

RCA Review



Optical Storage and Display

March 1972 Volume 33 No. 1

RCARCI 33(1) 1-324 (1972)

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.

Special Issue

Optical Storage and Display Media

- 3** Foreword
Juan J. Amodè
- 5** Holographic Information Storage
E. G. Ramberg
- 54** Materials for Magneto-Optic Memories
R. W. Cohen and R. S. Mezrich
- 71** Holographic Recording in Lithium Niobate
J. J. Amodè and D. L. Staebler
- 94** Optical and Holographic Storage Properties of Transition Metal Doped Lithium Niobate
W. Phillips, J. J. Amodè, and D. L. Staebler
- 110** Phase Holograms in Dichromated Gelatin
D. Meyerhofer
- 131** Redundant Holograms
A. H. Firester, E. C. Fox, T. Gayeski, W. J. Hannan, and M. Lurie
- 154** Wavelength Dependent Distortion in Fraunhofer Holograms and Applications to RCA Holotape
R. A. Bartolini, D. Karlsons, and M. Lurie
- 170** Recording Considerations for RCA Holotape
R. A. Bartolini, J. Bordogna, and D. Karlsons
- 206** Thermoplastic Media for Holographic Recording
T. L. Credelle and F. W. Spong
- 227** Recyclable Holographic Storage Media
J. Bordogna, S. A. Keneman, and J. J. Amodè
- 248** Erase-Mode Recording Characteristics of Photochromic CaF_2 , SrTiO_3 , and CaTiO_3 Crystals
R. C. Duncan, Jr.
- 273** High Contrast, High Sensitivity Cathodochromic Sodalite for Storage and Display Applications
B. W. Faughnan and I. Shidlovsky
- 296** Liquid Crystals for Electro-Optical Application
J. A. Castellano
- 311** Technical Papers
- 313** Patents
- 317** Authors

RCA Corporation

Robert W. Sarnoff Chairman of the Board and Chief Executive Officer
A. L. Conrad President and Chief Operating Officer

Editorial Advisory Board

Chairman, J. A. Rajchman RCA Laboratories

E. D. Becken RCA Global Communications
G. H. Brown RCA Patents and Licensing
G. D. Cody RCA Laboratories
H. L. Cooke RCA Research and Engineering
A. N. Goldsmith Honorary Vice President, RCA
N. L. Gordon RCA Laboratories
G. B. Herzog RCA Laboratories
J. Hillier RCA Research and Engineering
E. O. Johnson RCA Patents and Licensing
H. W. Leverenz RCA Patents and Licensing
D. S. McCoy RCA Laboratories
H. F. Olson RCA Laboratories
K. H. Powers RCA Laboratories
P. Rappaport RCA Laboratories
L. A. Shottliff RCA International Licensing
T. O. Stanley RCA Laboratories
J. J. Tietjen RCA Laboratories
W. M. Webster RCA Laboratories

Secretary, Charles C. Foster RCA Laboratories

Editor Ralph F. Clafone

Associate Editors

W. A. Chisholm RCA Limited
M. G. Gander RCA Service Company
T. G. Greene Missile and Surface Radar Division
W. O. Hadlock RCA Research and Engineering
W. A. Howard National Broadcasting System
C. Hoyt Consumer Electronic Systems Division
E. McElwee RCA Solid-State Division
C. A. Meyer RCA Electronic Components
M. G. Pietz Defense Engineering
C. W. Sall RCA Laboratories
I. M. Seideman Astro-Electronics Division
A. H. Lind Commercial Electronic Systems Division

© RCA Corporation 1972 All Rights Reserved Printed in USA

Foreword

Optical storage and retrieval of information for the purpose of processing or display is an extremely old technique, particularly if one remembers that hand drawings, writings, and paintings represent some of its most basic early forms. This art received its first major boost with the invention and evolution of camera imaging in the 16th and 17th centuries, followed by the successful practical harnessing of the long known optical sensitivity of silver halides in the 19th century. The more than three hundred years that elapsed between the discovery of light-induced changes in absorption in silver salts, and the development of a process with sufficient sensitivity and fixing capabilities for commercial use, should stand as a reminder of the progress made in the rate at which new knowledge is exploited. Over the years, optical storage and display technologies continued to advance at an evolutionary pace, utilizing improved recording and imaging techniques with incoherent light.

Lasers have provided the next large impetus to this field, with holography, coherent processing, and fast optical deflection representing the most important techniques that became practical with the availability of coherent monochromatic light sources. The inherent capability for parallel access and for transmission and processing of large quantities of information is perhaps the most fundamental key to eventual success of coherent optical techniques. The degree to which these attributes can be harnessed economically depends on the state of development of the components and devices needed for the specific task. Thus, much of the research effort of the past few years has been focused on developing better lasers and the control and storage devices needed for efficient optical information handling.

This issue of RCA Review is devoted to optical storage and display media, a keystone for nearly all optical systems. The contents include discussions of theoretical considerations in information storage, a brief comparative survey of storage media, and detailed articles on a representative number of the most promising media for holographic and direct-image storage. It is hoped that assembling this information under one cover will help workers in the field get a broader view of the present state of the art and will stimulate progress towards harnessing the potentials of this new technology at an ever increasing rate.

Juan J. Amodei

Head, Quantum Electronics Research



Holographic Information Storage

E. G. Ramberg

RCA Laboratories, Princeton, N.J.

Abstract—This paper presents a survey of the different types of holograms, their distinguishing characteristics, and the physical processes employed in preparing them, referring to other papers in this issue of RCA Review for detail. Factors limiting the storage capacity of plane holograms and volume holograms are indicated and, in part, evaluated in semi-quantitative fashion.

1. Nature of Stored Information

Since the advent of the laser, stress has been placed on the ability of the hologram to store and reproduce three-dimensional information about objects and scenes. In the present treatment I shall depart from this emphasis and shall consider the information to be stored to be two-dimensional—e.g., in the form of a half-tone picture, a line drawing, or a geometric array of dots. The last form is most commonly considered in computer applications, with, e.g., a luminous dot at an integer coordinate position (i,j) of a “memory page” representing a logical 1 and its absence in the same location representing a logical 0. Information in this form is particularly well suited for a quantitative comparison of the storage capacity of different holographic media. On this basis, the storage capacity in bits of a holographic medium is

given simply by the product of the number of "memory pages" that can be stored and the number of dot positions (i,j) on every page. To make a comparison of different media valid, care must be taken, of course, that the same signal-to-noise ratio, or the same certainty of discrimination between a 0 and a 1, prevails for the compared media.

Among the factors besides storage capacity that affect the usefulness of a holographic storage medium for a given application are:

- Possibility and speed of erasure and rewriting of information
- Access time to a particular memory location
- Permanence of storage
- Sensitivity for recording images ("holographic sensitivity")
- Efficiency of utilization of reconstructing beam for forming image ("holographic efficiency").

These factors must be considered in the selection of holographic media for information storage. They do not, however, lend themselves to the establishment of a comparative figure of merit for different media.

2. Classification of Holograms

Holograms can be classified in a number of different ways, depending on their action on an incident light wave, their thickness, the method of recording and reconstruction, and the hologram medium. These criteria, and the resulting subdivisions, will now be discussed.

2.1 Effect of Hologram on Reconstructing Wave

(a) Hologram producing amplitude changes.

1. Amplitude changes produced by variation in absorption index (absorption hologram).
2. Amplitude changes produced by variations in reflectance or magnetization of reflecting medium (magnetic Kerr effect in reflection).

(b) Hologram producing phase changes (phase hologram).

1. Phase changes produced by variations in refractive index or magnetization (Faraday effect in transmission).
2. Phase changes produced by surface relief.

In 1(a)1 and 1(b)1 the interior of the holographic medium is responsible for the observed effects, whereas 1(a)2 and 1(b)2 represent surface effects. Phase holograms have the basic advantage over amplitude holograms of no (necessary) energy dissipation within the hologram and potentially higher holographic efficiency.

2.2 Hologram Thickness

a. Plane Hologram

Plane holograms can be regarded as a superposition of plane gratings differing in grating constant and orientation. Accordingly, plane holograms direct radiation into several different orders: (1) the zero order (undisturbed reconstructing beam); (2) the first order of diffraction (corresponding to a virtual image in the position of the object if the reconstructing beam is identical with the reference beam in recording and to a real image in the same location if it is identical with the reference beam reversed in direction); (3) a minus first order equal in intensity to the first order and corresponding to the "conjugate image"; and (4) in general also into any number of higher orders of decreasing intensity. These higher orders, which occur particularly at high hologram efficiencies, are light distributions bearing no resemblance to the object. All of these several orders also appear if the reconstructing beam deviates from the reference beam in direction, convergence, or wave-length. However, such deviations result in aberrations of the image, as well as in changes in size and location.

b. Volume Hologram

Volume holograms, on the other hand, can be regarded as superpositions of three-dimensional gratings (Fig. 1). As is well known from x-ray diffraction by crystals, diffraction by a three-dimensional grating is subject to the Bragg condition,

$$2d \sin \theta = n\lambda \quad [1]$$

where d is the spacing of a set of parallel grating planes (on which all scattering centers of the grating are distributed), θ is the glancing angle of the incident ray with respect to the grating planes, n is the order of diffraction, and λ is the wavelength of the diffracted radiation. The diffracted ray is in the direction of reflection at the grating planes for the incident ray; i.e., it forms a glancing angle $-\theta$ with the grating planes.

In a volume hologram, interference of radiation from a point of an (infinitely) distant object and a parallel reference beam (Fraunhofer or Fourier-transform hologram) produces a sequence of parallel planes of maximum field intensity that generates, through interaction with

the hologram medium, a similar set of planes with a maximum change in refractive and/or absorption index. The orientation of these planes is that of the bisecting plane for the angle 2θ between the directions of the reference beam and the particular object ray considered. The spacing d of the planes is given by Eq. [1] with $n = 1$. The Bragg

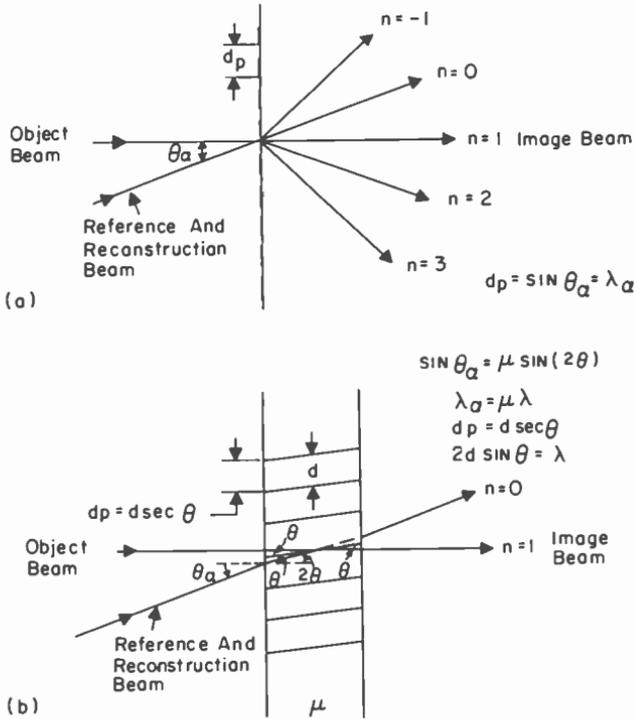


Fig. 1—Comparison of (a) plane and (b) volume holograms. The example of a "hologram grating" or Fraunhofer hologram of a point object is illustrated.

condition then tells us that a reconstructing ray (of the same wavelength λ) incident on the set of hologram planes considered is diffracted uniquely in the direction of the original object ray. Thus the volume hologram reconstructs a virtual image in the position of the object if the reconstructing beam coincides with the reference beam used in recording and a real image in the same location if the reconstructing beam is identical with the reversed reference beam. However, both the "conjugate image" and the higher-order diffractions are absent. Furthermore, with the exception of object points in specific

locations,* no light is diffracted, and image reconstruction does not take place if the direction of incidence of the reconstructing beam deviates from that of the reference beam. Similarly, a change in wavelength (without a change in the direction of the reconstructing beam relative to the reference beam) suppresses reconstruction of the image.**

These properties of the volume hologram make it possible to record in it a large number of images or pages in superposition, each read into it with a reference beam differing either in direction of incidence or in wave-length. These images can then be reconstructed without crosstalk by directing onto the volume hologram reconstructing beams with the same directions and wave-lengths as the reference beam. If the several reconstructing beams are directed at the hologram simultaneously, the corresponding images appear in superposition. This procedure has been used for color holography with white-light read-out.¹ The object (e.g., a color transparency) is illuminated by laser radiations corresponding to the three primary colors, and the transmitted or reflected radiation from the object is permitted to interfere in a volume hologram medium with three superposed reference beams derived from the same laser sources. If, then, a white-light reconstructing beam coinciding in direction with the reference beams is directed into the developed hologram, a virtual three-color image of the object is observed; only the spectral components of the white light that are used in recording contribute to this image. The remainder is not diffracted by the hologram.

The description of the behavior of the volume hologram here given is, of course, idealized. There is always a *range* of angle of incidence and a *range* of wavelength of the reconstructing beam for which image reconstruction takes place. These ranges decrease, and the properties

* The excepted object points, which are reconstructed by the rotated reconstructing beam, correspond to object rays on the cone containing the reference beam with axis normal to the plane bisecting the angle between reference beam and reconstructing beam. The reconstructed points are mirror images of these object points with respect to the bisecting plane. There is no diffraction or image reconstruction unless the solid angle of the object field overlaps the indicated object ray cone.

** Reconstructing wave-lengths, which are an integer fraction of the recording wavelength, will also lead to reconstruction if the variations in refractive and/or absorptive index recorded within the hologram medium are a nonlinear function of exposure. Since the visual spectrum embraces less than an octave, this is of no practical consequence. If both wavelength and direction of incidence are changed, again object points on a certain ray cone only are reconstructed. If, e.g. $\lambda'/\lambda = \cos \theta'$ (θ' is the angle of reconstructing beam with respect to reference beam), the object-point ray cone degenerates into the plane through the reference beam normal to the plane containing the reference beam and the reconstructing beam, and the image-point rays are displaced from this plane by the angle θ' .

of a hologram approach more closely those of an ideal volume hologram as the hologram thickness is increased. Both ranges also depend on the angle between the reference beam and the object beam. The internal angular range reaches a minimum when the angle between reference beam and object beam within the hologram approaches $\pi/2$ (as closely as permitted by refraction at the hologram surface),* and the wavelength range reaches a minimum when this angle approaches π (i.e., when object beam and reference beam are contragradient).²

In general, there is a gradual transition in the properties of a hologram from that of a plane hologram to that of a volume hologram as the medium thickness and the angle between object beam and reference beam are increased. Only surface-relief holograms or, more generally, the holograms listed under 1(a)2 and 1(b)2, are uniquely plane holograms. Klein³ has introduced the parameter Q ,

$$Q = \frac{2\pi\lambda_a T}{\mu d^2} \quad [2]$$

as criterion for distinguishing between a plane and volume hologram. Here λ_a is the wavelength in air, T the hologram thickness, μ the refractive index, and d the spacing between scattering planes. Klein notes that for $Q > 7$ or, if the scattering planes are parallel to the axis and θ_a is the external angle of incidence for Bragg reflection,

$$T > \frac{n\lambda_a}{4 \sin^2\theta_a}, \quad [3]$$

the hologram acts as a true volume hologram, being capable of over 95% holographic efficiency if the recording is in the form of a variation in refractive index (phase hologram). We might take as a criterion for a thin hologram that the conjugate image be no less than half as bright as the true reconstructed image. Applied to the case of an object and reference beam symmetrically placed with respect to the hologram normal, with angles of incidence $\pm\theta_a$, the intensity ratio becomes

* In a medium with refractive index $\mu = 1.5$ and with the object ray on-axis, the angular sensitivity referred to change in the external angle becomes a maximum for an angle of incidence of the reference beam of 48° ; the angular sensitivity for an angle of incidence approaching $\pi/2$ is reduced by the fact that, here, the change in internal angle θ_r corresponding to given change in external angle θ_i , $\Delta\theta_r = (\cos\theta_i/\cos\theta_r)\Delta\theta_i(1/\mu)$, becomes very small.

$$\frac{I_{-1}}{I_1} = \left(\frac{1}{T} \int_0^T \exp \left\{ j \frac{2\pi}{\lambda} t (\cos \theta - \cos \alpha) \right\} dt \right)^2$$

$$= \left[\frac{\sin \left(\frac{\pi T}{\lambda} (\cos \theta - \cos \alpha) \right)}{\frac{\pi T}{\lambda} (\cos \theta - \cos \alpha)} \right]^2 \quad [4]$$

Here θ is the Bragg angle (within the hologram medium of index μ) and α is the angle (within the medium) that the conjugate image beam makes with the hologram normal. For small θ , $\alpha \cong 3\theta$ and $\cos \theta - \cos \alpha \cong 4\theta^2$. Since $(\sin x)/x = 0.707$ for $x = 1.39$, the criterion of thinness, $I_{-1}/I_1 < 0.5$, takes the form

$$\frac{4\pi T\theta^2}{\lambda} < 1.39, T < \frac{\lambda}{9.0\theta^2} = \frac{\mu\lambda_a}{9\theta_a^2}, \text{ or } Q < 3. \quad [5]$$

2.3 Convergence of Object and Reference Beams with Respect to Hologram Medium

a. Fresnel Hologram. (Fig. 2)

A Fresnel hologram is recorded when the object is at some finite distance from the hologram medium and the reference source is at a different distance from it. Commonly the reference source is placed at infinity, i.e., a parallel reference beam is employed. For a single object point, the interference maxima then lie on paraboloids of revolution, with the object point as common focus. Their intersection with

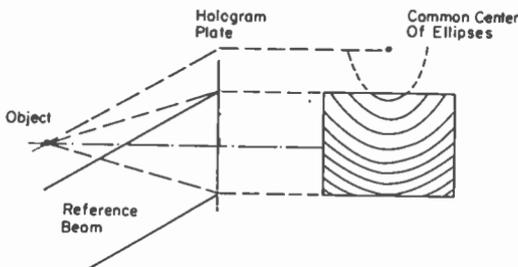


Fig. 2—Fresnel hologram of point object.

a plane at the hologram medium is a system of ellipses forming a zone plate pattern, with fringe spacing decreasing with increasing distance from the axis of the paraboloids, i.e., the line parallel to the reference beam passing through the object point in Fig. 2. If a Fresnel hologram, read out with a parallel reconstructing beam, is displaced, the image moves with it.

b. Fraunhofer Hologram* (Fig. 3).

A Fraunhofer hologram is formed when the (two-dimensional) object is placed in the focal plane of a lens interposed between the object and hologram and a parallel reference beam is employed. The interference

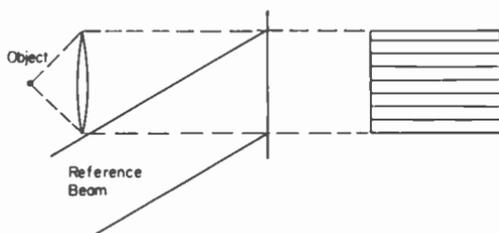


Fig. 3—Fraunhofer hologram of point object.

pattern for a single object point is now a system of parallel planes, intersecting a plane in the hologram medium as a line grating with sinusoidal intensity variation. Orientation and spacing of a particular hologram grating is uniquely correlated with the position of the object point relative to the reference-beam source, which is also located in the object plane if object beam and reference beam pass through the same (perfect) lens. If the hologram is displaced, the image does not move; with respect to the hologram, the effective object position is at infinity and its displacement (by the hologram displacement) is reduced to zero when the infinite object is reduced to finite size by the lens. The amplitude distribution at the image is, however, altered by a phase factor, with the phase difference increasing with angular displacement of the object point from the reference source.

* We adhere to the definition of a Fraunhofer hologram given by Leith and Upatnieks.⁴ Thompson et al⁵ use the same term for a hologram formed by the interference of a reference beam with light scattered from it by small objects located a distance $L \gg D^2\lambda$ from the hologram plane, placing the hologram in the far field of the diffraction pattern of the objects. D is here the diameter of the object.

b₁. "Lensless Fourier-Transform Hologram"* (Fig. 4)

Such a hologram is produced when the object and (point) reference source are equidistant from the hologram medium. The reinforcement surfaces of the interference pattern for a single object point are now hyperboloids of revolution with the object point and reference point

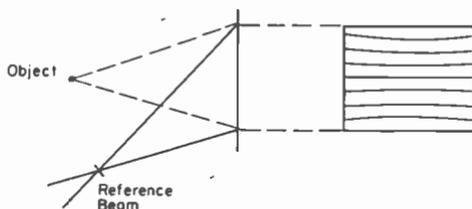


Fig. 4—"Lensless Fourier transform hologram" of point object.

as common foci. For small angular separation of these two points referred to the center of the hologram, this system of hyperboloids, at the hologram, can be closely approximated by a system of equidistant planes (as for the Fraunhofer hologram) and its intersection with the hologram plane by a sinusoidal line grating. This hologram is thus properly regarded as a modification of the Fraunhofer hologram and shares its properties to a first approximation. Motion of the hologram through the reconstructing beam leaves the image approximately immobilized, but alters the image phase by an amount proportional to the separation of image point and reconstructing source focus.

c. Fourier-Transform Hologram (Fig. 5).

The Fourier-transform hologram is a specialization of the Fraunhofer hologram realized when the distance between lens and hologram is made equal to the focal length of the lens. Then, apart from the effect of the aperture limitation of the lens, the amplitude distribution of the object beam in the hologram plane is the Fourier transform of the amplitude distribution in the object plane. The change in the reconstructed image with displacement of the hologram is the same for a Fourier-transform hologram as for any other Fraunhofer hologram; a displacement (Δx , Δy) of the hologram results, if the reconstructing beam coincides with the optical axis of the system, in a phase change

* This terminology was introduced by G. W. Stroke et al.⁶

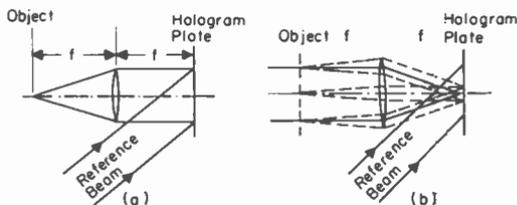


Fig. 5—Fourier transform hologram of (a) point object and (b) transparency illuminated by collimated beam. Amplitude distribution of object beam in hologram plane is Fourier transform of amplitude distribution in object plane.

$$\frac{2\pi}{\lambda} \frac{(x_o \Delta x + y_o \Delta y)}{f} \quad [6]$$

if the object point coordinates are (x_o, y_o) and f is the focal length of the lens, leaving the intensity distribution in the image unaltered.

d. Image hologram (Fig. 6).

An image hologram is formed by imaging the object on the hologram medium, which is simultaneously illuminated by a reference beam coherent with the object illumination. Such a hologram has properties intermediate between those of an ordinary photographic image and a conventional hologram. Viewed directly, the image is visible with low contrast and exhibits a fine grating structure. An image with normal contrast is obtained if the hologram is inserted in a Schlieren viewer. Here the hologram is illuminated by a crudely collimated beam of white light and is imaged onto a viewing screen in such fashion that the undiffracted light is intercepted by a mask. A phase image hologram thus has the advantage, over other phase holograms, of permit-

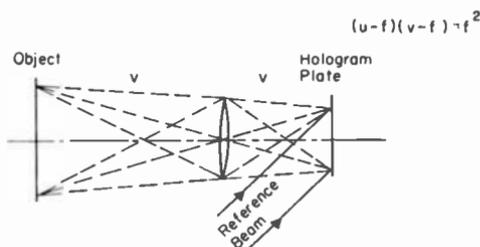


Fig. 6—Image hologram.

ting reconstruction with ordinary white light and shares with them the advantage of permitting optical projection with minimal thermal loading (no absorption required within the hologram frame!). It does not have the advantage of redundancy (i.e. diffusing the effect of hologram defects over an extended area of the image) nor that of image immobility with hologram displacement, which is characteristic of Fraunhofer holograms.

A phase image hologram can, in fact, be prepared without reference beam and without coherent radiation. It is only necessary to vary (e.g., by a molding or evaporation process) the thickness of the photosensitive material of the hologram plate, e.g., in proportion to

$$1 + k \cos \left(\frac{2\pi x}{d} \right), \quad d = \lambda / \sin \alpha_r \quad [7]$$

Exposure of such a plate to the intensity distribution $I(x,y)$ of the image and development such as to leave a transparent residue proportional to the thickness and the square root of the exposure ($\gamma = 0.5$) will result in a hologram imparting on the reconstructing beam a phase delay

$$\Delta\theta = \text{const.} \left(1 + k \cos \left(\frac{2\pi x}{d} \right) \right) \cdot |A_o(x,y)|, \quad A_o(x,y) = \sqrt{I(x,y)} \quad [8]$$

If, on the other hand, the hologram is prepared by imaging the object transparency illuminated with a plane wave with wavelength λ onto the hologram plate (with a flat photosensitive coating), exposing the same simultaneously to a parallel reference beam with an angle of incidence α_r and an amplitude A_r , the phase delay imparted by the developed hologram becomes,

$$\begin{aligned} \Delta\theta &= \text{const.} \left(I + |A_r|^2 + 2|A_r| |A_o| \cos \left(\frac{2\pi x}{d} \right) \right)^\gamma \\ &= \text{const.}' \left(1 + k' |A_o| \cos \left(\frac{2\pi x}{d} \right) + \dots \right) \end{aligned} \quad [9]$$

where $k' = 2 \left(\frac{\gamma |A_r|}{I + |A_r|^2} \right) \cong \frac{2\gamma}{|A_r|}$ for $I \ll |A_r|^2$.

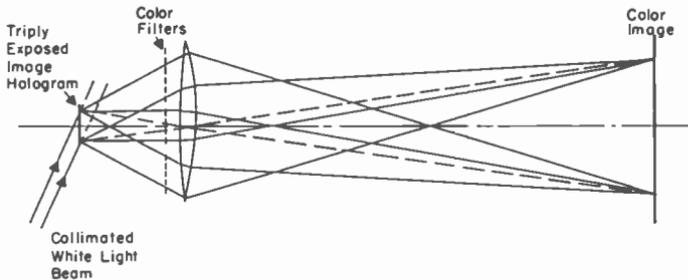


Fig. 7—Principle of reconstruction with white light of color image from plane image hologram.

