1) the quantum yield is higher, (2) the quantum yield is more uniform, and (3) the spectral response covers a range that previously required the use of two tubes with different cathode materials. The operational stability of GaAs cathodes is still inferior to that of conventional cathodes, but sufficient for spectroscopic applications where excessive light levels are never encountered. Improved stability and higher sensitivity for semitransparent cathodes are in the process of development and will further extend the usefulness of GaAs cathodes.

GaInAs photocathodes have the great advantage of longer threshold wavelength, but the activation and stability problems increase with increasing In content. Photomultipliers with GaInAs (6% In) are now becoming available. Although this small In content extends the threshold by only a few hundreds angstroms, this material is useful because it is sensitive to the radiation of the GaAs laser.

References:

[1] G. A. Antypas, et al., **Proc. 4th Int. Symp. on GaAs,** Inst. of Phys. Conf. Ser. No. 17, 1973.

[2] R. L. Bell, "Thermionic Emission of the GaAs Photocathode," **Solid State Elec.,** Vol. 12, p. 475 (1969), and "Thermionic Emission from III-V Infrared Photocathodes," Vol. 13, p. 397 (1970).

[3] R. L. Bell and W. E. Spicer, "III-V Compound Photocathodes: A New Family of Photoemitters with Greatly Improved Performance," **Proc. IEEE,** Vol. 58, p. 1788 (1970).

[4] N. P. Carleton and W. A. Traub, "Detection of Molecular Oxygen on Mars," **Science,** Vol. 177, p. 988 (1972).

[5] R. E. Enstrom, et. al., "Vapor Growth of $Ga_{1-x}In_xAs$ Alloys for Infrared Photocathode Applications," **Proc. 1970 Symp. on GaAs,** Inst. of Physics, London (1971).

[6] D. G. Fisher, R. E. Enstrom, and B. F. Williams, "Long Wavelength Photoemission from $Ga_{1-x}In_xAs$ Alloys,"**Appl. Phys. Letters,** Vol. 18, p. 371 (1971).

[7] S. K. Freeman, P. R. Reed, and D. O. Landon, **Raman News Letter** No. 41, p. 20 (1972).

[8] F. R. Hughes, E. D. Savoye, and D. L. Thoman, **J. Electronic Materials** (to be published).

[9] S. B. Hyder, "Thin Film GaAs Photocathodes Deposited on Simple Crystal Sapphire by a Modified RF Sputtering Technique," **J. Vac. Sci. and Tech.,** Vol. 8, p. 228 (1971).

[10] L. W. James, et al.,"Optimization of the $InAs_xP_{1-x}$-Cs_2O Photocathode," **J. Appl. Phys.,** Vol. 42, p. 580 (1971).

[11] H. R. Krall and D. E. Persyk, "Recent Work on Fast Photomultipliers Utilizing GaP (Cs) Dynodes," **IEEE Trans. Nucl. Sc.,** Vol. NS19, p. 45 (1972).

[12] B. Leskovar and C. C. Lo, "Performance Studies of Photomultipliers Having Dynodes with GaP(Cs) Secondary Emitting Surface," **IEEE Trans. Nucl. Sc.,** Vol. NS19, p. 50 (1972).

[13] J. Z. Liu, J. L. Moll, and W. E. Spicer, "Quantum Yield of GaAs Semitransparent Photocathodes," **Appl. Phys. Letters,** Vol. 17, p. 60 (1970).

[14] R U. Martinelli, "Infrared Photoemission from Silicon," **Appl. Phys. Letters,** Vol. 16, p. 261 (1970).

[15] G. A. Morton, H. M. Smith, and H. R. Krall, "Pulse Height Resolution of High-Gain First Dynode Photomultipliers," **Appl. Phys. Letters,** Vol. 13, p. 356 (1968).

[16] H. Nelson, "Epitaxial Growth from the Liquid State and Its Application to the Fabrication of Tunnel and Laser Diodes," **RCA Review,** Vol. 24, p. 603 (1963).

[17] S. K. Poultney, "Single Photon Detection and Timing in the Lunar Laser Ranging Experiment," **IEEE Trans. Nucl. Sc.,** Vol. NS19, p. 12 (1972).

[18] G. Present and D. B. Scarl, "Single Photon Time Resolution of Photomultipliers with Gallium Phosphide First Dynodes," **Rev. Sci. Instr.,** Vol. 41, p. 771 (1970).

[19] H. Schade, H. Nelson, and H. Kressel, "Novel GaAs-(AlGa) As Cold Cathode Structure and Factors Affecting Extended Operation," **Appl. Phys. Letters,** Vol. 20, p. 385 (1972).

[20] J. J. Scheer and J. van Laar, "GaAs-Cs: A New Type of Photoemitter," **Solid State Commun.,** Vol. 3, p. 189 (1965).

[21] R. E. Simon, A. H. Sommer, J. J. Tietjen, and B. F. Williams, "New High-Gain Dynode for Photomultipliers," **Appl. Phys. Letters,** Vol. 13, p. 355 (1968).

[22] R. E. Simon, A. H. Sommer, J. J. Tietjen, and B. F. Williams, "$GaAs_{1-x}P_x$ as a New High Quantum Yield Photoemissive Material for the Visible Spectrum," **Appl. Phys. Letters,** Vol. 15, p. 43 (1969).

[23] R. E. Simon and B. F. Williams, "Secondary Electron Emission," **IEEE Trans. Nucl. Sci.,** Vol. NS15, No. 3, p. 167 (1968).

[24] A. H. Sommer, **Photoemissive Materials,** John Wiley and Sons, New York (1968).

[25] A. H. Sommer, "Stability of Photocathodes," **Appl. Optics,** Vol. 12, p. 90 (1973)

[26] W. E. Spicer, "The Influence of Defect Levels on Photoemission," **RCA Review,** Vol. 19, p. 555 (1958).

[27] J. J. Tietjen and J. A. Amick, "The Preparation and Properties of Vapor-Deposited Epitaxial $GaAs_{1-x}$ P_x using Arsine and Phosphine," **J. Electrochem Soc.,** Vol. 113, p. 724 (1966).

[28] A. A. Turnbull and G. B. Evans, "Photoemission from GaAs-Cs-O," **J. Phys. D. (Brit. J. Appl. Phys.)** Vol. 1, p. 155 (1968).

[29] B. F. Williams and J. J. Tietjen, "Current Status of Negative Electron Affinity Devices," **Proc. IEEE,** Vol. 59, p. 1589 (1971).

Magnetoelectric Printing

E. C. Giaimo, Jr.

RCA Laboratories, Princeton, N. J.

Abstract—Magnetoelectric printing has successfully yielded good quality prints. The method uses the principle of counter-balanced forces on the toning material. The toning materials (iron, nickel, cobalt, etc.) can be acted upon simultaneously by both electric and magnetic fields. Light is used to modify the electric fields holding the toner to a charged photoconductive "surface" so that an applied concurrent magnetic field can physically remove the toner particles in areas where the electric field is reduced. The energy required to move the toner particles is derived from the external magnetic field and is not obtained solely from the stored electric energy, as is the case in some conventional electrophotographic systems.

Image-Formation Method

Magnetoelectric printing is a method for making images in a uniform layer of ferromagnetic particles carried on the surface of a charged photoconductive layer. The ferromagnetic particles are held to the photoconductive surface initially by electric forces alone. The ferromagnetic particles are capable of being acted upon simultaneously by both electric and magnetic forces. If the initial electric forces are

counterbalanced by magnetic forces and if the electric forces can be modified in selected areas by shining a light image on the photoconductive layer, then the magnetic forces can cause the ferromagnetic particles to move so that a pattern is produced in the particle layer which corresponds to the light image.

Images without edge effects (both black and white and continuous tone) have been made in layers composed of iron, nickel, and cobalt particles having dimensions in the range of 5 to 20 micrometers. A uniform layer of such particles can be produced on the surface of a photoconductive layer by first charging the photoconductor with a corona discharge in air from fine wires or points connected to a high-voltage power supply (3 to 10 kV) and then cascading a mass of the ferromagnetic particles over the corona-exposed surface. Good prints have been obtained on a particulate photoconductor–resin layer where the photoconductive crystallites were about 0.3 micrometer average diameter zinc oxide crystals. Such layers are commonly used as coatings for electrophotographic papers.

Scanning electron micrographs of the particulate zinc oxide–resin layers reveal the surface to be highly nonplanar and composed of a random distribution of photoconductor crystallites, insulating resin regions, cavities, and voids. Such a configuration on the microscopic scale of the 5 to 20 micrometer ferromagnetic particles should give rise to highly nonuniform electric field patterns at the "surface" of the photoconductive layer from one microscopic region to another. The fact that conductive iron, nickel, and cobalt powders adhere to the charged surface as they cascade over it as a loose stream of disconnected particles poured from an insulated polyethylene, glass, or grounded-metal container indicates that both positive and negative regions exist in the surface and that the particles become attached to one or the other of these regions without conductively connecting them. If intimate contact were to occur, current would flow through the conductive particles and thus provide discharge paths between such regions. If such discharge occurred, the conductive particles would fall off by gravity, for the electric fields introduced during corona charging would no longer exist.

The foregoing implies then that the charged regions introduced by corona charging are for the most part inaccessible to the conductive ferromagnetic material. When one also introduces the scanning electron microphotographic data into this framework, it is easily conceivable that the charged inaccessible regions are regions other than the high projections of the layer where conductive particles can contact and upon which they come to rest.

In order to determine the charge on the metal particles, a layer composed of photoconductive zinc oxide and resin was charged with a negative corona discharge on the photoconductive side and a positive corona discharge was simultaneously directed on the opposite side of its supporting paper substrate. A uniform layer of 5-micrometer iron was deposited on the photoconductive layer by passing the photoconductive strip through a trough containing the iron particles. A small region (¼ inch diameter area) was illuminated and the magnetic particles that were ejected when a magnetic field was applied were caught in a shielded metal container. The polarity of the collected particles was measured with an electrostatic voltmeter by measuring the charge on the metal receptacle containing the ejected particles. It was found they had a net positive charge. This appears to indicate that the conductive particles contacted positive regions and that the negatively charged regions were not accessible to the metal particles. Further, the conductive particles would also obtain a positive charge if they were inductively charged by momentarily being connected to earth while in the vicinity of a negatively charged body and they would then be attracted to the negative region.

When such particles are poured over the charged photoconductive layer, the microscopic structural details of the resulting ferromagnetic particle layer is dictated primarily by the local electric fields. These fields exist within the particle mass, between the conductive particle mass and the photoconductor-resin layer, and between one region of the photoconductor-resin layer and another. The particle layer structural details are not those the particles would assume in an externally applied magnetic field. Therefore, if the ferromagnetic particle layer first established by the electric fields is subsequently subjected to a magnetic field, the magnitude and orientation of which cannot appreciably modify the structure established by the electric fields, the particle layer will be stable. On the other hand, since the electric fields at the surface of the photoconductive layer can be modified with light, then one is able to create structural instabilities within the ferromagnetic particle layer by shining light on the photoconductive layer while impressing an appropriate magnetic field on the ferromagnetic layer. The resulting particle instabilities caused by changes in the electric forces can be made sufficiently large to cause the magnetic particles to leave the layer region, thus generating a pattern within the ferromagnetic particle layer corresponding to the light image impressed on the photoconductive layer.

As the electric forces on the ferromagnetic particles decrease during illumination of the photoconductive layer, the local anisotropy of

the particle mass covering the illuminated region varies as the light exposure proceeds. The magnetic particles line up along field lines and form chain-like structures extending in the direction of the nearest magnetic pole or toward a region of increasing magnetic flux density. Gyromagnetic effects also appear to be present during rapid transitions in the magnetic field when it is applied and removed or when the magnetic field is applied in a series of pulses, thus giving rise to particle motion and torques.

The energy utilized to create changes in the magnetic particle layer is derived from the magnetic field and not solely from the electric fields existing in the photoconductive layer. In normal electrophotographic systems using electroscopic toner particles, the energy needed for image development is derived solely from that stored in the photoconductive layer during the electric charging of it and from that which remains after exposure to light. In the magnetoelectric case, the work done on the particles is that resulting from the net force on the particles that is magnetic. That the particles come off with a net positive charge indicates that work is being done magnetically to overcome the electric forces trying to retain the particles to the surface. This effect is also exemplified by the fact that a visible image is not produced in the magnetic particle layer if the magnetic field is not applied.

Experimental Processor

To implement the magnetoelectric image formation method described, the processor shown in Fig. 1 was designed and constructed. Moveable-air-cored coils were chosen for the experimental device, because of the ease with which the magnetic field can be adjusted with coils as compared to a permanent magnet structure. The air-cored coils also provide means for viewing image formation and the ferromagnetic particle motion during the exposure and processing. In the processor the following steps are performed in the order listed:

[1] The photoconductive layer, carried on a paper substrate strip, is passed through a high-voltage double-corona charge unit.

[2] A uniform layer of ferromagnetic particles is applied to the charged photoconductor from a wedged shaped trough containing the particles.

[3] An optical image is projected on the rear side of the photoconductive layer, and, either concurrent with or immediately subsequent to the exposure, a pulse magnetic field is applied. The

density of particles is reduced in the particle layer adjacent to the illuminated areas by the action of the impressed magnetic field.

The processor was designed for experimental modification and flexibility. It consists of several main parts including a supply roll carrying a strip of photoconductive paper, a trough containing a mass

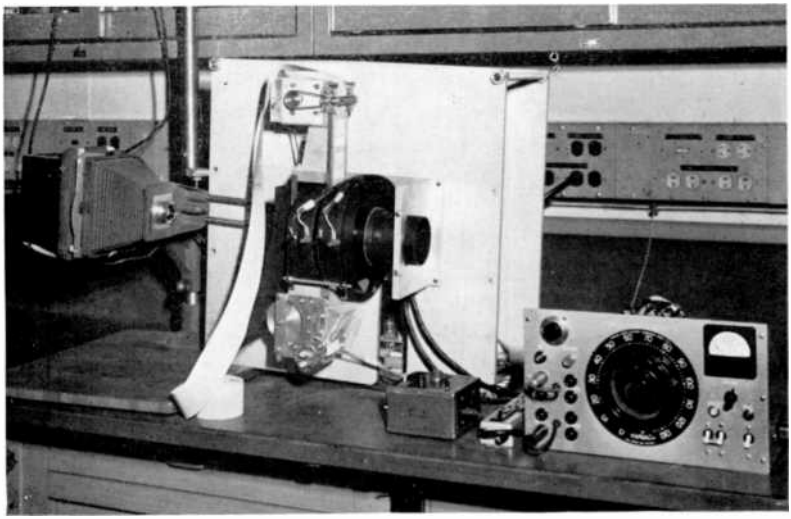


Fig. 1—Experimental magnetoelectric image processor.

of magnetic toner particles, means for guiding and moving a photoconductive strip along a predetermined path, a pair of air-cored coils, and a timed variable power source for energizing the coils to produce a controlled pulsed magnetic field. The magnetic field can be controlled with respect to magnitude, time of application, and time duration. The position of the coils with respect to the photoconductive strip passing between them is adjustable. Image exposure of the photoconductive strip is done using a timed projector; a positive transparency or other exposure means can also be used.

Fig. 1 is an overall view of the apparatus, and Fig. 2 is a close-up photograph of the output copy on the paper strip as it emerges from between the coils. Fig. 3 is a close-up photograph of the double corona charger and the toning trough. The picture shows charged paper entering the trough, and the strip, uniformly coated with the ferromagnetic particles, leaving the trough. The particles are prevented from leaking out of the bottom of the trough by a magnetic seal con-

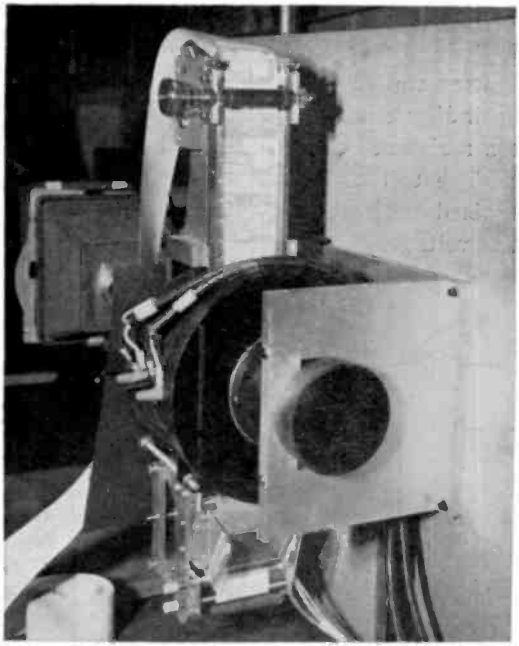


Fig. 2—Closeup of output copy on paper strip.

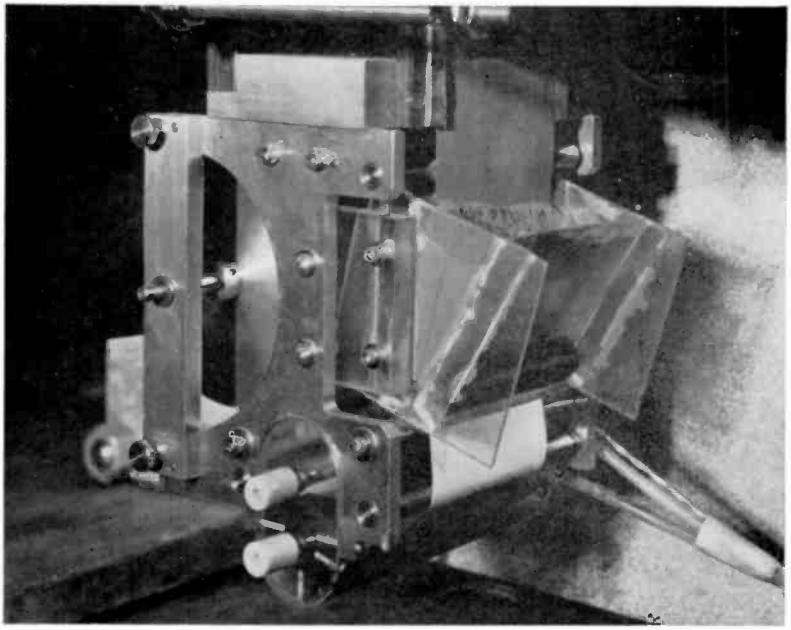


Fig. 3—Closeup of double-corona charger and toning trough.

sisting of two magnetic strips, one on each side of the lower end of the opening through which the paper enters the trough.

The magnet coils were wound with #17 copper wire insulated with H.D. Formex. The winding cross section is 1.4 × 1.4 inches for each coil. The coils have a resistance of 5.7 ohms and an inductance of approximately 80 mh. The coil turns are approximately 700 per coil. The coils were energized from full-wave rectified 60 Hz at between 70 and 100 volts peak voltage.

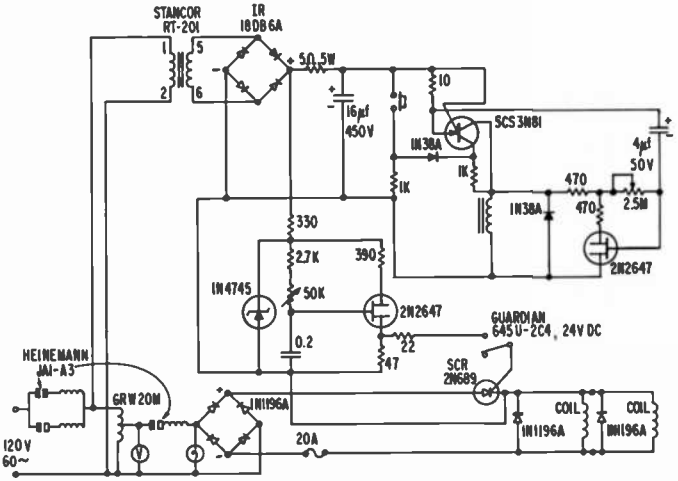


Fig. 4—Schematic diagram of experimental power supply.

The experimental power supply is shown in Figs. 1 and 4. An SCR and solid-state timer is used to control the time of application and the duration of the current applied to the coils. Typically the coils are energized for approximately ½ second at the end of the light-exposure interval. Light exposures through the paper substrate are similar to those used for conventional electrophotography.

Fig. 5 shows typical copy made with the experimental device showing continuous-tone and large-area fill-in lacking edge effects. The continuous tone prints were made using nickel powder (Fig. 5(a) and (b)) and the black-and-white print was made with iron powder (Fig. 5(c)). The prints were fixed with a clear lacquer spray applied in the dark and immediately after the prints were formed. Other methods of fixing, such as heat or solvent, might be applicable to other magnetic toners than the metal powders used for the prints of Fig. 5.

Conclusions

Magnetoelectric printing method provides a means for coupling energy to the toner particles without physically contacting the photosensitive medium. Various techniques utilizing electric fields have been proposed that require contact to the photoconductive layer and that the substrate upon which the photosensitive layer is coated have high conductivity. In some schemes conductive transparent electrodes have also been used. The magnetic coupling to the toner can be accomplished at large distances as compared to the close spacings required of external electric-field coupling methods of altering or enhancing development and toner motion.

A

B

C

Fig. 5—Typical copy made from experimental device: (a) and (b) continuous tone prints using nickel powder and (c) black-and-white print using iron powder.

In addition, the magnetoelectric method has given prints with large filled-in areas lacking edge effects and with good continuous-tone capabilities. The application of the basic method to color printing has been experimentally pursued but will not be elaborated upon at this time.

Since the prints shown in Fig. 5 are composed of metallic conductive magnetic particles, various image readout and reproduction methods utilizing these toner properties are possible, such as magnetic head reading.

The magnetoelectric printing method has several advantages. It is dry. Unfixed images can be stored for long periods on the inexpensive photoconductive layer. The unfixed images can be easily erased using the same magnetic-field-generating means employed for forming the image. Erasure is accomplished by reenergizing the coils if the electric field has decayed sufficiently with time or the unfixed image can be flood lighted to cause complete decay of the electric field. It has been found that the photoconductor can be repeatedly used after magnetic erasure of the unfixed images.

The fixed or unfixed images can be used as intermediates for optical print out on a less sensitive print-out medium. A magnetoelectric unfixed print on a transparent photoconductive layer has been printed out on an inexpensive low-sensitivity photochromic material. The transparent layer could then be erased and reused. The unfixed transparencies can also be projected for display by appropriate floodlighting.

The Bivicon† Camera Tube—A New Double Vidicon

R. L. Spalding, S. A. Ochs, and E. Luedicke*

RCA Electronic Components, Lancaster, Pennsylvania

Abstract—The RCA Bivicon television camera tube, a 1½ inch vidicon with two separate guns and targets, was developed for use in two-frame color systems. Its special features include a non-parallel twin-gun structure, a novel mesh-support design, and a faceplate with a double target structure and embedded signal contact buttons. The Bivicon produces excellent color signals with the Holotape† system. It can also be used in film chains, live pickup television cameras, and in a variety of other schemes based on a two-frame television system.

Introduction

Several methods have been developed for generating color-television signals by means of two separate camera tubes. In the RCA Holotape video playback system, for instance, two optical frames are imaged onto the two vidicon targets for each single television frame of the output color signal. The two input frames are in monochrome, one

* RCA Laboratories, Princeton, N.J.

† Registered trade name of RCA.

containing the luminance portion of the television signal, while the other frame contains all the information necessary for generating the chrominance portion.[1] Such a system requires excellent registry, i.e., spatial alignment, between the two scanned rasters so that the two simultaneous signals will be accurately in phase with each other. This requirement suggests that several advantages can be gained from combining the two tubes within a single enevelope:

(1) The relative position of the two sets of components of a double tube is established within its structure. If the tube has been properly designed as to rigidity and thermal stability, only the initial mechanical adjustment is needed for obtaining the required superposition of the output signals.

(2) A smaller and lower-cost camera results because the double tube can be operated in a single focus coil and deflection yoke.

(3) With the two camera-tube targets located on the same faceplate, the two optical images can be positioned close together, which can lead to considerable simplification of the camera optics. For instance, it would be possible to operate with a single lens.

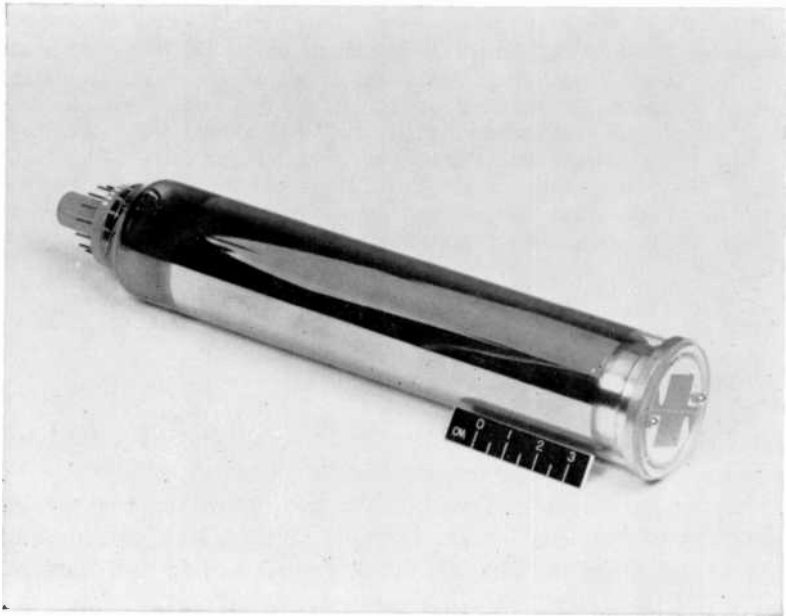


Fig. 1—Bivicon camera tube.

A double vidicon, of course, is only useful if it yields a satisfactory degree of coincidence of the two output signals and if it is economically practical. These conditions are satisfied by the Bivicon camera tube[2] (developmental type C23244), which is described in this paper. Its external appearance is shown in Fig. 1. The following sections discuss the special design features and the performance of this tube.

Design

One of the design goals was to produce a tube that can achieve a color signal comparable to that obtained from two one-inch vidicons. This goal suggested that each of the two rasters have the same dimensions as that used with the one-inch tube, i.e., 0.50×0.375 inch. With a small separation between the scanned regions, the total area required for the two rasters then became 0.50×0.875 inch, requiring a diagonal of slightly more than one inch. An outside diameter of 1.5 inch therefore was chosen for the tube. In order to achieve satisfactory resolution performance, it was decided to base the electro-optical system on that of the 8521 vidicon, a 1.5-inch-diameter vidicon of high resolution capability. This decision implied the use of magnetic focus and deflection and the inclusion of a separately connected mesh electrode near the two signal plates.

Because the two electron optical systems should yield identical scanned rasters, they were placed in symmetrical locations relative to the tube axis. For proper performance, as well as in the interest of economy, the tube has several unusual features which are described below.

Envelope

In order to minimize parts costs, it was decided to use soft (potash soda lead) glass for the envelope of the tube instead of the usual hard (borosilicate) glass. Use of soft glass not only reduces the envelope cost, but also permits the use of an inexpensive stem (the lower end of the envelope, which contains the contact pins and the exhaust tubulation) used in large quantities in the manufacture of picture tubes. Since the faceplate is attached to the envelope by an indium seal, as used in other vidicons, the decision to make the envelope of soft glass did not affect the choice of glass for the faceplate.

Guns

The guns developed for the Bivicon, shown in Fig. 2, have a "triode" arrangement, similar in concept to that of most vidicons, but their actual construction is based on that of the gun used in the RCA small-neck 110°-deflection-angle color picture tubes. This design was chosen because of its adaptability to large-quantity production. Fig. 3 is a sketch of one of the two (identical) guns developed for the Bivicon tube. Each gun consists of a cathode structure, a control (G_1) grid, an accelerator (G_2) assembly, and a focus (G_3) cylinder. These electrodes are held in place by two glass beads. This arrangement avoids the need for a brazed cathode–G_1 assembly, as is used in most other camera-tube and picture-tube guns.

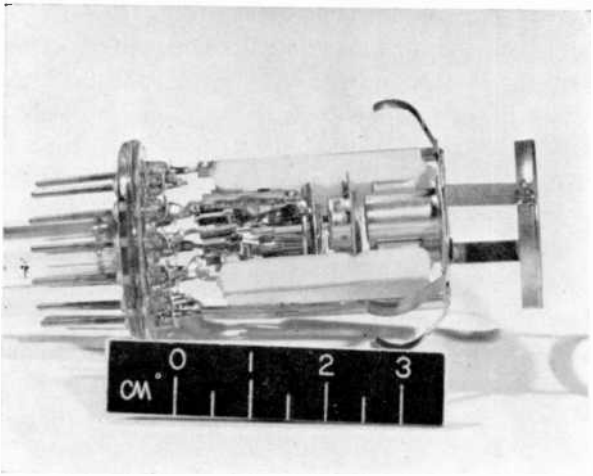


Fig. 2—Double-gun for Bivicon tube.