In either case, the light diffracted into the first order (at an angle α , with the incident radiation) and hence contributing to the image is proportional to the object intensity $I(x,y) = |A_o|^2$; the object detail is assumed to be coarse in comparison with the grating constant d .

The grating pattern impressed on the plate emulsion need not, of course, have the contours prescribed by Eq. [7] to impart the desired properties to the "image hologram". Groove shape and depth affect only holographic efficiency.

On the other hand, the use of an image hologram for color reproduction, with three separation transparencies recorded in superposition with reference beams of different orientation and wavelength, requires conventional holographic methods using coherent radiation in recording. Playback can then be with white light (Fig. 7).

2.4 Relative Direction of Incidence of Reconstructing Beam and Emergence of Image-Forming Beam

a. Incidence and emergence on opposite sides of hologram (transmission hologram) (Fig. 8)

Such holograms are prepared by illumination of the hologram medium with the object beam and the reference beam from the same side and embrace both volume and plane holograms and, among the latter,

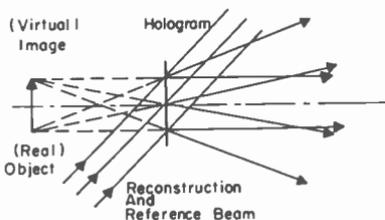


Fig. 8—Hologram with incidence of reconstruction beam and emergence of image beam on opposite sides.

surface relief holograms, holograms with internal refractive and absorptive index variations, and magnetic holograms utilizing the Faraday effect.

b. Incidence and emergence on same side of hologram

In the case of volume holograms, these reflection holograms are (Fig. 9) prepared by incidence of the object beam and reference beam on opposite sides of the medium. In them, the grating spacings are invariably close to $\lambda/2$ in magnitude, and the sensitivity-to-wavelength change is maximized.

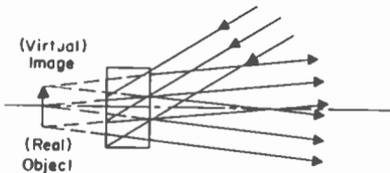


Fig. 9—Volume reflection hologram.

In plane holograms, emergence on the same side as incidence is realized through optical reflection at the surface of the hologram or a backing surface (Fig. 10). A relief hologram on a transparent substrate can serve either as a transmission or a reflection hologram. If the reflection at the back surface is suppressed (by a quarter-wave film or, in the case of the reflection hologram, an absorbing layer) and the depth modulation is substantially smaller than $\lambda/8$, we find that the ratio of the hologram efficiencies for the two types of play-back is

$$\frac{(\text{hologram eff.})_{\text{transm.}}}{(\text{hologram eff.})_{\text{refl.}}} = \frac{4n}{(n+1)^2} \left[\frac{(n-1)\delta t}{\lambda} \right]^2 = n. \quad [10]$$

The ratio of the holographic efficiencies is thus simply equal to the refractive index n of the hologram material, the effect of the larger phase change with reflection almost balancing out the ratio of the reflection coefficient to the transmission coefficient. If a thin metallic coating is deposited on the hologram surface, the holographic efficiency in reflection can, of course, greatly exceed the efficiency in transmission.

If reflection at the back surface is not suppressed, it may result in undesired interference effects in the image, for transmission as well as reflection holograms.* If a reflecting backing is employed to enhance the holographic efficiency of a reflection hologram, the entire hologram thickness T should fulfill a thinness criterion such as Eq. [5].

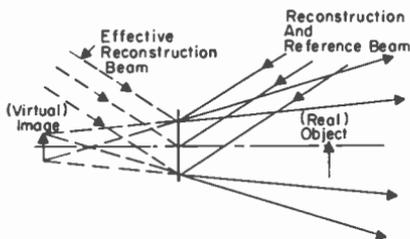


Fig. 10—Plane reflection hologram.

2.5 Object Illumination During Recording

a. Illumination by a monocentric beam.

A monocentric beam may, in particular, be a parallel beam, with the center at infinity. We assume the object to be a plane transparency. Then, unless the illuminating beam is focused at the hologram plane (as for a Fourier transform hologram of an object with parallel illumination), the object beam projects a shadow image at the hologram, with the object detail surrounded by diffraction fringes of a width proportional to the square root of the product of the wavelength and the object-hologram separation. As a consequence, the dynamic range of the object beam at the hologram is not far different from that at the object itself. Also the removal of a section of the hologram will prevent the reconstruction of a corresponding section of the image—there is essentially no redundancy. A further property of images reconstructed from holograms prepared with monocentric object illumination is the visibility of “cosmetic defects”, commonly in the form of circular interference fringes, resulting from dust specks and other defects on the object-illumination optics.

* See R. Bartolini, J. Bordogna, and D. Karlsons, “Recording Considerations for RCA Holotape”, in this issue of *RCA Review*.

b. Diffuse illumination.

The placing of a diffuser ahead of the object transparency overcomes these defects; the object beam at the hologram is now of macroscopically uniform intensity and there is practically unlimited redundancy—removal of any portion of the hologram has merely a minor effect on overall image resolution. Furthermore, cosmetic defects of the optics result, at most, in general contrast reduction. On the other hand, diffuse illumination introduces a serious deficiency in the image—speckle noise. As a result, any elementary area of the image with a linear dimension equal to or smaller than the diffraction-limited least-resolvable separation exhibits a mean-square intensity fluctuation equal to the mean intensity in that region of the image. For larger elements of area, the ratio of the mean intensity to the root-mean-square fluctuation in intensity increases in proportion to the linear dimension of the sample area. Thus the attainment of a reasonably good signal-to-noise ratio in the image demands a hologram area that is very large compared to that needed to provide the desired image resolution.

It may be noted that, with diffuse object illumination, the intensity distribution in the object beam at the hologram also exhibits speckle noise, with the quantity corresponding to the least resolvable separation given roughly by the wavelength divided by the angle at the hologram intercepted by the object diameter.

c. Polycentric illumination

The use of two-dimensional gratings in close juxtaposition to the object transparency or of aperture or lens-array plates ahead of it makes possible the distribution of the light from any one element of the object over different portions of the hologram, achieving both a large measure of redundancy and relative uniformity of the object-beam intensity near the hologram.⁷ At the same time, the grating or aperture plate introduces a regular spurious background pattern that can be rendered unobjectionable by making the pattern periods smaller than the smallest separation to be resolved. As with diffuse illumination, the visibility of cosmetic defects of the optics is greatly reduced.

With polycentric illumination, as with diffuse illumination, redundancy is purchased by the superposition of a spurious pattern unrelated to the transmission variation of the object transparency. With polycentric illumination, this pattern is regular; with diffuse illumination, it is random. In either case, the spurious pattern can be rendered

innocuous by increasing the hologram size or decreasing the density of information stored in the hologram. However, for a prescribed signal-to-noise ratio in the image, the "purchase price" in terms of increased hologram area can be made much less with polycentric illumination than with diffuse illumination.

For further discussion of holograms prepared with polycentric illumination we refer to the paper in this issue by Firester, Fox, Gayeski, Hannan, and Lurie.*

2.6 Physical Process Used to Record Hologram

a. Thermal evaporation.

Holograms have been prepared by the interference of intense ruby pulsed laser beams at solid metal surfaces (copper, silicon)^{7a} as well as on thin metal films (Au, Bi) on a glass substrate. Amodei and Mezrich⁸ found $50 \mu\text{J}/\text{cm}^2$ sufficient for recording on a 100-Å Bismuth film; resolution of 1000 lp/mm on film and holographic efficiencies of 6% were obtained. The holograms could be used for reconstruction by reflection or transmission.

b. Recording in silver halides (photographic plates or film)

High resolution photographic plates (particularly Eastman Kodak 649-F⁹ and, later, the more sensitive Agfa-Gevaert Scientia Plates, such as the 10E56 and 10E70¹⁰) are the classic materials for recording holograms. Since the emulsion thicknesses of these materials range from 15 μm to 6 μm , they form true plane holograms only for small angles between the object beam and reference beam (e.g., 10° or less). With object and reference beam incident from opposite sides, they form very effective volume holograms with high wavelength discrimination (provided that appropriate precautions are taken to minimize emulsion shrinkage).

Photographic plate holograms are normally primarily absorption holograms. However, bleaching the developed silver converts them into phase holograms, arising primarily from surface relief¹¹ or internal refractive index modulation¹² (depending on the bleaching process employed). Refractive index modulation, resulting from the replacement of the developed silver by silver salts with index higher than that of the gelatine, leads to the highest resolution in a photographic phase hologram.

* A. Firester, E. Fox, T. Gayeski, W. Hannan, and M. Lurie, "Redundant Holograms", in this issue of *RCA Review*.

c. Photoresists.

Photoresists are, in general, organic materials containing photosensitizers that, upon exposure to light, are either rendered soluble (positive photoresist) or insoluble (negative photoresist).¹³ They are washed after exposure, leaving a residue of insoluble material that, in the case of a holographic exposure, constitutes a surface-relief phase hologram. Negative photoresists are, of course, preferably exposed through a transparent substrate. Photoresists generally have very low sensitivity but are superior to bleached photographic plates with surface relief with respect to both resolution and reduction in light scattering.

d. Dichromated gelatine.

In addition to being a negative photoresist when soft, dichromated gelatine when hardened exhibits a refractive index change upon exposure that makes possible the recording of a volume hologram without development. This hologram may be fixed and the holographic sensitivity greatly increased by washing the material with isopropanol alcohol to remove unreduced dichromate and subsequent dehydration. The interesting properties of this material are described and interpreted in the paper by D. Meyerhofer in this issue of *RCA Review*.*

The photosensitivity of dichromated gelatine is much lower than that of photographic plates; furthermore, it is limited to relatively short-wavelength radiation. This limitation can be overcome by converting a hologram formed in a photographic emulsion (such as the 649F) with, e.g., helium-neon laser radiation at 6328 Å into a hardened-gelatine hologram.¹⁴ The procedure is to bleach the developed and fixed hologram plate, treat it with an ammonium dichromate solution containing a wetting agent, expose the dichromated plate to argon laser radiation (at 4880 Å) (preferably in the Bragg direction for the recorded hologram) or to radiation from a mercury arc, remove residual dichromate with bisulfite solution and residual silver salts by fixing, and develop the hardened gelatine emulsion in hot isopropanol followed by a hot-water wash. The resultant hologram is a permanent phase hologram with holographic efficiencies of 50% or more and high signal-to-noise level.

e. Photopolymers

Photopolymers produce permanent, irreversible phase holograms with high holographic efficiency. The initial material contains, as its es-

* D. Meyerhofer, "Phase Holograms in Dichromated Gelatine" this issue of *RCA Review*.

sential component, a relatively unstable mixture of a monomer and a generally dye-sensitized catalyst. The initial step is the holographic exposure at a wavelength strongly absorbed by the dye in the catalyst, which produces localized polymerization. The final step is fixation by flooding with short-wave radiation (e.g., radiation from a mercury arc) to polymerize the remainder of the monomer and to convert the dye into a stable nonabsorbing substance. The essential intermediate event is, presumably, the diffusion of monomer into the already polymerized regions, increasing their optical density after fixation. One process, which uses a ruby laser for the initial exposure of an aqueous solution of monomer and catalyst between glass plates some 25 μm apart, results in holograms in which surface relief contributes the major part of the holographic diffraction.¹⁵ In a second process,¹⁶ in which a mixture of monomer, catalyst, and polymeric binder is spread on a glass substrate and is initially exposed to ultraviolet (or, with dye sensitization, blue or green) recording radiation, the surface modulation is negligible at the frequencies above 100 lines/mm (i.e., at the frequencies of interest for holographic storage) and the hologram is truly a volume phase hologram. In both, while the holographic sensitivity is low, the entire processing can be carried out in situ.

f. Photoconductive-thermoplastic sandwich*

By a sequence of steps involving corona charging of a thermoplastic surface (with respect to a common ground electrode applied to the other side of the sandwich) and exposure to the light pattern to be recorded, a field depending on the light exposure is impressed across the thermoplastic layer. Heating (to some temperature between 60° and 100°C) softens the thermoplastic and causes it to flow from the more highly illuminated areas (with the stronger field) to the less illuminated areas, producing a surface relief that is frozen in position as the thermoplastic cools. In this manner, efficient plane holograms with minimal light scattering may be formed. Heating to a higher temperature or for a more prolonged time erases the hologram, making the material ready for reuse. Both thermal development and erasure can be carried out in situ by passing current through the base electrode. The fact that the frequency response of the thermoplastic hologram is peaked at a value depending on the thickness of the thermoplastic can be used to advantage.

* See T. Credelle and F. Spong, "Thermoplastic Media for Holographic Recording", in this issue of *RCA Review*.

g. Photochromic materials

Photochromic materials change their transmission spectrum in response to exposure to light. Examples are a variety of organic materials; alkali halide crystals, such as KBr; the photochromic glasses, containing a dispersion of silver halide crystals; and a long series of doped inorganic crystalline materials, among which CaF_2 , doped with the rare earths La or Ce, and SrTiO_3 doubly doped with Mo and Fe or Ni, are most prominent. They are, in general, darkened by exposure to short-wave visible or ultraviolet radiation and bleached by exposure to long-wave visible or infrared radiation, and are commonly practically unaffected by radiation at some intermediate wavelength. Hologram recording can take place either in the darkening or in the bleaching mode. Since the absorptive-index changes produced are rather small, photochromic holograms tend to be inherently (absorption) volume holograms. Since volume holograms must be read out at the recording wavelength, the reading process gradually wipes out the holographic recording by driving the darkening or bleaching process to saturation. On the other hand, recording, read-out, and erasure can take place in situ without development, and the material (with the exception of the organic photochromics) can be reused indefinitely. The properties of photochromics are discussed in more detail in the paper by R. Duncan.*

h. Electro-optic materials

Electro-optic materials are suitable for holographic recording if, by photoexcitation, local variations in the intensity of illumination result in corresponding variations in space-charge density. The electro-optic effect then converts accompanying field variations into a modulation of the refractive index. Since these index changes are small, the holographic material must be thick to diffract the reconstructing wave efficiently. Electro-optic materials thus form phase volume holograms. As with photochromic materials, the read-out process, necessarily carried out with radiation of the wavelength used in recording, gradually erases the hologram. With appropriate polarity of the crystal, this erasure process may be retarded, however, by an enhancement process consisting of the rerecording of the hologram through the interaction of the reconstructing beam with the image beam generated by it. Furthermore, in pure lithium niobate (LiNbO_3), in

* See R. Duncan, "Erase-Mode Recording Characteristics of Photochromic CaF_2 , SrTiO_3 , and CaTiO_3 Crystals", in this issue of *RCA Review*.

particular, it has been found possible to fix the hologram by a heat treatment that cancels the internal field produced by trapped electron concentrations by the displacement of impurity ions. Apart from lithium niobate (pure or doped with Fe to achieve greatly increased holographic sensitivity), many other materials, such as $\text{Ba}_2\text{NaNb}_5\text{O}_{15}$, LiTaO_3 , BaTiO_3 , and $\text{Bi}_4\text{Ti}_3\text{O}_{16}$, have been shown capable of recording volume phase holograms. The properties of such materials are discussed in papers by J. Amodei, D. L. Staebler, and W. Phillips in this issue of RCA Review.*

i. Magneto-optic Materials

If a film of MnBi, uniformly magnetized in a direction normal to film, is exposed to an interference pattern produced by pulsed laser radiation of sufficient intensity to raise the temperature at the antinodes of the pattern above the Curie point (360°C), the demagnetizing field from the remainder of the field produces a reversal of magnetization at these points. In this manner holograms can be recorded on the film and can be read out by the transmission or reflection of a reconstructing beam. In transmission, the local fields in the film rotate the plane of polarization by the Faraday effect, producing a first-order reconstruction with a light-field component polarized at right angles to the incident wave. In reflection, the magnetic Kerr effect similarly produces a rotation of the plane of polarization of the reflected wave, also resulting in a reconstruction polarized at right angles to the incident wave. Since the films are very thin, the magnetized film acts like a plane (phase) hologram in transmission and like a plane (amplitude) hologram in reflection. Magnetic holograms have also been recorded on EuO films, with a Curie point at 69°K , obviously requiring cryogenic cooling to prevent erasure of the hologram.¹⁷

Magnetic holograms can be erased thermally, permitting replacement of the record with a new record. Once recorded, MnBi holograms have indefinite shelf life. Apart from low holographic sensitivity and low efficiency in reconstruction, they have thus the most important properties desired for holographic storage with plane holograms. They are discussed in detail in the paper by Mezrich and Cohen in this issue.†

* J. Amodei and D. Staebler, "Holographic Recording in Lithium Niobate", and W. Phillips, J. Amodei and D. Staebler, "Optical and Holographic Storage Properties of Transition Metal Doped Lithium Niobate", in this issue of *RCA Review*.

† R. Mezrich and R. Cohen, "Magneto-Optic Recording Materials", in this issue.

j. Recording by phase transformation

In some materials, such as a thin VO_2 film sputtered on a glass substrate, phase transformations involving substantial changes in absorptive and refractive index can be produced locally (in volumes of the order of $1 \mu\text{m}^3$) by a momentary rise in temperature, such as may be produced by a pulsed-laser interference pattern.¹⁸ If the substrate is maintained at a temperature close to the transition temperature (63°C for VO_2), a hologram recorded in this manner is relatively permanent owing to the hysteresis of the transformation process; it may be erased by a momentary decrease in temperature. Since the primary optical effect of the (insulator-to-metal) transformation in VO_2 is a change in reflectivity, the VO_2 hologram is most efficient as a plane reflection hologram. It shares with the materials listed in section 2.6(e) to 2.6(g) the desirable property of permitting recording, read-out, and erasure in situ.

Much higher holographic efficiencies (up to 80% for $10 \mu\text{m}$ films read out at 6328 \AA) have been achieved with evaporated As_2S_3 films.^{18a} The phase transformation, which has not been uniquely identified, is accompanied by a 200 \AA absorption-edge shift and induced by exposure to argon laser radiation of wavelengths 4880 \AA or 5145 \AA . The record is erased by thermal cycling between 100°C and room temperature. The process appears related to holographic storage in arsenic-sulfur glasses with high sulfur content, observed in much thicker layers (e.g., 0.045 cm) and tentatively attributed to a devitrification process.¹⁹

k. Photoactivated Liquid Crystals*

Nematic liquid crystals exhibiting the dynamic scattering mode (DSM) and operating below the DSM threshold, as well as liquid crystals showing alignment effects, change their effective refractive index in response to applied electric fields. Thus a sandwich of a photoconductive layer of, for example, CdS of about $5 \mu\text{m}$ and a liquid crystal layer of $13 \mu\text{m}$ in a cell between transparent electrodes constitutes a sensitive system for producing transient holograms with limited angular range, the resolution being only about 100 lines/mm.²⁰ Radiant inputs of the order of $1 \mu\text{J}/\text{cm}^2$ suffice to create a hologram that has a decay time of the order of a second.

* For the optical properties of liquid crystals, see J. Castellano, "Liquid Crystals for Electro-optical Applications", in this issue of *RCA Review*.

1. Modulation of Color Center Alignment

Alkali halide crystals, such as KCl, doped with cations of smaller diameter, such as Na, may be discolored, for example, by heating in potassium vapor so as to form F -centers, or anion vacancies containing an electron. They may then be partly bleached with F -center-absorbed radiation, with the resultant formation of F'_a centers, or the combination of an electron in a vacancy with a foreign ion. These F'_a centers are randomly aligned with the crystal axes. When the electric vector coincides with their principal axis, they are excited by a lower frequency (e.g., 6328 Å light) than when it is normal thereto. Such excitation may result in ionic reorientation and a consequent reduction in absorption coefficient at the wavelength in question. Thus amplitude volume holograms may be recorded in the crystals, and when the crystal is kept at 77°K the record is permanent. Since 15 mJ/cm² were found to produce a density change of 0.4 in a crystal 0.2-cm thick, their sensitivity is about four orders of magnitude lower than that of photographic plates, but some two orders higher than that of KCl storing holograms by F -center formation.²¹

3. Holographic Efficiency

The holographic efficiency is defined as the ratio of the radiant flux diffracted into the reconstructed image beam to the incident flux of the reconstructing beam. In comparing the performance of different hologram materials, it is common practice to assume the simplest type of hologram—a hologram grating formed by the interference of a plane object beam and a plane reference beam or a Fraunhofer (or Fourier-transform) hologram of a point source. For a plane hologram formed by an object beam with angle of incidence θ_o and reference beam with angle of incidence θ_r , both in the xz -plane, the interference pattern is a plane grating with grating constant d ,

$$d = \frac{\lambda}{\sin \theta_r - \sin \theta_o} \quad [11]$$

If the hologram material has a photographic gamma γ and a modulation transfer function $M(l/d)$, an angle between the electric vectors of the object wave and reference wave of θ_p , and a ratio of reference-beam intensity to object-beam intensity at the hologram of R , the amplitude transmission of an absorption hologram can be written

$$T_A = T_{A0} \left(1 - m \cos \frac{2\pi x}{d} + \frac{\gamma + 2}{2\gamma} m^2 \cos^2 \left(\frac{2\pi x}{d} \right) \dots \right). \quad [12]$$

Here T_{A0} is the amplitude transmission in the absence of coherence between the object beam and the reference beam and

$$m = \frac{\gamma MR}{R + 1} \cos \theta_p \quad [13]$$

For $\gamma = -2$, only the first two terms in Eq. [12] are different from zero, and the three Fourier terms corresponding to the amplitude transmission expression give rise to the zero-order wave and two first-order diffracted waves of equal amplitude $m/2$. The maximum value of m is then realized for $\theta_p = 0^\circ$ (electric field normal to plane defined by the two wave normals), $M = 1$ (100% modulation transfer), and $R = 1$ (object and reference beam of equal intensity), and is given by $m = 1$. Furthermore, since $T_A \leq 1$, $T_{A0} \leq 1/(1 + m)$. Hence we find for the ratio of the amplitudes A_1 and A_0 of the first-order beam and the incident beam in reconstruction

$$\frac{A_1}{A_0} \leq \frac{m}{2(1 + m)} \leq \frac{1}{4}, \quad \eta = \frac{I_1}{I_0} \leq \frac{1}{16} \quad [14]$$

$\eta = 0.0625$ is thus the upper limit of the holographic efficiency of a plane amplitude hologram with a linear relationship between exposure and amplitude transmission. For small m , the relative diffracted-wave amplitude is quite generally $m/2$ (with m given by Eq. [13]), even if the response is nonlinear.

The extreme case of nonlinearity is the binary hologram, for which the amplitude transmission is T_{Amax} if the exposure exceeds some threshold value and is T_{Amin} if the exposure is less than this threshold value. If the exposure exceeds the threshold value a fraction p of the time (for the single-point Fraunhofer hologram), the hologram amplitude transmission is given by

$$T_A = T_{Amin} + (T_{Amax} - T_{Amin}) p \left\{ 1 + 2 \sum_{n=1}^{\infty} \frac{\sin(n\pi p)}{n\pi p} \cos \left(\frac{2\pi n x}{d} \right) \right\}. \quad [15]$$

The holographic efficiency in this case is

$$\frac{I_1}{I_0} = \left\{ (T_{A_{\max}} - T_{A_{\min}}) \frac{\sin(\pi p)}{\pi} \right\}^2. \quad [16]$$

This is clearly maximized for $T_{A_{\max}} - T_{A_{\min}} = 1$, $p = 1/2$, so that, for a binary amplitude hologram,

$$\eta = \frac{I_1}{I_0} \leq \frac{1}{\pi^2} = 0.1013. \quad [17]$$

The maximum holographic efficiency is realized when the threshold exposure equals the average exposure.

For a plane phase hologram, we obtain for the Fraunhofer hologram of a single object point

$$T_A = e^{i\phi}, \phi = \phi_0 \left(1 - m \cos \frac{2\pi x}{d} + \frac{\gamma + 2}{2\gamma} m^2 \cos^2 \left(\frac{2\pi x}{d} \right) \dots \right), \quad [18]$$

where m is again given by Eq. [13]. For $\gamma = -2$, only the first two terms in the expansion of ϕ differ from zero, and an expansion of T_A as a Fourier series in $(2\pi x/d)$ leads to a holographic efficiency

$$\eta = \frac{I_1}{I_0} = [J_1(m\phi_0)]^2 \leq 0.338, \quad [19]$$

the maximum efficiency being attained for $m\phi_0 = 1.84$. ϕ_0 increases, within limits, with exposure. For small values of m , the first-order diffraction efficiency is given by Eq. [19] quite generally; also for $\gamma \neq -2$. On the other hand, higher-order beams, with relative amplitudes given by Bessel functions of corresponding order, are generated even for $\gamma = -2$.

For a binary plane phase hologram grating we obtain, for phase step $\Delta\phi$,

$$\eta = \frac{I_1}{I_0} = \frac{4}{\pi^2} \sin^2(\pi p) \sin^2\left(\frac{\Delta\phi}{2}\right) \leq \frac{4}{\pi^2} = 0.4053. \quad [20]$$

The holographic efficiency for volume holograms with a sinusoidal variation of the absorption coefficient and index of refraction (i.e. for

idealized Fraunhofer holograms of a single object point) have been derived in greatest generality by H. Kogelnik,²² using the coupled-wave theory. In a volume hologram, it is necessary to consider the diffraction by the hologram grating of the image wave, as well as that of the reconstructing wave. Fortunately, for a sufficiently thick hologram, all of the scattering of the image wave can be regarded as serving to enhance the reconstructing wave, just as the scattering of the reconstructing wave can be regarded as contributing exclusively to the image wave. Thus, with suitable assumptions concerning the smallness of the changes in refractive index (n_1) and absorption constant (α_1)

$$n = n_o + n_1 \cos \frac{2\pi}{d} (x \sin \phi + z \cos \phi) \quad [21]$$

$$\alpha = \alpha_o + \alpha_1 \cos \frac{2\pi}{d} [x \sin \phi + z \cos \phi].$$

Maxwell's equations lead to a simple pair of coupled differential equations for the field amplitudes of the reconstructing and image waves; ϕ here designates the angle which the grating plane normal makes with the hologram normal (the z -axis).

Kogelnik derives the following expressions for the holographic efficiency η for the special case $\phi = \pi/2$ (object and reference beam symmetrical with respect to yz -plane):

Absorptive transmission hologram:

$$\eta = \exp \left\{ -\frac{2\alpha_o t}{\cos \theta_o} \right\} \sinh^2 \left(\frac{\alpha_1 t}{2 \cos \theta_o} \right) \quad [22]$$

$$\eta_{\max} = \frac{1}{27} = 0.037 \text{ for } \frac{\alpha_1 t}{\cos \theta_o} = \ln 3 \text{ and } \alpha_1 = \alpha_o \quad [23]$$

Absorptive reflection hologram:

$$\eta = \frac{\alpha_1^2}{\left[2\alpha_o + \sqrt{4\alpha_o^2 - \alpha_1^2} \coth \left(\frac{t}{2 \cos \theta_o} \sqrt{4\alpha_o^2 - \alpha_1^2} \right) \right]^2} \quad [24]$$

$$\eta_{\max} = \frac{1}{(2 + \sqrt{3})^2} = 0.072 \text{ for } \frac{\alpha_1 t}{\cos \theta_o} \rightarrow \infty \text{ and } \alpha_1 = \alpha_o \quad [25]$$

Phase transmission hologram:

$$\eta = \sin^2 \left(\frac{\pi n_1 t}{\lambda \cos \theta_o} \right) \quad [26]$$

$$\eta_{\max} = 1 \text{ for } \frac{2n_1 t}{\lambda \cos \theta_o} = 1, 3, 5, \dots \text{ (odd integer)} \quad [27]$$

Phase reflection hologram:

$$\eta = \tanh^2 \left(\frac{\pi n_1 t}{\lambda \cos \theta_o} \right) \quad [28]$$

$$\eta_{\max} = 1 \text{ for } \frac{2n_1 t}{\lambda \cos \theta_o} \rightarrow \infty \quad [29]$$

Here t is the thickness of the hologram, $2\alpha_o$ is the average absorption coefficient, λ is the wavelength of the radiation within the medium, θ_o is the glancing angle of the object and reference rays with respect to the plane of constant refractive index or absorption constant. The derivation requires that the hologram be very thick ($Q \gg 7$, with Q defined by Eq. [2]) and that $n_1 \ll n_o$, $\alpha_1 \ll 2\pi n_o/\lambda$, $\alpha_o \ll 2\pi n_o/\lambda$. Surface reflection effects, which could be nullified by the application of thin films of appropriate refractive index and thickness, are left out of account. Table 1 gives a summary of the upper limits to the holographic efficiencies of different types of holograms.

Table 1—Upper Limits to Holographic Efficiency

<u>Plane Holograms</u>				<u>Volume Holograms</u>			
<u>Amplitude</u>		<u>Phase</u>		<u>Absorption Modul.</u>	<u>Refractive Index Modul.</u>	<u>Absorption Modul.</u>	<u>Refractive Index Modul.</u>
<u>Linear</u>	<u>Binary</u>	<u>Linear</u>	<u>Binary</u>	<u>Transmission</u>	<u>Reflection</u>	<u>Transm.</u>	<u>Reflect.</u>
0.0625	0.1013	0.338	0.405	0.037	0.072	1.00	1.00

The indicated theoretical upper limits to hologram efficiency have been quite closely approached in practice with the simplest type of hologram—the hologram grating or Fraunhofer hologram for a single object point. However, the upper limit to the holographic efficiency of volume phase holograms may be shown to be 1 for complex objects as well. In particular, the reflection volume hologram with pure refractive-index modulation (and zero absorption) should have 100%

holographic efficiency for any picture content, provided only that the hologram is sufficiently thick.

It should also be emphasized that the various types of holograms listed in the table exhibit basic differences with respect to storage capability apart from their efficiency. We have already noted the advantage of volume holograms (as compared with plane holograms) resulting from their angular and wavelength sensitivity and the suppression of the conjugate image and higher-order diffractions. We should also note that, with binary plane holograms, the inoperativeness of the superposition principle affects the faithfulness in the reproduction of a complex image and that here the production of undesired diffracted beams is an even more serious problem than with the linear plane hologram.

4. Holographic Sensitivity

It is reasonable to define holographic sensitivity as the reciprocal of the hologram exposure required to produce a hologram with a prescribed efficiency. For any given mean exposure E_o the hologram efficiency η is proportional within a certain range—the “linear” range for which there is proportionality in the intensity of object and image—to the square of the fringe modulation or “visibility” V , defined by

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \quad [30]$$

For a simple interference pattern between two coherent parallel beams with an angle θ_p between their electric vectors,

$$V = \frac{2\sqrt{R}}{1 + R} \cos \theta_p, \quad R = \frac{I_r}{I_o}, \quad [31]$$

where I_r and I_o are the intensities at the hologram of the reference and object waves, respectively.

Lin²³ has appropriately defined the holographic sensitivity S of a hologram medium by the quantity

$$S = \frac{1}{E_o} \left[\frac{\partial \sqrt{\eta}}{\partial V} \right]_{E_o}. \quad [32]$$

Here, the derivative is taken in the linear range of $\sqrt{\eta}$ as a function of V . For media without an exposure threshold and an initial gamma equal to -2 , S is a constant for small exposures. For media with a threshold, such as photographic materials, S will be maximized for a certain exposure, which may be not too far from that yielding maximum linear range.

Through the holographic efficiency η , the holographic sensitivity S is also proportional to the modulation transfer function of the medium as applied to the space frequency of the fringe system, which in turn depends on the angle between the interfering beams.

5. Storage Capacity of a Hologram

It is possible to describe the information-storage capacity of a hologram by the general Hartley-Tuller-Shannon expression for the information capacity of a communication channel,²⁴

$$C = 2WT \log_2 \sqrt{1 + \frac{P_s}{P_n}} \quad [33]$$

Here, W is the frequency width of the communication channel and T is the transmission time; thus, $2WT$ is the number of independent samples that define the signal. P_s is the mean signal power and P_n the mean noise power.

If, as with a hologram, the signal transmission consists in the recording and reconstruction of a two-dimensional pattern, the expression for the storage capacity can be rewritten

$$C = n \log_2 \sqrt{1 + \left(\frac{S}{N}\right)^2} \quad [34]$$

where n is the number of picture elements and S/N the signal-to-noise ratio of the reconstruction of the individual picture elements.

The logarithmic term can be regarded as the logarithm of the number of steps of intensity that can be distinguished for each picture element.

In general, the number n is more readily determined than the signal-to-noise ratio S/N . In computer applications in particular, where the patterns normally are luminous dots at locations in a regular geometric array (presence of dot = 1, absence of dot = 0), n becomes simply the number of dot locations in an array or, if several

arrays are recorded in superposition, the product of the number of arrays and the number of dot locations in each array. Furthermore, the storage capacity in bits becomes at most equal to n ; the signal-to-noise ratio must simply be chosen high enough that the probability of error in the recovery of the information is less than some prescribed limit.

The simultaneous storage of many images or dot arrays is normally practiced only in volume holograms. It is thus appropriate to consider single-image (or dot-array) storage in a plane hologram and multiple-image storage in volume holograms separately.

5.1 Storage Capacity of a Plane Hologram

Consider a square area A on a photographic plate on which a square array of dots and vacancies is recorded. To distinguish dots and vacancies, we make the center-to-center separation of array locations equal to the limit of resolution of the plate, d_m . We then have the storage capacity of the plate for photographic imaging,

$$C = \frac{A}{d_m^2}. \quad [35]$$

As an alternative, we could obtain a similar array of dots by reconstructing a point-source array of C dots spaced d_{\min} apart with a hologram having a diffraction-limited resolution d_{\min} . Let us assume that we accomplish this with a square Fraunhofer hologram of area A_h , with the reference source adjoining the center of one of the sides of the object array (Fig. 11). If f is the focal length of the Fraunhofer lens, diffraction theory tells us that

$$d_{\min} = \frac{\lambda f}{\sqrt{A_h}}. \quad [36]$$

At the same time, the resolution limit of the hologram plate must be at most equal to the grating spacing corresponding to the object point farthest removed from the reference source:

$$d_m = \frac{\lambda}{\sin \theta_{\max}} = \frac{f\lambda}{\sqrt{\frac{4}{5} C d_{\min}^2}}. \quad [37]$$

Substitution of Eqs. [35] and [36] leads to the conclusion,

$$A_h = \frac{5}{4}A. \quad [38]$$

This rather crude analysis leads to the unquestionably correct conclusion that, with the identical recording material, the hologram area required to record a pattern with a given number of picture elements is somewhat larger than the area needed to record the pattern directly.

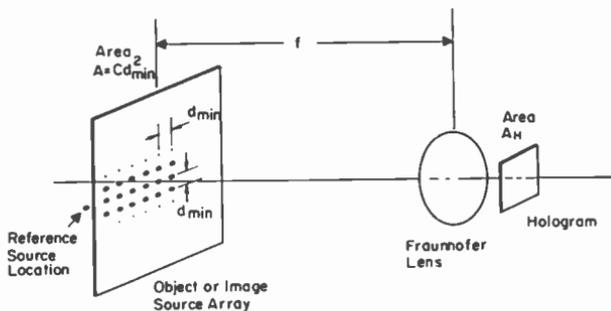


Fig. 11—Derivation of storage capacity of plane hologram.

A somewhat more convincing argument* would have led to the greater ratio of areas

$$\frac{A_h}{A} = \frac{5}{2}. \quad [39]$$

In practice, however, a reference beam offset just to the edge of the image field, as shown in Fig. 11, is almost never employed. The reason is that, with it, spurious intermodulation terms disturb the image. Consider the ideal case of an amplitude hologram, processed for $\gamma = -2$, recording a field on N point sources of equal amplitude A_0 and arbitrary phase φ_n . If the reference beam amplitude is A_r , the amplitude transmission of the hologram can be written

* This argument is based on requiring the same upper limits for the spatial spectrum of the direct image and of the hologram. The same result is obtained by the preceding argument if the array is oriented diagonally.

$$\begin{aligned}
T_A &= c \frac{|A_r + A_o \sum_{n=1}^N \exp \{j[k_n \cdot r + \varphi_n]\}|^2}{|A_r + A_o \sum_{n=1}^N \exp \{j[k_n \cdot r + \varphi_n]\}|^2} \quad [40] \\
&= c \left\{ 1 + \frac{A_r A_o}{A_r^2 + N A_o^2} \sum_{n=1}^N \left[\exp \{j[k_n \cdot r + \varphi_n]\} \right. \right. \\
&\quad \left. \left. + \exp \{-j[k_n \cdot r + \varphi_n]\} \right] + \frac{A_o^2}{A_r^2 + N A_o^2} \sum_{n=1}^N \sum_{m=1}^{n-1} \right. \\
&\quad \left. \left[\exp \{j[(k_n - k_m) \cdot r + \varphi_n - \varphi_m]\} \right. \right. \\
&\quad \left. \left. + \exp \{-j[(k_n - k_m) \cdot r + \varphi_n - \varphi_m]\} \right] \right\}. \quad [41]
\end{aligned}$$

Here k_n is the propagation vector corresponding to source n , with the components $(2\pi/\lambda) (x_n/f, y_n/f)$ and r is the position vector with components (x, y) . The last term indicates that, in the image, there will be spurious intensity contributions at the image points $\pm (x_n - x_m, y_n - y_m)$. These spurious intensity contributions can be quite large. Consider, e.g., a pattern made up of N dots located so that, pairwise, the vector displacement between them has the common value (x_k, y_k) . The intensity at the proper dot positions is then

$$I_o' = c' \frac{I_r I_o}{(I_r + N I_o)^2}, \quad [42]$$

whereas the expected value of the intensity at the vacant position (x_k, y_k) becomes

$$\bar{I}_k' = c' \frac{I_o^2}{(I_r + N I_o)^2} \left(\sum_{n=1}^{N/2} \exp \{j\varphi_n\} \right)^2 = c' \frac{N I_o^2}{2(I_r + N I_o)^2}. \quad [43]$$

The ratio of the expected value of the spurious intensity to the intensity at a recorded dot is here (for random phases φ_n)

$$\frac{\bar{I}_k'}{I_o'} = \frac{N I_o}{2 I_r} = \frac{R}{2}, \quad [44]$$

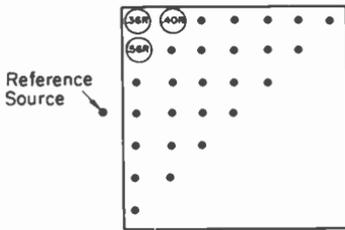


Fig. 12—Intensities at vacancies for a particular dot array with adjoining reference source. The ringed figures at the vacancies in the upper left-hand corner indicate the most probable values of the ratio of the intensity at the location in question to the intensity at dot positions.

or half the ratio of object beam intensity to the reference beam intensity. Fig. 12 illustrates the most probable intensities at the vacancies of a particular dot pattern.

More generally, Eq. [41] shows that the intermodulation terms extend over a square (centered on the reference source) with twice the side of the image field. Thus, to prevent overlap with the image field, the reference source is displaced a distance at least $3/2$ times the width of the image field from the center of the latter (Fig. 13). For an amplitude hologram not processed with $\gamma = -2$, and for phase holograms generally, the second-order diffraction terms extend over an area equal to that of the intermodulation terms beyond the image field (from the reference source) and the intensities of the intermodulation terms are modified. Third-order terms (in the power expansion of the amplitude transmittance) invariably lead to image distortions; these can however be minimized by making the (object

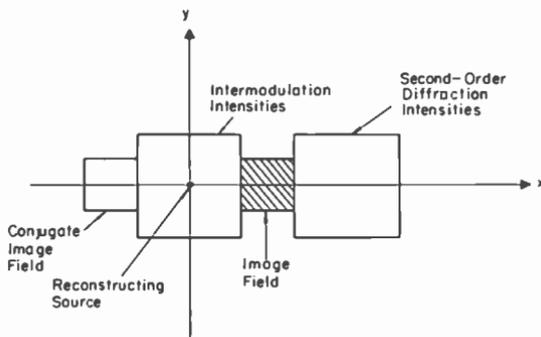


Fig. 13—Minimum spacing between image field and reconstructing source required to prevent image distortion by intermodulation.

beam)/(reference beam) intensity ratio R sufficiently small. It may be noted that a reference-source offset large compared to the image-field dimensions becomes mandatory, even apart from intermodulation term considerations, if the recording medium (such as thermoplastics) has a response range less in width than the space frequency of peak response. With the larger offset of the reference source ($3/2$ the image-field width), the requirement of an equal spatial spectrum range for the hologram and direct recording leads to an area ratio

$$\frac{A_h}{A} = \frac{17}{2}, \quad [45]$$

We thus come to the conclusion that we require approximately eight times the recording surface for recording a pattern of prescribed information content on a hologram as for direct photographic recording.

Clearly, holographic recording in plane holograms does not present an advantage over direct recording if maximum density of information storage is the only consideration. It presents a decided advantage, however, if certainty of full recovery of the information stored, from a densely stored record, is of primary importance. Furthermore, it can be far superior to ordinary photographic recording in faithfully recording a very wide dynamic range of light values on recording media with limited latitude.

With dense photographic information storage the settling of dust particles on the record or accidental scratches experienced in handling can completely obliterate important information. This hazard can be removed or diminished by introducing "redundancy" (e.g., by preparing duplicate records and viewing or sensing them, in some manner, in superposition). With holograms, redundancy is achieved more simply and more completely.

Not all types of holograms are equal in this respect. A hologram prepared with diffuse illumination of the object (diffuse hologram) possesses both the property of unlimited redundancy and that of recording a wide dynamic range on a material with small latitude. Damage to any part of the hologram affects all portions of the image equally, producing slight reductions in image brightness, contrast, and resolution. This applies irrespective of the light distribution at the object. If the light is concentrated at just a few object points, the dynamic range at the object and image will be enormous, while, at the same time, the light modulation at the hologram will be no greater than for

a flat object field with the same total luminous content. The clipping of the high peaks in the superposed interference patterns results, to a first approximation, simply in directing some of the reconstructing light into higher-order diffractions, outside of the image field.

As long as the object-beam/reference-beam intensity ratio R is kept reasonably small, the intensity distribution in the image is unaffected by the clipping produced by the nonlinearity of the recording material. This is demonstrated most strikingly by the reconstruction of quite faithful images with binary materials (i.e. materials capable of only two levels of amplitude or phase transmission) such as magnetic holograms.

Even so, the diffuse hologram is rarely the best choice in hologram recording. The reason for this is its unfavorable signal-to-noise properties. The source of noise (or spurious intensity fluctuations) in hologram images is partly purely statistical in character, partly "cosmetic" in nature (resulting from flaws and optical imperfections of recording and reconstruction system components), and partly incidental to system design (such as multiple reflections within the system). In photographic recording materials the granularity, which also determines the resolution and sine wave response of the material, is an important statistical source of noise.²⁵ However, there is no unique relationship between granularity-caused noise and resolution (or sine-wave response).²⁶

For essentially grainless materials, such as thermoplastics and photo-resist, such statistical scattering related to the recording process is much less important. In general, however, random scattering of the reconstructing beam in the hologram material and its support set a limit to the signal-to-noise ratio that can be obtained.

It is important to remember that light scattering, whatever its source, is far more important when the image is formed with coherent light than when it is formed with incoherent light. In the first case we are dealing with the addition of amplitudes, in the second with the addition of intensities. For example, let the intensity of scattered light I_{scatt} reach a portion of the image of intensity I . With incoherent light, the intensity will simply be increased by the fraction I_{scatt}/I . With coherent light, on the other hand, the intensity may vary in the range

$$(\sqrt{I} \pm \sqrt{I_{\text{scatt}}})^2 \cong I \left(1 \pm 2 \sqrt{\frac{I_{\text{scatt}}}{I}} \right). \quad [46]$$

Thus a scattering intensity that would simply increase the image brightness by a background brightness of 1% in incoherent imaging, may lead to the superposition of spurious fringe or granular pattern with 20% modulation in coherent imaging.

It has already been noted that, with diffuse illumination of the object transparency, these secondary sources of noise are overshadowed by speckle noise, which is inherent in the radiation pattern emanating from an object that is diffusely illuminated or is itself a diffuse reflector and hence cannot be removed or reduced by any subsequent manipulation. With speckle noise, the ratio of the root-mean-square fluctuation in light flux within an image area smaller than the limit of resolution of the imaging system is simply equal to the average light flux in this area ($S/N = 1$). As the hologram area A_h is increased so that the sampling area, or picture element, becomes substantially greater than the limit of resolution, the signal-to-noise ratio becomes equal to the ratio of the square root of the sampling area to the limit of resolution, or increases in proportion to the square root of the hologram area:*

$$\frac{S}{N} \cong \frac{\sqrt{A_h A_t}}{\lambda f} \quad \sqrt{A_t} > \frac{\lambda f}{\sqrt{A_h}} = d_{\text{min}} \quad [47]$$

Here, f is the focal length of the Fraunhofer lens for a Fraunhofer hologram and A_t the sampling area at the image; for a lensless hologram, f is simply the image-to-hologram separation.

* This applies if the object intensity over an area corresponding to A_t is substantially uniform or, generally, for continuous images. For a dot array, where the area A_o of the object sources, which here are in the form of diffusely illuminated apertures, is substantially less than the sampling area (corrected for an eventual magnification f/f_o between the recorded object and reconstructed image), we have instead

$$\frac{S}{N} \cong \frac{\sqrt{A_h A_o}}{\lambda f_o} \quad \sqrt{A_t} > \frac{f}{f_o} \sqrt{A_o}, \quad \sqrt{A_o} > \frac{\lambda f_o}{\sqrt{A_h}} \quad [48]$$

$$\cong 1 \quad \sqrt{A_t} > \frac{\lambda f}{\sqrt{A_h}}, \quad \sqrt{A_o} < \frac{\lambda f_o}{\sqrt{A_h}}$$

Here f_o is the Fraunhofer lens focal length or the object-to-hologram separation used in recording. The use of a diffusely illuminated aperture plate as object pattern is thus particularly disadvantageous with respect to the obtainable signal-to-noise ratio.

In summary, with diffuse illumination, the signal-to-noise ratio in the image is inherently small and is increased only very slowly as the hologram area is made larger than that required to resolve the individual picture element (defined by the acuity of the observer, the transmission capabilities of the remainder of the system, or the sampling aperture of the detector).

Fortunately, the high degree of redundancy, which is the major asset of the diffuse hologram, can be achieved without diffuse illumination of the object and the resulting introduction of speckle noise. An example is the Fourier-transform hologram shown in Fig. 5(b) or the closely analogous "lensless Fourier-transform hologram" obtained by illuminating the object with a beam converging at the hologram plane, shown in Fig. 14. We see immediately that, in either

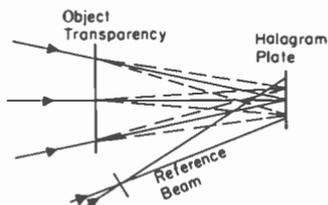


Fig. 14—Redundancy achieved with "lensless Fourier transform hologram" by letting object illuminating beam converge at hologram plane.

case, every part of the hologram receives light equally from all parts of the object, so that the effect of damage of any part of the hologram is distributed uniformly over the reconstructed image.

The holograms obtained as shown in Figs. 5(b) and 14 differ, however, from the diffuse hologram in that the hologram illumination is highly nonuniform; in particular, it exhibits a high intensity spike at the center, with intensities falling off at a rate depending on picture content. The limited latitude of the recording material will result in loss of low-frequency information in the image and, at the same time, very low holographic efficiency.

The advantageous properties of high redundancy and near-uniformity of illumination at the hologram can be achieved, however, without the introduction of speckle noise, by displacing the hologram in Figs. 5(b) and 14 from the position of lens focus and restoring the redundancy lost in the process by modifying the object illumination through the introduction of finely structured two-dimensional phase

gratings or aperture plates. Suitable procedures are described in the paper by Firester, Fox, Gayeski, Hannan, and Lurie.*

5.2 Storage Capacity of Volume Holograms

P. J. van Heerden²⁷ has given a general argument, based on the wavelength discrimination of volume holograms, to show that the ultimate storage capacity of such holograms is $V/(\lambda^3)$ picture elements, e.g., a crystal with a rectangular area ab and a thickness c can store, in superposition, c/λ pictures, each with ab/λ^2 resolvable picture elements. The same ultimate storage density, i.e. one element of information in a cell with a volume λ^3 , holds also for directly recorded, optically retrievable information, such as that stored in the emulsion of a nuclear-track plate. Furthermore, the optimal signal-to-noise ratio for each picture element resulting from the statistical distribution of light-responsive elements (such as color centers in a colored alkali halide crystal) is given in both cases by the square root of the number of elements in a volume equal to λ^3 .

The indicated storage density and signal-to-noise ratio must be regarded as upper limits that can at best only be approached in practice. Any specific model can be expected to lead to lower values for the storage density and signal-to-noise ratio. In the appendix**, I consider the example of n image fields consisting of picture elements of the same average intensity and random phase distribution and ask how many pictures can be recorded subject to the requirement that the ratio of the intensity of the desired image to the fluctuation in intensity induced by the superposition of amplitude components derived from the remaining stored images be equal to a prescribed signal-to-noise ratio $(S/N)_{np}$. I find for the maximum number n of stored pictures, (each with a number of picture elements of the order of (hologram area)/ λ^2),

$$n \cong \frac{c/\lambda}{4(S/N)_{np}^2} \quad [49]$$

Here c is, again, the crystal thickness.

* A. Firester, E. Fox, T. Gayeski, W. Hannan and M. Lurie, "Redundant Holograms", in this issue.

** The derivation disregards reconstructing wave attenuation and regeneration of the reconstructing wave by the diffracted wave and is thus strictly applicable only for hologram efficiencies well below the maximum; however, no qualitative change in the results is to be expected from a more complete, coupled-wave, analysis.

Clearly, a very high price has to be paid for increasing the required signal-to-noise ratio. As in the case of diffuse holograms where the number of picture elements in a picture reconstructed by the hologram decreases inversely as the square of the required signal-to-noise ratio (as determined by speckle noise), so here the number of storable pictures decreases with the inverse square of the prescribed signal-to-noise ratio.

It should be noted that the assumption of an array of essentially uniformly white or grey pictures is particularly unfavorable. Consider, instead, binary pictures, such as line drawings, in which a fraction f of the picture elements are white, the remainder black. Here we find, if the signal-to-noise ratio for the white elements is prescribed,

$$n = \frac{c/\lambda}{4f(S/N)_{np}^2} . \quad [50]$$

Since, for a white-on-black drawing, f may be a very small number, the number of such drawings that can be recorded as holograms in a crystal is much greater than that of essentially white pictures. Furthermore, in this instance, we are normally less interested in the signal-to-noise ratio for the individual white picture element than in the contrast C_{np} (again, as limited by the intensity superposed on the desired image by diffraction by the remaining stored holograms) between the "white" and "black" portions of the picture:

$$C_{np} = \frac{I_w - I_b}{I_b} . \quad [51]$$

In terms of this contrast we find,

$$n = \frac{c/\lambda}{2fC_{np}} . \quad [52]$$

The contrast for a binary black and white pattern (arising from hologram superposition alone!) is seen to enter the expression for the number of storable pictures linearly rather than quadratically. Doubling the required contrast for such patterns requires only doubling the hologram thickness; whereas, for substantially white or grey pictures, doubling the signal-to-noise ratio requires quadrupling the hologram thickness.

So far we have considered only the effect of picture crosstalk on the signal-to-noise ratio. If we consider pure phase volume holograms, which are both nonabsorbing and depend entirely on refractive index modulation for picture reconstruction, the statistical distribution of the photo-excited light scattering elements over individual holographic cells of the order of λ^3 in magnitude sets a further limit to the number of pictures with prescribed signal-to-noise ratio that can be stored.