A special feature of the electron gun is the method by which the limiting aperture (indicated in Fig. 3) is mounted within the G_2 electrode. When the gun is in operation, a cone-shaped stream of electrons issues from this aperture and is then focused, through the action of the electro-optical system of the tube and magnetic focus field, into a minute spot that scans the photoconductive target. If the aperture, whose diameter is 0.0015 inch, is off the gun axis by more than 0.0001 or 0.0002 inch, there will be an appreciable loss of resolving power of the beam and of its ability to properly discharge the

charge pattern stored on the photoconductor. A technique was therefore developed that results in accurate centering of the limiting aperture and is well suited to the mass production of tubes.

The two electron guns are mounted on either side of, and symmetrically with respect to, the tube axis. However, they are not parallel to the axis or to each other. To understand the actual location and orientation chosen for the guns, consider the path followed by an electron that starts from a source at a distance from the tube axis with an initial velocity parallel to this axis. An analysis of the motion of this electron shows that, with no deflection field applied, the electron arrives at the target (1) closer to the tube axis than the radial distance of its starting point, (2) in a different axial plane (i.e., at a different azimuthal angle) than that of its source, and (3) at an oblique angle of incidence.

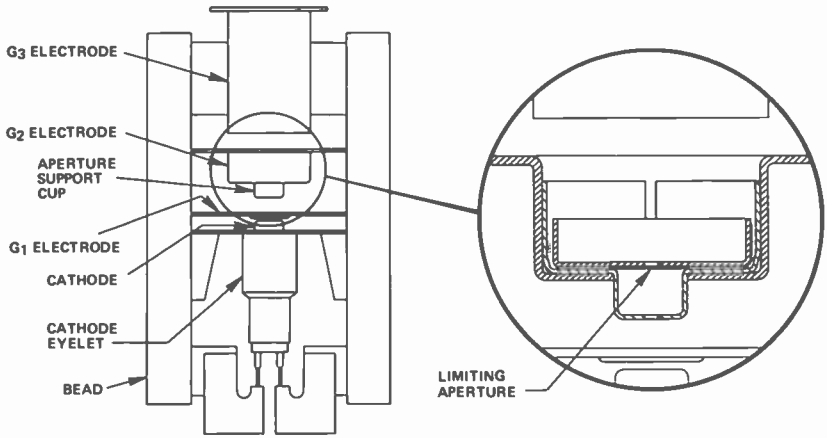


Fig. 3—Sketch of one of two Bivicon guns.

The first of these three items can be accommodated readily by determining the proper distance of the electron source (i.e., of the limiting aperture of the gun) from the tube axis, so as to yield the desired separation of the two electron spots on the target. Similarly, the second item results in an angle of rotation between the line joining the two limiting apertures and the common center line of the two scanned rasters that can be computed and measured. It is then taken into account by a suitable angular rotation of the electrode pattern on the tube faceplate relative to the position of the guns.

Item (3), however, cannot be dealt with quite so easily. It is complicated by the fact that when deflection fields are applied, the particular electrons considered above (those leaving the limiting apertures parallel to the tube axis) arrive at different angles of incidence depending on the amplitude and direction of the deflection fields at any given moment. In practice, this difference in the angle of incidence means that the tube would show objectionable shading if no correction were applied. This problem was solved by so aiming the two guns that the undeflected beams leave in such directions that they arrive at their respective targets parallel to the tube axis. When deflection fields are now applied so that the two electron spots scan their individual rasters, the angles of incidence will still vary slightly from the value at raster center, causing a certain amount of shading towards the corners of the scanned regions. However, the magnitude of this effect is no greater than it is in standard camera tubes, where it generally is found to be acceptable.

Upper G_3 Electrode and Mesh

In the interest of minimizing tube costs, a large portion of the G_3 electrode consists of a metallic coating on the inside wall of the glass envelope. This coating replaces the metal cylinder used in most other camera tubes that is connected to the beaded gun assembly by a funnel-shaped connecting piece. A considerable saving is achieved by eliminating the latter part, which would have a complicated shape to fit the double-gun structure, and the metal tubing connected to it. Also, due to the higher sheet resistance of the wall coating as compared to the resistance of a reasonably rigid metal cylinder, there is less loading of the deflection yoke (smaller eddy currents are induced). Several leaf springs are used to make contact to the wall coating at both ends, i.e., near the gun and near the mesh.

At the upper end of the tube is a structure, shown in Figs. 4(a) and 4(b), that contains two electrically isolated sections. One part is the upper end of the G_3 electrode and is connected to the conducting wall coating on the envelope. The other section consists of the mesh assembly. The tightly stretched nickel mesh contains 1000 lines per inch and is mounted parallel to the plane of the two targets. This arrangement, which basically is similar to that used in most camera tubes, provides an electrostatic lens (between G_3 electrode and mesh) followed by a region of uniform electric decelerating field (between mesh and faceplate). Its function is to bend the path of the electrons leaving the deflection region so as to cause them to approach the target

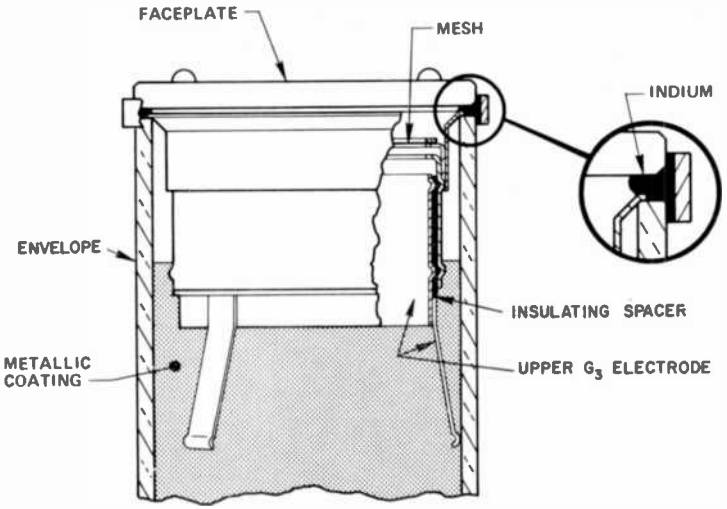


Fig. 4(a)—Sketch of mesh structure.

as nearly perpendicular to the target plane as possible. In other tubes that have continuous G_3 cylinders, as described above, the upper portion of the gun mount, including the mesh, is supported by the beaded lower gun assembly. The mesh electrode is then connected to one of the stem pins by a wire running the length of the tube next to the envelope wall. In the novel arrangement developed for the Bivicon tube, the mechanical support for the upper structure of the gun and

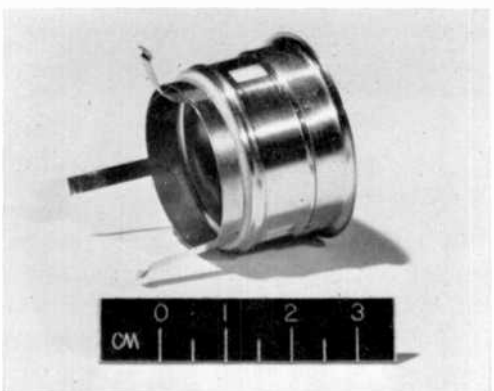


Fig. 4(b)—Photograph of mesh structure.

the electrical connection to the mesh are provided by a funnel-shaped ring that rests on the target end of the glass envelope and is rigidly secured in place by the indium faceplate-to-envelope seal. The external contact to the mesh is then made via the metal ring that surrounds the indium seal.

Faceplate

Fig. 5 is a view of a Bivicon faceplate as seen from the outside of the tube. Two conducting plugs set in the faceplate provide the external signal connections. On the inner surface of the faceplate are two separate transparent signal electrodes covered by a common layer of antimony trisulfide photoconductor. Each scanned raster area is surrounded, on three sides, by a "frame" of evaporated aluminum to

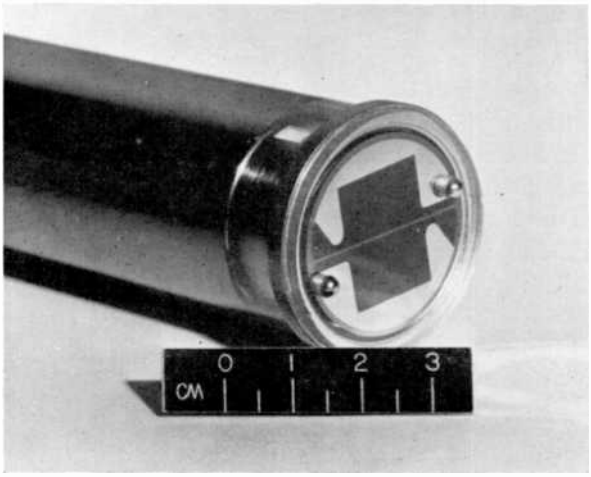


Fig. 5—Faceplate.

minimize variations in resistance from different regions of each raster to the corresponding conducting plug. (In conventional camera tubes this requirement is taken care of by the indium seal material that contacts the circular signal plate along its periphery.) Each conducting plug contacts one of the aluminum regions on the inner faceplate surface. On the other end of each plug, a convex metal button is provided for ease of making the signal connection.

Tube Assembly

The stem of the tube was chosen as the basic reference for precisely aligning the guns, envelope, and faceplate. Thus, in mounting the two-gun assembly on the stem, a fixture is used that ensures accurate positioning of the guns relative to the tube pins. The glass envelope is subsequently sealed to the stem, again lined up with reference to the pins such that the central axis of the gun assembly coincides with the center line of the envelope. Finally, when the faceplate is attached by an indium seal to the other end of the envelope, the pins in the tube stem again are used to provide the reference needed for proper alignment of the faceplate.

In this fashion the critical parts of the tube, such as the two guns, the G_3-mesh section, and the signal electrodes, are accurately positioned in relation to each other, and the user of the tube has a ready reference—the pin structure—for mounting the tube properly in the focus coil and deflection yoke of the television camera.

This method of tube assembly, in contrast to that used for other camera tubes, does not depend on optical alignment using human judgment. Also, it permits the use of relatively economical glass envelopes with reasonable wide dimensional tolerances because the inner envelope surface is not used for aligning the gun mount with the tube axis. However, the method described above requires that the structure that holds the gun mount to the stem pins be sufficiently stiff to ensure that it will not deform when it is handled or when it is inserted into the envelope. Therefore, several of the rather flexible stem leads that support the gun assembly were replaced with stiff support beams made by metal ribbons longitudinally bent so as to give them an L-shaped cross section.

Performance

In typical operation, the voltage of the signal electrodes (relative to the cathode) is adjusted so that the current for each target in the dark is 20 nA with a raster size 0.50 × 0.37 inch, an accelerating electrode voltage (E_{c2}) of 300 V, a focus electrode voltage (E_{c3}) of 900 V, and mesh voltage (E_{c4}) of 1400 V. Typical target voltage is 20 to 40 V. The required field strength then is about 46 gauss at the center of the focusing coil.

Because the photoconductor material used for the Bivicon tube is the same as that used for commercial vidicons such as the 8507A,

7735B, and 8521, the tube has essentially the same photoconductor-dependent performance characteristics (e.g., spectral response and sensitivity).

The resolution capability of the Bivicon is shown in Fig. 6 as a plot of the amplitude-response characteristic for each of its two outputs. The horizontal peak-to-peak response, without aperture correction, is given for a square-wave test pattern at the center of each 0.50×0.37 inch raster. The illumination is such that the high-light signal current is 300 nA. The abscissas are given in terms of television line number (total number of black-and-white lines per raster height) as well as in line-pairs per millimeter. Although not shown in Fig. 6, the limiting resolution in the center exceeds 1000 television lines, while near the corners it ranges from 500 to 700 television lines.

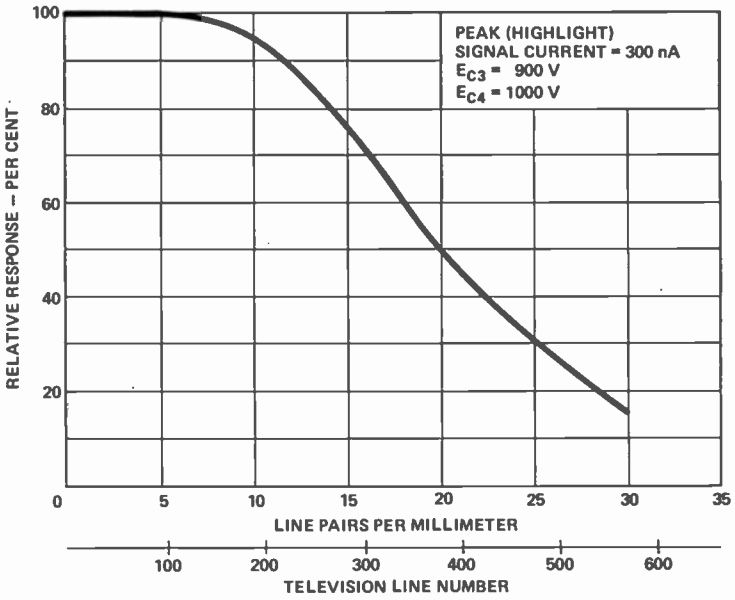


Fig. 6—Horizontal square-wave response.

When a Bivicon tube is first operated in a standard 1½ inch focus-coil/deflection-yoke, the output signals from its two signal plates overlap, but are not registered perfectly. To obtain scan registry, several parameters can now be adjusted such as the electrode voltages and rotation of the tube in the yoke. Without much difficulty the two

output signals can be registered to a high degree of perfection. In a typical case, with the two raster center points conicident, the greatest discrepancy in any region of the raster will be no more than 1% of the raster width.[1]

Conclusion

The RCA Bivicon television camera tube was designed to perform a quite demanding task and yet be suitable for use in low-cost systems. As a result of several innovations in construction methods, materials, and design concepts, these goals have been attained. The initial application of the tube was its use in two-frame color systems, and it has been demonstrated to perform its function in this application exceedingly well. Among the other applications in which such a "two-eyed" camera tube might be usefully employed are stereoscopic television, comparators, motion detection, surveillance with periodic switching between objects, and simultaneous pickup using wide and narrow-angle lenses with fast switching between the two.

Acknowledgment

The authors gratefully acknowledge the contributions made by J. D. Herrington and P. W. Rothweiler in the development work on various aspects of the tube and by J. F. Otto in making prototype tubes and valuable suggestions.

References:

[1] R. E. Flory, "Applications of the Bivicon Tube," **RCA Review**, Vol. 34, No. 1, p. 132, March 1973.

[2] R. L. Spalding et al, "Bivicon—A New Double Vidicon," **Proc. IEEE**, Vol. 60, p. 1236, Oct. 1972.

Applications of the Bivicon* Tube

R. E. Flory

RCA Laboratories, Princeton, N. J.

Abstract—The two-beam, two-raster Bivicon camera tube has some unique operating characteristics. These characteristics are described, as is a modular camera designed for the tube. Applications for the tube are also described; they are divided into three categories. First, are those applications using pre-encoded imagery for color pictures, such as Holotape.* Second are those applications that require color encoding in the camera, such as cameras for film or direct pickup. A general discussion of color encoding is included. The third category includes all those applications of the Bivicon that utilize the property of two rasters in some way other than for color television.

1. Introduction

The Bivicon tube was conceived and developed for the RCA Holotape video playback system. In this system it was recognized that limitations of bandwidth and spatial noise in the holograms precluded the application of color-encoding schemes using a single image. It was decided that a two-image format would be used, allowing substantial reduction of the spatial frequency of the color-encoding structure. In

* Registered trade name of RCA Corp.

color-television systems other than Holotape, the same performance limitations may lead to consideration of two-image formats. Some two-image formats allow luminance performance limited only by the camera tube, and place all color information in the second frame, on spatial carriers well below limits set by scanning-aperture losses and target-capacitance loading effects. In such cases, the use of the dual-raster Bivicon tube may make such a format more attractive than either single-image or three-image approaches.

2. Circuit Considerations for Bivicon Cameras

2.1 Registration

Accuracy of registration is, of course, of prime interest, since it is this attribute which gives the Bivicon its advantage over two vidicons. Horizontal and vertical center registration are achieved by different means. If the two images incident on the Bivicon targets are moveable, this is the preferred technique, as no compromise in electron optical performance is necessary. However, horizontal center registration may be effected by rotating the tube about its long axis, inside of the deflection-focus coil assembly. Generally, a raster shift of 10% of raster width can be achieved before reaching a limit set by defined target areas. To assure stable horizontal registration it is necessary to stabilize G_3 potential and focus-coil current. A change in G_3 potential of about 25 V will shift rasters by 1% of picture width. This shift in raster positions is noticed before defocusing takes place. Over a small range of horizontal movement, it is therefore possible to use G_3 potential for electrical horizontal registration control.

For vertical registration, adjustment of the optical image spacing is recommended for coarse adjustment. An electrical adjustment useful for a range of several percent is a change in the G_3-G_4 potential difference, effected by changing G_4 potential. As in any vidicon, a change in this potential affects the magnification of the electron optics near the target area, and the first-order effect is one of differential vertical movement. A change of 20 volts in G_4 potential changes vertical registration by about 1% of picture height. Image focus is not affected if G_3 is held at constant potential. The limit of several percent registration adjustment by this means is set by distortions of the rasters introduced by operating the G_3-G_4 lens outside of its design range.

When registered perfectly in the center, present Bivicon tubes, used with the recommended focus and deflection components, provide two-raster registration within 1.0% or less of picture width over the

entire raster. The character of the error is remarkably uniform over a large number of component assemblies and tubes made to the same specifications. Fig. 1 shows the registration errors in a typical Bivicon tube. A pair of cross-hatch test patterns was imaged on the two

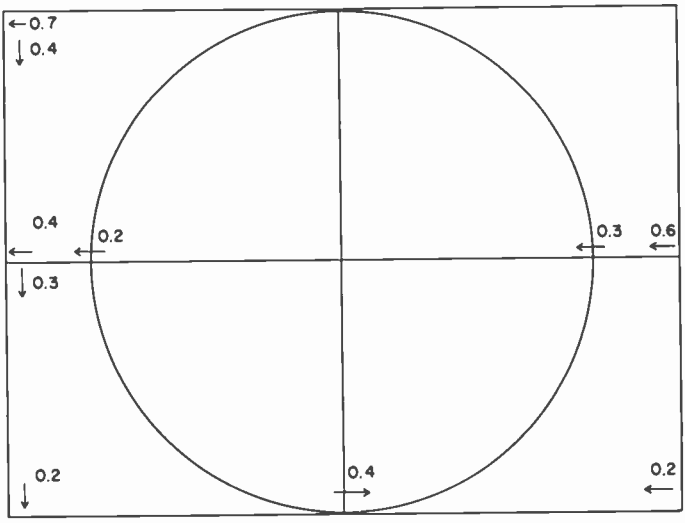


Fig. 1—Bivicon registration errors. Numbers are magnitudes in percent of picture width. Arrows indicate displacement of raster 1 relative to raster 2.

targets of the tube by an optical system previously proven to have negligible geometrical distortion. The two video outputs were superimposed with the centers of the two rasters adjusted for spatial coincidence on the monitor. The figure shows the magnitude and direction of registration errors, resolved in the horizontal and vertical directions. Absence of data indicates that the errors were smaller than 0.2% of picture width. To date, little effort has been applied to modification of the deflection and focus components to improve registration quality, for two reasons. First, we have been able to apply time-varying registration correction; this is discussed below. Second, for many applications the present registration quality is adequate. Subjective tests made in the laboratory indicate that if luminance information is registered accurately with chrominance information in the center of a picture, a gradually increasing error from center to edge of the picture is tolerable in a consumer product if the error at a circle touching raster top and bottom is less than 1%. Errors less than 0.25% are tolerable at the center of the picture. The present

Bivicon geometry achieves this without dynamic correction and with 0.5% stabilization of tube potentials.

In cases where the registration quality must exceed the figures cited above, such as applications not registering luminance and chrominance, dynamic correction has been used. The easiest dynamic registration mechanism to understand is modulation of G_4 voltage. Since changes in this voltage move the rasters vertically with respect to each other, it can be seen that vertical-registration errors in different parts of the raster can be corrected by application of appropriate time-varying voltages to G_4, superimposed on the required dc potential. As a practical matter, the close proximity of G_4 to the target results in too much capacitive coupling of these voltages to the signal plate to allow the use of other than vertical rate correction waveforms. As a consequence, the type of error correctable with G_4 modulation is that which requires vertical displacement at a vertical rate. This type of error is most clearly described as a vertical size mismatch.

Electromagnetic means are used to effect horizontal correction. The focus-deflection assembly used with the Bivicon includes beam-alignment coils. To achieve the usual translation of the beam for alignment purposes, these coils are connected with their fields adding. If the coils are connected so that the fields oppose, with flux lines originating parallel to a line drawn between the gun axes, the net result is to deflect one raster to the right, and the other to the left. This is shown in Fig. 2. If the guns are oriented instead in plane AA (see Fig. 2), it can be seen that the direction of flux is such as to provide differential vertical motion. Therefore, to provide dynamic vertical registration control, we install a second set of coils rotated 45° about the tube axis with respect to those used for horizontal registration. Substantial improvement in symmetry and independence of adjustment of vertical registration is achieved by using a quadripole coil assembly, with all coils connected in series.

It is possible to drive the horizontal and vertical registration coils with time-varying currents at both horizontal and vertical rates, resulting in a highly flexible registration system. The horizontal-registration coils are the same ones used for beam alignment. In the vertical case, new coils must be added, or the G_4 voltage modulation used instead, with the added constraint of using only vertical rate corrections.

2.2 Target and Preamplifiers

The target of the developmental type C23244 Bivicon is the standard RCA photoconductor, PCII.[1] Since each raster is the same size as in

a one-inch vidicon, the performance of the Bivicon with regard to sensitivity, lag, dark current, etc, is that of a standard RCA vidicon. A single test was made of a silicon-diode-target Bivicon, with satisfactory results.

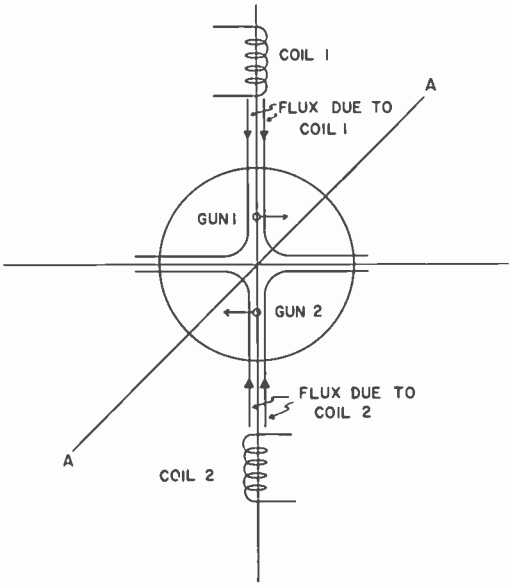


Fig. 2—Cross section of tube at alignment coil plane.

The location of two targets in such close proximity has resulted in some interaction. A cross-talk mechanism exists between the two targets. A high-contrast image on one target, having large energy components in the range of 0.5 to 3 MHz, will be evident in the video output from the second target. The cross talk usually takes the form of an outline ghost of one target's image on the other target output. The differentiated nature of the output suggests a capacitive coupling mechanism. This problem is particularly acute in applications of the Bivicon using a relatively low-frequency spatial color-encoding carrier image on one target and a luminance image on the other target. In this case, the luminance image is marred by the presence of coarse stripes cross-talking from the chrominance image.

Capacitance between targets was measured to be 1 pF, which yielded cross-talk magnitude consistent with experiment when using the simplified equivalent circuit shown in Fig. 3. I_{T1} is the signal current from target one. Current I_1 flows in R_1, which is the series

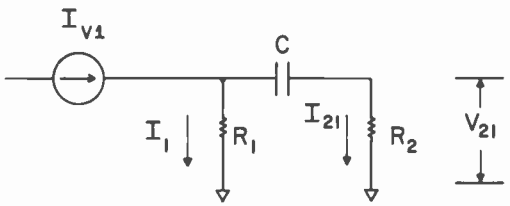


Fig. 3—Simplified equivalent circuit of capacitance between targets.

combination of preamplifier one input impedance and series impedance between target proper and the external target terminal. The capacitance between targets is C, and R_2 represents preamplifier two input impedance. The cross-talk voltage in preamplifier two due to signal on target one is then V_{21}. Of course the same mechanisms operate to produce reciprocal effects in preamplifier one.

Reduction of this cross talk to an acceptable level (50-60 dB below in-channel video) was achieved by developing a low-input-impedance preamplifier. The simplest of these uses a bipolar transistor in the grounded-base configuration. More satisfactory performance was achieved by use of feedback with a JFET input device. An input impedance of about 200 ohms is desirable. A simplified schematic is shown in Fig. 4. The low-impedance amplifier benefit can be negated

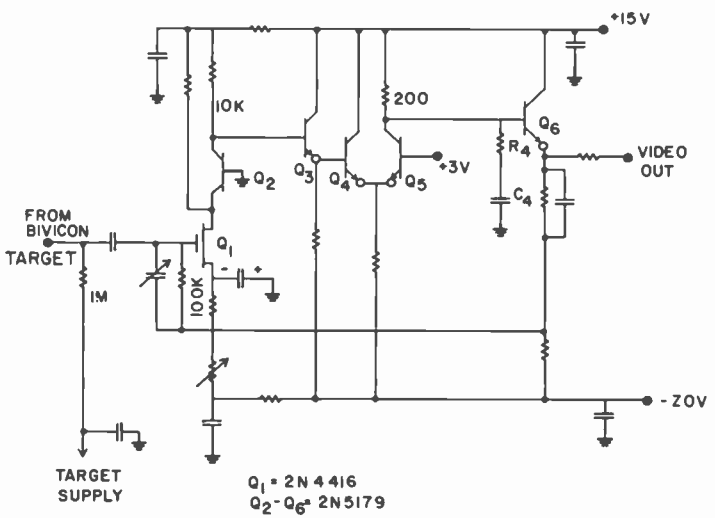


Fig. 4—Simplified schematic circuit of low-input-impedance preamplifier.

if the impedance between the target and outside target terminal is too high. Since this impedance is associated physically with a transparent conductor, it is not easy to ensure low impedance and high optical transmission. For this reason, we have specified an opaque conductor surrounding the active target area to shunt the transparent conductor outside of the picture area.

3. Basic Bivicon Camera

A flexible, modular camera has been constructed for the purpose of evaluating the Bivicon in a number of applications. We have not attempted to simplify or cost reduce this camera, and wherever possible have utilized modules from existing cameras or outside vendors. I shall describe the basic camera briefly and then review a number of applications.

A block diagram of the complete camera is shown in Fig. 5 and a photograph in Fig. 6. Extra space in the circuit-board rack was used in specific versions of this camera for color transcoding from the spatially encoded form to the composite NTSC signal. In the diagram and in the description that follows, I have minimized reference to many details of video technique in order to highlight practices unique to the Bivicon tube and camera.

The focus–align–deflection coil assembly is a standard product designed for the 1½-inch vidicon, RCA type 8521. The deflection currents are approximately 180 mA and 35 mA peak to peak for horizontal and vertical deflection, respectively. These are less than required for the 8521 tube in this yoke because each Bivicon raster is smaller than the usual raster in the 8521 vidicon.

The preamplifiers are the JFET input feedback amplifiers mentioned earlier. In spite of the low input impedance, they are located in close proximity to the target terminals in order to minimize capacitance, which would adversely affect signal-to-noise ratio. The low-impedance input does, however, eliminate the need for the usual capacitance compensation in vidicon amplifiers, since $R = X_c$ at about 50 MHz. Independent manual adjustment of target voltages is used, but the automatic sensitivity control adjusts both targets in response to video information from one channel only.