Let us consider a transmission volume-holographic medium storing n pictures with a prescribed signal-to-noise ratio owing to picture crosstalk $(S/N)_{np}$, and with refractive index modulation n_1 and a thickness c such as to lead to 100% holographic efficiency. If we neglect, for a moment, the quantum nature of the excitation giving rise to the index modulation and assume $(S/N)_{np} \gg 1$, this condition can always be approached very closely; Eq. [27] establishes the relation between the refractive index modulation (for the individual hologram) and the thickness c :

$$n_1 = \frac{\lambda \cos \theta_o}{2c}. \quad [53]$$

Substituting from Eq. [49] for $c = 4n\lambda(S/N)_{np}^2$ we find

$$n_1 = \frac{\mu \cos \theta_o}{8n \left(\frac{S}{N}\right)_{np}^2} \cong \frac{1}{8n \left(\frac{S}{N}\right)_{np}^2}. \quad [54]$$

For a thickness $c = 1$ cm, $\lambda = 5 \times 10^{-5}$ cm and relatively small Bragg angles θ_o , Eq. [53] yields

$$n_1 = 2.5 \times 10^{-5}.$$

Measurements by Turner²⁸ suggest that the field amplitude E in lithium niobate, which we shall consider as an example, is related to the index amplitude n_1 by

$$n_1 = 1.86 \times 10^{-8} E \text{ volts/cm} \quad [55]$$

and to the amplitude of the trapped charge density N , for a fringe period $d = 1 \mu\text{m}$, by

$$E = \frac{edN}{2\pi\epsilon} = 8.98 \times 10^{-13} N \text{ cm}^{-3}, \text{ with } \epsilon = 2.84 \times 10^{-10} \text{ F/m.} \quad [56]$$

With Eqs. [54] and [55] we thus find for the charge-concentration amplitude

$$N \cong \frac{7.5 \times 10^{18}}{n \left(\frac{S}{N} \right)_{np}}. \quad [57]$$

As the number of trapped charges in a cell with linear dimensions of the order of a half fringe period, which in the present example is approximately equal to

$$N\lambda^3 \cong \frac{0.9 \times 10^6}{n \left(\frac{S}{N} \right)_{np}} \text{ cm}^{-3}, \quad [58]$$

becomes very small (e.g., of the order of unity), we must expect considerable statistical variation in the position of the "scattering centers" (here, electro-optically induced index variations) relative to the mean scattering planes, as well as in their amplitudes. We can get some semiquantitative notion of the magnitude of these effects if we neglect the amplitude variation and put the statistical variation in position equal to the statistical variation in the position of the center of gravity of the excess charge in a cell of the indicated volume. We assume that the probability of trapping of a charge of one sign at a distance x from the scattering plane is given by

$$P(x) = \frac{\pi}{d} \cos \left(\frac{2\pi x}{d} \right), \quad -\frac{d}{4} < x < \frac{d}{4}, \quad [59]$$

where d is the spacing of the scattering planes. Then the mean-square deviation from the mean scattering plane of the center of gravity of the charge in a cell with linear dimension $d/2$ is

$$\overline{x_c^2} = 0.0371 d^2 \text{ if there is one charge in the cell}$$

$$\overline{x_c^2} = 0.0186 d^2 \text{ if there are two charges in the cell}$$

$$\overline{x_c^2} = 0.0048 d^2 \text{ if there are three charges in the cell, etc.}$$

We can infer the effect of such statistical displacements from Debye's theory of the effect of random thermal motions of the atoms of a lattice on x-ray diffraction.²⁹ Debye finds that the random displacement of the scattering centers reduces the diffracted radiation (or, in our case, the reconstructed image intensity) by a factor D^2 and, at the same time, increases the scattering background intensity in proportion to $1 - D^2$. Here

$$D^2 = \exp \{-4\pi^2 \overline{(x_c^2/d^2)}\}. \quad [60]$$

We thus find

$$D^2 = 0.231 \text{ for one charge in a cell}$$

$$D^2 = 0.480 \text{ for two charges in a cell}$$

$$D^2 = 0.825 \text{ for three charges in a cell, etc.}$$

Thus with one charge per cell of volume $(d/2)^3 = \lambda^3$ we see that the image intensity has been reduced, owing to the granular nature of the stored record, by a factor greater than 4. The rough computation given clearly underestimates the effect, since it considers only one aspect of the statistical disorganization of the diffracting lattice. The requirement of at least one charge or excitation process per cell with volume λ^3 for every stored image seems thus a reasonable requirement for multiple picture storage. Eq. [58] then suggests an upper limit of, for example, 100 halftone pictures with a signal-to-noise ratio of 100. For binary black-on-white line drawings it should be possible, based on Eq. [52], to store 2×10^4 pictures with a contrast ratio of 100 making use of Eq. [52]. The appropriate crystal thickness would be 20 cm. These are both definitely upper limits, since we have here limited our attention to the effect of picture crosstalk on signal-to-noise ratio and contrast, leaving out of account random scattering and other sources of a noise background. With a less thick crystal, the number of pictures that can be stored for a given signal-to-noise ratio or contrast is reduced in proportion to the thickness (as given, e.g., by Eqs. [49], [50], and [52]); at the same time, effects from the granular nature of the record are reduced.

The preceding rather crude and fragmentary consideration of factors affecting the storage capacity of volume holograms indicates the large potential of such holograms for multiple picture storage, a potential that has already been largely verified by experimental demonstration.

Appendix—Picture Crosstalk Limitation on Number of Pictures that Can Be Stored in a Volume Phase Hologram

Let a reconstructing beam be incident from a distant source at r_o' with direction cosines l_o', m_o', n_o' on a hologram in which the deviation of the dielectric constant from its mean value is given by $\Delta\epsilon(x, y, z)$. Let the hologram be a rectangular block with dimensions a, b, c centered on the origin. Consider the scattered amplitude $A(r_2)$ at a large distance r_2 from the hologram center, in the direction (l_2, m_2, n_2) . The expression for the radiation of an oscillating dipole³⁰ indicates that this amplitude is given by

$$A(r_2) = -\frac{\pi}{\lambda^2} A_o' \cos \theta' \frac{1}{\epsilon} \int \frac{\Delta\epsilon(r)}{|r_2 - r|} \exp \{j[k_2 \cdot (r_2 - r) + k' \cdot (r - r_o')]\} dr \quad [61]$$

where

$$\Delta\epsilon = B(A_o \sum A_j^* + A_o^* \sum A_j) = B|A_o| \left[\sum_j |A_j| (\exp \{j[(k_o - k_j) \cdot r + \phi_j]\} + \exp \{-j[(k_o - k_j) \cdot r + \phi_j]\}) \right] \quad [62]$$

In the preceding equations, the r 's and k 's are vector quantities and the integration is to be carried out over the volume of the hologram. The hologram is assumed to be a Fraunhofer hologram, with j enumerating the object points, with corresponding wave amplitudes A_j . A_o is the amplitude of the reference wave, A_o' that of the reconstructing wave, and B is a constant proportional to the exposure and the recording sensitivity of the hologram medium. θ' is the angle between the direction of polarization of the incident reconstructing wave and its projection on a plane normal to r_2 .

In cartesian coordinates we can write

$$A(x_2, y_2, z_2) = C \sum_j A_j \int_{\frac{a}{2}}^{\frac{a}{2}} dx \int_{\frac{b}{2}}^{\frac{b}{2}} dy \int_{\frac{c}{2}}^{\frac{c}{2}} dz \exp \{jk[l_2(x_2 - x) + m_2(y_2 - y) + n_2(z_2 - z)]\} \cdot \exp \{jk[l_o'(x - x_o') + m_o'(y - y_o') + n_o'(z - z_o')]\} \cdot [\exp \{jk[(l_o - l_j)x + (m_o - m_j)y + (n_o - n_j)z] + j\phi_j\} + \exp \{-jk[(l_o - l_j)x + (m_o - m_j)y + (n_o - n_j)z] - j\phi_j\}] \quad [63]$$

$$\begin{aligned}
&= CV \exp \{j\phi_2\} \sum_j A_j \left(\exp \{-j\phi_j\} \operatorname{sinc} \left[\frac{\pi a}{\lambda} (l_j - l_o + l_o' - l_2) \right] \right. \\
&\operatorname{sinc} \left[\frac{\pi b}{\lambda} (m_j - m_o + m_o' - m_2) \right] \cdot \operatorname{sinc} \left[\frac{\pi c}{\lambda} (n_j - n_o + n_o' - n_2) \right] \\
&+ \exp \{j\phi_j\} \operatorname{sinc} \left[\frac{\pi a}{\lambda} (-l_j + l_o + l_o' - l_2) \right] \cdot \\
&\operatorname{sinc} \left[\frac{\pi b}{\lambda} (-m_j + m_o + m_o' - m_2) \right] \cdot \\
&\left. \operatorname{sinc} \left[\frac{\pi c}{\lambda} (-n_j + n_o + n_o' - n_2) \right] \right) \Bigg) \quad [64]
\end{aligned}$$

Here ϕ_2 is a phase constant depending on the position of the reference point and

$$C = - \frac{bA_o A_o' \cos \theta'}{\lambda^2 \epsilon |r_2|} \quad [65]$$

If the phases ϕ_j are random, the average intensity distribution can be written

$$\begin{aligned}
I(x_2, y_2, z_2) &= C^2 V^2 \sum_j A_j^2 \left\{ \operatorname{sinc}^2 \left[\frac{\pi a}{\lambda} (l_j - l_o + l_o' - l_2) \right] \cdot \right. \\
&\operatorname{sinc}^2 \left[\frac{\pi b}{\lambda} (m_j - m_o + m_o' - m_2) \right] \cdot \operatorname{sinc}^2 \left[\frac{\pi c}{\lambda} (n_j - n_o + n_o' - n_2) \right] \\
&+ \operatorname{sinc}^2 \left[\frac{\pi a}{\lambda} (-l_j + l_o + l_o' - l_2) \right] \cdot \\
&\left. \operatorname{sinc}^2 \left[\frac{\pi b}{\lambda} (-m_j + m_o + m_o' - m_2) \right] \cdot \operatorname{sinc}^2 \left[\frac{\pi c}{\lambda} (-n_j + n_o + n_o' - n_2) \right] \right\} \quad [66]
\end{aligned}$$

with the usual notation $\operatorname{sinc}(x) = \sin(x)/x$.

Since the coefficients $(\pi a/\lambda)$, $(\pi b/\lambda)$, $(\pi c/\lambda)$ are very large for a thick hologram, only those terms contribute appreciably to the intensity, for which the multipliers of these three quantities are very close to zero. This condition is fulfilled for the first and second term, respectively, if the unit vectors $\pm(l_j, m_j, n_j)$, $\mp(l_o, m_o, n_o)$, (l_o', m_o', n_o') , $-(l_2, m_2, n_2)$ form very nearly a closed figure. For the first term this is fulfilled for

$$(l_o', m_o', n_o') = (l_o, m_o, n_o), (l_2, m_2, n_2) = (l_j, m_j, n_j) \quad [67]$$

corresponding to the formation of an image identical with the object when the reconstructing beam is identical with the reference beam. For the second term (corresponding to the conjugate image) it cannot be fulfilled, so that its contribution to the intensity is negligible.

We shall now assume that a series of image fields are recorded on a three-dimensional hologram by employing reference beams differing in direction (within the hologram medium) by intervals l_o, m_o . Since

$$dl_o dm_o = \sin \theta_o \cos \theta_o d\theta_o d\phi_o, \quad [68]$$

the distribution of the reference beam directions will (except for the factor $\cos \theta_o$, which lies between 1 and 0.7) be uniform in solid angle. In air, outside the medium, the density of the reference directions will be less by a factor $(1/\mu)^2$ (for not too large values of θ_o), where μ is the refractive index of the hologram (e.g., $\mu = 1.5$). We shall compare the intensity that we observe along the axis $(0,0,1)$ as the result of the recording for which the reconstructing beam and reference beam are identical with the intensity scattered in the same direction by all the other stored patterns, i.e., those recorded with different (l_o, m_o, n_o) . We shall assume that the object intensity is the same for all the patterns and is distributed in Lambertian fashion from $(l_1, m_1, n_1) = (0,0,1)$ to $(l_{1m} \cos \phi, l_{1m} \sin \phi, \sqrt{1-l_{1m}^2})$, where l_{1m} is the angular image field within the hologram medium and is necessarily much less than 1. With $|A_j|^2 = 1$ and replacing the sum over the first term in Eq. [66] by an integral over l_1, m_1 , with

$$(l_o', m_o', n_o') = (l_o, m_o, n_o), (l_2, m_2, n_2) = (0,0,1) \quad [69]$$

we find for the desired intensity

$$I = C^2 V^2 \int_{-l_{1m}}^{l_{1m}} \int_{-\sqrt{l_{1m}^2-l_1^2}}^{\sqrt{l_{1m}^2-l_1^2}} \text{sinc}^2 \left(\frac{\pi a l_1}{\lambda} \right) \text{sinc}^2 \left(\frac{\pi b m_1}{\lambda} \right) \times$$

$$\times \operatorname{sinc}^2 \left(\frac{\pi c \{ \sqrt{1 - l_1^2 - m_1^2} - 1 \}}{\lambda} \right) dl_1 dm_1. \quad [70]$$

For a thickness c comparable to or smaller than the lateral dimensions of the hologram, a and b , $\pi c(l_1^2 + m_1^2)/2\lambda$ is very much smaller than 1 for all values of $(\pi a l_1/\lambda)$, $(\pi b m_1/\lambda)$ for which the first two factors in Eq. [70] make an appreciable contribution to the integral. Hence the last factor can be replaced by 1 and we obtain

$$I \cong C^2 V^2 \frac{\lambda^2}{ab} \quad [71]$$

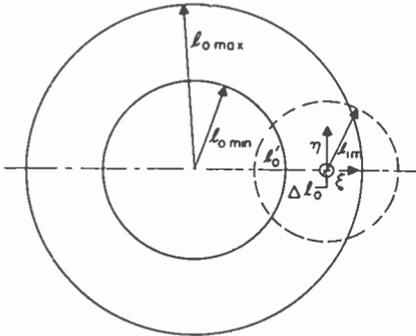


Fig. 15—Construction for integration over (l_o, m_o) .

We shall now endeavor to compute the contribution of the other stored patterns, I' , to the intensity observed in the forward direction. For convenience, we shall replace the summation over all the unwanted stored patterns by an integration over the (l_o, m_o) plane extending from the circle of radius $l_{o \min}$ to that of radius $l_{o \max}$, with the exclusion of a circle of radius Δl_o about the reference beam $(l_o', 0)$ as indicated in Fig. 15. We assume here that there is one reference beam position for pattern recording for an area $\pi \Delta l_o^2$ in the (l_o, m_o) plane. We then obtain for the undesired illumination in a forward direction

$$I' = \frac{C^2 V^2}{\pi \Delta l_o^2} \int d\xi \int d\eta \int_{-l_{1m}}^{l_{1m}} dl_1 \int_{-\sqrt{l_{1m}^2 - l_1^2}}^{\sqrt{l_{1m}^2 - l_1^2}} dm_1 \operatorname{sinc}^2 \left(\frac{\pi c}{\lambda} (l_1 - \xi) \right).$$

$$\text{sinc}^2 \left(\frac{\pi b}{\lambda} (m_1 - \eta) \right). \quad [72]$$

$$\text{sinc}^2 \left\{ \frac{\pi c}{\lambda} \left[\sqrt{1 - l_1^2 - m_1^2} - 1 + n_o' - \sqrt{n_o'^2 - 2\xi l_o' - \xi^2 - \eta^2} \right] \right\}.$$

Here $\xi = l_o - l_o'$, $\eta = m_o - m_o' = m_o$. With [73]

$$\Delta = l_1 - \xi, \Delta' = m_o - \eta \quad [74]$$

the argument of the last term can be written

$$\frac{\pi c}{\lambda} \left\{ \frac{\xi l_o'}{n_o'} + \frac{\xi^2(1 - n_o'^3)}{2n_o'^3} + \frac{\eta^2(1 - n_o')}{2n_o'} - \xi\Delta - \eta\Delta - \frac{\Delta^2}{2} - \frac{\Delta'^2}{2} \right\} \quad [75]$$

$$\approx \frac{\pi c}{\lambda} \left\{ \frac{\xi l_o'}{n_o'} + \frac{\xi^2(1 - n_o'^3)}{2n_o'^3} + \frac{\eta^2(1 - n_o')}{2n_o'} \right\},$$

since $\frac{\pi c \xi}{\lambda} \ll \frac{\pi a}{\lambda}$, $\frac{\pi c \eta}{\lambda} \ll \frac{\pi b}{\lambda}$. [76]

We can now integrate over Δ and Δ' , obtaining

$$I' = C^2 V^2 \frac{\lambda^2}{ab} \frac{1}{\pi \Delta l_o^2} \int d\xi \int d\eta \text{sinc}^2 \left\{ \frac{\pi c}{\lambda} \left[\frac{\xi l_o'}{n_o'} + \frac{\xi^2(1 - n_o'^3)}{2n_o'^3} + \frac{\eta^2(1 - n_o')}{2n_o'} \right] \right\}. \quad [77]$$

Here the integration over ξ and η is now limited by a circle of radius l_{1m} about the reconstructing beam direction as well as by the circles of radius l_{omin} and l_{omax} about the viewing direction (z -axis).

The argument of the sinc function vanishes for

$$\xi = -\frac{l_o' n_o'^2}{1 - n_o'^3} + \sqrt{\frac{l_o'^2 n_o'^4}{(1 - n_o'^3)^2} - \eta^2 \frac{n_o'^2(1 - n_o')}{1 - n_o'^3}} \quad [78]$$

$$-\zeta \cong \frac{\eta^2 (1 - n_o')}{2} \frac{l_o'}{l_o'} \cong \frac{l_o'}{4} \eta^2.$$

In a direction transverse to the line along which the sinc function argument vanishes, the argument is given by the product of the transverse displacement by

$$\frac{\pi c}{\lambda} \sqrt{\left[\frac{l_o'}{n_o'} + \frac{1 + n_o'^3}{n_o'^3} \xi \right]^2 + \frac{\eta^2 (1 - n_o')^2}{n_o'^2}} \cong \frac{\pi c}{\lambda} \sqrt{\left(\frac{l_o'}{n_o'} \right)^2 + \frac{\eta^2 (1 - n_o')^2}{n_o'^2}} \quad [79]$$

Integration in the transverse direction thus yields

$$\frac{\lambda}{c} \left\{ \left(\frac{l_o'}{n_o'} \right)^2 + \frac{\eta^2 (1 - n_o')^2}{n_o'^2} \right\}^{-1/2} \quad [80]$$

We hence obtain

$$\begin{aligned} \frac{I'}{I} &= \frac{2}{\pi (\Delta l_o)^2} \frac{\lambda}{c} \int d\xi \left[\left(\frac{l_o'}{n_o'} \right)^2 - 2\xi \frac{l_o' (1 - n_o')}{n_o'^2} \right]^{-1/2} \left[1 - \frac{l_o'}{2\xi (1 - n_o')} \right]^{1/2} \\ \frac{I'}{I} &= \frac{1}{\pi (\Delta l_o)^2} \frac{\lambda}{c} \int_{\Delta l_o}^{\eta_{\max}} d\eta \left[\left(\frac{l_o'}{n_o'} \right)^2 + \frac{\eta^2 (1 - n_o')^2}{n_o'^2} \right]^{-1/2} \left\{ 1 + \frac{\eta^2 (1 - n_o'^2)}{l_o'^2} \right\}^{1/2} \\ &= \frac{2}{\pi} \frac{n_o' \lambda}{l_o' c (\Delta l_o)^2} \left[\frac{l_{1m} - \Delta l_o}{\sqrt{l_{o\max}^2 - l_o'^2} - \Delta l_o} \right] \text{ whichever is smaller} \quad [82] \end{aligned}$$

The additive term $-\Delta l_o$ may here be omitted as negligible.

In the first instance, $l_{1m} < \sqrt{l_{o\max}^2 - l_o'^2}$, we give Δl_o^2 a value sufficiently large to lead to a prescribed signal-to-noise ratio, $S/N = \sqrt{I/(2I')}$, (in view of Eq. [46]) for values of l_o' down to $l_{o\min}$:

$$\Delta l_o^2 = \frac{4}{\pi} \frac{\lambda}{c} \left(\frac{S}{N} \right)^2 l_{1m} \frac{\sqrt{1 - l_{o\min}^2}}{l_{o\min}} \quad [83]$$

Thus for relatively small l_{1m} , the number of pictures that can be stored in the hologram becomes

$$n = \frac{l_{\text{omax}}^2 - l_{\text{omin}}^2}{\Delta l_o^2} = \frac{\pi c}{4 \lambda} \left(\frac{N}{S} \right)^2 \frac{1}{l_{1m}} \left[l_{\text{omin}} \frac{(1/\mu)^2 - l_{\text{omin}}^2}{\sqrt{1 - l_{\text{omin}}^2}} \right]_{\text{max}} \quad [84]$$

$$= 0.098 \frac{1}{l_{1m}} \frac{c}{\lambda} \left(\frac{N}{S} \right)^2, \text{ for } l_{\text{omin}} = 0.41, \mu = 1.5. \quad [85]$$

For $\sqrt{l_{\text{omax}}^2 - l_{\text{omin}}^2} < l_{1m}$,

$$\Delta l_o^2 = \frac{4 \lambda}{\pi c} \left(\frac{S}{N} \right)^2 \frac{\sqrt{l_{\text{omax}}^2 - l_{\text{omin}}^2} \sqrt{1 - l_{\text{omin}}^2}}{l_{\text{omin}}}, \quad [86]$$

and the number of pictures that can be stored in the hologram becomes

$$n = \frac{l_{\text{omax}}^2 - l_{\text{omin}}^2}{\Delta l_o^2} = \frac{\pi c}{4 \lambda} \left(\frac{N}{S} \right)^2 \left\{ l_{\text{omin}} \sqrt{\frac{(1/\mu)^2 - l_{\text{omin}}^2}{1 - l_{\text{omin}}^2}} \right\}_{\text{max}} \quad [87]$$

$$= 0.2 \frac{c}{\lambda} \left(\frac{N}{S} \right)^2 \text{ for } l_{\text{omin}} = 0.5, \mu = 1.5, l_{1m} > 0.433. \quad [88]$$

Eq. [88] is the basis for Eq. [49] of the text.

$$n \cong \frac{c/\lambda}{4(S/N)_{np}^2}$$

The rounding off of the numerical coefficient to $1/4$ seems reasonable insofar as the number of pictures that could be stored would be increased slightly by a more advantageous, nonuniform spacing of the reference beam orientations during recording.

References:

- 1 L. H. Lin, K. S. Pennington, G. W. Stroke and A. E. Labeyrie, "Multicolor Holographic Image Reconstruction with White-Light Illumination," *Bell Syst. Tech. J.*, Vol. 45, p. 659, April 1966.
- 2 E. G. Ramberg, "The Hologram—Properties and Applications," *RCA Rev.*, Vol. 27, p. 467, Dec. 1966.
- 3 W. R. Klein, "Theoretical Efficiency of Bragg Devices," *Proc. IEEE*, Vol. 54, p. 803, May 1966.
- 4 E. N. Leith and J. Upatnieks, "Wavefront Reconstruction with Diffused Illumination and Three-Dimensional Objects," *Jour. Opt. Soc. Amer.*, Vol. 54, p. 1295, Nov. 1964.
- 5 B. J. Thompson, J. H. Ward, and W. R. Zinky, "Application of Hologram Techniques for Particle Size Analysis," *Appl. Optics*, Vol. 6, p. 519, Mar. 1967.

- ⁶ G. W. Stroke, D. Brumm, and A. Funkhouser, "Three-Dimensional Holography with Lensless Fourier Transform Holograms and Coarse P/N Polaroid Film," *Jour. Opt. Soc. Amer.*, Vol. 55, p. 1327, Oct. 1965.
- ⁷ H. J. Gerritsen, W. J. Hannan, and E. G. Ramberg, "Elimination of Speckle Noise in Holograms with Redundancy," *Appl. Optics*, Vol. 7, p. 2301, Nov. 1968.
- ^{7a} H. J. Gerritsen and M. E. Heller, "Thermally Engraved Gratings Using a Giant Pulse Laser," *J. Appl. Physics*, Vol. 38, p. 2054, April 1967.
- ⁸ J. J. Amodei and R. S. Mezrich, "Holograms in Thin Bismuth Films," *Appl. Phys. Letters*, Vol. 15, p. 45, July 1969.
- ⁹ "Kodak Plates and Films for Science and Industry," Eastman Kodak Co., Rochester, N. Y., 1967.
- ¹⁰ H. Nassenstein, H. Dedden, H. J. Metz, H. E. Rieck, and D. Schultze, "Physical Properties of Holographic Materials," *Photog. Sci. Eng.*, Vol. 13, p. 194, July 1969.
- ¹¹ V. Russo and S. Sottini, "Bleached Holograms," *Appl. Optics*, Vol. 7, p. 202, Jan. 1968.
- ¹² C. B. Burckhardt and E. T. Doherty, "A Bleach Process for High-Efficiency, Low-Noise Holograms," *Appl. Optics*, Vol. 8, p. 2479, Dec. 1969.
- ¹³ R. J. Pressley, Ed., *Handbook of Lasers*, The Chemical Rubber Co., Cleveland, O., 1971, p. 563.
- ¹⁴ K. S. Pennington, J. S. Harper and F. P. Laming, "New Phototechnology Suitable for Recording Phase Holograms and Similar Information in Hardened Gelatine," *Appl. Phys. Letters*, Vol. 18, p. 80, Feb. 1971.
- ¹⁵ J. A. Jenney, "Holographic Recording with Photopolymers," *Jour. Opt. Soc. Amer.*, Vol. 60, p. 1155, Sept. 1970.
- ¹⁶ W. Coburn and K. Haines, "Volume Hologram Formation in Photopolymer Materials," *Appl. Optics*, Vol. 10, p. 1636, 1971.
- ¹⁷ G. Fan, K. Pennington, and J. H. Greiner, "Magneto-Optic Hologram," *J. Appl. Phys.*, Vol. 40, p. 974, March 1969.
- ¹⁸ W. R. Roach, "Holographic Storage in VO₂," *Appl. Phys. Letters*, Vol. 19, p. 453, Dec. 1971.
- ^{18a} S. A. Keneman, "Hologram Storage in Arsenic Trisulfide Thin Films," *Appl. Phys. Letters*, Vol. 19, p. 205, Sept. 1971.
- ¹⁹ R. G. Brandes, F. P. Laming and A. D. Pearson, "Optically Formed Dielectric Gratings in Thick Films of Arsenic Sulfur Glass," *Appl. Optics*, Vol. 9, p. 1712, July 1970.
- ²⁰ J. D. Margerum, T. D. Beard, W. P. Bleha, Jr. and S. Y. Wong, "Transparent Phase Images in Photoactivated Liquid Crystals," *Appl. Phys. Letters*, Vol. 19, p. 216, Oct. 1971.
- ²¹ F. Lanzl, U. Röder and W. Waidelich, "Hologram Recording by Aligning of Anisotropic Color Centers," *Appl. Phys. Letters*, Vol. 18, p. 56, Jan. 1971.
- ²² H. Kogelnik, "Coupled Wave Theory for Thick Hologram Grating," *Bell System Tech. J.*, Vol. 48, p. 2909, Nov. 1969.
- ²³ L. H. Lin, "Universal Method of Characterizing Hologram Recording Materials," *Jour. Opt. Soc. Amer.*, Vol. 60, p. 1562, Nov. 1970.
- ²⁴ M. Leifer and W. F. Schreiber, "Communication Theory," *Advances in Electronics*, Vol. 3, p. 305, 1951.
- ²⁵ J. W. Goodman, "Film Grain Noise and Wavefront Reconstruction Imaging," *Jour. Opt. Soc. Amer.*, Vol. 57, p. 493, April 1967.
- ²⁶ R. C. Jones, "Information Capacity of Photographic Film," *Jour. Opt. Soc. Amer.*, Vol. 51, p. 1159, Nov. 1961.
- ²⁷ P. J. van Heerden, "Theory of Information Storage in Solids," *Appl. Optics*, Vol. 2, p. 393, April 1963.
- ²⁸ E. H. Turner, "High-Frequency Electro-Optic Coefficients of Lithium Niobate," *Appl. Phys. Letters*, Vol. 8, p. 303, June 1966.
- ²⁹ P. Debye, "The Intensity Distribution in the Interference Produced by X-Rays," *Verh. dtsh. physik. Ges.*, Vol. 15, p. 738, 1913.
- ³⁰ A. Sommerfeld, "Electrodynamics," *Acad. Press*, N. Y., 1952, p. 150.

Materials for Magneto-Optic Memories

Roger W. Cohen and Reuben S. Mezrich

RCA Laboratories, Princeton, N. J.

Abstract—We review the capabilities of magneto-optic materials and indicate the possibilities and limitations of their use in data storage and retrieval systems. On the basis of ideal material properties, a lower limit of the required laser power for Curie-point writing is obtained, and the material systems in which this lower limit can be approached are discussed. We propose a new figure of merit which reflects both heat sensitivity and read-out requirements. This figure of merit allows the rapid evaluation of the relative qualities of materials for magneto-optic memories. It is found that, while system requirements for resolution and read-out efficiency can be met with presently available materials, there is a serious discrepancy between the write sensitivity demanded by a holographic system and material capabilities, either present or anticipated. We conclude that advances in laser technology are necessary in order to achieve the advantages of a holographic magneto-optic memory.

1. Introduction

Magneto-optic materials have recently become the subject of great interest to workers concerned with computer storage. Combining the best features of optical recording (high bit density and the possibility of redundancy) and of magnetic storage (permanence coupled with erasability and freedom from fatigue), successful magneto-optic storage would seem to offer the most promising method to satisfy the future needs of computer systems.

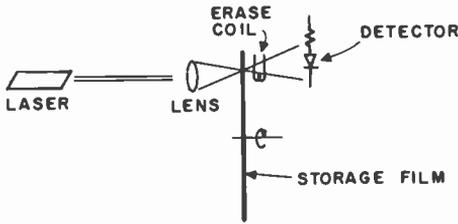


Fig. 1—Schematic representation of bit-by-bit storage.

Two main techniques for magneto-optic storage have been developed: bit-by-bit and holographic. In the bit-by-bit technique, schematically illustrated in Fig. 1, light is focused to a small spot ($1\text{-}5\mu$) at the surface of a thin magnetic film. Different spots are accessed by either moving the film or deflecting the spot of light. In the holographic approach,¹ illustrated in Fig. 2, two beams, one called the reference beam and the other the object beam (which contains light scattered from thousands of bits), interfere at a spot on the surface of a magnetic film ($\approx 1\text{ mm}$ spot size). Separate locations are usually accessed by deflecting the light beams. Read-out with both techniques is accomplished by simply illuminating the proper location on the film.

In both methods, information is stored through the mechanism of Curie-point writing,¹ in which incident radiation is absorbed in a magnetically saturated film, thereby heating the film. Strongly illuminated regions of the film are heated above the Curie-point, forcing these regions into the paramagnetic state. Upon cooling either in an applied magnetic field or the dipole fields of the surrounding magnetic material, the film will return to the magnetic state, but with a variation

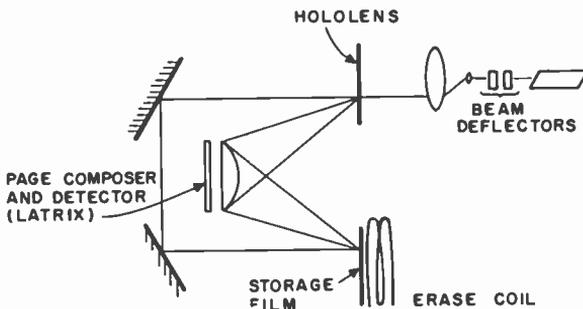


Fig. 2—Schematic representation of holographic storage.

in the direction of magnetization that is related to the incident light pattern. [We note in passing that the need to resort to Curie-point writing is due to the lack of any strong direct action of light on the magnetic state.]

The stored information is read either by light passing through the film (the Faraday effect) or by light reflected from the film (the Kerr effect). With the bit-by-bit method, polarizers and analyzers must be used to sense the different directions of magnetization; the holographic approach does not demand polarized light, although its use does enhance the detected signal-to-noise characteristics.

In this paper we discuss the features of magneto-optic storage, the material properties desired, the available materials, and the possibilities of successfully obtaining all the desired magneto-optic properties in one viable storage medium.

2. Material Requirements for Magneto-Optic Memories

The optimum magneto-optic storage medium must meet system requirements for high-bit-density capabilities, high write sensitivity, and large read-out efficiency. In this section, we discuss the effects of material parameters on these requirements and indicate the steps that can be taken to optimize the performance of magneto-optic media.

2.1 Material Requirements for High-Density Storage

The smallest resolvable spot in magneto-optic storage is determined by the minimum stable domain size. Although, in the general case, particular features such as grain size and an externally applied magnetic field can determine this limit, for the important class of materials that have one axis of easy magnetization perpendicular to the surface, the domain size can be related to the magnetic characteristics. The minimum stable domain size in this class of materials can be found by minimizing the total magnetic energy:¹

$$E = \left(\frac{16M_s^2 D}{\pi^2} \right) \sum_{n \text{ odd}}^{\infty} \frac{1}{n^3} [1 - \exp(-n\pi z/D)] + \left(\frac{2\pi z}{D} \right) \left(\frac{kT_c K_1}{2ra} \right)^2$$

In this expression M_s is the saturation magnetization, D is the domain size, z is the film thickness, T_c is the Curie temperature, K_1 is the magneto-crystalline anisotropy constant, r is the number of interacting neighbors, and a is the lattice constant. The first term on the right of

Eq. [1] represents the contribution due to the demagnetizing fields and the second term is the contribution due to the domain wall energy. Fig. 3 shows a plot of D versus film thickness obtained by minimizing E with respect to D (the parameters chosen are those for MnBi). We note that for thick films $z \cong D$ (e.g., MnBi films greater than one

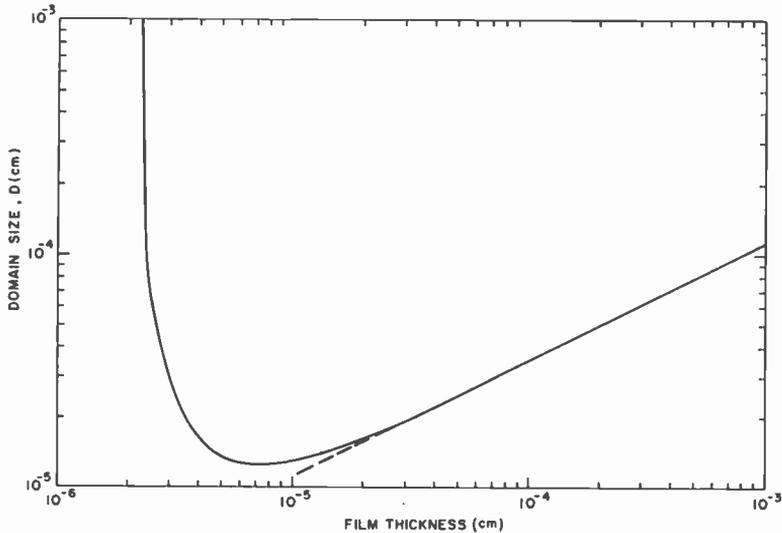


Fig. 3—Minimum domain size D as a function of film thickness z .

micron), the minimum stable domain size is given by the approximate expression

$$D = \frac{(2\pi^3 z)^{1/2}}{4M_s} \left[\frac{kT_c K_1}{2ra} \right]^{1/4} \quad [2]$$

Thus, the domain size is inversely proportional to the saturation magnetization and directly proportional to the square root of the film thickness.

The anisotropy constant K_1 influences the limiting resolution in two ways. Through its effect on the domain wall energy, Eq. [1], it helps determine the domain size. Its other effect is more significant; namely, it determines the direction of easy magnetization. For a thin film to sustain an easy axis of magnetization perpendicular to the film surface, the condition

$$K_1 > 2\pi M_s^2 \quad [3]$$

must be satisfied. From Eqs. [2] and [3], it is seen that, for the film to be capable of supporting small domains (i.e. have a large magnetization), it must have a very large crystalline anisotropy. The Curie temperature is seen to have only a slight influence on the determination of the domain size; its dominant effect is on the energy needed to achieve storage, as discussed in the following section.

2.2 Material Requirements for Curie-Point Writing

As described above, information contained in a laser beam can be stored in a magnetic film by heating selected portions of the material above the Curie temperature T_c and allowing them to cool, thereby locally reversing the direction of the magnetization. In this section, we shall discuss the important material properties that determine the laser power requirements for this process. We shall also estimate the minimum power that would be required to write information into a hypothetical material with ideal physical properties.

The energy per unit area that must be provided in order to raise a film to its Curie temperature is

$$Q = \frac{C(T_c - T_o)z}{V_M}, \quad [4]$$

where C is the average molar specific heat, T_o is the ambient or bias temperature, z is the film thickness, and V_M is the molar volume. Heat-flow considerations¹ show that the heating must be accomplished in a time less than the thermal diffusion time τ in order to avoid distortion of the incident intensity pattern. Assuming that the thermal match between the magnetic film and the substrate is sufficiently poor so that heat flow into the substrate can be neglected, the diffusion time τ is given by¹

$$\tau = \frac{Cd^2}{4\pi^2KV_M}, \quad [5]$$

where d is the required resolution or spot size, and K is the average thermal conductivity. Thus, the minimum power density which must be supplied in order to write information into a medium thermo-

magnetically is approximately

$$P = \frac{Q}{\tau} = \frac{4\pi^2 Kz(T_c - T_o)}{d^2}. \quad [6]$$

The power density P must be provided by a light beam through the mechanism of optical absorption. If τ is sufficiently long so that all heat generated by the light beam as it passes through the film is distributed throughout the thickness of film,* the relation between P and the intensity of the light beam is

$$P = I_o - I_z. \quad [7]$$

Here I_z is the light intensity at the back surface of the film and I_o is the intensity at the illuminated surface. Using Eqs. [6] and [7] along with Lambert's law $I_z = I_o \exp(-\alpha z)$, where α is the optical absorption coefficient, the required light intensity is

$$I_o = \frac{4\pi^2 Kz(T_c - T_o)}{d^2} [1 - \exp(-\alpha z)]. \quad [8]$$

From Eq. [8], it is seen that the important material parameters that determine the power requirement are the thermal conductivity, the Curie point, and the optical absorption coefficient. Low thermal conductivity is advantageous because heat is not able to diffuse quickly from an illuminated spot, and thus the heat is concentrated in the spot where it can be converted into an increase in temperature. A low Curie point is desired because it reduces the required temperature excursion that is to be provided by the laser. The film thickness should be small in order to minimize the volume of material that is to be heated. However, minimum values of z are set by the minimum domain size, as discussed in the preceding section, and by the optical absorption constant of the material. If $\alpha z \ll 1$, most of the light intensity passes through the film and does not contribute to heating of the material. In this limit, I_o is independent of z but is inversely proportional to α . This result demonstrates the advantage of high optical absorption to reduce laser power requirements.

Since the bit packing density ρ_B is inversely proportional to d^2 ,

* This condition requires that $Cz^2/KV_M \ll \tau$, or from Eq. [5], $z^2 \ll d^2/(4\pi^2)$.

Eq. [8] indicates that $I_o \propto \rho_B$. In a bit-by-bit system, a laser beam may be focused down to dimensions of order d , so that the required power output of the laser P_L is independent of ρ_B . However, in a holographic system $P_L \propto \rho_B$. Thus, high-bit-packing density increases the power requirements in a holographic system. This point is discussed in more detail in Section 4.

We now estimate the minimum laser intensity required to raise a hypothetical optimum magnetic material to its Curie point. We shall ignore the question of whether the combination of material properties that will be assumed can be achieved in a practical material. The object is to provide an absolute lower limit on the required laser power imposed on us by the restrictions of material properties.

To estimate the minimum thermal conductivity K_{\min} , we employ the simple kinetic theory result²

$$K = \frac{1}{3} C s l / V_M, \quad [9]$$

where s is the sound velocity, averaged over all phonon modes and frequencies, and l is the phonon mean free path. We assume that the minimum K corresponds to an l of atomic dimensions, approximately 3 Å. This low value of l implies a disordered or possibly amorphous material.* The remaining factors in Eq. [9] are expected to be minimized in materials with heavy, large atoms and, therefore, large molar volumes and low sound velocities. We choose $V_M = 20 \text{ cm}^3$ and $s = 1 \times 10^5 \text{ cm/sec}$, values appropriate for a material with a Debye temperature of about 100° K. Then for operations at and above 300° K, $C \cong 3R$, the Dulong-Petit value. Using Eq. [9], the thermal conductivity corresponding to the above values of l , V_M , s , and C is

$$K_{\min} \cong 1 \times 10^{-3} \text{ watt/cm-}^\circ\text{K.}$$

From Eq. [5], the corresponding maximum heating time allowed for a resolution $d = l \mu$ is

$$\tau_{\max} \cong 0.3 \times 10^{-6} \text{ sec.}$$

We next assume that the magneto-optical properties of the material

* It should be noted that such a material may not have sufficient magneto-crystalline anisotropy for a film to sustain an easy direction of magnetization perpendicular to the plane of the film.

are sufficiently large so that efficient reading can be accomplished on a film operated $(T_c - T_o) = 25^\circ \text{K}$ from its Curie temperature. We choose a high absorption constant $\alpha = 5 \times 10^5 \text{ cm}^{-1}$, appropriate for a metallic medium at visible wavelengths, and take $z = 1/\alpha = 200 \text{ \AA}$. Using the above values of K_{min} , z , α , $(T_c - T_o)$, and d , the minimum required intensity for our hypothetical material is obtained from Eq. [8]:

$$(I_o)_{\text{min}} \cong 300 \text{ watt/cm}^2.$$

The above figure would represent an improvement of more than three orders of magnitude over MnBi, most of the advantage coming from the lower thermal conductivity. Whether even this hypothetical material would lead to practical holographic memory systems will be discussed in Section 4.

2.3 Material Requirements for Magneto-Optic Read-Out

To retrieve information that is stored in a magnetic film in the form of an array of domains with various spin orientations, the magneto-optic properties of the material can be utilized. The read-out can be accomplished either by transmission (Faraday rotation or circular dichroism) or by reflection (the various Kerr effects). In the following, we examine the material properties that determine the read-out efficiency in the transmission mode and discuss the manner in which these properties conflict with those determining the heating power requirements discussed in Section 2.2. We introduce a new material figure of merit that includes aspects of both heating and read-out requirements.

Consider a linearly polarized beam of intensity I_R incident on a magnetic film. In passing through the film, the polarization vector of the beam is rotated in a direction that depends on the orientation of the magnetization of the film. For the case where the magnetization is perpendicular to the plane of the film, the intensity passing through an analyzer placed in back of the film and set for maximum contrast between contributions from "spin-up" and "spin-down" regions is

$$\Delta I = I_R \sin^2(2Fz) \exp(-\alpha z). \quad [10]$$

In Eq. [10], F is the specific Faraday rotation of the magnetic material. If I_R is sufficiently small compared to the intensity I_o required to heat the film to its Curie temperature, the value of F is appropriate to

that of the material at the temperature T_o . Since we are dealing with very thin films, the total rotation Fz is likely to be small, and we may replace the sin in Eq. [10] by its argument. The maximum value of $\Delta I/I_R$ defines the material efficiency Z_o ; it is obtained for $z = 2/\alpha$ and is given by

$$Z_o = 0.54 (2F/\alpha)^2. \quad [11]$$

The quantity in the parentheses in Eq. [11] is the Faraday figure of merit η_F . Letting the temperature dependence of F be governed by that of the spontaneous magnetization of the material, η_F has the form

$$\eta_F = \frac{2F_o \left(1 - \frac{T_o}{T_c}\right)^\beta}{\alpha}, \quad [12]$$

where F_o is the low-temperature specific Faraday rotation and the exponent β depends on whether the magnetic material has a first or second order transition at T_c :

$\beta = 0$, 1st order transition

$$\cong \frac{1}{3} \text{ to } \frac{1}{2}, \text{ 2nd order transition.} \quad [13]$$

The maximization of η_F naturally requires materials with large Faraday rotation F_o . A more general calculation¹ shows that the effect of circular dichroism $\Delta\alpha$ is to retain the form of Eq. [12], but with F_o replaced by a generalized effective rotation F_{eff} , where

$$F_{\text{eff}} = \left[F_o^2 + \frac{1}{4} (\Delta\alpha)^2 \right]^{1/2}. \quad [14]$$

Thus, large circular dichroism is also desired. Qualitative theoretical guidelines for the preparation of materials with large magneto-optic effects have been established.³ These guidelines predict that large Faraday rotations are favored by the large spin-orbit interactions present in materials with heavy atoms of high atomic number. To

eliminate the possibility of deleterious birefringence, structures with sufficient symmetry perpendicular to the direction of light propagation are required. Such structures include cubic, tetragonal with magnetization parallel to the c -axis, and amorphous materials.

From Eq. [12], one concludes that small values of the optical absorption coefficient α and low operating temperatures T_o (for $\beta > 0$) are required for efficient read-out. Both of these requirements conflict with those for low heating power, as discussed in the previous section. Reducing α to zero allows 100% read-out efficiency, but requires infinite heating power (Eq. [8]). Similarly, lowering T_o produces a larger Faraday effect, but raises the temperature excursion that must be produced by the laser.

We propose that a more realistic figure of merit for a magneto-optic memory material, one that reflects both heating and read-out requirements, is the ratio $\Delta I/I_o$ of the read-out signal to the heating power. Employing Eqs. [8] and [10] with $F_e \ll 1$, we find

$$\frac{\Delta I}{I_o} = A \left(\frac{F_{\text{eff}}^2}{KT_c} \right) \left(1 - \frac{T_o}{T_c} \right)^{2\beta-1} z \exp(-\alpha z) [1 - \exp(-\alpha z)], \quad [15]$$

where A is a constant that is independent of material properties. The maximum value of $\Delta I/I_o$ is obtained from a film thickness $z = 1.44/\alpha$. Thus, the combination of material properties which determines $\Delta I/I_o$ is a new figure of merit:

$$\eta = \frac{F_{\text{eff}}^2}{\alpha} (T_c K)^{-1} \left(1 - \frac{T_o}{T_c} \right)^{2\beta-1}. \quad [16]$$

In connection with Eq. [16], it should be noted that for materials with values of β in the range $1/3$ to $1/2$ (those which undergo second-order transitions), η is relatively insensitive to $(T_c - T_o)$ but is inversely proportional to T_c . On the other hand, for materials such as MnBi with $\beta = 0$ (those which undergo first-order transitions), η is inversely proportional to $(T_c - T_o)$ so that it is advantageous to operate the material close to its Curie temperature. However, the tolerance condition $I_R \ll I_o$ places a practical lower limit on the permissible value of $(T_c - T_o)$. We also note that the figure of merit Eq. [16] stresses the advantage of a low K , while, compared to η_F , it de-emphasizes the importance of small values of α .

3. Available Magneto-Optic Materials

In Section 2, we found that materials for magneto-optic memory applications should have sufficiently large magneto-crystalline anisotropy to overcome the demagnetizing field in the thin-film configuration. Large, heavy atoms favor low thermal conductivity and large Faraday rotation. The structure should have sufficient disorder or small grain size in order to minimize the phonon mean free path and, therefore, the thermal conductivity. Curie temperatures near room temperature are desirable, but the possibility of thermal biasing reduces the urgency of this requirement. If laser power requirements are more severe than those for read-out efficiency, metallic or intermetallic systems with high optical absorption are favored. On the other hand, if sufficient laser power is readily available, insulating materials with lower optical absorption and higher read-out efficiencies are appropriate. In this section we examine the properties of available materials and their utilization in magneto-optic memories.

3.1 "Conventional" Materials

The materials that have been the subject of most of the research on magneto-optic storage are the intermetallic compounds MnBi and MnAlGe, the magnetic semiconductor EuO, and the insulating ferrimagnets $Gd_3Fe_5O_{12}$ (GdIG) and $Y_3Fe_5O_{12}$ (YIG). The pertinent properties of these (and several other) materials are listed in Table 1. In connection with Table 1, we note that while the garnets and other insulators have values of η_F equal to or larger than metals, metals have far larger values of F_{eff}^2/α . This observation illustrates the inherent advantage of magnetic metals and inter-metallic compounds over insulators for Curie-point storage.

Bit-by-bit storage has been demonstrated with most of these materials and holographic storage with three of them (MnBi, EuO, MnAlGe). Of these materials, EuO and MnBi appear to be the most interesting. The major drawback of EuO is its very low Curie temperature; it must be operated in a cryogenic environment which, while it does not preclude use of the material in a practical system, does add to the complexity of the total system.

3.2 Granular Ferromagnets

Granular ferromagnets⁵⁻⁷ represent a new class of ferromagnetic material that shows promise as a potential storage media for optical

Table 1—Properties of Magneto-Optic Materials

Material	α (cm ⁻¹)	F_{rot} (deg/cm)	η_p (deg)	F^2_{rot}/α (deg ² /cm)	λ (Å)	T_c (°K)
MnBi	4.7×10^5	9.4×10^5	4.0	1.9×10^6	6300	630
MnAlGe	1×10^6	1×10^5 (†)	0.2(†)	1×10^4 (†)	5500	518
Co	8.5×10^5	3.8×10^5	0.89	1.7×10^5	5460	1390
Permalloy (Ni _{0.85} Fe _{0.15})	6×10^5	1.2×10^5 (†)	0.4(†)	2.4×10^4 (†)	5000	803
EuO(††)	9.7×10^4	3.6×10^5 (†)	7.4(†)	1.3×10^6 (†)	6000	69
GdIG	9×10^2	-1×10^3	2.2	1.1×10^3	6000	564 (Néel pt.)
YIG	1.5×10^3	2.4×10^3	3.2	3.8×10^3	5550	560 (Néel pt.)
FeBO ₃	40	2.3×10^3	115	1.3×10^5	5250	348
Bi _{0.7} Ca _{0.3} Fe _{0.85} V _{1.15} O ₁₂	1.4×10^3 60	9.3×10^3 1.9×10^3	13 63	6.2×10^4 6.0×10^4	6000 8000	520 (Néel pt.)

The data presented in the table were obtained from Reference [4]. Except where indicated, all optical properties were determined at 300°K.

(†) Circular dichroism not included.

(††) Optical properties determined at 5°K.

memories. These materials consist of an array of submicroscopic ($\leq 100 \text{ \AA}$) ferromagnetic grains embedded in an insulating matrix. Films of these materials are conveniently produced by the technique of co-sputtering⁸ a metallic element such as Ni, Co or Fe with a stable insulating substance such as SiO_2 . Either magnetic or nonmagnetic insulators may be used.

Measurements⁵⁻⁷ show that by simply varying the relative amounts of metal and insulator, the Curie temperature of a granular ferromagnet can be tailored to any desired value between 0° K and that of the pure metal. Thus, granular ferromagnets with Curie points conveniently near room temperature can be synthesized. Furthermore, due to the very large density of internal grain boundaries and the acoustic mismatch between the metallic and insulating components of the granular system, the room temperature thermal conductivity K is estimated* to be of order $10^{-2} \text{ watt/cm-}^\circ\text{K}$. This value of K is approximately an order of magnitude less than that¹ for MnBi, although still significantly larger than the minimum K estimated in Section 2.2. The combination of lower T_c 's and small K 's may reduce the required laser power for Curie-point writing by two orders of magnitude over MnBi.

We have also found experimentally that the Faraday figures of merit η_F of various granular ferromagnets are significantly enhanced over those of the pure metals. As can be seen from Table 2, observed values of η_F at 6328 \AA for the granular system $\text{Co}_{1-x}(\text{SiO}_2)_x$ are increased by over a factor of three over pure Co and are nearly as high as MnBi. Note, however, that the values of F_{eff}^2/α are still much smaller than that for MnBi. Similar enhancements of η_F have been observed⁹ in granular Ni and Fe. The enhancement can be understood in terms of a dielectric anomaly induced by the Lorentz dipole fields of the metal grains. It is found⁹ that a generalized Lorentz-Lorenz equation for the complex dielectric tensor reproduces the observed behavior in a semiquantitative manner and predicts that the enhancement is a general phenomenon for dispersed small metal grains in a dielectric matrix at frequencies sufficiently below the metal plasma frequency.

The most serious drawback of the granular ferromagnets synthesized to date is the lack of sufficiently large magnetic anisotropy (either crystalline or shape anisotropy) to produce the required irreversible magnetization loops for magnetic fields perpendicular to the plane

* Based on Eq. [9], using the observed particle size (Ref. [7]) and the calculated transmission coefficient for phonons passing through a Ni-SiO₂-Ni boundary.

of the film. Unless materials with such large anisotropy can be prepared, the magnetic moment must lie in the plane of the film in memory applications, thereby significantly reducing the read-out efficiency.

Table 2—Optical Properties of MnBi and Granular Co.

Material	α (cm ⁻¹)	F (deg/cm)	$\Delta\alpha$ (cm ⁻¹)	ηF (deg)	F^2_{rot}/α (deg ² /cm)
Pure Co	8.5×10^5	3.0×10^5	8.0×10^3	0.89	1.7×10^5
Co _{0.51} (SiO ₂) _{0.49}	1.15×10^5	$< 2 \times 10^4$	5.9×10^3	3.0	2.5×10^5
Co _{0.44} (SiO ₂) _{0.56}	1.05×10^5	$< 2 \times 10^4$	5.7×10^3	3.1	2.5×10^5
MnBi	4.7×10^5	6.1×10^5	25.0×10^3	4.0	19×10^5

Note: The granular Co properties were determined at 300°K using a He-Ne laser (6328Å). The applied magnetic field was 6kOe directed perpendicular to the plane of the film. This field was not sufficient to saturate the films, so that the values of $\Delta\alpha$ and ηF given in the table are estimated to be approximately 10 to 20% smaller than the saturation values. The values for pure Co were taken from the source given in Ref. [4]. The values for MnBi are for representative samples described in Ref. [10].