Video in both channels is clamped, blanked, and amplified in processing amplifiers. It should be noted at this point that for some applications, the video from the chrominance channel will undergo various band-filtering operations, obviating need for clamping, aperture correction, or sync adding in the chrominance channel. In the other

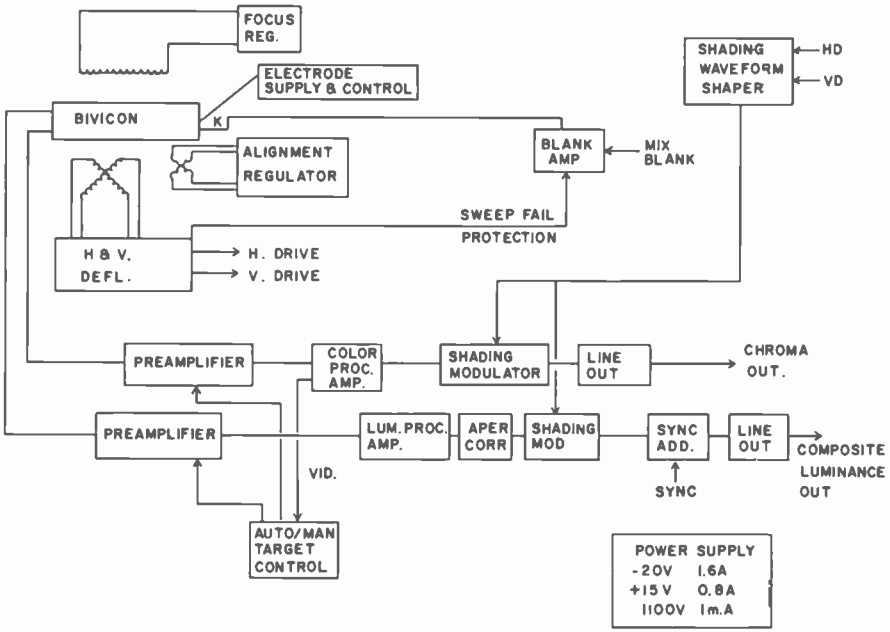


Fig. 5—Block diagram of camera.

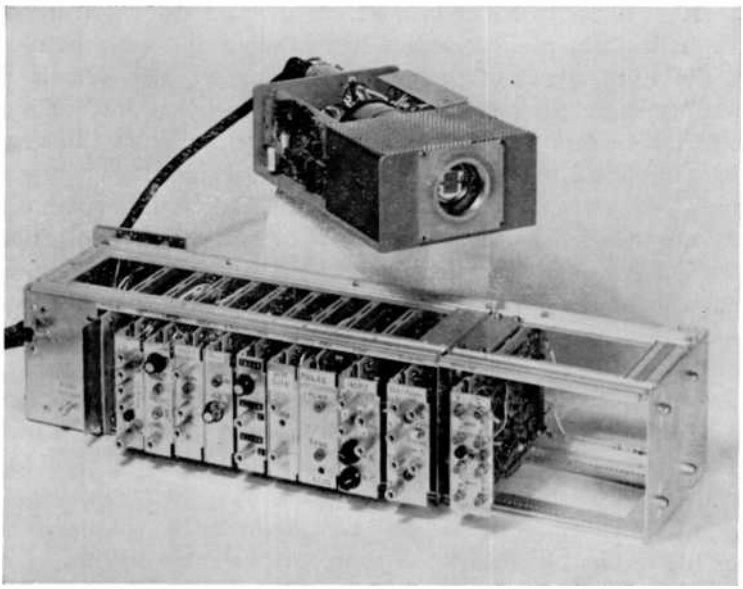


Fig. 6—Bivicon camera and circuit-board rack

extreme, the two channels may be treated exactly alike. In our camera, we have used a processing amplifier in the chrominance channel in order to have blanked, clamped video to aid in monitoring the performance of experimental color-encoding systems. Shading modulators are used in both channels if the flattest possible fields are desired. Aperture correction and sync addition functions are provided in the luminance channel, and line drivers providing 0.7 volt video above black level complete the video chains. The total power input is under 50 watts.

4. Cameras for Pre-Encoded Imaging

The Bivicon tube was developed as part of the Holotape project to provide a pickup device of high reliability for the two-image format required in Holotape. In the Holotape system, images are reconstructed from holograms embossed on low-cost, mass-replicated plastic tape. The system optics are designed to reconstruct in space two images of exactly the right size to fit the targets of the Bivicon tube. One image conveys the luminance of the television signal, the other image the chrominance.

Since the holograms are reconstructed with monochromatic light from an HeNe laser, color information is encoded spatially on the chrominance frame. Therefore, there is no need for an encoding structure in the optics or in the Bivicon tube. The holographic images are made from a silver halide film bearing an image representation of the video signal of the program material. This film is recorded in a commercial electron-beam recorder designed for making video-to-film transfers. By using an electronic recording process rather than an optical process, we can process the video signal in ways that would otherwise be difficult. This flexibility is apparent in the appearance of the encoded film. Fig. 7 shows a typical frame pair. The luminance image is easily recognized as the monochrome representation of the television signal. The form of the chrominance image is less obvious. It consists of a 2-MHz phase- and amplitude-modulated carrier and an unmodulated 1-MHz phase reference.[2] The 2-MHz carrier modulation has the same form as the NTSC chrominance carrier to facilitate decoding in NTSC receivers. A unique characteristic of electronic encoding is that it permits the elimination of all luminance components from the chrominance image. It is then possible to increase the dynamic range of the chrominance, with no allowance required for the residual luminance. A simple, trouble-free method of decoding this image is used in the Holotape player. The video output derived

from the chrominance frame is immediately band-passed at 1 MHz and 2 MHz separately. These two signals, reference and chrominance, are combined with an NTSC 3.579545-MHz reference in a two-mixer system to "transcode" directly from the Holotape code to the NTSC code.

The Holotape player consists of three major components—laser, tape transport, and Bivicon camera. Fig. 8 shows the prototype Holotape cassette player. The laser and Bivicon camera head are fitted in within the constraints of the optical and mechanical design of the transport.

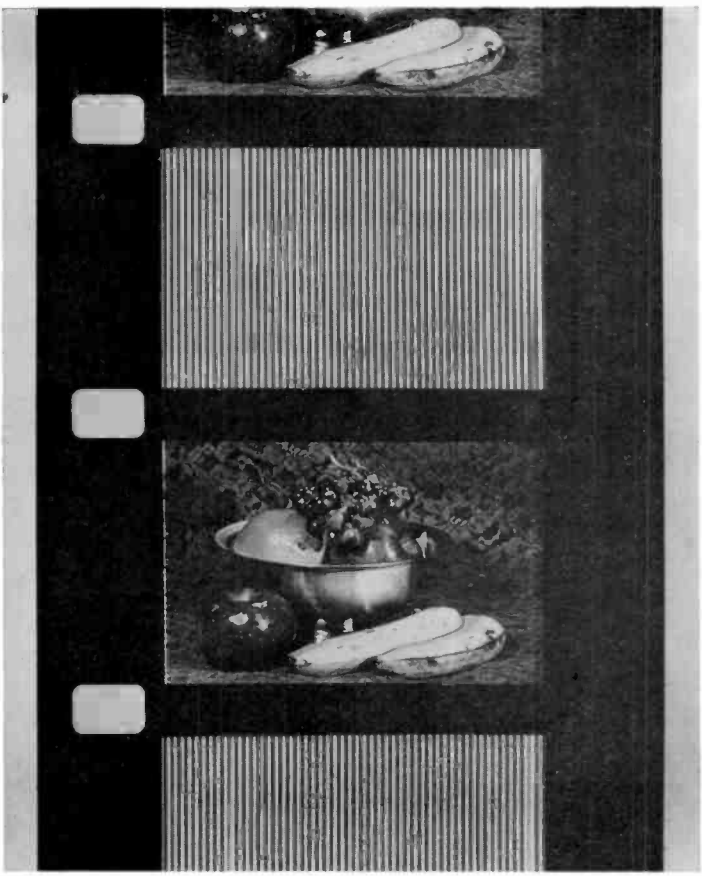
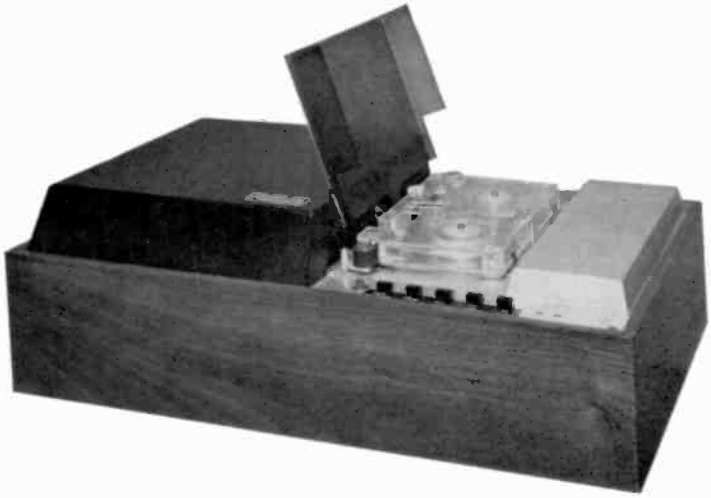


Fig. 7—Typical frame pair on monochrome silver halide film used to convey luminence and chrominance information to the Bivicon target. (The chrominance information is on the encoded frames.)



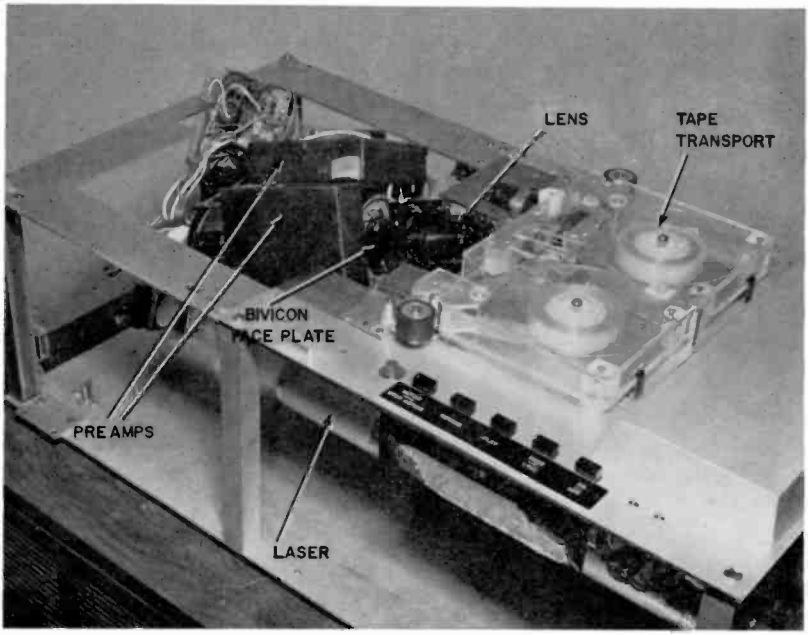


Fig. 8—Prototype Holotape cassette player (cabinet removed in bottom photograph).

In the process of making and testing the encoded silver halide film used to make the holographic tapes, it was found convenient to construct a Bivicon film reader that could decode and display the encoded film in motion. Our breadboard player is shown in Fig. 9. Continuous

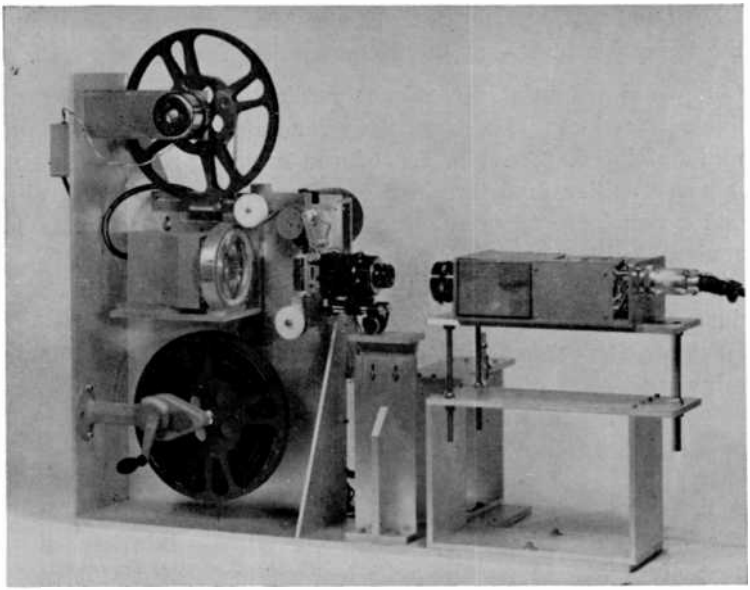


Fig. 9—Laboratory breadboard model of Bivicon film reader.

film motion was used for simplicity. A frame rate of 30 frame/sec was used. A xenon strobe lamp was used for illumination. The flash duration was about 10 μsec, short enough to immobilize the image. Since the film was recorded in an intermittent pull-down film transport, the images were registered with the perforation in the standard 16-mm film stock. For this reason, the flash lamp was triggered by a photodetector sensing the perforations. Excellent image stability is obtained, with no wear on perforations during playback. In fact, the film could be driven by friction, and image framing achieved by a sync pulse image without the presence of an actual hole.

While using this film reader as a tool in the production of Holotapes, the following performance features were noted for the system of encoding television and reproducing it in color from monochrome silver halide film:

(1) Good color quality. The input color video signal was handled as a video signal throughout the process until reproduced on the color

kinescope. No dyes or color-detection processes are involved.

(2) Good luminance resolution and signal/noise. The extremely fine-grain film has resolution well beyond the capability of the television system, unlike color print stock which, in 16 mm, has some influence on the system performance.

(3) Good motion rendition. Use of 30 frame/sec recording matches the television input–output better than 24 frame/sec motion pictures.

(4) Good potential for volume replication and an inexpensive, reliable reproducer. The technology for monochrome film duplication is well developed. The film reproducer has no precision moving parts, image stability is determined by the electronics of the photosensor and xenon lamp trigger. Simple pickup of the two-frame format is achieved with the Bivicon tube.

It is well recognized that a new format would be difficult to introduce, whatever performance advantages might accrue from its use. However, the performance, reliability, and economics of such a system might suit some of the low- to medium-volume market segments.

5. Color-Encoding Cameras

A more generally useful form of Bivicon camera is one that encodes a color input image. The two major subclassifications are cameras that are designed for use with color film input and those that are designed for direct pickup from any live scene. Generally, these differ only in the form of the optical system used to form the two images on the two targets of the Bivicon tube. It is in the area of optical systems that Bivicon camera design departs from other color-camera systems.

Previous multi-image camera designs have used separate camera tubes for each image. As a result, the location of these image planes could be chosen flexibly to fit the requirements of the desired optical designs. In the Bivicon tube, the image locations are inflexibly defined. Thus the price paid for the registration advantage of the Bivicon tube is that the two optical images must be located in one plane and in close proximity. Several solutions are described in succeeding sections.

5.1 Color Encoding for Two-Image Formats

In describing Bivicon color cameras, little will be said specifically about the color-encoding means used except for some discussion here of the characteristics of encoders and resulting cameras that are

unique to two-image systems. Perhaps the single most desirable characteristic of two-image formats is the ability to achieve luminance quality unlimited by the presence of spatial encoding structures. In single-image systems other than field-sequential, the luminance resolution achieved is always less than the capability of the camera tube, as some of the spatial spectrum must be allocated to the rendition of chrominance. To achieve reliable, low-noise chrominance, the location of color-encoding carriers in single-image cameras is often low enough to intrude on the desired luminance band.

In two-image cameras it is possible to use one image entirely for luminance and the other for chrominance. It is then possible to have luminance limited only by the camera-tube response and to locate the chrominance encoding structure at spatial frequencies lower than those usually selected in one-image systems. Separating the luminance and chrominance functions has enabled the designer to achieve performance of both components in optimum portions of the camera-tube spatial spectrum, while eliminating the interactions.

Color-encoding systems that best achieve the desired function of encoding all chrominance are those that encode with structures containing all three color primaries. One such system used in a two-tube camera has a color stripe array of red, green, blue, and black stripes. With bias light superimposed on the scene image, this encoder produces a black reference pulse irrespective of scene content. This reference pulse is used to generate sampling pulses that allow separate decoding of red, green, and blue video outputs at bandwidths adequate for chrominance. Note that the chrominance camera tube must resolve each color stripe to avoid hue errors. The second camera tube is then used to produce wide-band luminance.[3]

In our own work, we have favored encoding systems that have been directly decodable as a chrominance carrier, i.e., containing just hue and saturation information. Among the advantages of these systems are inherently stable white balance and direct translation to an NTSC-encoded chrominance signal. Such an encoder consists of groups of red, green, and blue stripes and a reference carrier structure. The color stripes are made with transmissions adjusted to be equal in white light, so that the carrier vanishes in a colorless image. Spatial frequency for the stripe groups should be such that a two to three megahertz carrier results in camera video output in order to achieve reasonable chrominance bandwidth. A spatial reference carrier is needed to allow decoding in the presence of optical and scanning distortions. We have used a reference carrier at half the color-carrier

frequency; the reference carrier is formed in the image by spatially interleaving a neutral structure with the color structure.

The video output from such a coding filter is translated to an NTSC color carrier by means of the transcoder shown in Fig. 10. White balance is independent of focus over the entire image. Any carrier-amplitude variation produces only a color-saturation error. Note also that the camera video output is band-passed at the color group frequency. It is not necessary to resolve each color stripe to achieve

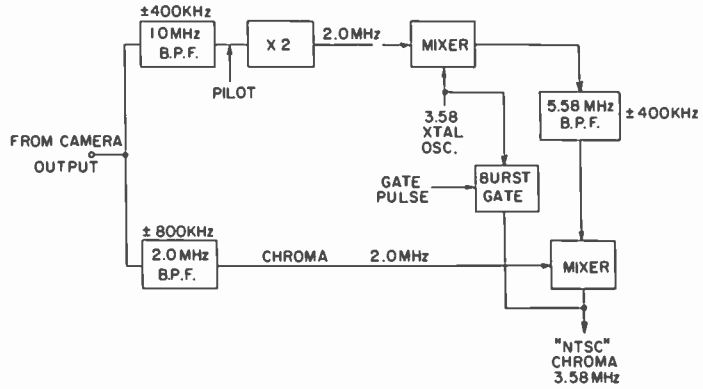


Fig. 10—Block diagram of transcoder to translate to NTSC carrier.

good color performance, and response is only required at all to the extent of the sideband above the color carrier.

There are other encoding systems that can make good use of the Bivicon format with simpler encoding structures. Several of these resemble three-image formats. One such system uses one image to encode green (or luminance) and the other image to encode red and blue on a frame-sequential basis.[4] This system has no spatial encoding artifacts, but suffers from some motion breakup and color flicker. This system is usually improved by using a one-field store to present simultaneous outputs in blue and red. There are then no flicker effects on stationary objects, but break-up on moving objects remains.

Another system uses a simple spatial encoding filter to achieve red and blue output from one tube.[6] The spatial portion of the filter is a simple array of yellow and clear stripes at about a 2-MHz repetition rate. A beam splitter is used to route a green image to one target and magenta to the other target. The yellow stripes in the magenta image encode blue on a carrier, and red is the larger portion of the low-passed content of that frame. Straightforward matrixing will

result in red, green, and blue outputs of chrominance bandwidth, and luminance having detail from the green channel only.

It is possible therefore to select a coding system for stable colorimetry at the expense of encoding filter complexity, or conversely, simple encoders with other performance sacrifices. The two-image format probably offers the most flexibility in choosing a color-encoding system.

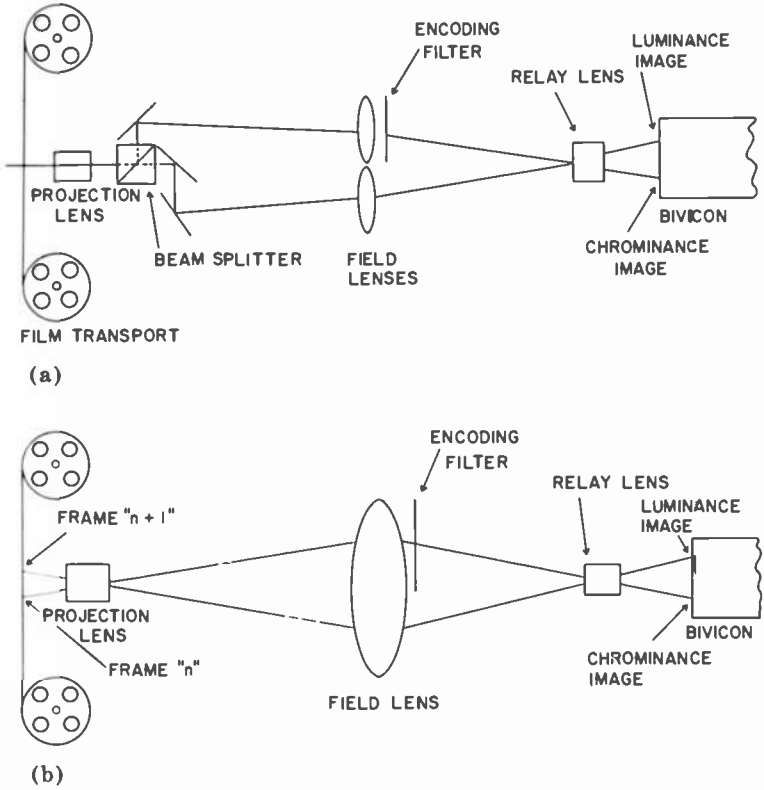


Fig. 11—Optical systems for film camera. System in (b) uses a single optical axis for both targets.

5.2 Film Cameras

The design of cameras made specifically for televising motion picture film or transparencies is facilitated by the possibility of a high light level and the fact that the object televised is of fixed size and location. Two optical systems are shown for film cameras in Fig. 11. Here the use of a large field lens allows the spatial encoding filter to be

physically of such a large size that fabrication is made easier. Configuration (b) is novel in its use of a single optical axis to achieve an image on each target of the Bivicon tube.[5] Use is made here of the fact that adjacent images on a motion picture film are very similar. If there is no motion, they are, in fact, identical. It is possible with this optical system to use one frame for luminance and the adjacent frame for chrominance at the same time. When film is advanced one frame at a time, the result is that the encoding of luminance and chrominance proceed separated in time by one film frame. If the color-encoding system encodes all chrominance in one frame, this camera yields a surprisingly good picture. The defect resulting from the time lag between luminance and chrominance appears subjectively as a "smear" in the direction of the motion. It is most obvious in large-area movement such as camera panning.

Bivicon film cameras appear to offer high performance at low cost and may be of interest in the educational and other closed-circuit markets.

5.3 "Live" Cameras

The design of cameras for direct pickup is complicated by the desire for high light efficiency and the need for a wide range of focal lengths and focus settings for the prime image-forming lens. Perhaps the most serious consequence of these boundary conditions is the stringent requirements placed on the lens used to image the color-encoding structure on the chrominance target of the Bivicon tube. This lens must image the encoding structure faithfully, but must do this at a large enough aperture to achieve the desired camera sensitivity. The problem of faithfully reproducing the coding structure is particularly severe because of the degree of color correction required at a relatively high spatial frequency. This becomes an additionally challenging role for this lens, in the configuration used in the film cameras, shown in Fig. 11, where the same lens is used to relay both luminance and chrominance images.

The configuration of one design of live camera optics is shown in Fig. 12. Here the single relay lens is replaced by two lenses, each working on the axis of its own field lens. Rhomboid prisms are used to separate the two image centers sufficiently to allow reasonable dimensions for relay lenses and the encoding filter. The coding filter dimensions are comparable to a 35-mm camera frame, and the picture-taking lens may be chosen from a variety of lenses designed for that application. A higher-efficiency, lower-cost optical system would re-

sult from the substitution of mirrors for all prisms in this system. However, keeping all of these mirror surfaces in good optical condition would be troublesome. Prisms were used in the experimental systems for this reason and because of the ease of supporting them in experimental setups.

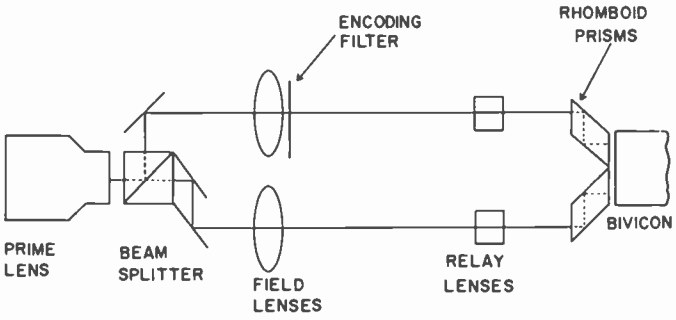


Fig. 12—Optical system for "live" camera.

The highest-quality and lowest-cost optics in a spatial encoding system for a color television camera is achieved when the spatial encoding filter is in effective contact with the camera-tube photoconductor. In this case, the encoding structure is very effectively superimposed on the input image with no need for relay lenses. Light losses are lower, and performance is generally superior. It is possible to use low-cost lenses designed for small-image formats provided their image plane is located far enough behind the lens to allow for the beam-splitting function. Lenses designed for 8-mm reflex viewfinder cameras are adaptable for this application.

Although the optical design is far simpler, the fabrication of the Bivicon camera tube has, of course, been made more complicated when integrating in the spatial encoding function. The economics in the choice of approach for a given camera design must be carefully weighed.

6. Other Applications for the Bivicon Tube

There are other potential applications for a dual raster pick-up device. None have been tested, but some are described and these will undoubtedly suggest others.

Stereo television is easily possible with the Bivicon camera. In

certain closed-circuit applications, this feature is desirable. The Bivicon offers a simple camera with good image match in spite of scan linearity or size errors. Rhomboid prisms or mirrors can be used to achieve the desired inter-lens spacing. Images can be displayed by two monitors optically superimposed, using linear polarizers to route images to the proper eye. An alternative display, which offers the same size and linearity match advantange as the Bivicon, would be the use of a shadow-mask kinescope as a two-color display, using one color phosphor for one image and the other two for the remaining image. Colored glasses would allow the observer to view the stereo image.

Pattern-matching or measurement may be facilitated by use of the Bivicon camera. A standard pattern, or measuring scale, in one raster can be compared with the image on the other raster. Such measurements would be reliable in spite of deflection-current nonlinearities or size changes, as both rasters would experience the same changes.

Panoramic television may be useful in certain applications. It may be useful to scan two adjacent segments of a scene with a Bivicon camera and to display on a monitor with the two images "spliced" together by some optical means.

It has been suggested that an inaccessible, remotely controlled camera would gain in reliability by use of a Bivicon tube. One lens covering both targets would center the object of interest on one target. In the event of target failure or video amplifier failure, the second target and video chain could be used.

7. Conclusions

The Bivicon camera tube can be used for virtually any application previously requiring two vidicons. Because of the ease of achieving a two-image camera system with this device, the two-image color formats become more attractive.

The success of Bivicon color cameras depends upon close cooperation between the designers of the optical system, the spatial encoder, and the rest of the camera. A variety of optical configurations and color encoding systems can fit a variety of requirements.

Acknowledgments

The invaluable role played by RCA Electronic Components is obvious to the reader of the companion paper on the Bivicon tube contained in this issue of *RCA Review*. W. J. Hannan shared in the conception

of the Bivicon tube and provided valuable leadership and encouragement in our work at RCA Laboratories and at the Electronic Components Division.

The camera was designed by W. G. Gibson with W. E. Barnette and executed with the skilled assistance of J. M. Ferguson and P. Williams. Encoding filters were fabricated in Ektachrome film by R. Watson.

References:

[1] RCA Vidicon Data Manual CAM-700, RCA Electronic Components, 415 So. Fifth St., Harrison, N. J. 07029.

[2] P. Goldmark et al, "Color Electronic Video Recording," **J. SMPTE**, Vol. 79, p. 677, Aug. 1970.

[3] Sony Corp., Service Manual, Color Camera DXC-5000, 1970.

[4] M. Berry and J. Poole, "A Portable Color TV Camera System," **J. SMPTE**, Vol. 78, p. 938, Nov. 1969.

[5] R. Flory, "A Simple Color Telecine Camera," **J. SMPTE**, Vol. 82, 1973 (in press).

[6] Y. Takemura, I. Sato, and H. Tajiri, "Compact Color Television Camera," **J. SMPTE**, Vol. 82, p. 12, Jan. 1973.

Information Processing With Transferred-Electron Devices

Fred Sterzer

RCA Laboratories, Princeton, N. J.

Abstract—The principles of operation of information processing circuits using transferred electron devices are reviewed, and several new circuits are proposed. These new circuits are expected to prove useful in processing information at multi-gigabit rates.

Introduction

The transferred-electron effect is one of the most interesting and useful new effects in semiconductors. This effect, which was theoretically predicted in the early 1960's[1,2] and experimentally demonstrated in 1963,[3] occurs in several III-V semiconductors such as GaAs and InP. The effect is a bulk negative differential mobility that is due to the transfer of electrons from high- to low-mobility energy bands. The time constants involved in the transfer of electrons between these energy bands are of the order of picoseconds, and the negative differential mobility is therefore retained at frequencies well into the millimeter-wave range.