4. System Considerations and Conclusions

In the preceding sections we have reviewed the capabilities of magneto-optic materials and have indicated the possibilities and limitations of their use in data storage and retrieval systems. Among some of the important results are the determination of a lower limit of required laser power for Curie-point writing and the derivation of a new figure of merit that allows a practical evaluation of the relative qualities of magneto-optical materials. These results aid in the selection of suitable media for use in storage systems and should prove useful in determining goals for future research on magneto-optical media.

The knowledge of the characteristics of magneto-optic media or the demonstration of which material, among a collection of media, will have favorable properties is a necessary but not a sufficient goal of research. The goal of materials research is to supply elements capable of operating within specified system limitations, which are set, in turn, by the capabilities of auxiliary state of the art components. It is, therefore, of interest to compare the possible and actual magneto-optic media characteristics with those required in practical systems. As we

have stressed before, the key parameters are resolution, read-out efficiency, and write (heating) sensitivity.

4.1 Resolution

To achieve a limiting resolution determined by the wavelength of light (i.e. effective bit dimensions of the order of one micron), the medium must have resolution capabilities in excess of 10^3 lines per millimeter. The intrinsic resolution of magnetic materials is determined by the minimum stable domain size, which is proportional to the square root of the film thickness (Eq. [2]). By using thin films (of the order of 1000 Å), domain sizes less than 0.25μ (resolutions in excess of 2000 lines/mm) can be realized. Another limitation on the resolution is set by thermal diffusion; if heat is applied to one region of the film for a long time, the temperature in remote parts of the film will rise to the Curie temperature. The nature of the effects of thermal diffusion are seen by examining the temperature response of the film to a spatially periodic input of the form $I(x) = 1 + \cos(2\pi x/d)$, where d is the spatial period. The temperature rise produced in the film is found to be of the form¹

$$\Delta T(x,t) = \Delta T(O,\infty) [1 + (\tau/t) \cos(2\pi x/d)], \quad [17]$$

where t is the exposure time and τ is the diffusion time defined in Eq. [5]. The term τ/t can be viewed as a transfer function between the heat input and the film. Since $\tau \propto d^2$ (Eq. [5]), we see that thermal response, the fidelity with which the film can store high spatial frequency information (corresponding to high bit intensity), decreases as the square root of the exposure time. With optical pulse durations of the order of $0.1 \mu\text{sec}$, spatial resolutions of the order of 10^3 lines/mm have been achieved. It is noteworthy that, although the energy needed to bring the average temperature of the film to the Curie temperature is modest, the demand of high resolution requires short exposure times $t \leq \tau$ and, therefore, high peak power. There exists no analog to "integration" such as is found in photographic film or photoconduction systems.

4.2 Read-Out Efficiency

As we have stated, to be of practical importance, the magneto-optic storage medium must perform within limits set by other state-of-the-art components. To this end, we can estimate the efficiency required in a practical system and compare it to the values determined previously.

At the present state of the art, the optical signal at a solid-state photodetector must be greater than¹¹ 2×10^{-14} joules to achieve an error rate less than 10^{-8} . Present-day continuous-wave lasers or pulsed lasers capable of high repetition rate (greater than 10^5 Hz) have less than 10^2 watts output power capability. Optical systems efficiencies Z_o , which including surface reflections, deflector inefficiencies, and the Gaussian nature of a laser beam, are less than 10%; practical overall efficiencies are of the order of 1%. Within these limits we can calculate the required storage medium read-out efficiency Z_o .

The energy per bit reconstructed from a hologram is

$$E_o = P_L Z_o Z_s t_r / N, \quad [18]$$

where N is the number of stored bits, P_L is the laser power, and t_r is the read-out time. If we assume* $P_L = 10$ watts, $E_o \cong 2 \times 10^{-14}$ joules, $Z_r = 10^{-2}$, $N = 10^4$ and $t_r = 10^{-6}$ seconds, we find

$$Z_o \cong 2 \times 10^{-3}.$$

The energy available during read-out from a bit-by-bit system is N times greater than the above value. In this case, the required efficiency is

$$Z_o \cong 2 \times 10^{-7}.$$

The difference in the two values of Z_o lies in the fact that the bit-transfer rate of the holographic method is N times that of the bit-by-bit method.

From Eq. [11] and the material parameters given in Table 1, it is evident that the capabilities of several materials are well within the limit set by systems considerations for bit-by-bit operation (for example, MnBi gives $Z_o \cong 2.7 \times 10^{-3}$ in the maximum contrast configuration). Operation in the holographic mode is seen to stretch the capabilities of available materials to this limit.

4.3 Write Sensitivity

In previous sections we have shown that due to the material and thermal properties of potential storage media, the minimum power

* These values for P_L , Z_s and N correspond to a light intensity incident on the film $I_R = Z_s \rho_B P_L / N = 10^{-5} \rho_B$ (watt/cm²), where ρ_B is the bit density. Thus, for $\rho_B \leq 10^9$ cm⁻², the tolerance condition $I_R \ll I_o$ is well satisfied even for the "optimum" material discussed in Section 2.2.

needed to achieve high density storage is $I_0 \cong 300$ watts/cm² (present-day materials such as MnBi and EuO require power densities of the order of 10⁵ watts/cm²). If we assume, as above, that practical system efficiencies Z_s are of the order of 10⁻², and the available laser power is 10 watts, then the maximum area of a storage element (either a hologram or bit) is 3×10^{-4} cm². In the typical bit-by-bit method (for example, in an optical disk) spot sizes can be less than 10⁻⁶ cm². From optical considerations, the optimum hologram size for proposed¹² memories of 10⁸ bit capacity is 10⁻² cm².

The preceding discussions have attempted to show the possibilities and limitations of magneto-optic media used in data storage and retrieval systems. The requirements of such systems with respect to bit density and read-out efficiency can be met with presently available materials. However, with regard to the write sensitivity, there is a serious discrepancy between the needs of the system and the capabilities of the materials in the case of holographic operation. Even assuming the theoretical minimum power required in an optimum material, as was done in the above discussion, the required sensitivity is two orders of magnitude beyond attainable levels. To realize the advantages of a holographic magneto-optic memory, the improvements must come from progress in laser technology. For bit-by-bit applications, improvements in material characteristics could lead to simplified systems using low-power solid-state light sources.

References:

- ¹ See for example R. S. Mezrich, **Magnetic Holography** (Thesis), University Microfilms Inc., Ann Arbor, Mich.; R. S. Mezrich, **App. Opt.**, Vol. 9, p. 2275 (1970).
- ² See for example, R. D. Present, **Kinetic Theory of Gases**, p. 42, McGraw Hill Book Co., Inc., New York, 1958.
- ³ P. J. Wojtowicz, "Magnetic Materials", Chap. 12 in **Solid State and Quantum Electronics**, D. W. Herschberger, ed., J. Wiley and Sons, 1972.
- ⁴ For the magneto-optic properties of MnAlGe, see R. Sherwood et al, **J. Appl. Phys.**, Vol. 42, p. 1704 (1971); the properties of Bi_{0.7}Ca_{2.3}Fe_{3.85}V_{1.15}O₁₂ are given in A. Akselrad, **Proc. Seventeenth Conf. Magnetism and Magnetic Materials**, to be published. The properties of the other materials can be found in Di Chen, "Magneto-Optic Materials", **Handbook of Lasers**, p. 460, CRC Press, Cleveland, Ohio, 1971.
- ⁵ Y. Goldstein and J. I. Gittleman, **Solid State Comm.**, Vol. 9, p. 11971 (1971).
- ⁶ J. I. Gittleman, Y. Goldstein, and S. Bozowski, **Phys. Rev.** (in press).
- ⁷ M. Rayl, P. J. Wojtowicz, M. S. Abrahams, R. L. Harvey, and C. J. Buiochi, **Phys. Letters**, Vol. 36A, p. 477 (1971); **Proc. Seventeenth Conf. on Magnetism and Magnetic Materials**, Chicago, Nov. 1971, to be published.
- ⁸ J. J. Hanak, "The 'Multiple-Sample Concept' in Materials Research: Synthesis, Compositional Analysis, and Testing of Entire Multi-component Systems," **J. Mat. Science**, Vol. 5, p. 964 (1970).
- ⁹ S. B. Berger, A. Amith, and R. W. Cohen, to be published.
- ¹⁰ I. Gordon, S. B. Berger, E. A. Cufone, and J. Cserevits, **Proc. Seventeenth Conf. on Magnetism and Magnetic Materials**, Chicago, Nov. 1971, to be published.
- ¹¹ D. H. R. Vilkomerson, R. S. Mezrich, and J. M. Assour, **IEEE Trans. Solid State Circuits**, p. 317, Dec. 1969.
- ¹² W. C. Stewart and L. S. Cosentino, "Optics for a Read-Write Holographic Memory," **App. Opt.**, Vol. 9, p. 2271 (1970).

Holographic Recording in Lithium Niobate

J. J. Amodei and D. L. Staebler

RCA Laboratories, Princeton, N. J.

Abstract—This article provides an up to date comprehensive review of the physics and performance properties of undoped LiNbO_3 in holographic storage applications. It describes the physical model and shows the quantitative and qualitative agreement between the transport theory developed earlier and the available experimental evidence. Some of the practical aspects of performance are discussed and methods for improving the multiple-hologram recording properties of the crystals are explained.

1. Introduction

In the few years since the early experiments of Chen^{1,2} and his co-workers, lithium niobate storage effects have developed from a laboratory curiosity to the point where the material has become a strong potential candidate for many holographic recording applications. Mainly responsible for the increase in importance of this approach to optical storage are the improvements in theoretical understanding and in the performance of the material, most of which have been covered in the scattered literature of the past few years. At this stage, the process of recording and thermal fixing of holograms in undoped and in γ -irradiated lithium niobate, as well as the practical performance parameters of the materials, are understood and characterized well enough to warrant an effort at tying together what is

known about the behavior and performance of this medium. It is, therefore, the purpose of this article to give an up-to-date review of LiNbO_3 as a holographic recording material, with sufficient detail to permit physical understanding and evaluation of its present and future potentials by researchers and prospective users alike.

After a brief discussion of the pertinent physical properties of LiNbO_3 , this paper will review the transport model of holographic recording and will discuss the evidence showing that this model gives a simple self-consistent and quantitatively correct picture of the behavior of the undoped crystals. We will then proceed to discuss the practical performance of the material as a medium for read-write or optically nonerasable storage. Discussion of the performance of doped materials, some aspects which are not as well understood, will be left for an accompanying paper³ in this issue.

2. Physical Properties of LiNbO_3

Lithium niobate is a rhombohedral ferroelectric crystal with point group symmetry $3m$.⁴ It is transparent in the visible and infrared range from 3800 Å to 5 μm .

The crystal is uniaxial and its ordinary and extraordinary indices of refraction at 4500 Å are $n_o = 2.37$ and $n_e = 2.27$.⁵ The electro-optic tensor for lithium niobate has the following components:⁶

$$r_{33} = 30.8 \times 10^{-10} \text{ cm/V}$$

$$r_{13} = 8.6 \times 10^{-10} \text{ cm/V}$$

$$r_{22} = 3.4 \times 10^{-10} \text{ cm/V}$$

$$r_{42} = 2.8 \times 10^{-10} \text{ cm/V}$$

The dielectric constant of LiNbO_3 for directions perpendicular to the c axis is 78, along the c axis it is 32.⁷

The dark conductivity of LiNbO_3 at room temperature is extremely low. There has been some controversy on this point leading to questioning of the ability of these materials to sustain internal fields.⁸ Many measurements have now been carried out over a wide range of temperatures, however, and the consistency of the results suggests that extremely high resistivities are indeed typical of this material.

Generally, oxidation treatments produce crystals with the lowest conductivity. Values as low as $10^{-18} \text{ (ohm-cm)}^{-1}$ have been measured in such crystals using dielectric relaxation techniques.⁹ Larger conductivities are observed under optical illumination or at elevated tem-

peratures. The optically induced conductivity is the basis of holographic storage in these materials and arises from the photo-excitation of electrons out of traps deep within the ~ 3.3 -eV forbidden gap. The thermally activated conductivity is believed to be predominantly ionic, and has a ~ 1.1 -eV activation energy.⁹ Thermal activation of the trapped electrons requires a larger activation energy, as evidenced by the success of thermal fixing techniques in this material.¹⁰

The Curie temperature is quite large, 1210°C,¹¹ only ~ 50 °C below the melting temperature. Near room temperature, therefore, this material is a stable ferroelectric material; fields as large as ~ 3000 KV/cm are required for polarization reversal.

The fields required for holographic storage are much lower than the coercive fields, so that in this application the crystals remain single domain (i.e., holograms are created through linear electro-optic effects rather than by polarization reversal).

Pyroelectric effects are readily observable in LiNbO₃, and the pyroelectric coefficient is $\sim 10^{-2}$ $\mu\text{C}/(\text{m}^2\text{deg})$ ¹² at temperatures in the neighborhood of 100°C. Because of the strong temperature dependence of the ionic conductivity and the large pyroelectric effect, an interesting phenomenon occurs in these crystals when they are heat-cycled.⁹ As the temperature rises above 100°C, the conductivity becomes high enough for the resulting pyroelectric fields to relax in a matter of minutes. When the crystal is cooled, however, the pyroelectric effect is reversed and, since the conductivity drops rapidly, a substantial portion of the field remains for many weeks if the material is not exposed to light. This effect appears to provide the most logical explanation of the internal fields of unknown origin postulated by Chen^{1,2} to explain optical damage by drift of electrons. It also explains the inconsistency that often is observed in the behavior of a given sample during holographic recording. Because of the above effect, the degree to which the recording process is governed by drift or thermal diffusion^{13,14} depends on the previous heating and exposure history of the sample.

3. Physical Model and Transport Theory of Holographic Recording

3.1 Qualitative Description

The qualitative aspects of this type of recording can be explained by referring to Fig. 1. One begins with a substantially transparent medium that possesses a high electro-optic coefficient and a given concentration of localized traps containing electrons that can be excited

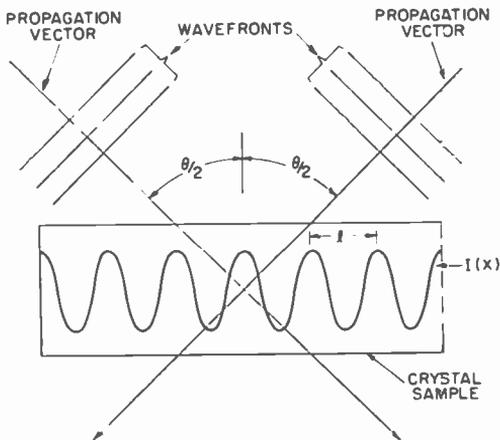


Fig. 1—Schematic of plane wave hologram recording.

by light of the appropriate wavelength. To allow for redistribution of charge, one must also assume that a certain percentage of the traps are empty, as shown schematically in Fig. 2(a), or that a given trap containing an electron can also trap an additional electron.

Exposure of such a medium to the light-interference pattern constituting the hologram excites electrons from the traps to the conduction band at rates that are proportional to the light intensity at any given point. This gives rise to an inhomogeneous concentration of free

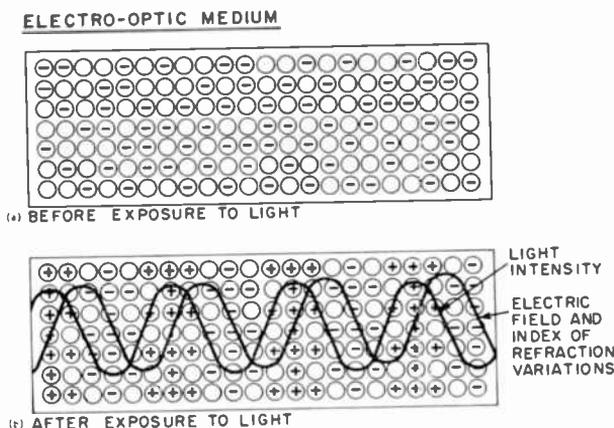


Fig. 2—Electro-optic storage medium: (a) trap distribution before exposure and (b) charge distribution after exposure.

carriers that diffuse thermally, or drift under applied or internal electric fields, and become retrapped preferentially in regions of low-intensity light. The end result is a net space-charge pattern that is positive in regions of high intensity and negative in regions of low intensity, as shown schematically in Fig. 2(b) for a sinusoidal interference pattern. The space charge generates a field that modulates the index of refraction and gives rise to a phase hologram.

Upon exposure to uniform light, as during readout, the electrons are reexcited out of the traps and redistributed evenly throughout the volume, thus eventually removing the field and erasing the hologram.

3.2 Quantitative Model

The details of the dynamics of recording and erasure, as well as the shape and magnitude of the index of refraction changes, depend on the physical model and the assumptions that are made for the calculations. There have been several different qualitative models proposed in the literature as responsible for, or contributors to, the photo-induced index changes occurring in electro-optic crystals. The early work of Chen^{1,2} led him to conclude that the index variations in lithium niobate and lithium tantalate, upon exposure of these materials to a laser beam or to a holographic interference pattern, were caused by photoexcited electrons drifting under a field and getting retrapped. Where the phenomenon occurs in the absence of externally applied fields, as in LiNbO_3 , Chen postulated the presence of internal fields due to unneutralized polarization charges. In a later paper, Johnston⁸ proposed an alternate model in which the drift fields were photo-generated by polarization changes that occurred as a result of a given number of electrons being in the conduction band during exposure to light. His model assumes electron migration due to these fields, giving rise to changes in the density of impurity dipoles in the regions of high intensity, with opposite changes in the regions of low intensity. Johnston proposed that these polarization variations somehow persist, causing the index changes, while the fields generated by the space charge (due to the electron migration) relax in a few minutes through intrinsic conductivity. Because of the low value of resistivity that he had measured ($\sim 10^{13}$ ohm-cm), Johnston assumed that neither internal drift fields nor space-charge-generated fields can exist in LiNbO_3 for more than a few minutes. While the model proposed by Johnston has some merit, it requires the transfer of a much larger number of electrons to give the same index variation as the space-

charge model; also, the assumptions that fields cannot exist in the material are based on what appears to have been an incorrect determination of the dielectric relaxation. At the present time, therefore, evidence points to drift under internal fields (as proposed by Chen) and diffusion of electrons due to thermal energy as proposed and analyzed by Amodei¹³ as the two mechanisms responsible for charge separation and consequent index changes in LiNbO₃. The fields generated by both of these transport processes were calculated and the relative importance (and differences) between the drift and diffusion processes in holographic recording were assessed in a recent paper.¹⁴ We will briefly review the mathematical model and the analysis covered in that work, the results of which are supported by our experimental observations.

The essential features of the model that will be used to analyze the process of holographic storage can be summarized as follows:

(a) The electro-optic material contains localized centers with trapped electrons that can be excited into the conduction band by light of the proper wavelength.

(b) Exposure to a holographic interference pattern results in an inhomogeneous density of free carriers that drift and/or diffuse and become retrapped. This gives rise to a net accumulation of negative space charge trapped in regions of lower intensity, balanced by positive charge in regions of high intensity.

(c) The space charge generates an electric field that modulates the index of refraction through the electro-optic effect, thus producing a hologram.

The theory is developed for the simplest form of interference pattern, the sinusoidal distribution given by the superposition of two plane waves, as shown in Fig. 1.

$$I(x) = I_0 (1 + m \cos Kx) \quad [1]$$

where m = modulation ratio

$$K = \frac{2\pi}{l} \quad [2]$$

$$l = \frac{\lambda}{2 \sin \theta/2} = \text{wavelength of the interference pattern.} \quad [3]$$

The model assumes that the trapping time is very short, so that the electrons will not be able to diffuse or drift very far before they are retrapped. Under these conditions, the concentration of free electrons in the conduction band will remain a faithful replica of the free-electron generation rate. If one further assumes that the redistribution of electrons only affects a small percentage of the total number of trapped electrons, so that the generation rate throughout the volume varies only as the light intensity, the free electron distribution $n(x)$ will remain directly proportional to the intensity distribution of the light interference pattern,

$$n(x) \approx n_0 (1 + m \cos Kx) \quad [4]$$

Here, n_0 is a constant proportional to the peak intensity I_0 , the trapping time, and the absorption coefficient due to trapped electrons.

Under the influence of an electric field or through thermal diffusion, the above-mentioned free-carrier concentration would produce a spatially inhomogeneous current given by

$$J(x) = qD_n \frac{dn}{dx} + qn\mu_n E(x), \quad [5]$$

where q is the electronic charge,

D_n is the diffusion constant for electrons,

μ_n is the mobility for electrons, and

$E(x)$ is the total electric field, assumed in the x -direction.

The spatially varying current would give rise to a trapped space-charge accumulation which, at any point, would increase at a rate given by

$$\frac{\delta\rho}{\delta t} = -\nabla \cdot J. \quad [6]$$

Thus, one may always calculate the space charge resulting from a certain exposure by integrating the divergence of the current density as given by Eq. [5]. Since the local field $E(x)$ depends on the space charge, numerical techniques are usually required except in special cases where drift or diffusion dominate the process. For these

limiting cases, one can compare the relative contribution of the two transport mechanisms and observe the fundamental differences in their resulting patterns.

In practice, it is fairly common to have a situation where transport occurs primarily via drift or through diffusion. The critical value of applied field E_o at which drift and diffusion contribute equally has been shown to be given by¹⁴

$$E_o = \frac{kT}{q} K, \quad [7]$$

where k is the Boltzmann constant. At room temperature Eq. [7] can be written as $E_o > 0.16/l$ in units of V/cm, where l , as defined before, is the wavelength of the interference pattern.

Thus, if the applied or internal field in the material is much greater than the above value, one can justifiably neglect diffusion and calculate the space-charge field recorded in the material, including only the drift component or current. Alternately, if no fields are present, diffusion is the only active mechanism.

For large applied fields and within exposure times that yield space-charge fields much smaller than E_o , the above model predicts a space-charge field $E_p(x)$ that grows linearly with time and remains a faithful replica of the intensity distribution of the light:

$$E_p(x) \approx -E_m \cos kx, \quad [8]$$

where $E_m = q\mu_n E_o m / \epsilon$ and ϵ is the dielectric constant of the material.

It is important to note from Eq. [8] that the peaks of the drift-generated field occur at the high-intensity minima of the light interference pattern, and that the magnitude of the field is independent of l . These features are characteristic of the drift-induced pattern only and do not apply to the diffusion process, as will be seen later.

If the exposure is allowed to continue until an equilibrium space-charge pattern is established, the resultant field is no longer linear. Neglecting diffusion, and assuming that only a small percentage of the total trapped-electron concentration has been redistributed (so that $n(x)$ is still given by Eq. [4]), the equilibrium field would be given by

$$E(x) = \frac{J_o}{q\mu_n n_o (1 + m \cos Kx)} \quad [9]$$

where J_0 is the current density when equilibrium is reached.

That this expression diverges for 100% modulation ($m = 1$) is a consequence of the assumption of negligible diffusion length and unlimited trap density. A finite diffusion length would have an effect equivalent to reducing the modulation ratio, and the finite trap density would put a limit on the maximum field for a given fringe spacing. In general, since either applied or inherent fields can be much larger than the space-charge fields required for holographic recording, one could operate in the linear range where Eq. [8] is valid. As was mentioned in the discussion of pyroelectric properties, high fields can be frozen into these materials any time the sample is heated and later cooled rapidly to room temperature.

For the case where there are no applied or internal fields, the diffusion mechanism governs the formation of the space-charge fields. Under these conditions, the equilibrium value of the space charge field can be shown to be¹³

$$E_p(x) = \left(\frac{kT}{q} \right) \left(\frac{Km \sin Kx}{1 + m \cos Kx} \right). \quad [10]$$

If this expression is expanded in a Fourier series, the fundamental term, which contributes to holographic reconstruction, would be given by

$$E_1(x) = \left(\frac{kTKm}{q\sqrt{1-m^2}} \right) \sin Kx. \quad [11]$$

The peaks of this fundamental are seen to occur 90° out of phase relative to those of the drift-generated pattern shown in Eq. [8]. This phase shift is most significant in that it gives rise to completely different behavior of the interactive coupling between the object and reference beams during recording, and will be discussed in more detail in what follows.

3.3 Diffraction Efficiency, Resolution, and Sensitivity

So far we have only discussed the formation of the electric field during exposure of the material to the holographic interference pattern. The actual diffraction of the light occurs as a result of the variation in the index of refraction $n_1(x)$, which is caused by the electric field through the electro-optic effect. The determination of the index of re-

fraction variations seen by light propagating at an arbitrary angle and arbitrary polarization in an anisotropic crystal, as caused by a given field pattern, is a fairly involved problem. It requires the use of the electro-optic tensor for the particular material to determine the changes in the index ellipsoid and the effect of these changes on the beam in question. However, the maximum effect in LiNbO_3 occurs for the special case when the c -axis of the material is perpendicular to the bisector of the two beams and the polarization is in the plane defined by the c -axis and this bisector. For small angles of incidence, the changes in index of refraction can then be calculated for this condition rather simply and to a good degree of approximation using the formula

$$n_1(x) \approx 1.8 \times 10^{-8} E(x). \quad [12]$$

The relationship between the diffraction efficiency of a sinusoidal grating hologram η and the index variation, as given by Kogelnik,¹⁸ is

$$\eta = \sin^2 \left[\frac{\pi n_1 d}{\lambda \cos \theta/2} \right], \quad [13]$$

where d is the thickness of the medium.

Thus, the index variation amplitude needed to produce a hologram with 100% diffraction efficiency for light of 5000 Å wavelength for small angles θ in a 1-cm-thick crystal is $\sim 2.5 \times 10^{-5}$, corresponding to fields of about 1400 V/cm.

Fields of this magnitude can easily be achieved through either process, drift or diffusion.

The amount of incident light energy that is required to give a certain diffraction efficiency is a measure of the sensitivity of the material. It determines the exposure time required to store a single hologram with a given power density, and it is thus a most important practical parameter in most applications. It is interesting to consider the theoretical limit of sensitivity in this process as well as the physical mechanism that controls the sensitivity in a practical situation.