Transferred-electron devices (TED's), i.e., devices whose operation depends on the transferred-electron effect, can be used to amplify and generate rf power, and to process analog and digital information. TED's are now widely used in commercially produced microwave amplifiers and oscillators,[4] but TED's for information processing are still in the early laboratory stage. However, the potential of TED's for ultra-high-speed information processing is very great.[5-8] For example, dynamic TED frequency-dividing circuits have been operated with input frequencies as high as 16 GHz,[9] and an integrated digital four-stage TED circuit has been operated with a delay per stage of only 40 ps.[10] The present paper briefly reviews the basic principles of operation of TED's for information processing (these principles being fundamentally different from those governing the operation of conventional silicon IC's) and proposes several new information-processing TED circuits such as variable delay lines, phase coding circuits, and memory cells. These new circuits are expected to prove useful in processing information at multi-gigabit rates.

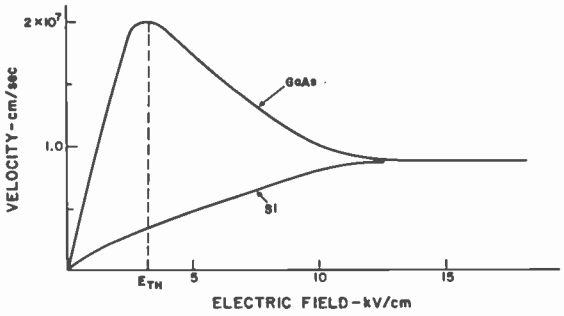


Fig. 1—Curves of electron drift velocity versus electric field for silicon and gallium arsenide.

Principles of Operation of TED's Designed for Information Processing

Fig. 1 shows graphs of electron drift velocity versus electric field for silicon and GaAs. The drift velocity for silicon exhibits a "normal" behavior in that it increases monotonically with increasing electric field. In GaAs, however, the drift velocity decreases with increasing field above about 3 kV/cm, i.e., for electric fields above about 3 kV/cm, the differential mobility (dv/dE) is negative in GaAs.

Transferred-electron materials biased into their negative differential mobility region exhibit a behavior strikingly different from that of materials with positive differential mobilities.[4] In the latter materials, charges of like polarity repel one another, and any accumulation of space charge decays exponentially with a time constant equal to the dielectric relaxation time

$$\tau_r = \frac{\epsilon}{en\mu},$$

where ϵ is dielectric constant, e is electronic charge, n is carrier concentration, and μ is differential mobility. In transferred-electron materials biased into their negative differential mobility region, the situation is the opposite. Charges of like polarity attract one another and any accumulation of space charge grows with time at the rate $\exp(t/|\tau_r|)$.

Consider now a bar of transferred-electron material of length L with ohmic cathode and anode contacts. The anode is biased positive with respect to the cathode at a magnitude larger than the threshold voltage $V_{TH} = E_{TH} \times L$ (see Fig. 1 for the definition of E_{TH}). Therefore, part of the device is biased into the negative-mobility region. Excess electron space charge introduced at the cathode moves through the device under the influence of the applied field. The charge grows exponentially in the negative-mobility region at the rate

$$\exp\left(\frac{t}{|\tau_r|}\right) = \exp\left(\frac{len|\bar{\mu}|}{\bar{v}_\epsilon}\right),$$

where l is the length of the negative mobility region, v is the velocity of the electrons, and a bar over the symbol indicates an average value. If the exponential growth rate is sufficiently large ($nl \geq 7.6 \times 10^{11}/\text{cm}^2$ for GaAs), then high-field space-charge domains will form in the transferred-electron material. A typical domain nucleates at the cathode of the TED. It grows exponentially while moving toward the anode until the voltage across it is so large that the field in the parts of the TED outside the domain falls below threshold. The fully formed or "mature" domain disappears when it is collected at the anode. As the domain is being collected, the electric field throughout the TED rises above threshold, a new domain is nucleated at the cathode, and the cycle repeats itself.

Mature domains moving through a TED have the following important properties:

1. Domains move with nearly constant velocity. In GaAs, this velocity is about 10^7 cm/sec, or about 1/3000 the velocity of electromagnetic radiation in free space.
2. While the domain is moving through the device, the device current is lower than when no domain is present.[11] Thus each domain produces a negative-going current pulse as it moves through the TED.
3. A mature domain moving through a uniformly doped TED with gradually varying nonuniform cross section is equivalent to a wall moving with constant velocity through which constant density of current flows.[11] Therefore the current through a TED containing a moving mature domain is at every instant approximately proportional to the cross-sectional area of the TED at the location of the domain at that instant.

In most signal-processing applications of TEDs, one must be able to precisely control the launching of domains. Control can be accomplished with the three-terminal TED shown in Fig. 2.[12,13] The

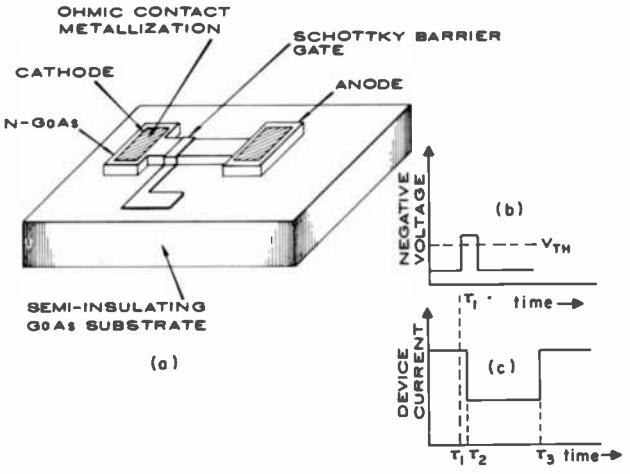


Fig. 2—Three-terminal planar TED: (a) device structure, (b) gate voltage, and (c) resulting device current. When the domain reaches the anode, a narrow current spike is usually generated. Such spikes do not in general affect the operation of the logic TEDs described in this paper. To simplify the presentation of basic concepts, these spikes are not shown in this or in subsequent figures.

cathode–anode voltage across the TED is adjusted to a value slightly below threshold. As a result, no domains are ordinarily formed in the device. If, however, the Schottky gate is biased sufficiently negative with respect to the cathode, a depletion layer will form underneath the gate electrode, the electric field lines in the vicinity of the cathode are compressed, and the electric field at the cathode becomes high enough to nucleate domains. Thus, a narrow negative voltage pulse of sufficient height applied to the gate (Fig. 2b) will start a domain moving across the device (Fig. 2c). The time difference between the application of the gate pulse and the drop in device current due to the formation of the domain, $\tau_2 - \tau_1$ (Fig. 2), can be as short as 30 ps.[10]

The basic circuit for using the three-terminal TED of Fig. 2 is shown schematically in Fig. 3. This circuit has the following four major characteristics:

(1) Whenever the input gate pulse exceeds a certain threshold value (Fig. 2c), an output pulse is generated. The shape of the output pulse is independent of the shape of the input pulse; the shape of the output pulse is determined solely by the geometry and doping of the TED.

(2) The TED can support only one domain at a time. This means that if an input pulse is applied when a domain is moving through the device, this input pulse does not create another domain. In other words, the TED does not respond to an input signal as long as a domain is traveling through it.

(3) The TED circuit has gain, i.e., the output pulse can be significantly larger than the input pulse. Fan-outs of 10:1 appear feasible.[10]*

(4) The Schottky gate is highly effective in blocking the propagation of signals traveling in the reverse direction, i.e., from output to input. Reverse attenuations of 15 dB have been demonstrated.[10]

The above characteristics together with the inherent subnanosecond speed of TED's make three-terminal Schottky-gate TED's attractive building blocks for ultra-high-speed digital computing applications. Since three-terminal TED's are threshold devices, all threshold logic circuits can be implemented in a straightforward fashion.[16,17] Memory elements can be constructed by placing a capacitor of appropriate size

* Three-terminal Schottky gate TEDs possess all of the properties of a neuristor:[14,15] (1) attenuationless propagation of signals, (2) uniform velocity of propagation, (3) a refractory period, i.e., a neuristor has a 'dead time' after a signal propagates through it, and (4) a threshold of stimulation.

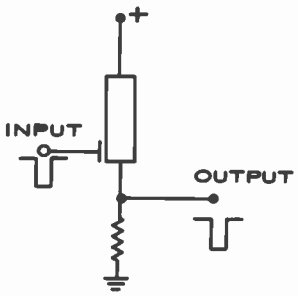


Fig. 3—Basic signal-processing circuit using three-terminal TED.

across the load resistance of the TED.[17,18] In such a memory element, once the TED is triggered by a negative pulse to its gate, output pulses are generated until a positive-going erase pulse is applied to the gate.

New TED Information-Processing Circuits

The "Notched" Three-Terminal TED

Most of the circuits described in this section use the notched three-terminal TED shown in Fig. 4. This TED is similar to the TED of Fig. 2 except for a constriction, or notch, located between the gate and the anode. A domain moving through this device is constricted as it passes the notch, and there is an additional drop in device current as shown in Fig. 4b. The time delay between the nucleation of the domain and the start of the additional current drop due to the notch is given by

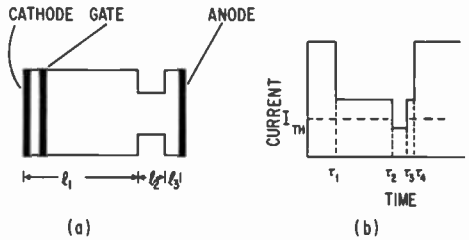


Fig. 4—Three-terminal planar TED with notch near anode: (a) device geometry and (b) device current as a function of time. The domain is nucleated at time τ_1, reaches the notch at time τ_2, leaves the notch at time τ_3, is absorbed at the anode at time τ_4.

$$(\tau_2 - \tau_1) = \frac{l_1}{v_D},$$

where v_D is the domain velocity.

Fixed and Tapped Delay Lines

Fig. 5 shows a delay line using three notched TED's of the type shown in Fig. 4. The input pulse triggers a domain in the first device. The current through the first device drops; only when the domain reaches the notch in the device, however, does the current drop to a sufficiently low value to trigger a domain in the second device, and so on. If l_1 in Fig. 4a is, for example, 100 μm, then the signal delay per stage will be approximately 1 nanosecond in a GaAs device ($v_D \sim 10^7$cm/sec). Note that pulses moving through the delay line get regenerated in both amplitude and shape at every stage of the line. Therefore, in contrast to passive delay lines such as acoustic lines, pulses move unattenuated through TED delay lines of the type shown in Fig. 5. It is also possible of course, to tap signals off at every element, so that tapped delay lines can be realized in a straightforward manner.

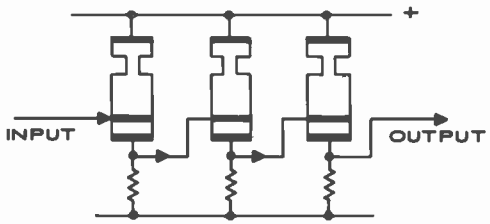


Fig. 5—Three-stage delay line using the three-terminal TED of Fig. 4.

Continuously Variable Delay Lines

Fig. 6 shows a TED circuit for obtaining continuously variable delay. The input pulse is applied to the gate of a TED with tapered cross-section. (To prevent damage to the device when the domain is discharged at the anode, the taper of the device is not continued all the way to the anode.)[19] A domain moving through this device produces a tapered voltage pulse $V(t)$ across the load resistor of the device as

shown in Fig. 6b. This voltage pulse is superimposed on a variable dc bias voltage V_B, and the sum of these two voltages is applied to the gate of the second TED. The larger V_B, the greater the voltage drop in $V(t)$ necessary to trigger a domain in the second TED, and the longer the time delay between input and output pulse. For example, for bias voltage V_{B1}, a domain is triggered in the second TED at time τ_1, while for bias voltage V_{B2} ($V_{B2} > V_{B1}$) a domain is triggered at time τ_2 (Fig. 6b).

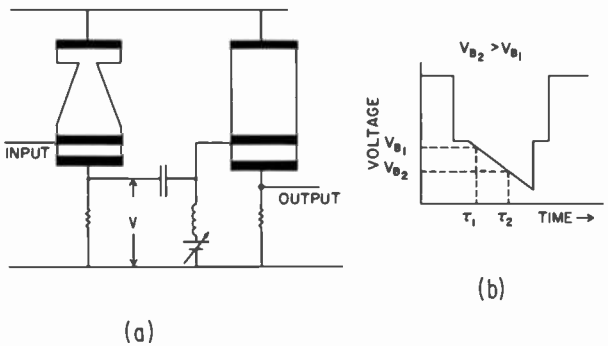


Fig. 6—TED circuit for obtaining continuously variable delay: (a) schematic of circuit and (b) voltage across the load resistor of the first device as a function of time.

Phase Code Modulators

Fig. 7a shows a biphase modulator using two TEDs. The input pulses (Fig. 7c) are applied to the gate of a notched TED. The output of this TED and a control voltage V_c are applied to the gate of a second, shorter TED. With no applied control voltage, the second TED is triggered at time τ_1 (Fig. 7b); when a positive control voltage of appropriate magnitude is applied, the second TED is triggered at time τ_2. Biphase modulation is obtained if the time difference $\tau_2 - \tau_1$ is made equal to one half the period of the input pulse train. In this case, as shown in Fig. 7c, output pulse train 1 ($V_c = 0$) is 180° out of phase with output pulse train 2 ($V_c = V_B$).

Higher-order modulators, such as quadriphase and octiphase modulators, can also be readily synthesized with TEDs. For example, a quadriphase modulator can be constructed by cascading the biphase modulator of Fig. 7 with a similar modulator that introduces differential phase shifts of 90°.

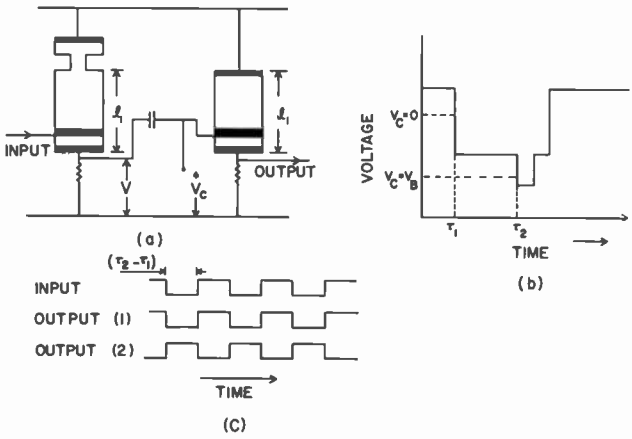


Fig. 7—Biphase modulator using two TEDs: (a) schematic of circuit; (b) voltage across the load resistor of the first device as a function of time, and (c) input and output pulse trains.

Memory Element

Fig. 8 shows the circuit for a TED memory element that uses two TED's rather than a TED and a capacitor as described in Ref. [17]. The first TED contains a notch near its anode and the second a notch near its gate. The memory element is turned "on" by applying a negative set pulse to the gate of the first TED. This set pulse caused

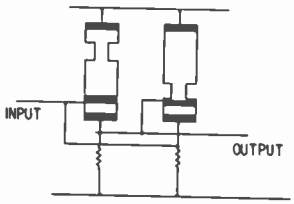


Fig. 8—Schematic diagram of a TED memory element.

a domain to be nucleated in the first TED. When this domain reaches the notch in the TED, the voltage in its load resistor drops to a sufficiently low value to trigger a domain in the second TED. When the domain traveling through the second device reaches its notch, a negative pulse of sufficient amplitude is generated in its load resistor to trigger a second domain in the first device, and so on. Thus a

negative input (set) pulse creates a continuing train of output pulses. To turn the memory element "off," a positive-going reset pulse is applied to the gate of the first TED. This reset pulse prevents the first TED from being triggered by the second TED and the memory element will not generate pulses until a set pulse is again applied to the input.

Comparator

Fig. 9 shows the circuit for a TED comparator.* The input pulses are applied to the gate of a notched input TED and to the gate of the lower of the two series connected TED's. The input TED is triggered if either or both of the negative input pulses are present, while the lower series-connected TED is triggered only if both input pulses are simultaneously present. The gate of the upper series-connected TED

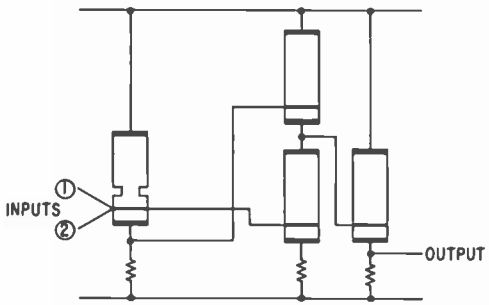


Fig. 9—Schematic diagram of a TED comparator.

is connected to the cathode of the input TED. The upper TED is triggered whenever a domain passes the notch of the input TED, provided there are no domains traversing the lower TED. If the lower TED carries a domain, its resistance becomes so large that the voltage across the upper TED falls below the threshold value required for domain formation. The anode of the lower TED and the cathode of the upper TED are connected to the gate of an output TED. This output TED is triggered whenever its gate is driven negative.

* The truth table for a comparator or "Exclusive OR" is as follows:

Input		Output
(1)	(2)	
0	0	0
0	1	1
1	0	1
1	1	0

If there are no input pulses, no domains are triggered in any of the TED's and no output pulse is generated. If one input pulse is present, the input and the upper TED are triggered, the gate of the output TED is driven negative, and an output pulse is produced. If both input pulses are present, the input TED and the lower TED are triggered, the gate of the output TED is driven positive, and no output pulse is generated.

Conclusion

The negative differential mobility exhibited by several III-V semiconductors such as GaAs can be utilized to create a powerful new technology for processing of information. This technology has several unique and desirable features—speeds consistent with operation at multi-gigabit rates, neuristor-type logic behavior, and the possibility of generating specialized information processing functions by shaping device geometries. To put these technologies into practice will require much additional effort, but expending this effort seems well worth while in view of the potential "order of magnitude" improvements over present technologies.

Acknowledgment

The author thanks S. Y. Narayan and L. C. Upadhyayula for many helpful discussions.

References:

[1] B. K. Ridley and T. B. Watkins, "The Possibility of Negative Resistance Effects in Semiconductors," **Proc. Phys. Soc.**, Vol. 78, p. 293, Aug. 1961

[2] C. Hilsum, "Transferred Electron Amplifiers and Oscillators," **Proc. IRE,** Vol. 50, p. 185, Feb. 1962.

[3] J. B. Gunn, "Microwave Oscillations of Current in III-V Semiconductors," **Solid State Commun.**, Vol. 1, p. 88, Sept. 1963.

[4] F. Sterzer, "Transferred Electron (Gunn) Amplifiers and Oscillators for Microwave Applications," **Proc. IEEE,** Vol. 59, p. 1155, Aug. 1971.

[5] J. A. Copeland, T. Hayashi, and M. Uehohara, "Logic and Memory Elements Using Two-Valley Semiconductors," **Proc. IEEE,** Vol. 55, p. 584, 1967.

[6] R. S. Engelbrecht, "Bulk Effect Devices for Future Transmission Systems," **Bell Lab. Rec.**, p. 192, 1967.

[7] H. L. Hartnagel, "Theory of Gunn-Effect Logic," **Solid-State Electronics,** Vol. 12, p. 19, 1969.

[8] F. Sterzer, "Power-Delay Product of Digital Transferred-Electron Devices," **Electronic Letters,** Vol. 7, Nov. 23, 1971.

[9] L. C. Upadhyayula and S. Y. Narayan, "Microwave Frequency Division Using Transferred Electron Devices," 1973 Workshop in Compound Semiconductors for Microwave Devices, Feb. 12-13, 1973.

[10] K. Mause et al, "Circuit Integration with Gate-Controlled Gunn Devices," 4th International Symp. on GaAs and Related Compounds, Boulder, Colo., Sept. 1972.

[11] M. Shoji, "Functional Bulk Semiconductor Oscillators," **IEEE Trans. Electron Devices,** Vol. ED-14, p. 535, Sept. 1967.

[12] T. Sugeta et al, "Schottky-Gate Bulk Effect Digital Devices," **Proc. IEEE,** Vol. 59, p. 1629, Nov. 1971.

[13] M. Takeuchi, et al, "GaAs Bulk-Effect Pulse Regenerator with a Schottky Barrier Control Gate," **Proc. IEEE,** Vol. 60, p. 740, June 1972.

[14] H. D. Crane, "The Neuristor," **IRE Trans. Electron. Comput.,** Vol. EC-9, p. 370, Sept. 1960.

[15] T Sugeta, et al, "Bulk Neuristor Using the Gunn Effect," **Proc. IEEE** (Lett.), Vol. 56, p. 239, Feb. 1968.

[16] D. Hampel and R. O. Winder, "Threshold Logic," **IEEE Spectrum,** p. 32-39, May 1971.

[17] T. Sugeta and H. Yanai, "Logic and Memory Applications of the Schottky-Gate Gunn-Effect Digital Device," **Proc. IEEE,** Vol. 60, p. 238, Feb. 1972.

[18] H. Imada, et al, "Computer Simulation of a Memory Device Using a Gunn Effect Diode," **Proc. IEEE,** Vol. 60, p. 761, June 1972.

[19] M. Takeuchi, et al, "GaAs Planar Gunn Diodes for DC-Biased Operation," **IEEE Trans. Electron Devices,** Vol. ED-19, p. 125, Jan. 1972.

Two-Phase Charge-Coupled Devices with Overlapping Polysilicon and Aluminum Gates*

W. F. Kosonocky and J. E. Carnes

RCA Laboratories, Princeton, N. J.

Abstract—Two-phase charge-coupled shift registers were developed in the form of polysilicon gates overlapped by aluminum gates. The devices were made on 1.0 ohm-cm n-type substrates with (111) and (100) orientations. Charge-transfer loss of less than 10^{-4} per transfer was obtained in the fat-zero mode of operation of 64 and 128-stage shift registers. The same registers operating in the bias-charge or bucket-brigade mode had transfer loss in excess of 10^{-3} per transfer. Experimental results are also presented on the transfer loss as the function of fat zero level for registers with 5-, 1- and 0.5-mil wide channels.

1. Introduction

Charge-coupled semiconductor devices[1] consist of closely-spaced MOS capacitors that are pulsed into deep depletion by the clock phase voltages. For times much shorter than that required to form an inversion layer of minority carriers by thermal generation, potential wells are formed at the silicon surface. These potential wells can be

* This work was partially supported by NASA Contracts 10983 and 11861, Langley Research Center, Hampton, Va.

used to store and transport minority carriers representing information. The signal charge can be introduced either electrically or optically. The propagation of the information is accomplished by clock pulses applied to the electrodes of the successive MOS capacitors (i.e., charge-coupled elements), resulting in a motion or spilling of charges from shallow potential wells to deeper potential wells. Such propagation of signal into the successive minima of the surface potential produces a shift register for analog signals that has a signal transfer efficiency approaching unity. Such analog shift registers may be used for various signal processing applications such as electronically-variable delay lines or self-scanning photosensor arrays. By adding simple charge-refreshing stages,[2,3] a charge-coupled shift register for digital signals can be constructed.

If the charge-coupled structures are formed with symmetrical potential wells, at least three clock phases are required to determine the directionality of the signal flow. The use of more than three phases may be dictated by either the construction design symmetry, as in the case of a four-phase silicon gate overlapped by aluminum structures, or special signal coding schemes in which more than one bit may be propogated in one clock cycle.

Two-phase operation requires that the charge-coupled structures be formed so that the potential wells induced by the phase voltage pulses are deeper in the direction of the signal flow. In this case, as one phase voltage is lowered, the resulting potential barriers force a unidirectional signal flow. It should be noted that a one-clock operation can also be obtained in a two-phase charge-coupled shift register if a proper dc voltage is applied to one of the phases. Furthermore, a true single-phase or a uniphase charge-coupled structure[4] can be formed by replacing the dc-biased phase by a structure involving a fixed charge in the oxide. Such structures could be formed by the use of ion-implanted barriers[5] or by variation of the fixed oxide charge for the formation of the asymmetrical potential wells. The first charge-coupled devices were made by the thick-oxide p-MOS process.[2,3,6,7] The major limitation of this process is the etching of the separation between the gates, which should be no larger than about 0.1 mil in order to control the surface potential in the resulting gap. On the other hand, the operation of n-MOS CCD's that are made with diffused channel stops can be less sensitive in the interelectrode spacings because of the positive charge usually present in thermally grown SiO_2 channel oxides.

The sealed-channel polysilicon-aluminum structures described in Sec. 2 of this paper are the most compact structures that can be

fabricated with more or less conventional layout rules. The self-aligning-gate construction of these devices allows fabrication of charge-coupled structures with gate separation comparable to the thickness of the channel oxides, as well as having the channel oxide always covered by one of the metallizations. The important advantage of the silicon-gate process is that it provides a very simple method for the construction of two-phase CCD's. The two-phase CCD's described in this paper employ two thicknesses of channel oxide for the formation of the asymmetrical potential wells needed for the unidirectional flow of signal.

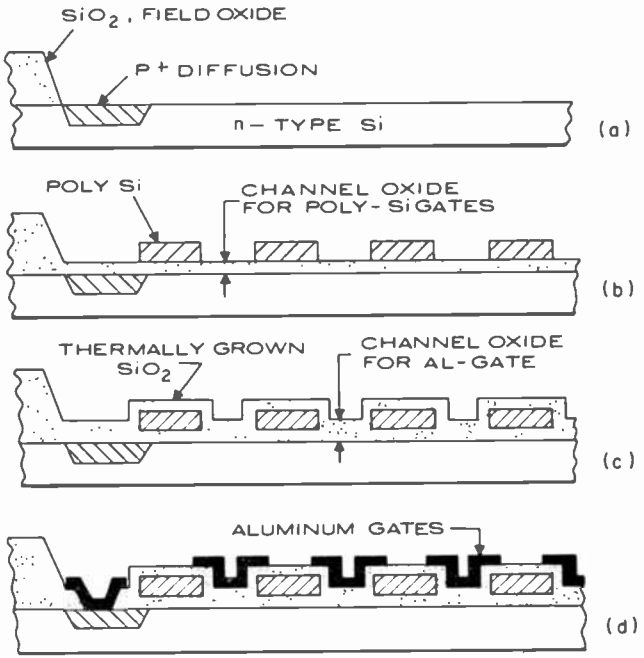


Fig. 1—Construction of 2-phase charge-coupled devices in the form of polysilicon gates overlapped by aluminum gates.

2. Device Construction

The fabrication of the test devices is illustrated in Fig. 1. The substrate used was 1.0 to 0.5 ohm-cm n-type silicon with (111) and (100) orientations. As shown in (a) the p+ diffusion and the field oxide for these devices were prepared following a standard thick-oxide p-MOS process. Boron nitride deposition at 1000°C was used as the

doping source for the p+ diffusions. The field oxide was made as a combination of 7000 Å steam SiO_2 grown at 1100°C followed by 5000 Å deposited SiO_2. The next sequence of the process step, as shown in (b), consisted of thermally growing approximately 1000 Å-thick channel oxide, depositing the polysilicon film, and defining it into the polysilicon gates. Then, as shown in (c), as the polysilicon gates were insulated by a thermally-grown oxide, the channel oxide for the aluminum gates was simultaneously grown to the desired thickness. The initial devices having substrates with (111) orientation were made with an aluminium-gate channel oxide of 3000 Å, in which case the insulating oxide over the polysilicon gates was about 2500Å. However, all of the devices reported here that were made on substrates with (100) orientation were prepared with 2400Å thick channel oxide for the aluminum gates. For these devices the insulating oxide for the polysilicon gates was about 2000Å. Finally, as shown in (d), the device structure was completed by opening contacts to the p+ diffusions and the polysilicon gates, depositing about 10,000Å thick aluminum, and defining it into the aluminum gates.

A photomicrograph of one of the CCD chips used for our experiments is shown in Fig. 2. The data reported here concern the operation of the two middle registers on the chip. They are 128 and 64-stage registers with 1.2-mil-long stages. The construction of these registers is illustrated in detail in Fig. 3. As is shown, the input structure consists of a source diffusion S-1 and input gates G-1 and G-2. Separate electrical access has also been provided to the polysilicon and aluminum electrodes of each phase; i.e., ϕ-1(poly), ϕ-1(Al), ϕ-2 (poly), and ϕ-2(Al). The output can be detected as the current flow out of the drain diffusion D-1 or as a voltage change resulting from the charge signal introduced on the floating diffusion that, in turn, controls the gate voltage of a 3-mil-wide output MOS device with a source S-2 and drain D-2. The electrodes G-3 and G-4 are externally available for controlling the signal flow in and out of the floating diffusion. The center-to-center spacing of 1.2 mil/stage represents the minimum that can be achieved with 0.2 mil spaces between lines and 0.1 mil overlap between the polysilicon and aluminum gates.

The channel confinement, or the channel stops, along the width of the CCD structure was obtained by means of 12,000 Å field oxide (See Appendix 1). Unless otherwise designated, most of the devices studied were 5.5 mil wide. However, to study the effect of the width of the CCD channel on the performance of the registers, special CCD arrays were made having basically the same layout as is shown in Fig. 2 and 3, but with 1.0- and 0.5-mil-wide channels. In this case,

the 128-stage registers were made 0.5 mil wide and the 64-stage registers 1.0 mil wide. Finally, the tested devices were bonded on 28-lead dual-in-line ceramic packages.

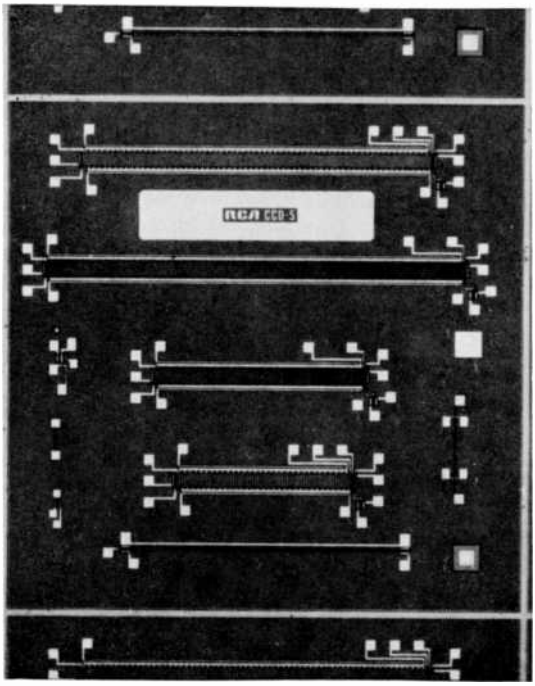


Fig. 2—Photomicrograph of CCD-5.

At the bottom of the photomicrograph in Fig. 2 is a test device consisting of two MOS devices driven by a common polysilicon gate. This test device was used to verify the expected delay time due to the fact that the long polysilicon gate behaves as an *RC* transmission line. The analysis and experimental verification of this delay time is contained in Appendix 3.

3. Operation and Design of Two-Phase CCD's

3.1 Complete Charge Transfer Mode

Assuming essentially zero fixed charge in the channel oxide, substrates with relatively large doping concentrations are required to obtain a substantial difference between the surface potential under the polysilicon gates and the potential under the aluminum gates powered

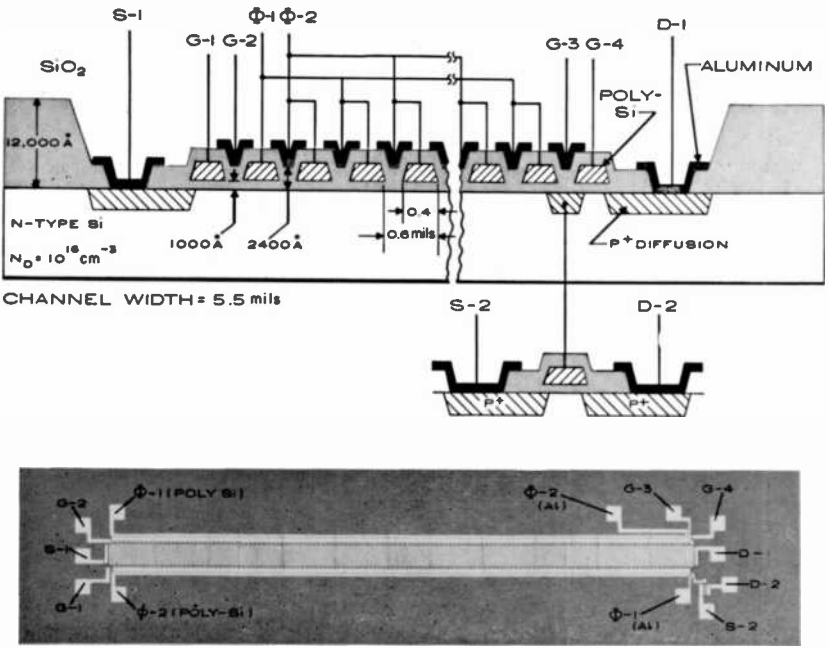


Fig. 3—Cross-sectional view and a labeled photograph of CCD-5 128-stage shift register.

by the same phase voltage. The operation of such a device is illustrated in Fig. 4, showing a computer solution for a two-phase structure with substrate doping concentration of $N_D = 10^{16}\text{cm}^{-3}$ and channel oxide thicknesses of 1000 Å and 3000Å under the polysilicon and aluminum gates, respectively. Since, in this case, the potential barrier formed under the aluminum gate with respect to the surface potential under the polysilicon gate of the same phase is not constant as the phase voltage changes, the maximum amount of charge signal that can be stored and transferred will depend to some degree on the waveshapes of the phase voltages. These waveshapes may be symmetrical with equal rise time and fall time (See Fig. 5(c)), nonoverlapping, or overlapping. In the overlapping case, the transfer of charge is preceded by the surface potential profile shown in the dotted line in Fig. 4. If we can assume that the signal charge will be originally contained in the potential well indicated by the phase-1 gate, then as the phase voltage ϕ_1 is changed from $\phi_1 = -25\text{V}$ to $\phi_1 = -10\text{V}$ and the surface potential is raised from the dotted line to the solid line, the charge signal is transferred to a new potential well on the right side under

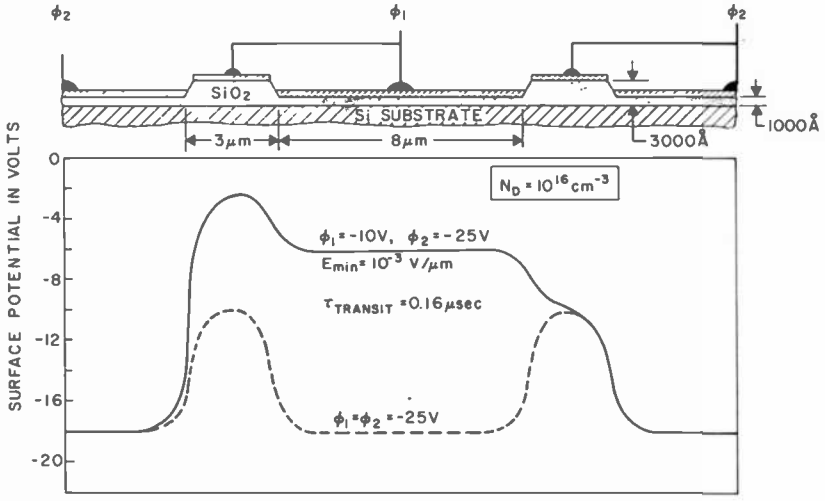


Fig. 4—Computer solutions for the surface potential of two-phase CCD using 1000 and 3000 Å channel oxides.

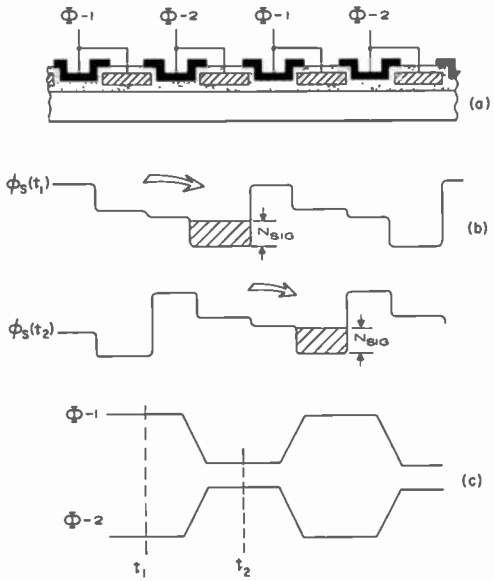


Fig. 5—Complete charge-transfer operation of two-phase CCD.

the phase-2 gate. This mode of operation for two-phase charge-coupled devices will be referred to as the "complete charge-transfer" mode, or the C-C mode.

This conventional mode of CCD operation is again illustrated in Fig. 5 by two profiles of the surface potential. In this case at time t_1, the charge signal is accumulating in the potential wells under the phase-2 gates, and in the second half-cycle the charge signal will be transferred to the potential wells under the phase-1 gates.

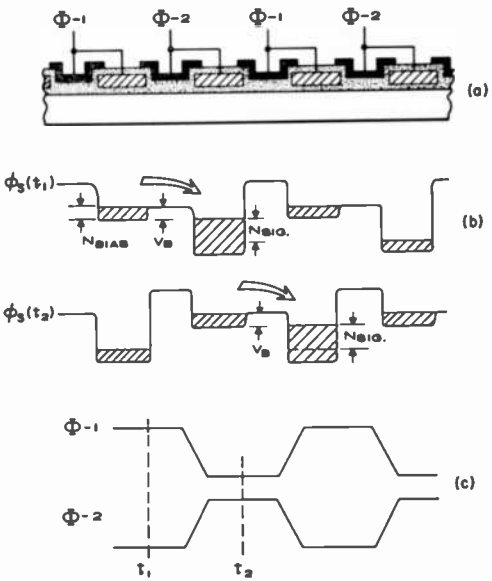


Fig. 6—Bias-charge (bucket-brigade) mode of operation of two-phase CCD.

3.2 Bias-Charge or Bucket-Brigade Mode

Another mode of operation of two-phase charge-coupled devices is illustrated in Fig. 6. In this case a barrier in the surface potential under the aluminum gates is so high that the potential wells under the polysilicon gates can never be completely emptied. In other words, the surface potential under the polysilicon gates powered by the minimum phase voltage is lower than the surface potential under the aluminum gates powered by the maximum phase voltage. This mode of operation will be referred to as the "bias-charge" or "bucket-

brigade" mode. We will refer to this mode as the B-B mode. The bias charge N_{BIAS} illustrated in Fig. 6(b) represents the background charge always present under the gates accumulating or storing the charge signal. In the steady-state operation of such charge-coupled shift registers, the bias charge will be maintained by the thermal generation of carriers.

The transition from the C-C mode to the B-B mode of operation can take place by increasing the dc bias level on the phase voltages or by providing larger dc voltages for the polysilicon gates than for the aluminum gates. The presence of positive charge in the channel oxide in the case of n-channel devices will also increase the barrier under the aluminum gates above the values indicated for the example shown in Fig. 4. The B-B mode of operation of two-phase charge-coupled shift registers is very similar to the operation of bucket-brigade (8-9) shift registers. In the bucket-brigade device, the bias-charge regions are replaced by floating diffusions; otherwise, the operation of these two types of shift registers is very similar.

3.3 One-Clock Operation

As is generally known in connection with the operation of a bucket-brigade shift registers,[21,22] a two-phase charge-coupled shift register can also be operated with one of the phases dc biased at a voltage halfway between the minimum and the maximum voltage applied to the other phase.[10] This "one-phase" operation, or rather one-clock, is illustrated in Fig. 7. A uniphase charge-coupled structure[4] requiring only one set of externally controlled gates requires that the dc-biased phase shown in Fig. 7 be replaced by a built-in bias in the structure which may be obtained by the presence of fixed charge in the channel oxide.

4. Testing Procedures

4.1 Test Setup

The circuit used for testing the experimental two-phase shift registers is shown in Fig. 8. The input signal is introduced into the shift register under the control of dc bias voltages E_S and $E_{G\text{-}S}$ and an input phase V_{in}. The phase voltage clocks ϕ-1 and ϕ-2 are applied with dc bias E_ϕ. Not shown in this figure is the fact that the phase voltages, as well as all of the input and the output control signals were supplied by separate clamping circuits. This permitted us to use different dc

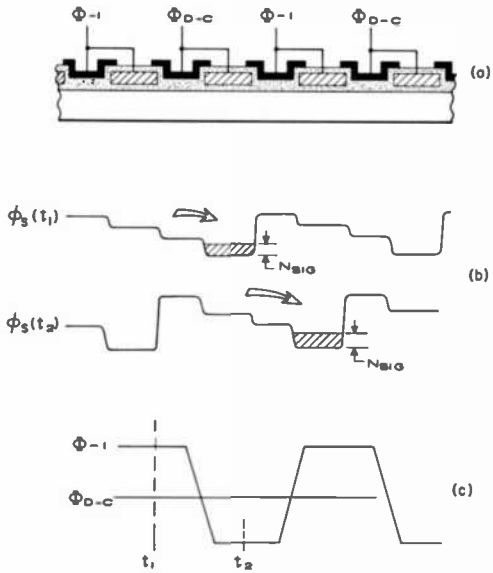


Fig. 7—One clock operation of two-phase CCD.

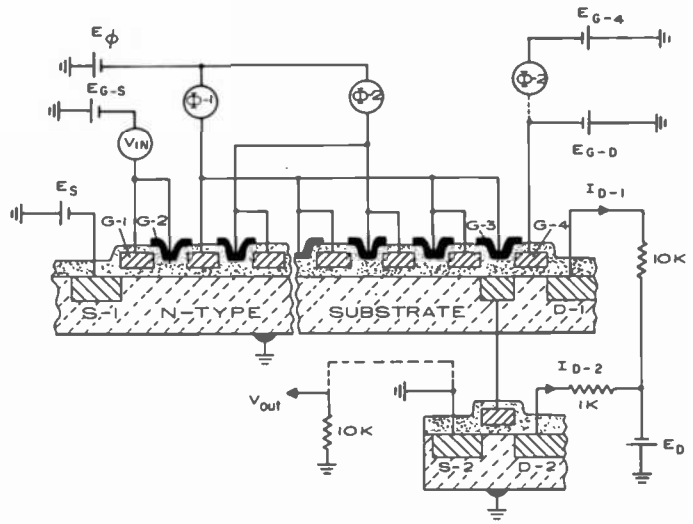


Fig. 8—Circuit diagram for the tests of two-phase charge-coupled shift registers.

levels for the polysilicon and the aluminum gates. However, for the tests in which polysilicon and aliuminum gates were operated with the same dc levels, the phase clocks could be applied directly without the use of clamping circuits, in this case the voltage E_ϕ was applied as a positive substrate bias. In all of the tests the output gate G-3 was connected to the phase voltage ϕ-1. The output gate G-4, in the early tests, was dc biased by $E_{G\text{-}D}$ voltage with a typical value of -15 V while the drain voltage was typically $E_D = -25$ V. In the tests for which most of the data are reported, the output gate G-4 was driven by the ϕ-2 clock with $E_{G\text{-}4}$ as the dc bias. In typical operation the output circuit was biased with $E_{G\text{-}4} = 17$ V and $E_D = 2$ V. Such resetting of the floating diffusion to the drain potential E_D by periodic pulsing of the gate G-4 was found to improve the operation of the narrow channel devices, i.e., the registers with 0.5- and 1.0-wide channels. This was due to the higher loading of the fixed external capacitance on the floating diffusion, which required more positive resetting.

In some of the tests the output signal was sensed directly as current $I_{D\text{-}1}$. But we found it more convenient to measure the output signal by observing the potential of the floating diffusion under the G-3 gate, which is proportional to the charge signal. Such charge, or voltage, sensing was made by measuring the output either as current $I_{D\text{-}2}$ or a source follower voltage V_{out}. A measured transfer curve of the source follower circuit for typical 5-mil-wide devices is shown in Fig. 9. This transfer curve was obtained by biasing the gate G-4 with a -10 V relative to the drain D-1 and measuring the source follower output voltage V_{out} as a function of a d-c voltage applied to the drain D-1. Drain D-2 of the source follower circuit in this test was maintained at -25 V.

4.2 Transfer Loss Measurements

In order to measure fractional loss per transfer, one must know how much charge is lost, how much charge was originally in the signal pulse, and how many transfers occurred. When only electrical inputs are possible, i.e., optical inputs are not available, the number of transfers is fixed by the length of the device, and the determination of the original amount of charge in the signal pulse is not always straightforward. For example, if a signal "1" pulse of amplitude p_o is put in the input and the ϵN_g product (ϵ is the fractional loss per gate and N_g is the number of gates) approaches unity, by the time the pulse reaches the output it will be spread over many different pulses, all

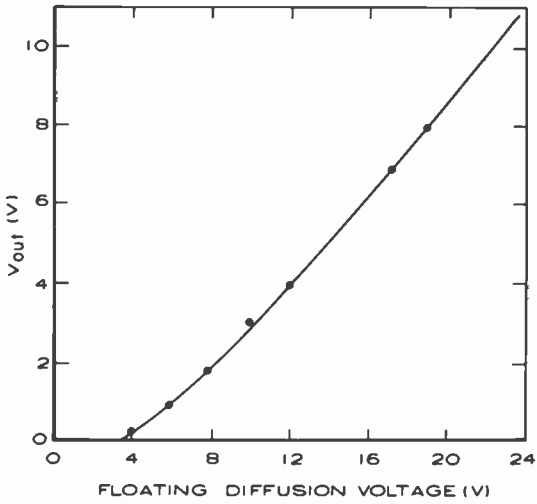


Fig. 9—Measured voltage transfer curve of the source follower output, V_{OUT}, as the function of the voltage of the floating diffusion.

with reduced amplitude, making it very difficult to determine the original signal amplitude. The procedure we have used to circumvent these difficulties consists of applying a string of consecutive "1's" to measure the step or pulse response to the shift register. Even in the face of sizable losses with ϵN_g products exceeding unity, provided the strings of "1's" is long enough, the output signal amplitude will reach the original output value. This occurs because a pulse well removed from the leading one receives the losses from the preceding pulses, which are just enough to match the losses it suffers. Thus, as the number of "1's" in the string increases, the output amplitude increases until it saturates at the amplitude at the input. Then the sum of the trailing pulse amplitudes that follow the last "1" constitutes the total amount of charge lost by that last "1" as it passes through all the stages of the shift register. Since the last "1" is always maintained at full amplitude (by losses from previous pulses), the losses it suffers are easily analyzed in the case where the loss is a fixed fraction of the total charge, i.e., for free-charge loss. This analysis is shown in Appendix 2.

Thus, for free charge losses, provided the string of "1's" is long enough to provide full amplitude, the total loss $\epsilon N_g p_o$ is found by adding all the trailing pulse amplitudes. Alternately, the attenuation of the leading "1's", which will be symmetrical with the trailing edge, can be totalled. Thus,

$$\epsilon = \frac{\Sigma p_i(N_g)}{N_g p_o}$$

(see Appendix 2 for notation). As shown in Appendix 2, since $\Sigma p_i(N_g)$ is proportional to p_o, ϵ does not depend upon p_o and the relative shape of the output pulse stays the same as the pulse amplitude is changed.

On the other hand, the fractional loss into fast interface states ϵ_s is inversely proportional to the signal amplitude.[11] Another feature of fast state losses is that the attenuation of the leading edge of the pulse train is not symmetrical with the trailing pulse amplitude. Since the release times of the fast states vary over many orders of magnitude, the charge lost into fast states is released over a large number of trailing pulses, with the jth trailing pulse having an amplitude of

$$p_j = \frac{kT}{q} N_{ss} N_g \ln\left(\frac{2j+1}{2j-1}\right).$$

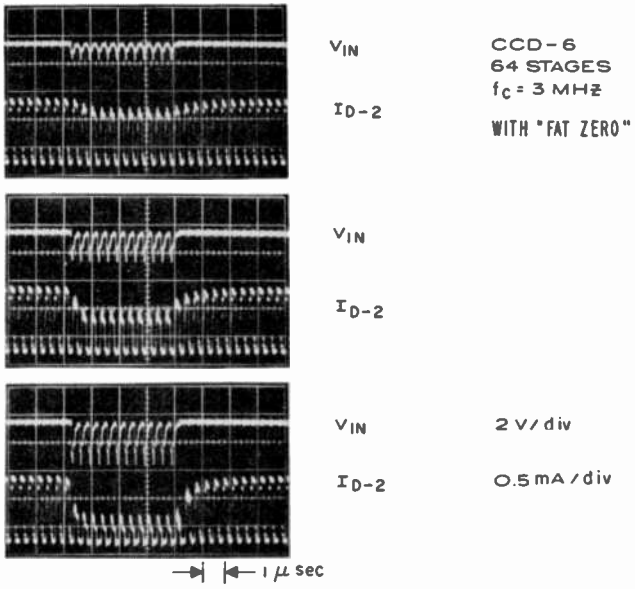


Fig. 10—Oscilloscope photographs of input pulses and I_{D-2} current (traces are offset to align with one another) for 64-stage register at 3 MHz with fat zero. Three different photographs show how the free-charge losses scale with signal amplitude.

However, the leading edge is attenuated by an amount required to fill the states at each gate. Therefore, when measuring fast-state losses, the attenuation of the leading edge is totalled and divided by $p_o N_g$ to find ϵ_s.

Figs. 10 and 11 demonstrate the behavior of free-charge loss and fast-state loss, respectively, as signal amplitude is varied. Note symmetrical appearance of leading and trailing pulses for free-charge loss and the manner in which the loss scales with signal amplitude. Fast-state loss, however, is not symmetrical (i.e., more loss is apparent in leading pulses than for trailing pulses) and has a total fixed loss independent of signal, so that as signal is reduced the fractional loss increases considerably.

5. Experimental Data

5.1 Dielectric Strength of the Oxide Insulating the Polysilicon Gates

The measurements of the breakdown voltage between the polysilicon and aluminum gates showed that the dielectric strength of the oxide grown on the polysilicon was about the same as that of the oxide grown on the silicon substrate. The breakdown voltage between the

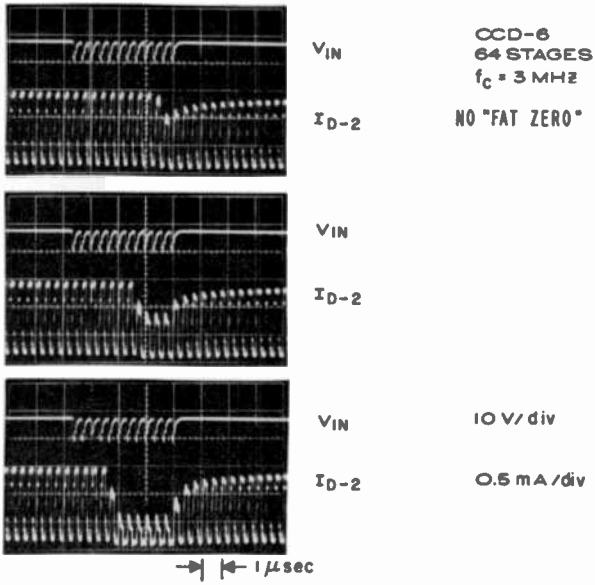


Fig. 11—Oscilloscope photographs of input pulses and I_{D-2} current for 64-stage device at 3 MHz without fat zero. As signal level decreases, more leading edge pulses are required to fill fast states.

aluminum gates and the polysilicon gates insulated by 2000 Å thick oxide was typically 70 V for negative potentials applied to the aluminum gates and 50 V for positive potentials applied to the aluminum gates. The breakdown for the case when the positive potential was applied to the aluminum gates was preceded by a leakage current on the order of 10 to 20 μA. In this case breakdown voltages of 60, 50, and 40 V were observed for insulating oxide thicknesses of 2,500, 2,000, and 1,200 Å, respectively.

5.2 Threshold Voltages

The threshold voltages of the gates could be measured by operating the register as a large single MOS device. However, the measurement of the threshold voltage that will be reported here consisted of the following procedure. To measure the threshold voltage under the polysilicon gates, the gate G-2 was connected to ϕ-1 (See Fig. 3). Then the threshold voltage $V_{th\ (\text{polysilicon})}$ was measured as the dc voltage applied between gate G-1 and the source S-1 that resulted in the introduction of a charge signal into the register on the order of 1% of the full well. During this test, the source S-1 was connected to the substrate. The threshold voltage for the aluminum gates was measured by a similar procedure, by forward biasing the gate G-1 by 10 V with respect to the source S-1.

Typical threshold voltages measured with 1.0 MHz clock frequency for the reported devices were as follows. For the devices made on substrates with (100) orientation, $V_{th\ (\text{polysilicon})}$ was typically between -0.6 and -0.9 V and $V_{th\ (\text{aluminum})}$ between -2.8 and -3.5 V. For the devices made on substrates with (111) orientation, $V_{th\ (\text{polysilicon})}$ was between -1.8 and -2.2 V and $V_{th\ (\text{aluminum})}$ between -7.5 and -9.0 V.

5.3 Thermally-Generated Background Charge

The thermally-generated background charge (dark current) was studied by operating the shift registers with interrupted pulse clocks. This type of operation is similar to the operation of the charge-coupled register as a line sensor but without the optical input. The clock waveforms used in this test are illustrated in Fig. 12. As is shown in this figure, the clock signal is interrupted periodically during the charge integration time T_i while the thermally generated charge is being collected in the potential wells resulting from a constant bias

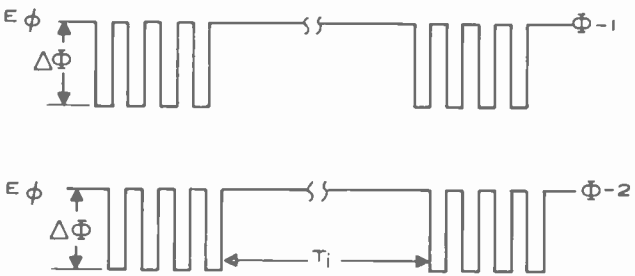


Fig. 12—Illustration of clock waveforms for measurements of thermally generated background charge (dark current).

voltage E_ϕ applied to both phases of the register. The burst of clock pulses that follows each charge integration time reads out the detected charge signal. Typical waveforms of the detected outputs are shown in Fig. 13 for devices made on (100) substrates. The waveform in Fig. 13(a) is for a unit with low and uniform dark current. The waveform in Fig. 13(b) illustrates the thermally-generated charge for a device with large local dark currents.

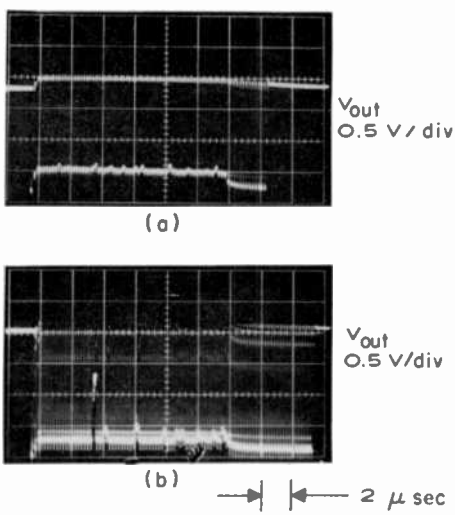


Fig. 13—Typical waveforms of thermally generated background charge for 64-stage registers operating with phase bias voltage $E_\phi = 5$V: (a) output for a unit with uniform dark current for $T_i = 0.1$ sec. and (b) output for a unit with large local dark currents for $T_i = 10$ msec.

The variation of the thermally-generated background charge as a function the phase bias (or rather substrate bias) voltage E_ϕ is shown in Fig. 14. Here we see that the uniform background charge (Curve

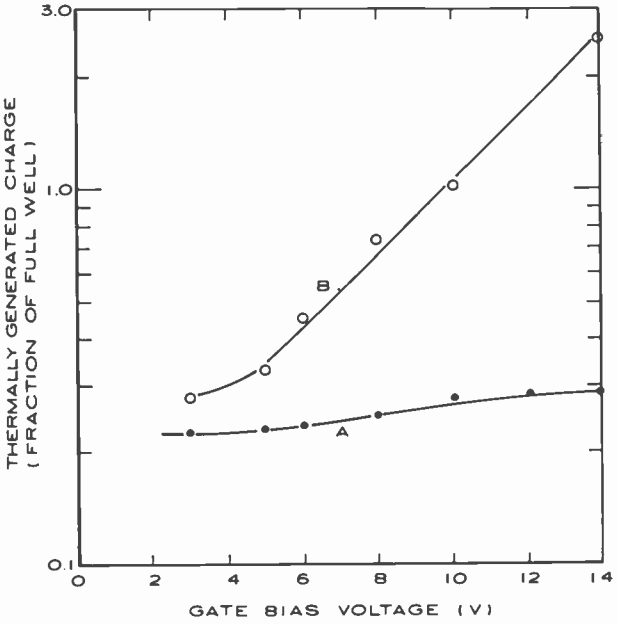


Fig. 14—Variation of the thermally generated charge with the gate bias voltage (E_ϕ) for 64-stage register with large local dark currents; curve A is for uniform background charge and curve B is for a local dark current spike.

A) changes very slowly with the substrate bias voltage applied during the integration time T_i. However, a local high-dark-current spike varies exponentially with the bias voltage, as is illustrated by Curve B. As the bias voltage changes from 5 to 14 volts the locally-generated charge changes from about ⅓ to 3 full wells.

5.4 Transfer Loss Measurements

(a) Typical Waveforms

Figs. 15 through 20 illustrate the waveforms obtained in the measurements of the transfer loss of the 2-phase registers. Figs. 15 through

18 are for the conventional C-C mode of operation. Typical waveforms obtained for the devices operating in the B-B mode are shown in Figs. 19 and 20.