It was pointed out above that the field required to achieve 100% diffraction efficiency in a 1-cm-thick crystal of LiNbO_3 is $E_m \approx 1400$ V/cm. The associated charge distribution is given by

$$\rho(x) = \rho_m \sin Kx = \epsilon K E_m \sin Kx \quad [14]$$

so that, for the above field, and assuming $\epsilon_3 = 2.9 \times 10^{-10}$ F/m for LiNbO_3 and a grating spacing $l = 1 \mu\text{m}$, we obtain a peak charge density of $\rho_m = 2.5 \times 10^{-4}$ coul/cm³. The average displaced charge can be calculated as

$$\rho_{\text{av}} = \frac{1}{l} \int_0^{l/2} \rho_m \sin Kx \, dx = \frac{\rho_m}{\pi} \quad [15]$$

The total number of displaced electrons N is then

$$N = \frac{\rho_m}{\pi q} \approx 0.5 \times 10^{15} \text{ electrons/cm}^3 \quad [16]$$

Normally, each electron would have to be reexcited many times in order to move to its final equilibrium position in the redistributed pattern. In fact, a rough estimate of the number of trips that it must make to the conduction band would be given by the ratio of the fringe spacing l to the mean drift or diffusion length \bar{x} . The highest quantum efficiency would result when this diffusion length is very large. Even for this case, however, the quantum efficiency would be something less than unity, because the electron would not always be trapped at the correct final location. Nevertheless, we can assume unity quantum efficiency for the sake of the argument and arrive at a number for the minimum absorbed energy that this process would require. For light of $0.5\text{-}\mu\text{m}$ wavelength, the energy per photon is $\hbar\omega = 4. \times 10^{-19}$ joules. Thus, the theoretical minimum absorbed energy that would be required to reach 100% diffraction efficiency in LiNbO_3 for a unity quantum efficiency process is

$$\mathcal{E}_{\text{min}} = \hbar\omega N = 2 \times 10^{-3} \text{ joules/cm}^2. \quad [17]$$

For 1%, one would require $\sim 10^{-4}$ J/cm². It should be mentioned that this number is independent of thickness. The highest sensitivity obtained so far from our Fe-doped crystals^{3,16} is a factor of about 1000 below the theoretical maximum.

As discussed above, we can obtain an order-of-magnitude estimate of the minimum absorbed energy required from an actual crystal by assuming that an electron must make an average number of l/\bar{x} trips to the conduction band before it moves the required distance. In this case,

the absorbed energy would be given by

$$\mathcal{E} \approx \frac{\hbar\omega Nl}{\bar{x}} = \frac{2\hbar\omega\epsilon}{q\bar{x}} E_m. \quad [18]$$

It should also be pointed out in passing that Eq. [14] also determines a minimum impurity concentration required to achieve a certain field for a given fringe spacing. Since the charge density required to sustain a given field is inversely proportional to the fringe spacing, this fact may explain the inability to obtain high diffraction efficiencies in thin samples of undoped LiNbO_3 , even though large index changes (10^{-3}) can be observed when laser damage is measured with beams of millimeter dimensions.

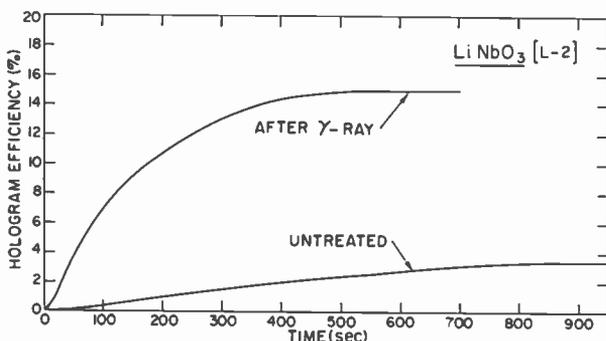


Fig. 3—Effect of γ -irradiation on sample sensitivity.

4. Experimental Results on Dynamic Behavior

4.1 Recording

Whether the dominant transport mechanism is drift or diffusion, the initial rate at which the space-charge field constituting the hologram builds up in these materials is linear. This implies a quadratic initial increase of diffraction efficiency with time, behavior that was indeed observed in nearly all of our experiments. The buildup of diffraction efficiency during exposure of a sample of LiNbO_3 before and after γ -irradiation is shown in Fig. 3. The change in absorption of the sample is shown in Fig. 4. The quantum efficiency, measured by the ratio of minimum to actual absorbed energy required to reach a given diffraction efficiency, is of the order of 10^{-5} . The minimum energy is calculated as explained earlier by assuming that a single absorbed photon is required to transfer an electron to its final location.

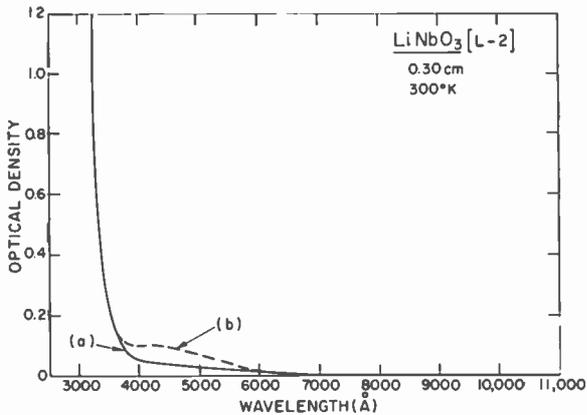


Fig. 4—Optical absorption of a LiNbO₃ sample before and after exposure to $\sim 10^6$ rads of γ -irradiation.

In the absence of applied or internally generated electric fields, the recording and erasure behavior of ~ 1 -cm-thick undoped crystals of LiNbO₃, whether γ -irradiated or not, is nearly symmetrical as predicted by the transport model.¹⁷ The erasure process is simply one of field relaxation by the photoconductivity generated in the material, while the recording process occurs through diffusion as explained above. This behavior is illustrated by curve A in Fig. 5, taken for a 1-cm sample of LiNbO₃ that had been exposed to light for a long period of time prior to recording to relax all internal fields.

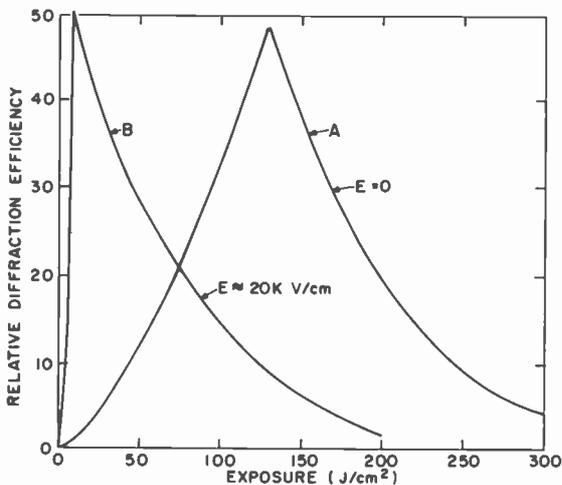


Fig. 5—Record-erase behavior of LiNbO₃ with and without applied fields.

An electric field applied externally (or generated internally) will increase the recording rate to a degree determined by the relative magnitude of the applied field and the equivalent diffusion field (Eq. [7]). This behavior is illustrated by curve B in Fig. 5. The increase in recording rate is in agreement with what is predicted by the transport model, since the equivalent thermal diffusion field (Eq. [7]) is 1600V/cm at $T = 300^\circ\text{K}$ and $l = 1\mu\text{m}$. The erasure rate is not affected strongly by the field, as one would expect in the case where the redistributed charge is a small percentage of the total trapped electron concentration.¹⁷

4.2 Effects of Beam Coupling During Recording

In the general discussion of hologram storage in these materials, one must also consider the effect that a hologram has on the writing beams. Because of the very high diffraction efficiencies that can be obtained, certain beam coupling effects that can usually be ignored during recording in other media become quite significant.¹⁸ In principle, a phase hologram can act as a periodic array that couples two interfering light beams. Although similar coupled-wave phenomena are well known in waveguide technologies, it is not generally realized that an analogous effect can play an important role in hologram storage. The point is that as the two intersecting beams begin to record a hologram, the hologram couples the two beams and thus rearranges their amplitude and phase as they travel through the crystal. The result is to limit or distort the hologram that is ultimately recorded.

We have used the above effects to verify that in LiNbO_3 , hologram storage is associated with charge transfer. The significant aspect of beam coupling is its dependence on the relative position of the phase grating, a parameter that depends on the storage mechanism. From Section 3, it can be seen that when drift dominates, the resulting index modulation is simply proportional to the interference pattern of the reference beams. That is, the position of the phase grating is such that the index modulation is symmetrical with respect to the light interference pattern. This occurs for all other storage media known to date, and produces very little coupling. The only effect that should be observable is a slight "bending" of the hologram caused by a rearrangement of the relative phase between the two light beams.¹⁸

However, the situation is quite different when diffusion dominates. Here, the index modulation is shifted so that it is $\pi/2$ out of phase with the interference pattern. In other words, a $\cos x$ light pattern produces a $\sin x$ index modulation. This shift is unique to electro-

optic materials and results in a highly efficient transfer of energy from one writing beam to the other. Here, the effect on storage is to limit the ultimate diffraction efficiency that can be achieved; transfer of all light to one of the beams creates regions of uniform light intensity that tend to erase the hologram.¹⁸

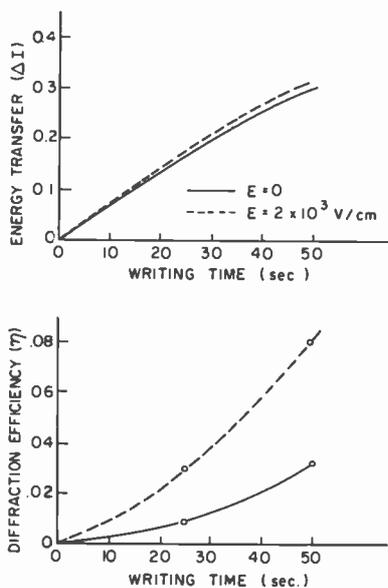


Fig. 6—Energy transfer and diffraction efficiency for recording by diffusion only and by combination of drift and diffusion.

Studies of undoped LiNbO_3 have demonstrated both diffusion and drift effects. Fig. 6 shows the diffraction efficiency, and the amount of light transferred, as a function of exposure. Normally, i.e., without an applied field, the storage is dominated by diffusion, as indicated by the large transfer shown in the top curve. The amount of transfer agrees well with theory. One can show that when diffusion is the only transport mechanism, the fractional amount of light, ΔI , transferred from one writing beam (e.g., the object beam) to the other, is related to the diffraction efficiency η by

$$\Delta I = \pm 2 [\eta - \eta^2]^{1/2} \quad [19]$$

This equation shows, as does Fig. 6, that a hologram with a fairly small diffraction efficiency can cause a sizable transfer. For example, when $\eta = 1\%$, $\Delta I = 20\%$.

An added feature of this coupled-wave analysis is that the sign of the photo-excited carriers is easily determined. This is done by merely observing the direction of transfer (the sign of Eq. [19]) relative to the orientation of the crystal. For all of our measurements on LiNbO_3 , electrons have been the dominant carriers.

Fig. 6 also shows what happens when one applies an electric field slightly larger than the thermal field for diffusion. The storage rate is enhanced by drift, but the interaction remains essentially unchanged. This occurs because the additional component to the phase grating, the component due to drift, does not contribute to the interaction. That is, the interaction depends only on the diffusion component, and is not affected by an electric field.

The results presented above, combined with the behavior observed in fixing experiments discussed in the next section, leaves little doubt that the transport model involving both drift and diffusion explains correctly all aspects of holographic storage in undoped LiNbO_3 .

5. Thermal Fixing of Holograms

A basic obstacle to the practical application of hologram storage in electro-optic crystals is the optical erasure of the hologram during read-out. This is a particularly severe problem in the use of these holograms for read-only applications that require repeated access to the stored patterns. Although it is possible to achieve nondestructive readout by using a different wavelength for readout and writing, this introduces distortions in the displayed image. The most desirable solution is to fix the holographic patterns once they are recorded, so as to prevent optical erasure of the information. The thermal fixing technique that was developed¹⁰ achieves this quite effectively.

The method for fixing can be described step by step as follows. First, a "normal" hologram is written by the interference of two intersecting laser beams. This optical interference pattern generates a modulated concentration of free carriers that migrate by diffusion or drift, producing a space-charge pattern of trapped electrons or holes, referred to hereafter as the electronic charge pattern. This charge displacement sets up an electric field that modulates the index of refraction. The fixing mechanism that is described here takes advantage of this electric field to create an optically stable charge pattern. It is accomplished by heating the sample to a temperature at which ions in the lattice become relatively mobile, while the trapped electronic charges remain thermally stable. The ions are hence allowed to drift in this field until they nearly completely compensate the trapped electronic

charge. Upon being cooled to room temperature, the resulting ionic charge pattern is "frozen" in; so that when the sample is exposed to light, the trapped electronic charges redistribute, leaving an electric field pattern that mirrors that of the original hologram. The initial electronic charge pattern is thus converted into an optically stable ionic charge pattern, and the hologram is fixed.

An example of the room-temperature readout of such a thermally fixed hologram, compared with the readout of a normal one, is shown

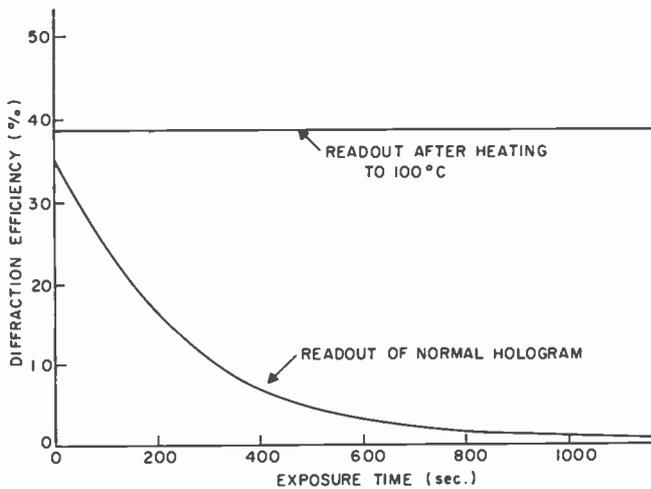


Fig. 7—Erasure of a normal and a fixed hologram in LiNbO_3 .

in Fig. 7 for an undoped LiNbO_3 sample. The first curve shows the readout of a normal hologram. It decays as the uniform readout beam reexcites the trapped electrons so that they may migrate back, thus "erasing" the holographic field. The second curve shows the room-temperature readout obtained after first recording a hologram in the same manner and then heating the crystal to 100°C for $\sim 1/2$ hour. The initial efficiency is seen to be quite low, because the ionic charge had compensated the electron charge and erased the field. During readout, however, the diffraction efficiency increases as the electronic charge is redistributed by the uniform readout beam, leaving the electric field pattern due to the fixed ionic charges.

If higher temperatures are used for fixing, it can be done in a shorter time than shown above because the ionic mobility is thermally activated.⁹ The temperature dependence of the fixing can be determined

from the following relation for thermally activated conductivity:

$$\sigma = \sigma_0 \exp \left\{ - \frac{E_a}{kT} \right\},$$

where σ is the conductivity, σ_0 is a constant, k is the Boltzmann constant, T is the temperature, and E_a is the thermal activation energy for ionic motion. This ionic conductivity relaxes the electric field pattern of a hologram during fixing. The time in which this occurs, i.e., the fixing time, is related to the dielectric relaxation time, ϵ/σ . Thus,

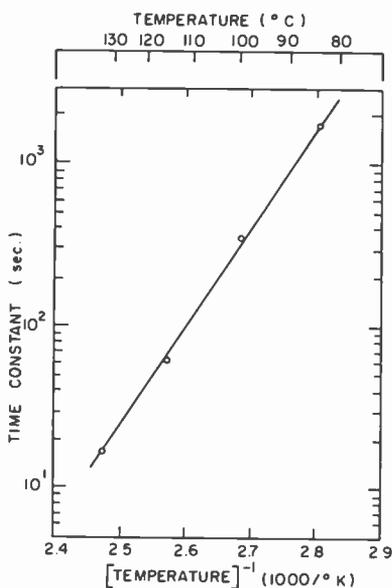


Fig. 8—Temperature dependence of fixing times in LiNbO₃.

by increasing the temperature, one can increase the ionic conductivity and shorten the time required for fixing. This temperature dependence for undoped LiNbO₃ is explicitly demonstrated in Fig. 8, which shows the measured relaxation time observed during fixing at a number of temperatures. The slope indicates an ~1.1-eV activation energy for ionic motion.

In principle, one can obtain a fixed hologram having a diffraction efficiency equal to that of the original hologram. In practice, however, the strength of the fixed hologram is limited by the relative concentra-

tion of ionic and electronic charges. The first requirement for complete fixing is that the concentration of ionic defects n_i is large enough to relax the electric field of the displaced electronic charge. In fact, if we assume that the density of trapped electronic charge n_e is sinusoidal, i.e.,

$$n_e = N_o + n_o \sin Kx, \quad [20]$$

where $K = 2\pi/l$ and l is the grating spacing, and that $n_i \gg n_o$, it can be shown that the equilibrium value of the amplitude of the relaxed field E is given by

$$E \approx \frac{E_o}{1 + \frac{q q n_i}{kT K^2 \epsilon}}. \quad [21]$$

Here q is the ionic charge, k is Boltzmann's constant, T is the temperature, ϵ is the c -axis dielectric constant, and E_o is the initial amplitude of the field pattern. Typically, the grating spacing is approximately $1\mu\text{m}$, so that the relaxed field becomes

$$E \approx \frac{E_o}{1 + \frac{n_i}{2 \times 10^{15} \text{ cm}^{-3}}}. \quad [22]$$

where we have assumed $T = 400^\circ\text{K}$ and ionic defects of unit charge. According to this equation, complete relaxation occurs if $n_i \gg 2 \times 10^{15} \text{ cm}^{-3}$, a reasonable condition for the concentration of ions.

The second condition for fixing is that the density of trapped electrons is small enough to allow reconstruction of the field pattern. This follows because the holographic field reappears only when the trapped electrons redistribute evenly by diffusion during exposure to uniform light. The final, or fixed, value of the field due to the partially unneutralized ionic distribution can be shown, by the same argument used above, to be

$$E_f \approx \frac{-E_o}{1 + \frac{N_o}{1.6 \times 10^{15} \text{ cm}^{-3}}}. \quad [23]$$

where N_0 is the average density of trapped electronic charges. In this case, we have assumed a small equilibrium modulation on the net density of trapped electrons n_e and that $T = 300^\circ\text{K}$. Thus, a significant fraction of the field is reconstructed only if the density of occupied traps is no greater than $\sim 10^{16}\text{ cm}^{-3}$. This places a stringent limitation on the dynamic range and sensitivity of the storage medium when good fixing performance is required as well. However, one can overcome the above problem to some extent by using a fairly thick crystal. This reduces the space-charge density required to obtain a significant diffraction efficiency, e.g., a space charge of only $\sim 4 \times 10^{14}\text{ cm}^{-3}$ will give a 50% diffraction efficiency in a 2-cm crystal of LiNbO_3 .

When a fixed hologram has a weak electric field pattern, because the ion concentration is too low or a trapped charge concentration is too large, one can still obtain large readout efficiencies by making use of a "self-enhancement" effect. This effect, brought about by the interference of the readout beam and the diffracted beam, occurs because the fixed holograms are in a thick, optically sensitive storage medium. Thus, when a readout beam is applied and part of it is diffracted by the fixed hologram, the two beams (the readout beam and diffracted beam) intersect within the crystal to form an interference pattern that writes a new hologram by the normal mechanism, i.e., the displacement of trapped charge. This new hologram is identical to the fixed one and can thus enhance the net diffraction efficiency, as long as the two holograms constructively interfere. Such an enhancement effect is shown in Fig. 9 for an undoped LiNbO_3 sample with the c -axis pointed in the proper direction. When the hologram is read out, the efficiency increases. When the sample is turned slightly, so that the Bragg condition for readout is no longer satisfied, the enhancement is erased.

6. Other Performance Considerations

6.1 Multiple Hologram Storage

While a single hologram stored in a given volume of these materials can achieve very high diffraction efficiency, recording of a large number of patterns can normally be accomplished only at the expense of overall diffraction efficiency. The illumination of the crystal for each successive recording erases a large fraction of the already recorded patterns unless they have already been fixed. In fact, in a 1-cm-thick crystal that could support nearly 100% diffraction efficiency for a single holo-

gram, storage of 10 patterns reduces the efficiency to a fraction of a percent. If one resorts to thermal fixing after recording each pattern, the results are improved. This is a very time consuming process, however, and is not completely effective because some erasure of the holograms during each heating cycle still takes place.

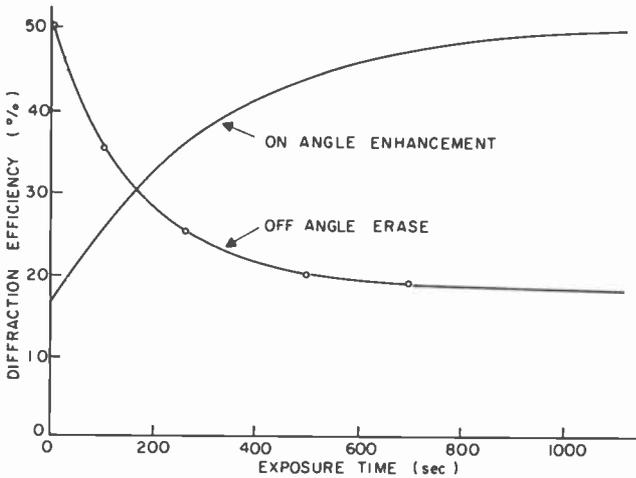


Fig. 9—Enhancement of fixed hologram.

Where multiple recording is desirable in order to take advantage of the increased storage capacity of a thick crystal, the best technique consists of taking advantage of the asymmetry in the record-erasure curves induced by electric fields (as illustrated in Fig. 5). It has been demonstrated¹⁷ that, by applying an electric field along the *c*-axis of the crystal during recording, it is possible to record ten patterns with more than 30% efficiency and over 100 holograms with efficiencies well above 1%. Using these techniques it may be possible to record up to 1000 patterns with usable levels of diffraction efficiency in the reconstructed image.

6.2 Optical Erasure

The optical erasure of holograms that have not been thermally fixed in LiNbO_3 is illustrated in Fig. 7. For the case where there is no electric field applied to the sample, the pattern is erased when readout exposure reaches approximately the same level required for recording. When an electric field is used, about 10 to 20 readouts can be made at

energy levels used for recording before losing the information. In many applications, it is possible to use considerably reduced power levels during readout, allowing the information to be retrieved many times nondestructively.

Where multiple recording for read-only type applications is envisioned, thermal fixing of the patterns constitutes the best solution. Fixed holograms show negligible optical erasure at all intensity levels where the crystal does not actually become substantially heated by the readout beam.

6.3 Thermal Erasure

The thermal decay of holographic fields in LiNbO_3 corresponds to a dielectric relaxation process governed by the intrinsic conductivity and dielectric constant of the material. As mentioned earlier, our measurements show resistivities that are of the order of 10^{18} ohm-cm and the conduction process is believed to be predominantly ionic at temperatures below 150°C . These values correspond to relaxation times of about a month for the holographic fields, which is what is found in practice.

The activation energy for ionic conductivity was measured at about 1.1 eV, so that there is a very strong temperature dependence of the relaxation times and slight cooling can increase the storage time by many orders of magnitude. At 0°C the estimated relaxation time is about 10 years, already in the range suitable for archival storage.

7. Conclusions

While some aspects of the behavior of doped LiNbO_3 are still not thoroughly understood, the performance of undoped and γ -irradiated samples of these crystals has been found to follow closely the predictions of theoretical models. Engineering design to begin incorporating these materials into storage devices can therefore proceed on a sound basis.

The lower sensitivity of these crystals can be partially offset by recording with an applied electric field, which was shown to narrow substantially the performance gap between doped and undoped materials. The excellent performance of the thermal fixing techniques, which makes these crystals good candidates for read-only, long-term storage applications, provides an additional and very significant advantage over other volume storage media, such as photochromics, where no known method exists for achieving high erasure resistance.

Finally, the ready availability at reasonable prices of high optical

quality LiNbO_3 crystals in boules of up to several inches in dimensions makes this material quite practical from an economic point of view as well.

Acknowledgments

The authors are indebted to W. Phillips for the growth of many of the samples used in this work and to G. Latham and L. Levin for their skillful assistance in the experimental work that was done. They also wish to acknowledge many helpful technical discussions with R. Williams who contributed important ideas and suggestions during the course of this work.

References:

- ¹ F. S. Chen, J. T. LaMacchia, and D. B. Fraser, "Holographic Storage in Lithium Niobate," *Appl. Phys. Letters*, Vol. 13, p. 223, 1968.
- ² F. S. Chen, "Optically Induced Change of Refractive Indices in LiNbO_3 and LiTaO_3 ," *J. Appl. Phys.*, Vol. 40, p. 3389, 1969.
- ³ W. Phillips, J. J. Amodei, and D. L. Staebler, "Optical and Holographic Storage Properties of Transition Metal Doped Lithium Niobate," *RCA Review*, this issue.
- ⁴ S. C. Abrahams, J. M. Reddy, and J. L. Bernstein, "Ferroelectric Lithium Niobate. 3-Single Crystal X-Ray Diffraction Study at 24°C," *J. Phys. Chem. Solids*, Vol. 27, p. 997, 1966.
- ⁵ G. D. Boyd et al., "Refractive Index as a Function of Temperature in LiNbO_3 ," *J. Appl. Phys.*, Vol. 38, p. 1941, 1967.
- ⁶ E. H. Turner, "High Frequency Electro-optic Coefficients of Lithium Niobate," *Appl. Phys. Letters*, Vol. 8, p. 303, 1966.
- ⁷ K. Nassau, H. J. Levinstein, and G. M. Loiacono, "Ferroelectric Lithium Niobate. 2.Preparation of Single Domain Crystals," *J. Phys. Chem. Solids*, Vol. 27, p. 989, 1966.
- ⁸ W. D. Johnston, Jr., "Optical Index Damage in LiNbO_3 and Other Pyroelectric Insulators," *J. Appl. Phys.*, Vol. 41, p. 3279, 1970.
- ⁹ D. L. Staebler and J. J. Amodei, "Thermally Fixed Hologram in LiNbO_3 ," *Ferroelectrics*, Vol. 3, p. 107 (1972).
- ¹⁰ J. J. Amodei and D. L. Staebler, "Holographic Pattern Fixing in Electro-Optic Crystals," *Appl. Phys. Letters*, Vol. 18, p. 540, 1971.
- ¹¹ K. Nassau and H. J. Levinstein, "Ferroelectric Behavior of Lithium Niobate," *Appl. Phys. Letters*, Vol. 7, p. 69 (1965).
- ¹² M. B. Roitberg, V. K. Novik, and N. D. Gavrilora, "Characteristic Features of the Pyroelectric Effect and Electrical Conductivity in Single Crystals of LiNbO_3 in the Range 20-250°C," *Sov. Phys. Crystallog.*, Vol. 14, p. 814 (1970).
- ¹³ J. J. Amodei, "Electron Diffusion Effects During Hologram Recording in Crystals," *Appl. Phys. Letters*, Vol. 18, p. 22 (1971).
- ¹⁴ J. J. Amodei, "Analysis of Transport Processes During Holographic Recording in Insulators," *RCA Review*, Vol. 32, p. 185 (1971).
- ¹⁵ H. Koyelnix, "Coupled Wave Theory of Thick Hologram Gratings," *Bell Sys. Tech. J.*, Vol. 48, p. 2909 (1969).
- ¹⁶ J. J. Amodei, W. Phillips, and D. L. Staebler, "Improved Electro-optic Materials and Fixing Techniques for Holographic Recording," *Proc. 1971 IEEE/OSA Conf. Laser Engineering and Applications*, June 8, 1971; also *Applied Optics*, Vol. 11, p. 390 (1972).
- ¹⁷ D. L. Staebler, J. J. Amodei, and W. Phillips, "Multiple Storage of Thick Phase Holograms in LiNbO_3 ," VII International Quantum Electronics Conference, Montreal, May 1972 (to be published).
- ¹⁸ D. L. Staebler and J. J. Amodei, "Coupled-Wave Analysis of Holographic Storage in LiNbO_3 ," *J. Appl. Phys.*, Vol. 43, p. 1042 (1972).