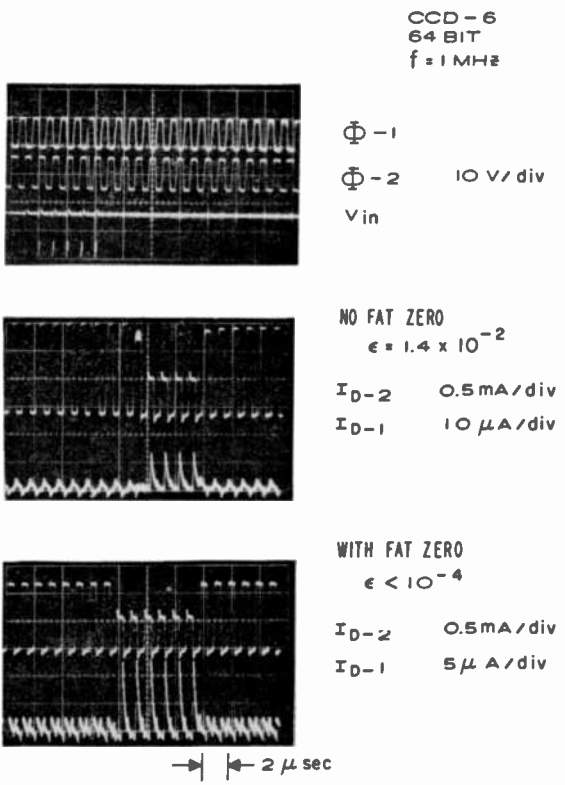


Fig. 15—Typical waveforms for 64-stage register at 1 MHz with (111) substrate orientation.

The waveforms in Fig. 15 illustrate the operation of a 64-stage register, in which case the output gate (See Fig. 8) was biased with a fixed d-c voltage $E_{G\text{-}D}$. The top oscilloscope photograph in Fig. 15 shows the two negative-going clocks along with V_{in} pulses. The negative V_{in} pulses occur when ϕ-1 is "on" or negative. The two detected ouputs $I_{D\text{-}1}$ and $I_{D\text{-}2}$ for the operation without any electrically introduced background charge, i.e., "no fat zero," are shown in the middle photo. In this case the transfer loss, ϵ, due to the charge trapping by the fast interface states is $\epsilon = 1.4 \times 10^{-2}$. However, as is shown in the bottom photo, the introduction of the electrically introduced

charge, or "fat zero", into the register reduces the transfer loss to $\epsilon < 10^{-4}$.

Our techniques for measurements of the transfer loss are further illustrated in Fig. 16 for a 64-stage 1.0-mil-wide register and in Fig. 17 for a 128-stage 0.5-mil-wide register. In these two cases, the output signal is detected as the source follower output, V_{out}, and the d-c voltage of the output gate G-4 was adjusted so that the clock-voltage pick-up is cancelled from the output signal.

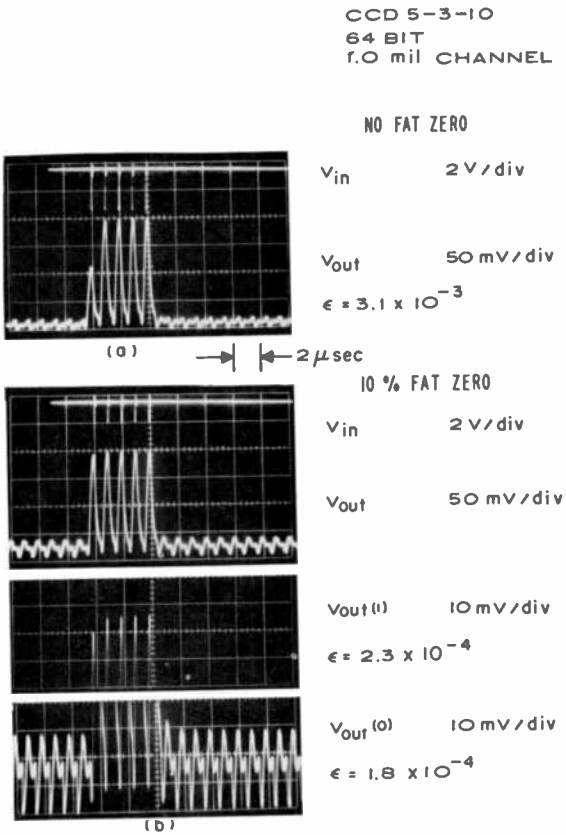


Fig. 16—Waveforms for a 1.0-mil-wide 64-stage register at 1 MHz. (100) substrate.

Inspection of Fig. 16 shows that for "no fat zero" operation, the transfer loss per transfer $\epsilon = 3.1 \times 10^{-3}$ due to the interface trapping is measured as the loss of the first "1" in the string of "1's". For the measurement of transfer loss for operation with 10% "fat zero", $\epsilon = 2.3 \times 10^{-4}$ is the loss per transfer measured as the attenuation of

the first "1" in the string of "1's" and $\epsilon = 1.8 \times 10^{-4}$ as the increase of the first "0" following the string of "1's". For this device, according to our definition of the transfer loss, we would chose $\epsilon = 2.3 \times 10^{-4}$ as the value for transfer loss. However, as the value of fat zero is

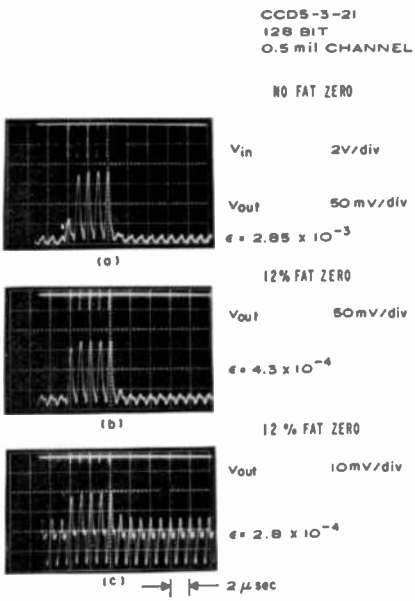


Fig. 17—Waveforms for a 0.5-mil-wide 128-stage register at 1 MHz. (100) substrate.

increased, the same transfer loss would be measured as either the reduction of the first "1" in the string of "1's" or as the increase of the first "0" following the string of "1's". Such measurement is illustrated in Fig. 17(c). Comparison of the waveforms in parts (b) and (c) of Fig. 17 illustrates the reduction in the transfer loss as the signal is reduced for a given value of fat zero.

The operation of the same 64-stage register (CCD-5-6-8) in the C-C and B-B modes is illustrated by Fig. 18, 19 and 20. The comparison of the waveforms shows that the operation in the B-B mode with larger barrier heights leads to larger transfer losses with and without "fat zero".

(b) Transfer Loss as Function of Fat Zero

The reduction of the transfer loss as the function of fat zero is illustrated in Fig. 21. The comparison of the curve A for C-C mode with curves B and C for B-B modes shows that about 10% of fat zero reduces the transfer loss to a minimum value in the case of B-B mode, but the minimum loss for the C-C mode requires larger value of fat zero.

Fig. 22 shows how the transfer loss as a function of fat zero varies with the channel width. The curves on this figure illustrate the edge-effect[12] that becomes more important in the case of the narrower devices.

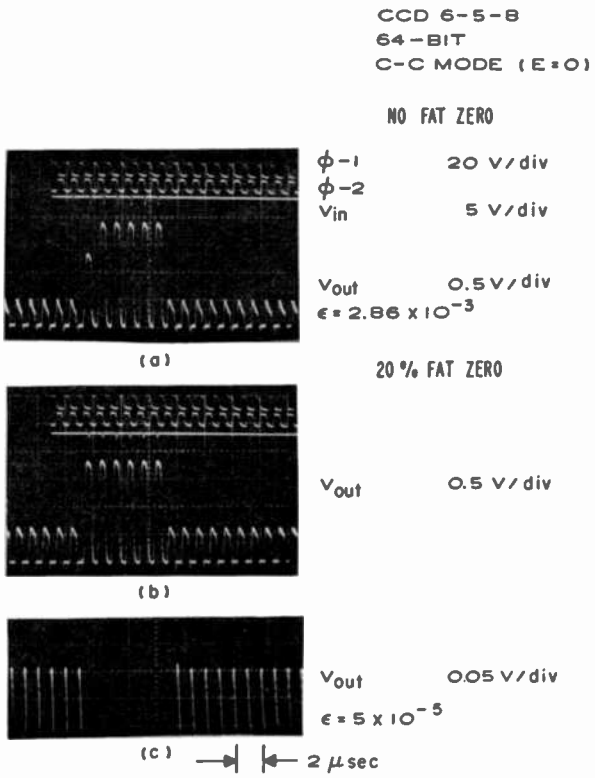


Fig. 18—Waveforms for a 5-mil-wide 64-stage register operating in C-C mode at 1.0 MHz for $E_\phi = 5$ V, $\Delta\phi = 10$ V, $E = 0$, $E_{G4} = 17$ V, and $E_D = 20$ V (substrate (100)).

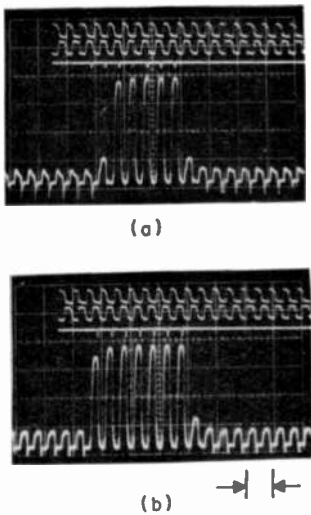


Fig. 19—Waveforms for a 64-stage register of Fig. 18, but operating in B-B mode at 1.0 MHz, with $E_\phi = 5$ V, $\Delta\phi = 10$ V, $E = 5$ V, $E_{G-4} = 17$, and $E_D = 20$ V.

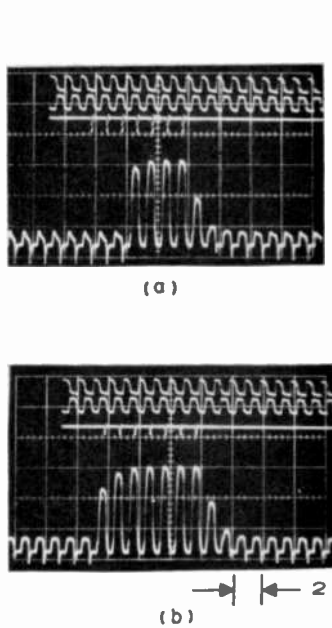


Fig. 20—Waveforms for the 64-stage register of Fig. 18, but operating in B-B mode at 1 MHz with $E_\phi = 5$ V, $\Delta\phi = 10$ V, $E = 10$ V, $E_{G-4} = 17$, and $E_D = 20$ V.

(c) Transfer loss as a function of clock frequency

The measurements of the transfer loss as function of clock frequency are shown in Fig. 23 for devices fabricated on (111) substrates. The dotted curve B represents the best fit to the data of the calculated fast interface state trapping losses according to the following equation,[11]

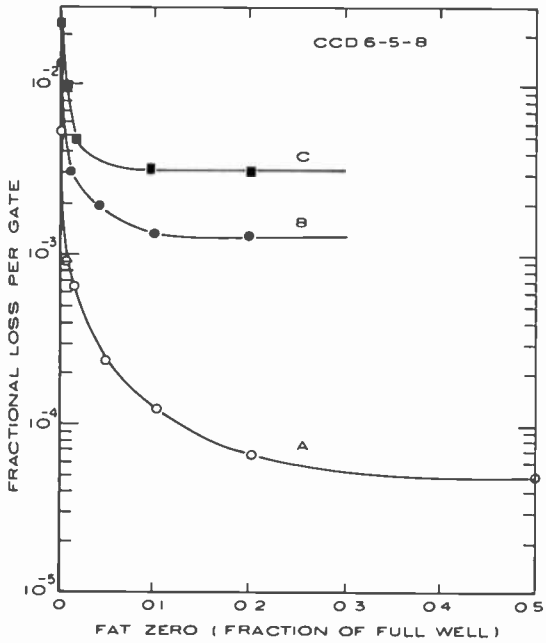


Fig. 21—Fractional loss per transfer versus amount of fat zero for 5 mil-wide 64-stage register. Substrate (100).
Curve A: C-C mode, $E_\phi = 5$ V, $E = 0$, and $\Delta\phi = 10$ V.
Curve B: B-B mode, $E_\phi = 5$ V, $E = 5$ V, and $\Delta\phi = 10$ V.
Curve C: B-B mode, $E_\phi = 5$ V, $E = 10$ V, and $\Delta\phi = 10$ V.

$$\epsilon_s = \frac{(kT/q)\, N_{ss}}{N_{SIG}} \ln\left(1 + \frac{f}{f_o}\right). \qquad [1]$$

From this data we estimate fast interface state densities of $N_{ss} = 1.2 \times 10^{11}$ $(\text{cm}^2 - \text{eV})^{-1}$ for the devices made on silicon with (111) orientation. According to these measurements, the transfer loss with fat zero also decreases as a function of clock frequency for both the

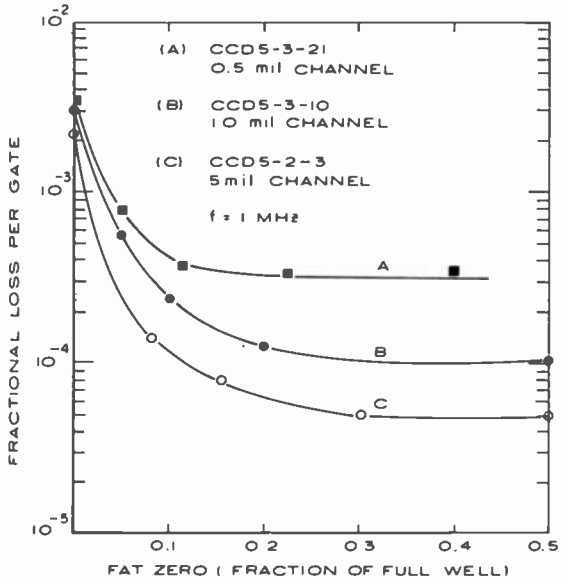


Fig. 22—Fractional loss per transfer at 1 MHz versus amount of fat zero for 0.5, 1.0, and 5.0-mil-wide register made on (100) substrates are shown as Curves A, B, and C, respectively.

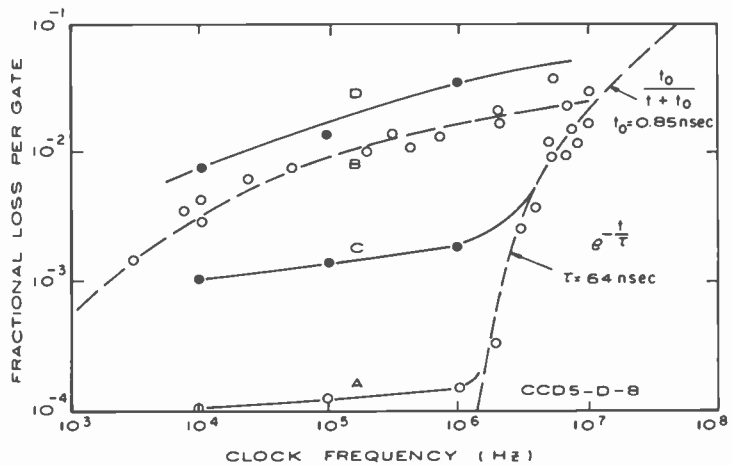


Fig. 23—Fractional loss per transfer versus clock frequency for 128-stage registers made on (111) substrates.
Curve A: C-C mode ($E_\phi = 20$ V, $E = -5$ V, $\Delta\phi = 10$ V) with 10% fat zero;
Curve B: C-C mode with no fat zero.
Curve C: B-B mode ($E_\phi = 10$ V, $E = 5$ V, $\Delta\phi = 10$ V) with 10% fat zero;
Curve D: B-B mode with no fat zero.

C-C and the B-B modes. The dotted curve on the right represents the calculated transfer loss for free-charge transfer for a 0.4-mil-long storage gates assuming self-induced drift dominates transfer for the first 99%, with characteristic time $t_o = 0.75$ nsec, and thermal diffusion dominates thereafter with time constant of 64 nsec appropriate for $L = 0.4$ mil.[13]

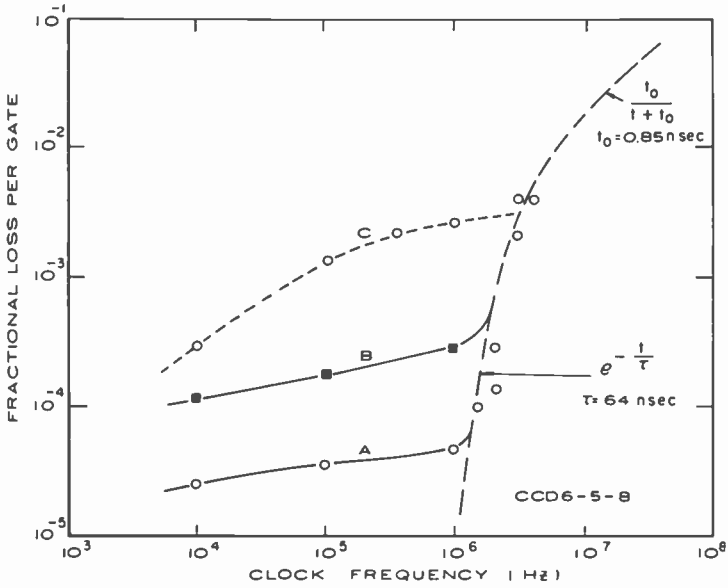


Fig. 24—Fractional loss versus clock frequency for registers made on (100) substrates operating in C-C mode ($E_\phi = 5$ V, $E = 0$, $\Delta\phi = 10$ V). Curve A: 5.0-mil-wide 64-stage register with 50% fat zero; Curve B: 0.5-mil-wide 128-stage register with 30% fat zero; Curve C: 5-mil-wide 64-stage register with no fat zero.

The variation of transfer loss curves for the C-C mode for registers made on (100) substrates are shown in Fig. 24. Here again, the best fit to data points for curve C indicates $N_{ss} = 2.9 \times 10^{10}$ $(\text{cm}^2 - \text{eV})^{-1}$. Curve A shows the transfer loss for a 5-mil-wide register and 50% fat zero. Curve B shows the transfer loss for a 0.5-mil register with 30% fat zero.

The transfer loss for registers made on (100) substrates operated in the B-B mode with two different barrier heights is shown in Fig. 25. The most interesting result of these measurements is that the registers operated in the B-B mode with relatively small barriers have transfer loss that decreased as the clock frequency is reduced.

6 Discussion of the Operation of the Registers in the C-C and B-B Modes

Our initial 2-phase CCD's were made on (111) substrates. Most of these devices operated in the B-B mode or near the transition between the C-C and B-B mode. To operate these devices in the C-C mode, a larger d-c bias had to be applied to the aluminum gates than to the polysilicon gates. To assure operation in the C-C mode with wide range of substrate bias voltages, the devices for which most experimental data is presented were made on (100) substrates with a channel-oxide thickness under the aluminum gates of 2400 Å. According to our data, these devices operated in the C-C mode for substrate bias voltages ranging from 1 V to about 18 or 20 V. Therefore, rather large substrate bias voltages had to be used to obtain the B-B mode of operation. To minimize the effect of the dark current in the comparison of the transfer losses between the C-C and B-B modes for operation with no fat zero, measurements were made on a device with low and uniform dark current. The B-B mode of operation was also obtained by applying a larger d-c bias voltage to the polysilicon gates than to the aluminum gates.

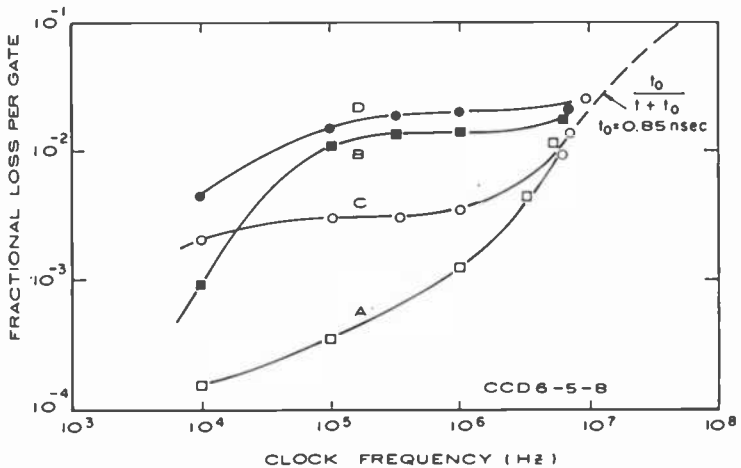


Fig. 25—Fractional loss versus clock frequency for 5-mil-wide 64-stage registers made on (100) substrate operating B-B modes.
Curve A: $E_\phi = 5$ V, $E = 5$ V, $\Delta\phi = 10$ V, and 10% fat zero.
Curve B: Same as A but with no fat zero.
Curve C: $E_\phi = 5$ V, $E = 10$ V, $\Delta\phi = 10$ V, and 10% fat zero.
Curve D: Same as C but with no fat zero.

A more quantitative illustration of the biases required for operation of the registers in the C-C and B-B modes is given in Figs. 26-28. The surface potential curves for the polysilicon and aluminum gates were computed for substrate doping of 10^{16} cm^{-3} and the flat-band voltages were adjusted to conform with the measured threshold voltages for the devices made on substrates with (100) orientation, i.e., $V_{th\,(\text{polysilicon})} = -0.6$V and $V_{th\,(\text{aluminum})} = -3$V. The C-C mode with dc bias $E_\phi = -5$V and clock voltage swing of $\Delta\Phi = 10$V is shown in

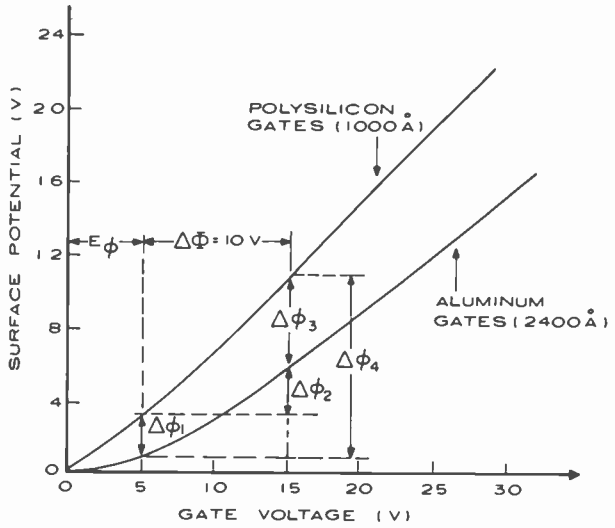


Fig. 26—Calculated surface potentials versus gate voltage for polysilicon gates and aluminum gates having channel oxides of 1000 Å and 2400 Å, respectively, and substrate doping $V_D = 10^{16}$cm^{-3}. Complete charge transfer mode is obtained for both gates with the same dc bias of 5 V and phase voltage swing of 10 V ($E_\phi = 5$ V, $E = 0$, and $\Delta\phi$-1 = $\Delta\phi$-2 = 10 V).

Figs. 26 and 28(b). In this case, $\Delta\phi_3$ corresponds to the maximum signal that can be transferred when the device is operated with overlapping pulses. With nonoverlapping clock pulses, the maximum transferable signal corresponds to $\Delta\phi_1$.

The operation of a 2-phase register in the B-B mode is shown in Figs. 27 and 28(c). Here, the transition to the B-B mode was accomplished by maintaining the aluminum gates at a d-c bias voltage $E_\phi = -5$V, while the d-c bias voltage applied to the polysilicon gates is increased to $E_\phi + E = 15$V. As a result of this new biasing condition, a barrier potential, $\Delta\phi_2'$ shown in Fig. 28(c), is formed under the aluminum transfer gate. Because of this potential barrier, during

the continuous operation of the register in the B-B mode, a conductive layer is maintained under the polysilicon gates that operationally resembles the diffusions of a bucket-brigade structure.

The transition from the C-C mode to the B-B mode by increasing the substrate bias, or the phase bias E_ϕ, while both the polysilicon and the aluminum gates are maintained at the same d-c potential, is also illustrated in Fig. 27. The cross-over point is obtained by shifting

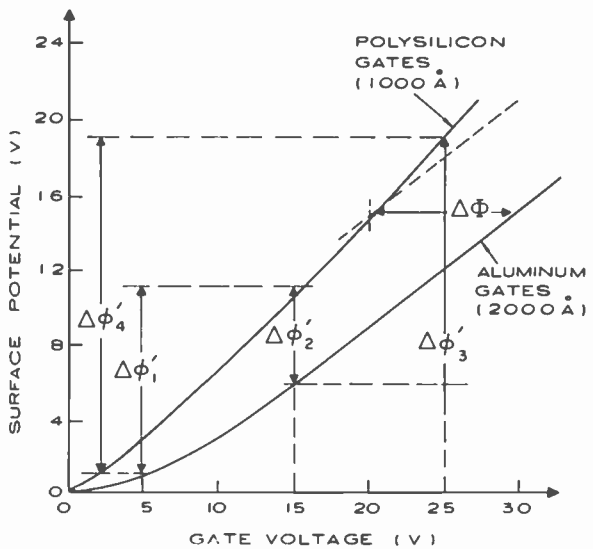


Fig. 27—Bucket-brigade mode of operation is illustrated for polysilicon gates biased with 15 V and aluminum gates with 5 V ($E_\phi = 5$ V, $E = 10$ V, and $\Delta\phi$-1 = $\Delta\phi$-2 = 10 V).

the aluminum gate surface potential by the clock voltage $\Delta\phi = 10$V, as shown by the dotted line. The intersection of the dotted line with the polysilicon gate surface potential is the cross-over bias voltage. The calculated value of this cross-over bias voltage of approximately 20V is in reasonably good agreement with the experimentally observed values of 18-20V.

Another good agreement between the experimental data and the calculated operation shown in Figs. 26-28 is the detected output voltage swing ΔV_{out}. According to the calculations in Figs. 26 and 28(b), full well signal charge should correspond to a surface potential charge, $\Delta\phi_3 \cong 5$V. The value is consistent with the maximum detected output signal $\Delta V_{out} = 1.5$ to 2V taking into account the total capacitance associated with the output diffusion $C_F \sim 1.2$ pF and the total capacitance of the storage gate $C_{os} + C_D = 0.4$ pF.

Finally, the difference between the expected charge transfer characteristics between the C-C mode and the B-B mode is illustrated in Fig. 29. As shown in Fig. 29(a) the final transfer of charge in the C-C mode is expected[13] to take place either under the influence of thermal diffusion with a time constant.

$$\tau_{\text{th}_{\text{Diff}}} = \frac{L_s^2}{2.5\text{D}} \qquad [2]$$

where D is the diffusion coefficient, or, when applicable, by fringing field[14] drift with estimated time constant

$$\tau_{\text{FR Drift}} = 1/3\, t_{\text{transit}} \,. \qquad [3]$$

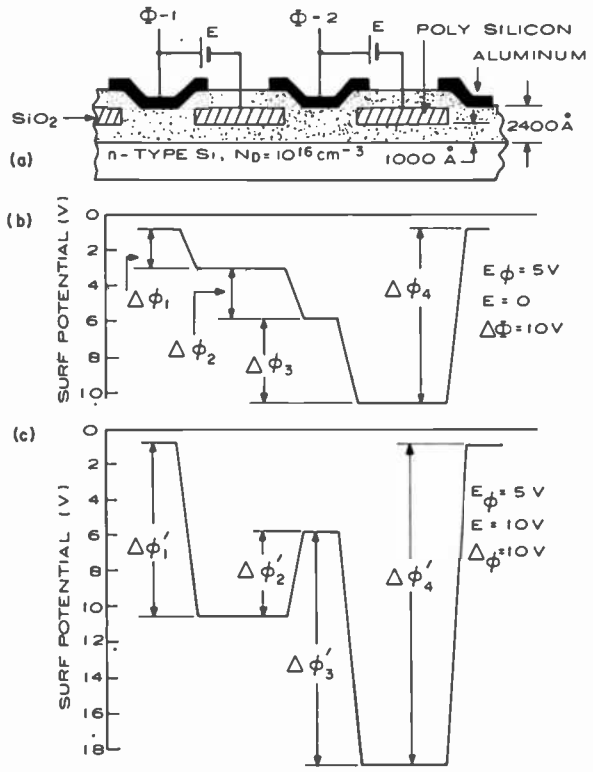


Fig. 28—(a) Surface potential profiles for 2-phase charge-coupled structure; (b) operating in the C-C mode; and (c) in the B-B mode.

Here, $t_{transit}$ represents the transit time of a single carrier in the fringing field.

In the case of operation of 2-phase CCD's in the B-B mode, the initial charge transfer process is expected to be similar to the C-C mode except that the critical length is the transfer gate L_t. But because of the presence of the potential barrier under the transfer gate, the final transfer of charge is expected to involve injection of the charge across the barrier as illustrated in Fig. 29(b).

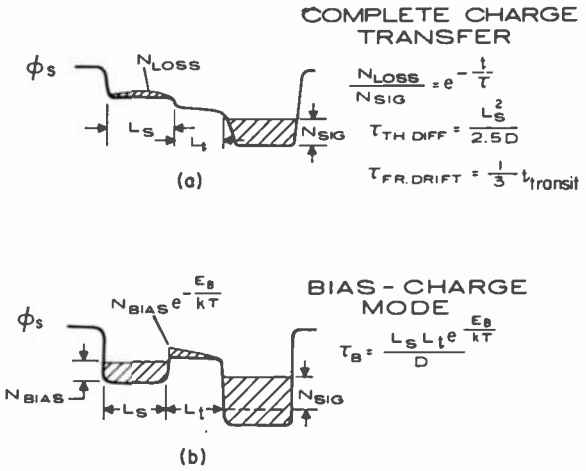


Fig. 29—Comparison of the charge transfer characteristics (a) for the C-C mode and (b) for the B-B mode.

After the charge in the first well is reduced below the barrier under the transfer gate, thermal diffusion will continue with time constant τ_B;

$$\tau_B = \frac{L_s L_t}{D} \exp\left\{\frac{E_B}{kT}\right\}, \qquad [4]$$

where E_B represents the difference between the quasi-Fermi level for N_{BIAS} and the flat barrier voltage under the transfer gate. The expression for τ_B assumes no fringing field under the transfer gate.*

* The experimental demonstration of the C-C and B-B modes of operation of 2-phase CCD's and the proposed model illustrated in Fig. 29(b) were originally described by the authors at the 1972 ISSCC in Philadelphia.

An interesting effect was observed concerning the fractional loss per gate as a function of frequency for devices operating in the B-B mode with a small barrier voltage, see curves A and B in Fig. 25. For operation with fat zero, Curve A, the transfer loss continued to decrease as the clock frequency was reduced. On the other hand, without fat zero, Curve B, a sharp decrease in the transfer loss was observed as the clock frequency was reduced below 100 KHz.

7. Conclusions

Two-phase charge-coupled devices were developed in the form of polysilicon gates overlapped by aluminum gates. Experimental data have been presented for the operation of 64- and 128-stage registers. The two-phase operation in the complete charge-transfer mode was accomplished over a wide range of substrate bias voltages for devices made on n-type substrates with (100) orientation and having resistivity of 0.5 to 1.0 ohm-cm and channel oxide thickness of 1000 Å for the polysilicon gates and 2400 Å for the aluminum gates.

Charge transfer loss of 5×10^{-5} per gate was achieved in the operation of the best 5-mil-wide registers operating with a constant circulating charge—the fat zero. These results were found to be in good agreement with the expected charge-transfer efficiency resulting from incomplete free-charge transfer. The tests of the experimental charge-coupled shift registers showed that without fat zero, the charge-transfer losses are mainly due to the charge trapping by the fast interface states of the $Si\text{-}SiO_2$ interface. The fast interface densities estimated from the experimental work for devices made on (111) substrate were $N_{ss} = 1.2 \times 10^{11}\ (cm^2 - eV)^{-1}$ and for (100) substrates $N_{ss} = 2.9 \times 10^{10}\ (cm^2 - eV)^{-1}$.

The tests showed a higher transfer loss in narrower registers. A test of one set of devices made with 5.0, 1.0, and 0.5 mil-wide channels operated with up to 50% fat zero at 100 KHz gave transfer losses of 5×10^{-5}, 10^{-4}, and 3×10^{-4}, respectively.

A comparison of the complete charge-transfer mode (C-C mode) and the bias-charge or bucket-brigade mode (B-B mode) of operation of the registers showed that at 1.0-MHz clock frequency the complete charge-transfer mode is greater than 10 times more efficient than the bias charge mode. At higher clock frequencies, such as 10 MHz, the two modes of operation were found to have about the same charge-transfer efficiency. Both of these results were obtained with fat-zero operation. For operation without fat zero, the charge-trapping-like transfer losses for the B-B mode were found to be larger than for the

C-C mode. This effect was clearly demonstrated in registers made on (100) substrates.

Acknowledgments

The authors wish to acknowledge the significant contributions of W. Romito who fabricated the devices, J. Sarace and W. French who provided processing assistance, E. Cuomo who helped in mask fabrication, and P. Levine who assisted with measurements. The work of E. Ramberg and L. Chen is also included in the appendices and is greatly appreciated.

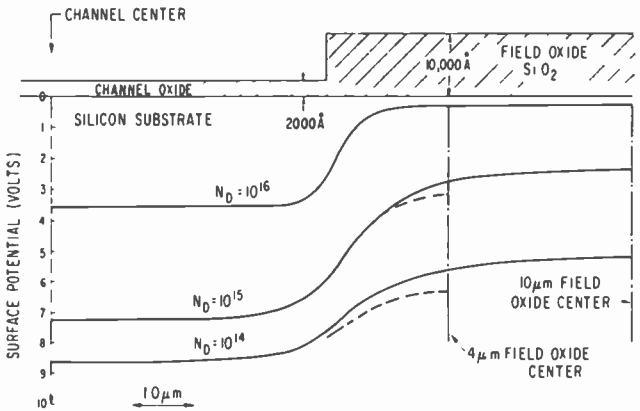


Fig. 30—Typical results obtained from computer analysis of channel confinement problem.

Appendix 1—Channel Confinement

In CCD operation the signal charge must be confined to a narrow potential well called the channel. This channel should provide a deep and abrupt potential well where the surface potential changes in response to the clock pulse voltages. The region outside of the channel, the field region, should be insensitive to clock voltage changes and should be in accumulation.

There are three general methods for channel confinement:

(1) two thicknesses of oxide or thick field oxide,
(2) guard ring diffusion on "channel stops", and
(3) electrostatic guard rings in the form of polysilicon layers.

Computer analyses of the channel confinement problem using channel stop diffusions and two thicknesses of oxide have been made* and typical results for two thicknesses of oxide are shown in Fig. 30.

* These computer studies were performed by Lish-Yann Chen.

The surface potential cross-section is shown for several different substrate dopings for 10-μm-wide channels and either 10-μm or 4-μm wide thick oxide regions. The wells are more abrupt, but shallower at higher doping. As illustrated, the two-oxide method works well for devices with relatively low resistivity substrates. This approach also is more effective for silicon substrates with (111) orientation rather than for the substrate with (100) orientation. (100) orientation has lower fixed oxide charge. On the other hand, the (100) substrates are preferable as they also have lower densities of fast interface states. However, probably the most serious limitation of using thick field oxide for channel confinement is the more difficult and lower yield definition of the metallization over the steps in the oxide. Similar computer investigations have been made for channel confinement using guard ring diffusions shown in Fig. 31. The diffusion guard ring approach

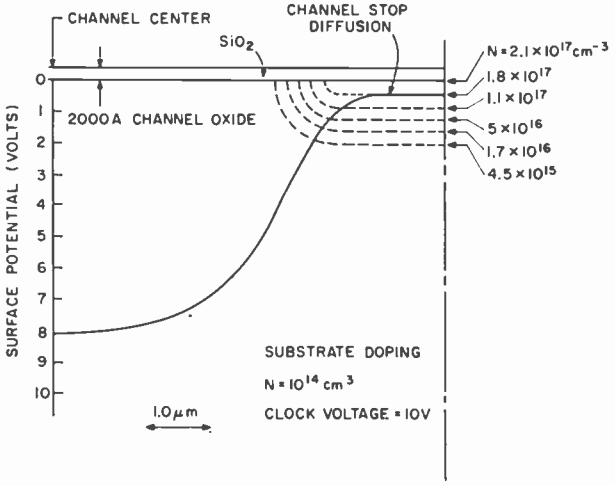


Fig. 31—Computer analysis of channel confinement using guard ring diffusions.

is applicable to high resistivity substrates and also provides for the simplest processing for large-area devices. Ideally the diffusion channel stops should be very abrupt and relatively low-doped (10^{17} to $10^{18}/cm^{-3}$).

The important difference between the first two methods of channel confinement and the polysilicon electrostatic guard rings, also referred to as polysilicon field shield, is that in the later case the surface potential at the channel stop can be determined by an externally controlled potential. Thus, the regions between the CCD channels can be accumulated or held at any other surface potential.

Appendix 2—Trailing Pulse Amplitude for Case of Free Charge Losses (ϵ Independent of Signal Level)

It is desired to find the general expression for the amplitude jth trailing pulse after the last "1" in a string of "1's", after being transferred to the ith gate, $p_j(i)$, when the loss at each gate is a fixed fraction ϵ of the charge transferred. It is assumed that the string of "1's" is long enough for the pulses to reach the full input amplitude, p_o. Consider the first trailing pulse after the last "1", $p_1(i)$. When the last "1" transfer from gate 1 to gate 2, it leaves behind ϵp_o; therefore,

$$p_1(1) = \epsilon p_o. \qquad [5]$$

At gate 2, the last "1" leaves behind another ϵp_o, and $(1-\epsilon)p_1(1)$ is transferred from gate 1 to gate 2 so that

$$p_1(2) = 2\epsilon p_o - \epsilon^2 p_o. \qquad [6]$$

Likewise,

$$p_1(3) = 3\epsilon^2 p_o + \epsilon^3 p_o. \qquad [7]$$

By proceeding in this manner, one can establish that the general expression for the first trailing pulse after i transfers is

$$p_1(i) = p_o \sum_{k=1}^{i} (-1)^{k+1} \binom{i}{k} \binom{k-1}{0} \epsilon^k, \qquad [8]$$

where $\binom{i}{k}$ is the binomial coefficient

$$\binom{i}{k} = \frac{i!}{(i-k)!k!} \qquad [9]$$

The generalized expression for the amplitude of the jth trailing pulse after i transfers is given by

$$p_j(i) = p_o \sum_{k=1}^{i} (-1)^{k+1} \epsilon^{k+j-1} \binom{i+j-1}{k+j-1} \binom{k+j-2}{j-1}. \qquad [10]$$

If this analysis is correct, then the sum of all trailing pulses at any instant in time should be equal to ϵp_o times the number of transfers made by the last "1".

$$\text{Total loss} = \epsilon N_g p_o = \sum_{j=1}^{N_g} p_j (N_g - j + 1). \qquad [11]$$

Note that when the last "1" has made N_g transfers, the first trailer is at $i = N_g$, the second at $i = N_g - 1$, etc., so the index is $N_g - j + 1$.

$$\text{Total Loss} = p_o \sum_{j=1}^{N_g} \sum_{k=1}^{N_g - j + 1} (-1)^{k+1} \epsilon^{k+j-1} \binom{N_g}{k+j-1} \binom{k+j-2}{j-1}$$

This double sum is best evaluated by examining the sum of terms in like powers of ϵ, say ϵ^r. If $k + j = r$, then the second sum is removed and only the $k = r + 1 - j$ term used.

$$\text{Coefficient of } \epsilon^r \text{ term} = p_o \sum_{j=1}^{N_g} (-1)^{r-j} \epsilon^r \binom{N_g}{r} \binom{r-1}{j-1} \qquad [13]$$

$$= p_o \binom{N_g}{r} \epsilon^r (-1)^r \sum_{j=1}^{N_g} (-1)^{j-1} \binom{r-1}{j-1}.$$

However, provided $u \neq 0$, the sum:

$$\sum_{v=0}^{u} (-1)^v \binom{u}{v} = 0. \qquad [14]$$

This is the same as the sum in Eq. [13] for all r except $r = 1$; therefore, the coefficients of all terms of ϵ^r for $r > 1$ are zero. Thus, as can be seen by evaluating Eq. [13] for $r = 1$ and $j = 1$, the total loss is given by $p_o \epsilon N_g$, as expected.

Appendix 3—Analysis of Voltage Buildup and Decay on RC-Line*

Consider a line with resistance per unit length ρ and with capacitance-to-ground per unit length c. For such a line the differential equation for the voltage is

* This analysis was prepared by E. G. Ramberg of RCA Laboratories.

$$\frac{\partial V}{\partial t} = \frac{1}{c\rho}\frac{\partial^2 V}{\partial x^2} \quad (c,\rho \text{ assumed constant}). \qquad [15]$$

For voltage buildup on a line terminated at $x = L$, the boundary conditions are

$$V(0,t) = V_o;\ \frac{\partial V}{\partial x}(L,t) = 0;\ V(x,0) = 0,\ 0 < x \leqslant L. \qquad [16]$$

With these boundary conditions the solution of Eq. [15] becomes

$$V = V_o + \sum_{n=0}^{\infty} A_n(t) \sin\left(\frac{(2n+1)\pi x}{2L}\right). \qquad [17]$$

$A_n(t)$ obeys the differential equation

$$\frac{dA_n}{dt} + \left(\frac{(2n+1)\pi}{2L}\right)^2 \frac{1}{c\rho} A_n = 0 \qquad [18]$$

with the boundary condition

$$A_n(0) = -\frac{2V_o}{L}\int_0^L \sin\left(\frac{(2n+1)\pi x}{2L}\right) dx = -\frac{4V_o}{\pi}\frac{1}{2n+1}. \qquad [19]$$

The solution of Eq. [18] is

$$A_n = -\frac{4V_o}{\pi}\frac{1}{2n+1}\exp\left\{-\left[\frac{(2n+1)\pi}{2L}\right]^2 \frac{t}{c\rho}\right\}. \qquad [20]$$

Substituting Eq. [20] in Eq. [17] leads to

$$V = V_o\left(1 - \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin\left(\frac{(2n+1)\pi x}{2L}\right)\exp\left\{-\left[\frac{(2n+1)\pi}{2L}\right]^2 \frac{t}{c\rho}\right\}\right) \qquad [21]$$

and

$$V(L,t) = V_o\left(1 - \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\exp\left\{-\left[\frac{(2n+1)\pi}{2L}\right]^2\frac{t}{c\rho}\right\}\right). \quad [22]$$

For voltage decay, the boundary conditions, also for the line terminated at L, are

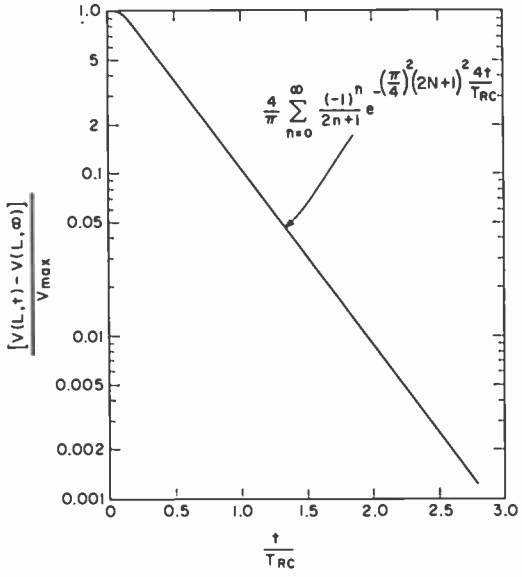


Fig. 32—Normalized transient response at end of RC transmission line.

$$V(0,t) = 0; \quad \frac{\partial V}{\partial x}(L,t) = 0; \quad V(x,0) = V_o, \quad 0 < x \leqslant L. \quad [23]$$

The procedure for solving Eq. [15] is similar to that for voltage build-up and leads simply to

$$V(x,t) = \frac{4V_o}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin\left(\frac{(2n+1)\pi x}{2L}\right)\exp\left\{-\left[\frac{(2n+1)\pi}{2L}\right]^2\frac{t}{c\rho}\right\} \quad [2$$

$$V(L,t) = \frac{4V_o}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\exp\left\{-\left[\frac{(2n+1)\pi}{2L}\right]^2\frac{t}{c\rho}\right\} \quad [25]$$

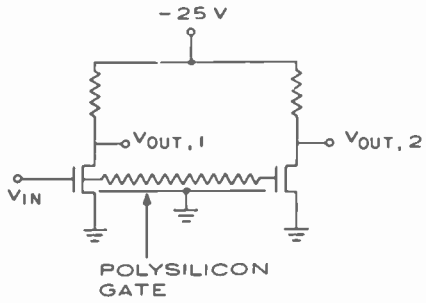


Fig. 33—Test circuit used to verify delay time.

The normalized transient response at the end of the *RC* transmission line of length *L* is plotted in Fig. 32.

The test circuit shown near the bottom of the chip in Fig. 2 was used to experimentally verify the delay time of a long, 100-mil-long polysilicon gate. Fig. 33 schematically shows the entire test circuit

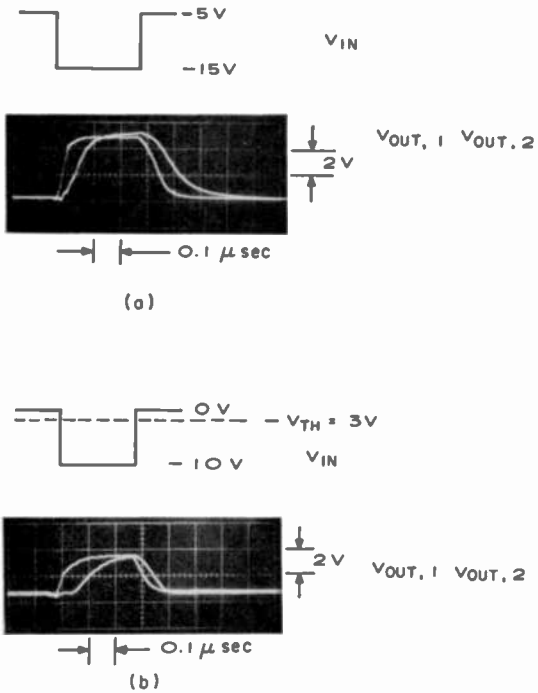


Fig. 34—Detected output (a) where the channel under the polysilicon gate is always in inversion and (b) where it is driven from accumulation into inversion.

for this experiment. The structure consisted of two MOS devices driven by a single, long polysilicon gate that was placed over 1000-Å-thick channel oxide. The gate was 100 mils long and 0.4 mils wide. The sheet resistance of the polysilicon used in this test was approximately 80 ohm per square so that the total resistance was 20 kilohms.

Fig. 34 shows actual oscilloscope traces of the detected output $V_{OUT,1}$ and $V_{OUT,2}$ for the case (a) where the channel under the polysilicon gate was always in inversion and (b) where the channel was driven from accumulation into inversion. The measured delay at 90% build-up in (a) of about 70 nsec is consistent with the predicted time constant of 80 nsec using the inversion layer capacitance of 0.1 pF/mil^2. (Total capacitance of the gate in this case is 4 pF.) In case (b), assuming the effective capacitance corresponds to the accumulation capacitance of 0.2 pF/mil^2, the predicted delay is approximately 160 nsec, which is also consistent with the observed delay of about 130 nsec.

References:

[1] W. S. Boyle and G. E. Smith, "Charge-Coupled Semiconductor Devices," **Bell System Tech. J.**, Vol. 49, p. 587, April 1970.

[2] W. F. Kosonocky and J. E. Carnes, "Charge-Coupled Digital Circuits," **Digest Tech. Papers IEEE Solid State Circuit Conf.**, p. 162, Feb. 19, 1971.

[3] W. F. Kosonocky and J. E. Carnes, "Charge-Coupled Digital Circuits," **IEEE J. Solid-State Circuits**, Vol. SC-6, p. 314 (1971).

[4] G. Amelio, "A Little Bit on Charge-Coupled Devices," International Electron Devices Meeting, Washington, D.C., Oct. 11-13, 1971.

[5] R. H. Krambeck, R. H. Walden, and K. A. Pikar ,"Implanted Barrier Two-Phase CCD," International Electron Devices Meeting, Washington, D.C., October 11-13, 1971.

[6] G. F. Amelio, M. F. Tompsett, and G. E. Smith, "Experimental Verification of the Charge Coupled Device Concept," **Bell System Tech. J.**, Briefs, Vol. 49, p. 593, April 1970.

[7] M. F. Tompsett, et al., "Charge-Coupled 8-Bit Shift Register," **Appl. Phys. Letters**, Vol. 17, p. 111 (1970).

[8] F. L. J. Sangster and K. Terrm, "Bucket-Brigade Electronics—New Possibilities of Delay, Time Axis Conversion, and Scanning," **IEEE J., Solid State Circuits**, Vol. SC-4, p. 3, June 1969.

[9] F. L. J. Sangster, "Integrated MOS and Bipolar Analog Delay Lines Using Bucket-Brigade Capacitor Storage," **ISSCC Digest of Technical Papers**, Vol. XIII, p. 74, 1970.

[10] R. D. Melen and J. D. Meindl, "One-Phase CCD: A New Approach to Charge-Coupled Device Clocking," **IEEE J. of Solid-State Circuits**, Vol. SC-7, Feb. 1972, p. 92.

[11] J. E. Carnes and W. F. Kosonocky, "Fast Interface State Losses in Charge-Coupled Devices," **Appl. Phys. Ltrs.**, Vol. 20, p. 261, April 1, 1972.

[12] M. F. Tompsett, "The Quantitative Effects of Interface States on the Performance of Charge-Coupled Devices," **IEEE Trans. on Electron Devices**, Vol. ED-20, p. 45, Jan. 1973.

[13] J. E. Carnes, W. F. Kosonocky and E. G. Ramberg, "Free Charge Transfer in Charge-Coupled Devices," **IEEE Trans. on Electron Devices**, Vol. ED-19, No. 6, p. 798, June 1972.

[14] J. E. Carnes, W. F. Kosonocky and E. G. Ramberg, "Drift-Aiding Fringing Fields in Charge-Coupled Devices," **IEEE J. Solid State Circuits**, Vol. SC-6, p. 322 (1971).

Recent Papers by RCA Authors

Listing is alphabetical by name of publication. For copies of reprints, the reader should contact the publication directly.

D. Vilkomerson, "Analysis of Various Ultrasonic Holographic Imaging Methods for Medical Diagnosis," **Acoustical Holography,** (Book) Vol. 4, p. 401

D. P. Bortfield and H. Meier, "Refractive Indices and Electro-Optic Coefficients of the Eulites $Bi_4Ge_3O_{12}$ and $Bi_4Si_3O_{12}$," **Applied Physics, J. of,** Vol. 43, p. 5110, Dec. 1972

M. Ettenberg, H. F. Lockwood, H. S. Sommers, Jr., "Radiation Trapping in Laser Diodes," **Applied Physics, J. of,** Vol. 43, p. 5047, Dec. 1972

R. U. Martinelli, M. L. Schultz, and H. F. Gossenberger, "Reflection and Transmission Secondary Emission from GaAs," **Applied Physics, J. of,** Nov. 1972

A. R. Moore and J. W. Davenport, "Acoustoelectric Domains in GaAs: On and Off Axis," **Applied Physics, J. of,** Vol. 43, p. 4513, Nov. 1972

W. E. Ham, "Surface-Charge Effects on the Resistivity and Hall Coefficient of Thin Silicon-on Sapphire Films," **Applied Physics Lett.,** Vol. 21, p. 440, Nov. 1972

A. C. Ipri, "Accumulation and Inversion-Layer Hall Mobilities in Silicon Films on Sapphire," **Applied Physics Lett.,** Vol. 22, p. 16, Jan. 1973

V. S. Ban, "Mass Spectrometric and Thermodynamic Studies of the CVD of Some III-V Compounds," **Crystal Growth, J. of,** Vol. 17, p. 19, 1972

G. W. Cullen and F. C. Dougherty, "Silicon on Spinel: The Interaction Between Deposition Constituents and the Substrate Surface," **Crystal Growth, J. of,** Vol. 17, p. 230, 1972

B. J. Curtis and J. P. Dismukes, "Effects of Natural and Forced Convection in Vapor Phase Growth Systems," **Crystal Growth, J. of,** Vol. 17, p. 128, 1972

G. E. Gottlieb and J. F. Corboy, Jr., "The Development of a Dual Rate Technique for the Growth of Silicon-on-Sapphire Films," **Crystal Growth, J. of,** Vol. 17, p. 261, 1972

J. L. Vossen and E. B. Davidson, "The Interaction of Photoresists with Metals and Oxides During RF Sputter-Etching," **Electrochemical Soc., J. of,** Vol. 119, p. 1708, Dec. 1972

H. A. Wittlinger, "Simple, Low-Cost Sample and Hold Circuit," **Electronic Design,** Oct. 12, 1972

M. V. Hoover, "Programmable Power Switch Amplifier," **Electronic Engineering Times,** Nov. 1973

K. Kugele, "Designing with Hybrid Regulators," **Electronic Engineering Times,** Dec. 1972

B. E. Berson, "Transferred-Electron Devices Take on More Roles in Microwave Systems," tems," **Electronics,** Dec. 1972

D. Morgan, "The RCA COS/MOS CD4046 Micro-Power Phase-Lock Loop," **Electronics,** Sept. 11, 1972

D. Morgan, "The RCA COS/MOS CD4046A Micro-Power Phase-Locked Loop," **Electronics,** Sept. 25, 1972

A. Sheng, "Build a Dual-Tracking Voltage Regulator for Less than $8," **Electronic Products,** Sept. 1972

H. A. Wittlinger and A. J. Visioli, "Super-Sweep Function Generator," **Electronic Products,** Aug. 1972

H. Kawamoto, "Gigahertz-Rate Hundred Volt Pulse Generator," **IEEE J. of Solid-State Circuits,** Vol. SC-8, p. 63, Feb. 1973

C. L. Upadhyayula and B. S. Perlman, "Design and Performance of Transferred Electron Amplifiers Using Distributed Equalizer Networks," **IEEE J. of Solid-State Circuits,** Vol. SC-8, p. 29, Feb. 1973

R. H. Dean and R. J. Matarese, "The GaAs Traveling-Wave Amplifier as a New Kind of Microwave Transistor," **IEEE Proc.,** Vol. 60, p. 1486, Dec. 1972

J. I. Pankove and J. Berkeyheiser, "Frequency Response of GaN Light-Emitting Diodes," **IEEE Proc.,** Vol. 60, p. 1456, Nov. 1972

R. L. Spalding, S. A. Ochs, E. Luedicke, and R. E. Flory, "Bivicon—A New Double Vidicon," **IEEE Proc.,** Oct. 1972

N. W. Patrick "Developments in Cathode Ray Tubes for the Schmidt Projector," **IEEE Conf. on Display Devices, Proc.,** Oct. 1972

R. E. Simon, "A Solid-State Boost for Electron-Emission Devices," **IEEE Spectrum,** Vol. 9, p. 74, Dec. 1972

J. Alves, "Power Bandwidth in Audio Amplifiers," **IEEE Trans. Broadcast and TV Receivers,** May 1973

W. F. Dietz, "A Boosted B+ Regulator for the SCR Deflection System," **IEEE Trans. Broadcast and TV Receivers,** Feb. 1973

G. W. Beakley and F. B. Tuteur, "Distribution-Free Pattern Verification Using Statistically Equivalent Blocks," **IEEE Trans. Computers,** Vol. C-21, p. 1337, Dec. 1972

R. H. Dean, "Reflection Amplification in Thin Layers of n-GaAs," **IEE Trans. Electron Devices,** Vol. ED-19, p. 1148, Nov. 1972

R. H. Dean, "A Practical Technique for Controlling Field Profile in Thin Layers of n-GaAs," **IEEE Trans. Electron Devices,** Vol. ED-19, p. 1144, Nov. 1972

H. Johnson, "A Unified Small-Signal Theory of Uniform-Carrier-Velocity Semiconductor Transit-Time Diodes," **IEEE Trans. Electron Devices,** Vol. ED-19, p. 1156, Nov. 1972

L. Schiff, "Statistical Suppression of Interference with Diversity in a Mobile-Radio Environment," **IEEE Trans. on Vehicular Technology,** Vol. VT-21, p. 121, Nov. 1972

J. Shefer and G. S. Kaplan, "An X-Band Vehicle-Location System," **IEEE Trans. Vehicular Technology,** Vol. VT-21, p. 117, Nov. 1972

H. Kressel, "Semiconductor Lasers: Devices," **Laser Handbook,** Chap. B5, p. 443, North-Holland Pub. Co., Amsterdam, 1972

J. Pankove, "UV DC Electroluminescence from GaN," **Luminescence, J. of,** Vol. 5, p. 482, 1972

G. Hodowanec, "High-Power Microwave Transistor Oscillators," **Microwave J.,** Oct. 1972

D. S. Jacobson, "Comparison of Microwave Power Transistor Structures," **Microwaves,** July 1972

A. Marnamm, "Automotive Industry," **Modern Electronik,** Sweden, Jan. 1973

M. T. McCaffrey and J. A. Castellano, "Liquid Crystals VII. The Mesomorphic Behavior of Homologous p-Alkoxy-p'-Acyloxyazoxybenzenes," **Molecular Crystals and Liquid Crystals,** Vol. 18, p. 209, 1972

D. A. deWolf, "Strong Irradiance Fluctuations in Turbulent Air: Plane Waves," **Optical Society of America, J. of,** Vol. 63, p. 171, Feb. 1973

W. Rehwald, "Dependence of the Ultrasonic Attenuation in Dielectric Materials on Structural Parameters," **Phonon Scattering in Solids,** Proc. Intl. Conf., Paris, July 3-6, p. 45, 1972

J. R. Sandercock, S. B. Palmer, R. J. Elliott, W. Hayes, S. R. Smith, and A. P. Young, "Brillouin Scattering, Ultrasonic and Theoretical Studies of Acoustic Anomalies in Crystals Showing Jahn-Teller Phase Transitions," **Physics, J. of,** Vol. 5C, p. 3126, Oct. 1972

R. C. Alig, "Comments on Superconducting Surface Resistance," **Physical Review B,** Vol. 7, p. 1188, Feb. 1973

J. R. Sandercock, "Structure in the Brillouin Spectra of Thin Films," **Physical Review Lett.,** Vol. 29, p. 1735, Dec. 25, 1972

H. Fujita, Y. Iguchi, Y. Okada, and T. Sasaki, "The Measurements of the Reflectivity Spectrum near the $L_{2,3}$ Edge of p-Type Silicon at Room Temperature," **Physical Society of Japan, J. of,** Vol. 33, p. 1494, Nov. 1972

R. Hirota, "Exact Solution of the Modified Korteweg-de Vries Equation for Multiple Collisions of Solitons," **Physical Society of Japan, J. of,** Vol. 33, p. 1456, Nov. 1972

R. Hirota, "Exact Solution of the Sine-Gordon Equation for Multiple Collisions of Solitons," **Physical Society of Japan, J. of,** Vol. 33, p. 1459, Nov. 1972

N. Feldstein, "Plating on Plastics—As a Method for Producing a New Type of Cathode-Ray-Tube Shadow Mask," **Plating,** Vol. 60, Jan. 1973

S. Alexander, "A Thyristor Horizontal Deflection System with Auxiliary Power Supplies for 110° Monochrome Receivers," **RCA Application Note** AN-6076

W. M. Austin, "A Flexible Integrated-Circuit Color Demodulator for Color Television," **RCA Application Note** ICAN-6724

P. A. Bothner, "Off-Line 100-Volt 1-Kilowatt Converter," **RCA Application Note** AN6054

L. R. Campbell and H. A. Wittlinger, "Some Applications of a Programmable Power Switch/Amplifier," **RCA Application Note** ICAN-6048

W. Dietz, "A Thyristor Line-Scan and EHT System for the RCA A67-150X Color Picture Tube," **RCA Application Note** AN6009

S. S. Eaton, "Timekeeping Advances Through COS/MOS Technology," **RCA Application Note** ICAN-6086

R. C. Heuner, "Liquid Crystals/COS Offer Minimal-Power Digital Displays," **RCA Application Note** ICAN-6100

R Heuner, J Litus, Jr, and A. Havasy, "COS/MOS MSI Counter and Register Design and Applications," **RCA Application Note** ICAN-6166

C. Kamnitisis, B. Maximow, and M. O'Molesky, "Building Blocks for Mobile Radio Design," **RCA Application Note** AN-6099

J. Locke, "10-, 16-, 30-, and 60-Watt Broadband (620-to-960-MHz) Power Amplifiers Using the RCA-2N6266 and 2N6267 Microwave Power Transistors," **RCA Application Note** AN-6118

B. Maximow, "RCA 2N6105 UHF Power Transistor—Its Characteristic and Application," **RCA Application Note** AN6010

B. Maximow, "60- and 100-Watt Push-Pull Broadband (225-400MHz) Amplifiers Using RCA-266105 Transistors," **RCA Application Note** AN-6126

T. J. Robe, "Measurement of Burst ("Popcorn") Noise in Linear Integrated Circuits," **RCA Application Note** ICAN-6732

O. H. Schade, Jr., "Digital-to-Analog Conversion Using the RCA-CD4007A COS/MOS IC," **RCA Application Note** ICAN-6080

H. A. Wittlinger, "Applications of the CA3080 and CA3080A High-Performance Operational Transconductance Amplifiers," **RCA Application Note** ICAN-6668

J. W. Tuska and M. H. Woods, "Degradation of MNOS Memory Transistor Characteristics and Failure Mechanism Model," **Reliability Physics, Proc. 10th Annual Symp.,** p. 120, 1972

H. S. Veloric, S. Lazar, R. Minton, H. Meisel, P. Schnitzler, and C. Kamnitsis, "Lumped-Element Integrated Circuits for UHF and L-Band Applications," **Seminar Notes of the First Regional Seminar on High-Frequency Applications of Hybrid Microelectronics,** Technical Applications and Possibilities, Cambridge, Mass., June 1972

T. Freund, "Less-Than-Bandgap Photosensitivity of ZnO and Dye Sensitization by Ferrocyanide Ion," **Surface Science,** Vol. 33, p. 295, 1972

T. M. VanLier, "Performance Test of Semi-Conductors," **Technical Review of University of Liege,** Oct. 1972

J. L. Vossen and E. S. Poliniak, "The Properties of Very Thin RF Sputtered Transparent Conducting Films of SnO_2:Sb and In_2O_3:Sn," **Thin Solid Films,** Vol. 12, p. 281, 1972

D. R. Carley, "An Engineering Assessment of Digital Integrated Circuits for Consumer Applications," **WESCON Conf. Proc.,** Sept. 1972

D. P. Del Frate, "The Solid State Quality-Cost Equation—Its Impact on the Specifier, Buyer, and Supplier," **WESCON Conf. Proc.,** Sept. 1972

Patents Issued to RCA Inventors

October

G. A. Alphonse Sonic Page Composer for Holographic Memory (3,698,794)

J. M. Assour Avalanche Diode (3,696,272)

J. Avins and J. Craft Automatic Frequency Control Circuits (3,697,885)

B. W. Beyers and L. L. Tretter Communications Among Computers (3,699,529)

E. J. Boleky III Semiconductor Device Including Fusible Elements (3,699,395)

E. J. Boleky III Fusible Semiconductor Device Including Means for Reducing the Required Fusing Current (3,699,403)

D. R. Bosomworth Character Generator Utilizing a Display with Photochromic Layer (3,700,791)

J. R. Burns Memory Subsytsem Array (3,701,984)

S. Caplan and M. F. Lamorte Semiconductor Laser Devices Utilizing Light Reflective Metallic Layers (3,701,047)

E. W. Christensen II and R. D. Eudaly Deflection Yoke Mount (3,697,909)

B. P. Clay Holographic Multicolor Technique (3,695,744)

L. A. Cochran Color Compensating Network for an Integrated Circuit Television Receiver (3,701,844)

J. Craft Peak-to-Peak Detector (3,701,022)

J. A. Dodd, Jr. and R. A. Geshner Method of Making a Graded Photoprinting Master (3,698,903)

W. J. Donoghue Signal Generating Circuit Including a Pair of Cascade Connected Field Effect Transistors (3,697,777)

E. C. Douglas Method of Projection Printing Photoresist Masking Layers, Including Elimination of Spurious Diffraction-Associated Patterns from the Print (3,697,178)

R. N. Epifano and E. L. Jordan Method of Making a Semiconductor Device Including a Polyimide Resist Film (3,700,497)

N. Feldstein Method of Metallizing an Electrically Insulating Surface (3,697,319)

E. C. Giaimo, Jr. Apparatus for Charging a Recording Element with an Electrostatic Charge of a Desired Amplitude (3,699,335)

E. A. Gronka Hight Voltage Processing of Cathode Ray Tubes (3,698,786)

P. E. Haferl Deflection and Pincushion Correction Circuit (3,700,958)

L. A. Harwood and E. J. Wittmann Amplifiier Circuits (3,699,257)

P. M. Heyman Method of Improving Cathodochromic Sensitivity (3,700,804)

A. Hildebrants Cathode Ray Tube Implosion Protection System and Method (3,697,686)

G. M. Hunter Memory System Including Buffer Memories (3,699,533)

G. E. Kelly Safety Apparatus for Hot-Chassis Electronic Instruments (3,699,562)

L. B. Kimbrough Apparatus and Method for Applying a Bead of Sealing Material to a Sealing Surface of a Cathode-Ray Tube (3,701,674)

L. R. Kirkwood, L. A. Cochran, and R. D. Altmanshofer Color Compensating Network with Range Limitation (3,701,842)

M. F. Lamorte Method of Making Semiconductor Devices (3,697,336)

N. R. Landry, R. J. Mason, W. H. Schaedla, and W. T. Patton Flexible and Slidable Waveguide Feed System for a Radiating Horn Antenna (3,698,000)

J. J. McCardle and R. L. Rauck High Voltage Hold Down Circuit (3,697,800)

C. F. Madrazo, E. M. Fulcher, and K. P. McDonagh Gate Circuit (3,699,355)

R. A. Mancini Crystal Controlled Digital Logic Gate Oscillator (3,699,476)

J. C. Marsh, Jr. Automatic Degaussing Apparatus for Minimizing Residual Curring During Steady State Operation (3,699,400)

M. E. Miller Phonograph Pickup with Self Formed Female Receptacles (3,699,268)

J. W. Mirsch Color Picture Tube Beam Convergence Apparatus (3,701,065)

L. A. Murray Method of Making Complementary Insulated Gate Field Effect Transistors (3,700,507)

K. J. Phillips Nutation Damping in Dual-Spin Spacecraft (3,695,554)

F. R. Ragland Method of Assembling and Mounting an Aperture Mask in a Mask-Panel Assembly of Cathode-Ray Tube Using a Full Surface Spacer (3,701,193)

F. E. Richter Fabricating Relatively Thick Ceramic Articles (3,695,960)

A. N. Saxena Schottky Barrier Diode and Method of Making the Same (3,700,979)

R. E. Simon and B. F. Williams Negative Effective Electron Affinity Emitters with Drift Fields Using Deep Acceptor Doping (3,699,404)

H. Sorkin and R. I. Klein Fabrication of Liquid Crystals Devices (3,698,057)

R. R. Speers Method for Forming Isolated Semiconductor Devices (3,695,956)

J. Stark, Jr., and D. H. Carpenter Hue Control Circuit for a Color Television Receiver (3,701,845)

S. A. Steckler Electronic Signal Amplifier (3,701,032)

H. A. Stern Fabrication of Liquid Crystal Devices (3,701,368)

R. G. Stoudenheimer and R. P. Dourte Photocathode Comprising Layers of Tin Oxide, Antimony Oxide, and Antimony (3,697,794)

J. J. Tietjen, B. F. Williams, and C. C. Wang Photoemissive Electron Tube Comprising a Thin Film Transmissive Semiconductor Photocathode Structure (3,699,401)

M. Van Renssen Method of Assembling a Mask with a Frame Assembly for Mounting in Cathode-Ray Tube Using a Remote Assembly Position (3,701,185)

C. G. Warner and J. H. Kade Reversible Fixture for Positioning Magnetic Memory Cores (3,698,057)

H. R. Warren Web Transport Apparatus (3,700,152)

P. K. Weimer Signal Transfer System for Panel Type Image Sensor (3,696,250)

J. G. Williams Computer with Probability Means to Transfer Pages from Large Memory to Fast Memory (3,701,107)

J. M. Yongue Mounting Attachment for a Modular Substrate (3,697,817)

November

C. J. Alduck and R. A. Mancini Zero Crossing Point Switching Circuit (3,702,941)

S. F. Burtis Process for Forming an Isolated Circuit Pattern on a Conductive Substrate (3,704,207)

J. A. Castellano Liquid Crystal Color Displays (3,703,329)

S. B. Deal and D. W. Bartch Method for Preparing the Viewing-Screen Structure of Cathode-Ray Tube (3,703,401)

A. G. F. Dingwall Shaped Riser on Substrate Step for Promoting Metal Film Continuity (3,703,667)

M. T. Duffy MOS Transistor with Aluminum Oxide Gate Dielectric (3,702,786)

J. E. Goldmacher and G. H. Heilmeier Liquid Crystal Display Element Having Storage (3,703,331)

W. J. Hannan, J. P. Russell, and A. H. Firester Relief Phase Holograms (3,703,407)

R. C. Heuner and R. P. Fillmore Field-Effect Transistor Circuit for Detecting Changes in Voltage Level (3,702,943)

H. Kawamoto and E. L. Allen, Jr. Broadband Microwave Apparatus Using Multiple Avalanche Diodes Operating in the Anomalous Mode (3,702,971)

A. Macovski Decoding of Color-Encoded Phase Grating (3,702,725)

E. C. Ross Variable Threshold Memory System Using Minimum Amplitude Signals (3,702,990)

R. R. Russo Process for Forming a Conductive Coating on a Substrate (3,704,208)

G. W. Steudel Voltage-Controlled Oscillator Using Complementary Symmetry MOSFET Devices (3,702,446)

G. W. Taylor Light Modulator and Display Device (3,703,332)

December

T. A. Bridgewater Motor Control Circuit (3,706,016)

R. H. Dean Semiconductor Device (3,706,014)

W. J. Gorkiewicz and J. A. van Raalte Achromatic Depth-of-Field Correction for Off-Axis Optical System (3,704,936)

M. L. Levene Light Beam Scanning (3,707,723)

A. L. Limberg FM Stereophonic Receiver Detection Apparatus and Disabling Means (3,707,603)

L. M. Lunn Video Blanking and Audio Muting Circuit (3,707,597)

F. I. Palmer and J. R. Fogleboch Wide Dynamic Range Product Detector (3,705,355)

R. D. Scott Aerial Photography (3,707,254)

I. Shidlovsky Cathodochromic Sodalite and Cathode Ray Tube Employing Same (3,705,323)

AUTHORS

James E. Carnes received the B.S. degree from Pennsylvania State University, University Park, in 1961, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, N. J., in 1967 and 1970, respectively. His Ph.D. dissertation was an investigation on photo-induced currents and charge transport in polyvinylcarbazole, an organic polymer. He was in the U. S. Navy from 1961 to 1965. During the summers of 1966 and 1967 he investigated metallic contacts and dc electroluminescence in strontium titanate at RCA Laboratories, David Sarnoff Research Center, Princeton, N. J., which he joined as a member of the technical staff in 1969. Since that time he has studied electrical breakdown, conduction, and interface properties of various insulating films on silicon and is currently involved in the investigation of charge-coupled devices.

Dr. Carnes is a member of the American Physical Society, Tau Beta Pi, Phi Kappa Phi, and IEEE.

Louis S. Cosentino received the B.E.E. degree from City College of New York, N. Y., in 1960 and the M.S. degree from Princeton University, Princeton, N. J. in 1962. Since 1960 he has been a Member of the Technical Staff at RCA Laboratories, Princeton, N. J. His past endeavors have included the design and development of tunnel-diode circuits for high-speed computers and several years of research in cryoelectrics, including switching devices, memories, and associated electronics. More recently he has worked on the optics and components for read-write optical memory systems utilizing holography.

Mr. Cosentino was co-recipient of an RCA Laboratories Outstanding Achievement Award in 1968 and 1971. He is a senior member of the IEEE, and a member of Eta Kappa Nu and AAAS.

Robert E. Flory received his B.E.E. degree from Cornell University in 1954. After graduation he joined RCA Laboratories. In 1955 he was awarded an RCA David Sarnoff Fellowship to attend the University of Pennsylvania for one academic year. Upon completion of his thesis on time division multiplex systems, the M.S. in EE was granted in 1962. At RCA Laboratories, Mr. Flory has worked on video systems, television standards conversion, ultra-high speed computers, and various TV projects, including digitally addressed electroluminescent displays and color TV displays. In 1959 he spent three months in the Soviet Union installing and operating color television equipment for RCA at the American Exhibition in Moscow. Mr. Flory is currently a member of the Consumer Electronics Research Laboratory. He was team leader for video and sound on the Holotape project and is currently involved in work on color TV cameras, and other imaging systems. In 1971, he received an RCA Laboratories Achievement Award for Electron Beam Recording of Color TV, and in 1972, he shared in a David Sarnoff Outstanding Achievement Award in Science for the Holotape project.

Mr. Flory is a Senior Member of the IEEE and a member of SMPTE.

Edward C. Glaimo, Jr. received his B.S.E. degree with Honors in Engineering from Princeton University and an M.S.E. from the same University. He subsequently studied Solid State Electronics at Princeton. From 1943-46 he served with the United States Navy, completed the United States Navy Radio Material Program and was assigned to the Naval Research Laboratory, Washington, D.C. In 1951 he joined RCA Laboratories where he made contributions to ruggedization of precision traveling-wave-tube helix structures, developed apparatus for the purification and testing of early semiconductor materials used for transistors, has made original contributions in many aspects of the Electrofax process and materials and related electrophotographic systems. From 1957 to 1959 he was a member of the technical staff of C-Stellarator Associates, an organization of RCA and Allis Chalmers engineers, charged with the task of designing and building a thermonuclear research reactor for fusion research at the Matterhorn Project, Princeton University. His staff engineering duties with this organization included system specification and coordination.

Mr. Glaimo is a member of the American Association for the Advancement of Science, Sigma Xi and a Senior Member of the IEEE.

E. O. Johnson received the BSEE from Pratt Institute in Brooklyn in 1948 and pursued graduate studies in physics and electrical engineering at Princeton University and at the Swiss Federal Institute of Technology in Zurich. He worked in the fields of plasma, solid-state-surface, and general-device physics at RCA Laboratories, Princeton, from 1950 to 1959. He was Manager of Advanced Development, and Chief Engineer of the RCA Solid State Division until 1964, and was Manager of Engineering of the RCA Electronic Components and Devices Division until 1970. He is presently Manager of Technical Liaison for RCA International Licensing. Mr. Johnson is a Fellow of the Institute of Electrical and Electronic Engineers.

Walter F. Kosonocky received the B.S. and M.S. degrees in electrical engineering from Newark College of Engineering, Newark, N. J., in 1955 and 1957, respectively, and the Sc.D. degree in engineering from Columbia University, New York, N. Y. in 1965. Since June 1955 he has been employed at RCA Laboratories, Princeton, N. J., where he has conducted research on application of new phenomena and new devices for information processing systems. This work has included ferrite memory systems, parametric digital devices, tunnel-diode circuits, tunnel-diode and transistor circuits, pattern-recognition systems, applications of lasers for digital systems (including a study of saturable absorbers for Q-switched lasers and semiconductor laser digital devices), optical hologram memory systems and page composition, optically controlled p-MOS circuits, and a liquid-crystal image converter. Presently he is working on the development and applications of charge-coupled devices.

Dr. Kosonocky is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and a senior member of IEEE.

Robert D. Lohman received a B.S. degree in Electrical Engineering from Norwich University in 1949 and an M.S. degree in Electrical Engineering from North Carolina State College in 1951. In June 1951, he joined the RCA Laboratories as a member of the technical staff. From 1951 to 1956, he was engaged in research in the areas of basic semiconductor noise phenomena, transistor circuit development, color television display systems, and information theory. In 1956, he joined the RCA Semiconductor and Materials Division as an applications engineer. In 1957, he became Manager of Applications for computer devices. In 1960, Mr. Lohman was promoted to Engineering Manager, Computer Products, and in 1963 was made Manager of Integrated Circuit Engineering. In June 1966, he returned to RCA Laboratories as Group Head, Integrated Electronics Research. His present position is Head, Optical Data Storage Research.

Mr. Lohman is a Fellow of the IEEE.

Eduard Luedicke received his degrees in Electrical Engineering from the Technical University at Berlin and at Karlsruhe, West Germany—the Diplom Engineer in 1950, the Doctor degree (summa cum laude) in 1955. From 1941 to 1945, he was a development engineer for radar and guidance techniques in the Telefunken Laboratories, Berlin, Germany and from 1951 to 1955 a development engineer in the field of industrial television at the Zentral Laboratorium of Siemens in Karlsruhe, West Germany. In 1955 he joined the RCA Victor Company Ltd., Montreal, Canada as a development engineer for television components. He was appointed manager of engineering of the RCA Victor Renfrew Plant in Canada in 1956. In 1959, Dr. Luedicke transferred to the RCA Solid State Division, Somerville, N. J., working initially on micromodules and later on a computer in the 1000 megacycle range. From 1962 to 1967, Dr. Luedicke was a member of the Astro-Electronics Division, working on television camera tubes for use in space applications and in studies of astronomy. In 1967, he transferred to the Electro-Optics Laboratory at the David Sarnoff Research Center in Princeton, N. J., where he has made extensive studies of the dynamic characteristics of camera tubes at various scan rates and contributed to the development of TV storage tubes and the double-gun vidicon. Most recently, he has been working on a cold cathode vidicon. He received an RCA Laboratories Outstanding Achievement Award and an IR100 Award in 1971, and is presently a member of the television systems research group at RCA Laboratories.

Dr. Luedicke is a member of Fernseh Technische Gesellschaft, Germany, and of SMPTE.

Reuben S. Mezrich received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from the Polytechnic Institute of Brooklyn. In 1961, while studying for the B.S., he received a National Science Foundation grant for research on materials capable of laser action. In 1963 he joined the RCA Laboratories and was associated with the cryoelectric devices group working on cryogenic logic elements. At the same time he was on the RCA Graduate Study program, which enabled him to fulfill the requirements for the master's degree. Dr. Mezrich received the RCA Laboratories Graduate Study award in 1964, allowing him to fulfill the course requirements for the Ph.D. in Electrical Engineering. Since his return to the Laboratories in 1965, Dr. Mezrich has been involved in research in holography and its applications to computer storage. These studies have included read-only memories, image storage and retrieval systems, and reversible memories. In 1968 he was co-recipient of an IR 100 award for the development of a holographic read-only memory. In 1969 he was the recipient of an IR 100 award for the development of holographic storage on magnetic films. He received an RCA Laboratories Outstanding Achievement Award in 1970 for his work on magnetic holography.

Dr. Mezrich is a member of Tau Beta Pi, Eta Kappa Nu and the IEEE.

Eugene M. Nagle graduated from RCA Institutes in 1952. From 1952 until 1956 was with Lavoie Laboratories, Morganville, N. J. developing communication and test equipment. From 1956 to 1960 he was with Stavid Engineering in Plainfield, N. J. working at Bell Telephone Laboratories, Murray Hill, N. J. on wideband transmission measurement equipment and circuit design for logic and memory equipment for large data processing systems. From 1960 until 1962 he worked for Electronic Associates Inc. at Long Branch, N. J. doing logic and memory circuit development for use in hybrid and digital computers. Mr. Nagle joined RCA Laboratories in 1962 where he has worked on the development and characterization of thin-film memory elements and development of large memory systems. Mr. Nagle was co-recipient of a David Sarnoff Outstanding Achievement Award in science in 1968 and is a member of IEEE.

Stefan A. Ochs received his Ph.D. Degree in Physics from Columbia University in 1953. He joined the RCA Laboratories in 1952 where he worked on television camera tubes, electrostatic signal storage, and solid-state devices. Since 1965 he has been a senior engineer at RCA Electronic Components where he has been active on development of camera tubes. Dr. Ochs is a member of the Physical Society and of Sigma Xi, and a senior member of the IEEE. He has been the recipient of two RCA Achievement Awards.

A. H. Sommer received the Dr. Phil. degree in Physical Chemistry from Berlin University in 1934. From 1935 to 1945, he worked with Baird Television Company, London, in the field of photoelectric and secondary emission. In 1946 he joined EMI Research Laboratories in Hayes, England, to work on photomultipliers and television camera tubes. In 1953 he joined RCA Laboratories where he is engaged in the study of photoemissive and secondary electron emitting materials. His work resulted in the development of two new photocathodes that have found widespread application. Since 1967 he has participated in the development of "negative electron affinity" materials that are used for photoelectric, secondary, and cold cathode electron emission.

Dr. Sommer is a Fellow of the Institute of Electrical and Electronics Engineers and a member of the Physical Society and of Sigma Xi.

Richard L. Spalding received the BSEE degree from Tri State College in 1948. His experience includes developmental and manufacturing engineering work in image converters and black-and-white picture tubes at Capehart Farnsworth, Fort Wayne, Indiana and electrometers and geiger counters at Victoreen Instrument Company, Cleveland, Ohio. He joined RCA in 1954 in color kinescope manufacturing and was associated with both the Lancaster and Scranton plants as an engineer and engineering manager until 1968. Design and developmental work in low cost vidicons and the Elvicon was undertaken until 1971. He is presently manager of Quality and Reliability Systems for the Industrial Tube Division.

Fred Sterzer received his B.S. degree in Physics from the College of the City of New York in 1951, and his M.S. and Ph.D. degrees in Physics from New York University in 1952 and 1955, respectively. From 1952 to 1953, he was employed by the Allied Control Corporation in New York. During 1953 and 1954, he was an instructor in Physics at the Newark College of Engineering, Newark N. J., and a research assistant at New York University. Dr. Sterzer joined RCA in 1954 and is now Director of the Microwave Technology Center of the RCA Laboratories in Princeton, N. J. His work has been in the field of microwave spectroscopy, microwave tubes, light modulators and demodulators, and microwave solid state devices.

Dr. Sterzer is a member of Phi Beta Kappa, Sigma Xi, the American Physical Society, and is a Fellow of the IEEE.

Wilber C. Stewart received the B.S., M.S., and Ph.D. degrees in electrical engineering at Duke University. Upon joining RCA Laboratories in 1964, his interests in superconducting devices included studies of field-induced switching phenomena, flux coupling in films, general inductance relations, signal propagation in large device arrays, and Josephson tunneling phenomena. His more recent activities in optical holography have included research on electrical and optical switching phenomena in ferroelectric ceramics, analyses of random phase data masks, and studies of the optical system requirements for holographic memories.

Dr. Stewart was the recipient of RCA Laboratories Outstanding Achievement Awards in 1967 and 1970. He is a member of the IEEE, Optical Society of America, Phi Beta Kappa, Tau Beta Pi, Eta Kappa Nu, Pi Mu Epsilon, and Sigma Xi.

Frank S. Wendt obtained the B.S. degree in physics from the University of Richmond, Virginia in 1949. In 1950 he became a member of the technical staff of the RCA Laboratories, Princeton, New Jersey where he spent three years in the acoustical research laboratory; part of this time being spent developing audio magnetic amplifiers. Mr. Wendt worked for the U. S. Government from 1953 to 1956 and in 1956 became a member of the Research Staff at General Motors Corporation developing electronic instrumentation for automated highways and automobile safety devices. After a short stay with an electrohydraulic servomechanism company, Mr. Wendt returned to RCA Laboratories in 1959 where he has designed and built instrumentation and hardware for many diversified projects, including seismic detection on the ocean floor, vacuum systems for thin-film evaporation, cryoelectric and holographic optical memories for computers, silicon storage tube circuitry, and low-noise amplifiers.

Mr. Wendt has held membership in the Acoustical Society of America and the Audio Engineering Society and is presently a member of the IEEE and Sigma Phi Sigma. In 1964 he was a co-recipient of an RCA Laboratories Outstanding Achievement Award for work on cryoelectric memories.

STATEMENT OF OWNERSHIP

Statement of Ownership, Management and Circulation (Act of October 23, 1962, Section 4369, Title 39, United States Code.)

1. Date of filing: October 9, 1972. 2. Title of publication: RCA REVIEW. 3. Frequency of issue: Quarterly. 4. Location of known office of publication: RCA Laboratories, Princeton, New Jersey, 08540. 5. Location of headquarters or general business offices of the publishers: RCA Laboratories, Princeton, New Jersey, 08540. 6. Names and addresses of Publisher and Editor: Publisher, RCA Laboratories, Princeton, New Jersey, 08540; Editor, R. F. Ciafone, RCA Laboratories, Princeton, New Jersey, 08540 7. Owner: RCA Corporation, 30 Rockefeller Plaza, New York, New York, 10020; Merrill, Lynch, Pierce, Fenner and Smith, Inc., 70 Pine St., New York, New York; Cudd and Co., c/o Chase Manhattan Bank, 1 Chase Manhattan Plaza, New York, New York; Martin B. Seretean, c/o Coronet Industries, Box 1248, Dalton, Georgia 30720; Burl Jackson Bandy, c/o Coronet Industries, Box 1248, Dalton, Georgia 30720.

I certify that the statements made by me above are correct and complete.

Ralph F. Ciafone, Editor