Optical and Holographic Storage Properties of Transition Metal Doped Lithium Niobate*

W. Phillips, J. J. Amodei, and D. L. Staebler
RCA Laboratories, Princeton, N.J.

Abstract—Measurements of recording sensitivity, maximum diffraction efficiency, and erase behavior have been carried out on LiNbO_3 crystals doped with various first-row transition elements. Manganese, iron, and copper doped crystals have distinctive absorption spectra and show large increases in sensitivity over undoped crystals, a factor of 500 in the case of iron. The erase sensitivity of crystals doped with iron during growth is an order of magnitude lower than the record sensitivity. When iron-doping is carried out by diffusion, record and erase behavior become nearly symmetrical.

1. Introduction

The storage of volume phase holograms in LiNbO_3 ¹ relies on the presence of localized centers containing electrons that can be optically excited into the conduction band. Attempts to improve the recording performance of this material have, therefore, focused on radiation treatments and on the incorporation of dopants² to allow experimentation with a variety of trapping centers in controlled concentrations. In this paper we will discuss the optical properties and holographic storage performance of transition metal-doped crystals of LiNbO_3 with particular emphasis on iron-doped materials whose sensitivity was found to be superior to all the others.

* This work was supported in part by the Naval Air Development Center, Naval Air Systems Command, Department of the Navy.

The storage mechanism in undoped LiNbO_3 crystals is described in detail in an accompanying paper³ in this issue of RCA Review. Gamma irradiation of undoped crystals produces substantial improvements in their holographic recording sensitivity by increasing the concentration of lattice defects available to act as electron traps. Impurity doping of LiNbO_3 makes possible the introduction of a new set of traps with properties differing from those of the radiation-induced defects, and perhaps more suited to specific applications.

The introduction of dopants into LiNbO_3 can be expected to affect its behavior in two ways: (1) The optical absorption may be increased and if the excitation produces a free electron, the recording sensitivity may be increased proportionally. (2) The transport parameters and other bulk properties of the material may be modified by sufficiently large concentrations of impurities, leading to concentration-dependent changes in the storage or erasure process.

In this paper we describe (1) the choice of suitable dopants for LiNbO_3 , (2) methods of material preparation, (3) the optical absorption characteristics of the doped crystals, and (4) their improved holographic storage properties. In addition we discuss several interesting, and as yet incompletely understood, phenomena that occur in the doped crystals.

2. Choice of Dopants

The characteristics required of a suitable dopant, i.e., the ability to give up an electron under visible light excitation and to later retrap an electron, are very similar to the properties of activators in inorganic photochromic materials.⁴ In fact the only difference between the two processes is that photochromic behavior relies on the retrapping of the excited electron in another center with different optical absorption than the first, whereas electro-optic storage requires charge transport and subsequent retrapping of the free electron in a different location, giving rise to a local electric field.

The useful activators in photochromic oxides are usually transition elements capable of giving up and recapturing d electrons. In particular, this is the case for alkaline earth titanate photochromic materials such as SrTiO_3 .⁵ Therefore, the transition metals were the first candidates for doping LiNbO_3 . These elements can be expected to enter the LiNbO_3 lattice substitutionally in either the Li or Nb sites, both of which are surrounded by a roughly octahedral array of oxygen atoms. This environment resembles that of the dopants in photochromic alkaline earth titanates, and it can be anticipated that some

transition metals will give rise to trapped electrons lying at optical energies below the conduction band in LiNbO_3 and capable of undergoing optically excited charge transfer.

One can expect large variations in effectiveness of different dopants, depending on the nature of the transition giving rise to the absorption at the excitation wavelength. For instance, in a material where this absorption occurs between a ground state and an excited state that is several kT in energy below the conduction band, the transition will seldom lead to a free electron and the quantum efficiency of the process will be accordingly very small. However, if the excited state is closely coupled to the conduction band, the quantum efficiency will be large.

During the course of the present study, the behavior of LiNbO_3 , doped with first-row transition elements has been examined. The elements, Cr, Co, and Ni produce only minor changes in the storage characteristics of LiNbO_3 . On the other hand Mn, Fe, or Cu doping produces large increases in both the sensitivity and maximum diffraction efficiency.

3. Crystal Preparation

Doped single crystals used in this work were grown by the Czochralski method using resistance heating of melts (usually having the congruently melting composition) contained in 100 cc platinum crucibles. The starting materials were Grade I lithium carbonate and niobium pentoxide obtained from United Mineral and Chemical Company. Dopants were added to the melt as the appropriate transition metal oxides. Concentrations given throughout this paper are the nominal values. Exploratory crystals were usually grown at rates of 5 mm per hour to a total length of 25 to 30 mm and a diameter of 12 to 14 mm. It was useful to maintain a small current (2 mA) through the crystal during growth, from melt to seed, because this resulted in reduced domain structure compared to crystals grown with no current or with current in the opposite direction. However, single-domain congruent crystals could not be obtained in this way, because, in the low vertical temperature gradient of the electric furnace, these crystals were above their Curie temperature of 1140°C for some time after growth was terminated. Following growth, the crystals were lowered into the central portion of the furnace, annealed at about 1150°C for 4 or 5 hours, then cooled at 100°C per hour to room temperature. The crystals were later poled in an oxygen atmosphere at about 1200°C . In spite of the oxygen atmosphere, however, the crystals are found to

become reduced during poling, producing coloration starting at the positive electrode and continuing down the crystal. This effect was minimized by using the lowest possible poling current; 1.5 or 2 mA was found sufficient to produce complete poling in crystals grown as described above.

The amount of reduction occurring during poling of Fe-doped LiNbO_3 is quite variable. Poled crystals usually had to be reoxidized in a separate operation before they could be used for measurements in 0.5-cm thicknesses. The amount of oxidation required was determined empirically in each case. Crystals containing large amounts of iron required higher oxidation temperatures than those only lightly doped with iron. Reduction of the crystals, when required, was accomplished by annealing in an argon atmosphere for a short period of time, e.g., one hour to 650°C for 1 or 2 mm thick samples.

Both doped and undoped crystals grown and processed as described above were consistently free of cracks and visible inhomogeneities. A series of light-scattering measurements was performed to detect short-range refractive-index variations. The results were the same for doped and undoped crystals; most of the scattering that could be detected appeared to arise from the polished surfaces of the crystals. Distortions of holographic patterns that would arise from long-range refractive index variations could not be detected.

4. Optical Properties of Transition-Metal-Doped LiNbO_3

Pure LiNbO_3 is transparent from about $5\mu\text{m}$ to 3500 Å, where the fundamental electronic absorption begins (Fig. 1). The introduction of the transition elements Cr through Cu produces visible optical absorption as shown in Figs. 2 and 3 (the curves are for 0.05 mole percent nominal doping in 0.5-cm thick samples). Two types of absorption spectra are apparent. For the elements Cr, Ni and Co, the absorption at 4880 Å and longer wavelengths is associated with well-defined absorption bands isolated from the fundamental edge. In the case of Mn, Fe and Cu, the absorption bands are not distinct from the absorption edge but appear to be a continuation of it. In fact, for Cu no independent absorption bands can be discerned.

These observations suggest that in LiNbO_3 doped with Cr, Co, and Ni, the optical absorption arises from localized electronic states, whereas for Mn, Fe, and Cu the absorption arises from states in which excited electrons are coupled to the conduction band and are only loosely bound to the transition metal. Further evidence to support this view is found in the behavior of doped crystals with respect to oxidiz-

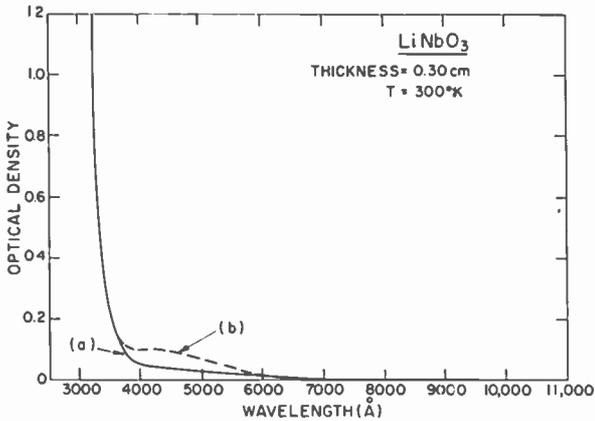


Fig. 1—Absorption spectrum of an undoped LiNbO₃ crystal (curve a) and absorption spectrum after gamma irradiation (curve b).

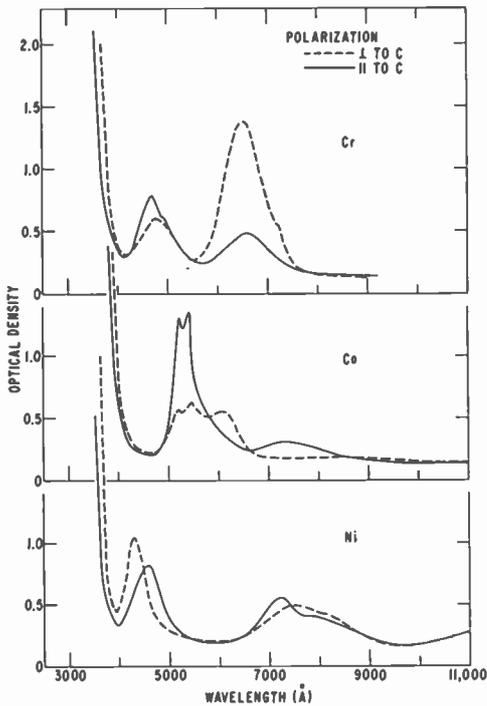


Fig. 2—Polarized absorption spectra of LiNbO₃ crystals doped with Cr, Ni, and Co. In each case the crystals are 0.5-cm thick and nominally contain 0.05 mole percent transition metal.

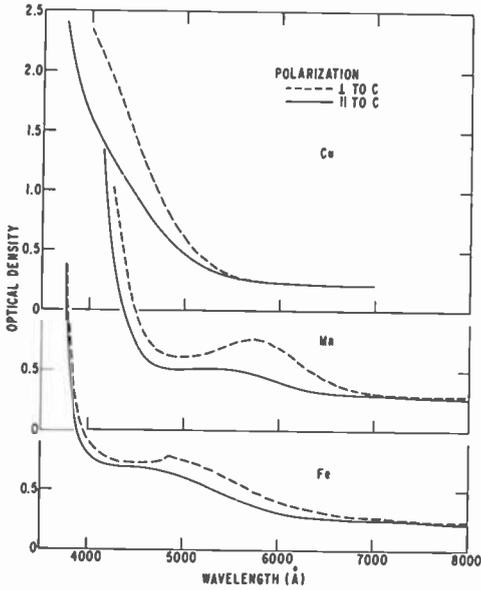


Fig. 3—Polarized absorption spectra of LiNbO_3 crystals doped with Mn, Fe, and Cu. In each case the crystals are 0.5 cm thick and normally contain 0.5 mole percent transition element.

ing and reducing heat treatments. In Fig. 4, we compare the optical absorption that is induced in doped crystals by reducing heat treatments to the absorption induced in an undoped crystal by a reducing heat treatment (each curve is the difference in absorption between a reduced crystal and the same crystal before reduction.) The absorption induced in Ni-doped crystals closely resembles that induced in

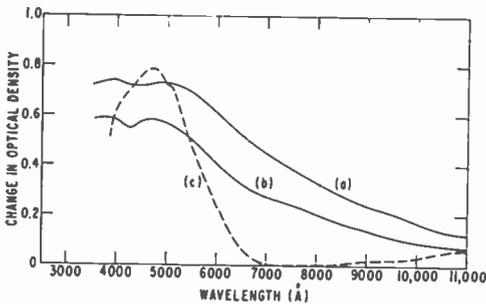


Fig. 4—The absorption spectra induced by the reduction of (a) an undoped LiNbO_3 crystal, (b) LiNbO_3 doped with Ni, and (c) LiNbO_3 doped with Fe. These curves were obtained by subtracting the absorption spectra of oxidized crystals from the spectra of the same crystals after reduction.

undoped crystals. No change in the intensity of the Ni band can be detected. The same behavior is found for Co- and Cr-doped crystals. The conclusion is that these three impurities have extremely stable valence states in LiNbO_3 . By comparing the absorption spectra of Fig. 2 to those obtained for the same ions in other materials,⁷ the probable valence states can be identified as Cr^{3+} , Ni^{2+} , and Co^{2+} .

On the other hand, the reduction of iron-doped crystals induces more of the characteristic iron absorption (see dotted curve of Fig. 4) rather than the absorption induced in similarly treated undoped crystals. The valence state of iron in fully oxidized Fe-doped crystals has been identified as Fe^{3+} by EPR techniques.⁸ The coloration of reduced Fe-doped crystals is, therefore, due to the formation of Fe^{2+} ions during reduction. The trap depth of electrons in these sites is evidently greater than that of electrons trapped at the sites responsible for the coloration of undoped LiNbO_3 crystals.

The behavior of copper-doped crystals upon reduction parallels that of iron-doped crystals. However, the intensity of the manganese absorption in the 4880 Å region decreases upon reduction, while another absorption band appears elsewhere, indicating that two valence states of Mn produce visible coloration.

In general, this group of impurities (Fe, Mn, and Cu) appears to change valence states in LiNbO_3 more easily than Cr, Co, and Ni. In fact, we have recently demonstrated reversible optically induced (photochromic) charge transfer between Fe and Mn ions and between Fe and Cu ions in double-doped LiNbO_3 crystals.⁹ This further confirms the capability of Fe, Cu, and Mn ions to fulfill the photo-induced charge transfer requirements of holographic storage.

5. Holographic Storage

5.1 Recording

As anticipated, Mn, Fe, and Cu doping produces large increases in the recording sensitivity and maximum diffraction efficiency achieved in LiNbO_3 crystals. Recording curves for samples containing these dopants are shown in Fig. 5, together with recording curves for an undoped sample and for a sample doped with Ni. These curves were obtained by measuring the intensity of the diffracted beam for brief intervals while grating holograms were being recorded in the crystals with a known power density of 4880 Å laser light and with the polarization vector aligned parallel to the *c* axis.

The largest increase in sensitivity is found in crystals doped with Fe. These crystals typically show a 500-fold increase in sensitivity

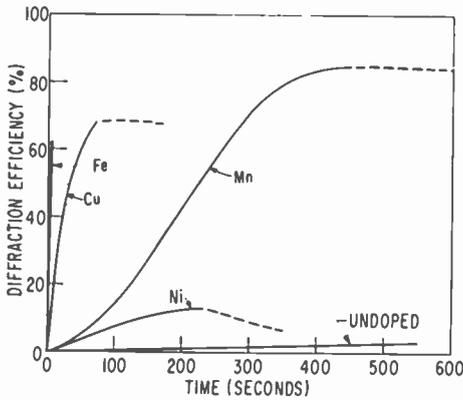


Fig. 5—Writing (solid) and erasure (dashed) curves for several LiNbO_3 crystals. Incident power density was approximately 400 mW/cm^2 .

over undoped crystals and achieve diffraction efficiencies approaching 100%. Cu- and Mn-doped crystals also exhibit increased sensitivity and high diffraction efficiencies. On the other hand, Ni, Cr, and Co doping produces only minor changes in the recording sensitivity of the crystals. The curve shown in Fig. 5 for Ni is representative of Cr and Co as well.

The sensitivity of iron-doped crystals has been found to depend approximately linearly on the intensity of the Fe^{2+} absorption at 4880 \AA for all iron concentrations in the range 0.015 to 0.1 mole percent. This is illustrated in Fig. 6 in which the inverse of the incident energy required to record a hologram with 40% diffraction

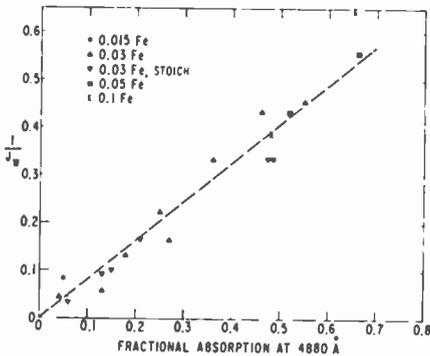


Fig. 6—The writing sensitivity, $1/J_w$, plotted against fractional absorption for several iron-doped crystals having undergone varying degrees of oxidation or reduction. The crystal marked "stoich" was grown from a stoichiometric melt, the rest from congruent melts. J_w is the incident energy density (J/cm^2) at 4880 \AA required to achieve 40% diffraction efficiency.

efficiency ($1/J_w$) is plotted against the fractional absorption (corrected for reflection losses) measured at 4880 Å for a number of different crystals (5-mm thick) that have been subjected to varying degrees of oxidation or reduction. The fact that the data points tend to cluster about a straight line through the origin, regardless of the concentration of Fe in the crystals, indicates that a constant diffraction efficiency is produced per unit of energy absorbed by Fe²⁺ ions. About 1.25 joules/cm² of absorbed energy is required to produce a 40% diffraction efficiency. This strongly supports the view that holographic storage in iron-doped crystals arises from the photoionization of Fe²⁺ ions. Presumably, the free electrons are subsequently retrapped by Fe³⁺ ions in regions of low light intensity. The storage mechanism in Cu- and Mn-doped crystals is believed to be similar to that in Fe-doped samples.

Crystals containing Mn, Cu, and Fe are sensitive to light throughout the visible. This is particularly true for the iron-doped crystals (see Fig. 3). Thus, one can record color holograms by using three (or more) separate laser lines covering the red, blue, and green regions of the spectrum. We have accomplished this with an argon-krypton mixed gas laser. The color hologram was composed of three holograms, recorded within the same volume by the 4880 Å (blue), 5145 Å (green), and 6471 Å (red) laser lines. During reconstruction with all three lines, a particular laser wavelength reads out only its corresponding hologram, and the combination of all three diffracted beams reproduces the color information.

Since the absorption of the doped samples generally decreases for larger wavelengths, there is a corresponding drop in the storage sensitivity toward the red. This is accentuated by the inverse dependence of the diffraction efficiency on the wavelength.¹⁰ Measurements on Fe-doped samples show that the sensitivity to incident light (polarized along the *c* axis) decreases by roughly a factor of 50 going from 4880 Å to 6328 Å. Of course, the red sensitivity can be increased by using crystals with higher absorption, particularly where good transmission in the blue is not required.

5.2 Erasure

Holograms stored in Fe-doped LiNbO₃ decay thermally at room temperature with characteristic lifetimes of several weeks. They can also be erased by visible light excitation. Optical erasure at 4880 Å shows a sensitivity roughly an order of magnitude lower than the writing sensitivity for a given crystal. This is illustrated in Fig. 7

which $1/J_e$, where J_e is the energy density (J/cm^2) required to erase a hologram to half its original intensity, is plotted against the absorption coefficient at 4880 \AA in a manner analogous to Fig. 6. To obtain meaningful optical erasure data, one must minimize the self-enhancement effect.¹¹ This has been done by taking the measurements in Fig. 7 with an off-angle coherent-light beam that was periodically scanned through the proper readout angle.

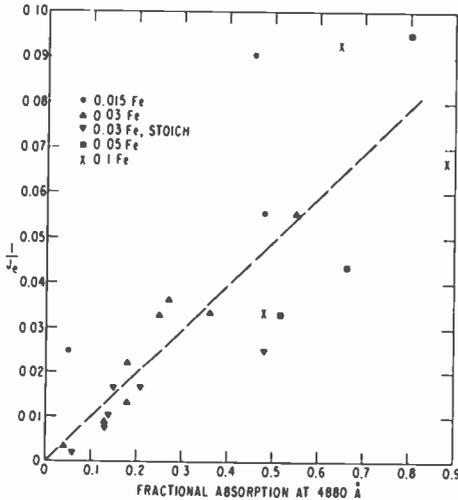


Fig. 7—The erasure sensitivity $1/J_e$ plotted against fractional absorption at 4880 \AA for several iron-doped crystals, as described for Fig. 6. J_e is the energy required to erase a hologram to half its original intensity.

On the basis of the storage mechanism discussed earlier, the erasure process would also involve photoexcitation of Fe^{2+} ions. However, the lower sensitivity for erasure as compared to writing may arise from an additional complexity. The charge-migration mechanism in iron-doped crystals involves drift in an internal electric field. It can be shown³ that this enhances the recording process but not the erasure process, thus making the record sensitivity greater than the erasure sensitivity.

The erasure sensitivity of Mn- and Cu-doped crystals is also smaller than the record sensitivity for these crystals, indicating that similar consideration may apply to these dopants, but this aspect of their behavior has not yet been studied in detail.

An important advantage of the relative insensitivity of doped crystals to erasure is that multiple holograms can be recorded more

easily (the erasure problem encountered in the multiple-storage process is minimized¹²). This erasure problem arises because each hologram is recorded in the same volume of the crystal, with only the angle of the crystal relative to the writing beams being changed. Thus when a new hologram is recorded, the writing beam tends to erase those holograms already recorded at other angles. The slow erasure inherent to the doped crystals significantly alleviates this problem.

5.3 Thermal Fixing

Although the erasure time is relatively long in doped crystals, continued readout will ultimately completely erase the hologram. This is not desirable in some applications, such as read-only memories. One solution to this problem is the thermal fixing technique previously demonstrated in undoped crystals.¹³ To fix a hologram or a set of holograms, one heats the crystal to $\sim 120^\circ\text{C}$ for 5 minutes or more in the dark, and then cools it to room temperature. This converts the electronic charge pattern of the original hologram into a charge pattern consisting of ionic defects. Since the latter is insensitive to light, the hologram is fixed and does not decay during readout. With this technique, high-diffraction-efficiency holograms of detailed patterns have been successfully fixed in undoped crystals.

However, the performance of doped crystals where the hologram has been fixed in this manner has been less than satisfactory, since the diffraction efficiency of the fixed hologram is typically much smaller than that of the original hologram. Reductions of one to two orders of magnitude are common for samples doped with 0.01% or more of iron. One way to circumvent this problem is to use samples with lower doping. This reduces the density of occupied traps and allows more complete reconstruction of the holographic field after fixing.¹³ We have found that in samples doped with only 0.001% of iron, more than $\frac{1}{3}$ of the diffraction efficiency is retained after the fixing treatment. However, this improved performance is obtained at the sacrifice of sensitivity, because lightly doped crystals have less optical absorption. Within practical limits, the sensitivity can be increased by using thicker samples.

5.4 Optically Induced Scattering

Another problem encountered in transition-metal-doped samples of LiNbO_3 exhibiting high sensitivity is the appearance of optical scattering induced by exposure to coherent light. This phenomenon occurs

during exposure to a single or to two superimposed beams. The scattered light pattern takes the form of two lobes with their axes aligned along the c axis. Uniform incoherent light does not give rise to scattering and, in fact, erases it. It was originally thought that the scattering arises from interferences due to internal reflections in the crystal, but it was later found that antireflection coatings on the

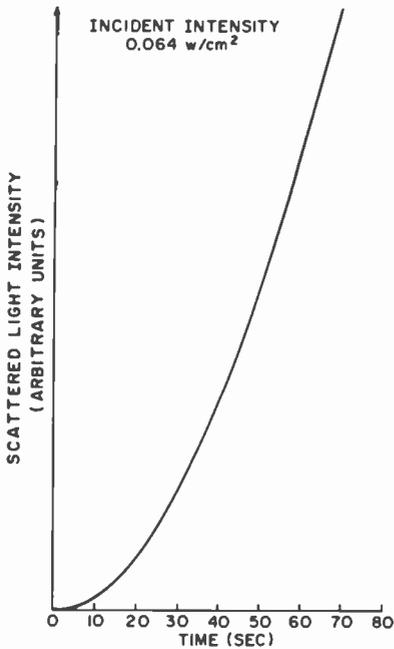


Fig. 8—Scattered light intensity as a function of time for illumination with a uniform beam at an angle of 15° to the surface.

samples do not alleviate the problem. Fig. 8 illustrates the increase in scattered light intensity as a function of time for illumination with a nominally uniform beam having an intensity of 0.064 W/cm^2 in a sample doped with 0.05% Fe. The incident angle was 15° with the normal to the surface of the crystal, and the polarization was in the plane containing the c -axis. The initial build-up follows a square law as does the build-up of stored holograms. The fact that the scattering appears only during exposure to coherent light, and that it shows pronounced angular selectivity (see Fig. 9) suggests strongly that the process is at least triggered by interference fringes present in the beam.

The angular dependence shown in Fig. 9 is consistent with the presence of fringes with an approximate $10\ \mu\text{m}$ spacing. The fact that such widely spaced fringes can be recorded is consistent with other indications that in doped crystals the storage process is associated with an internal electric field (see Sec. 5.2). When diffusion is the transport mechanism involved in storage, as is the case for γ -irradiated undoped crystals, the thermal field associated with intensity variations over a $10\ \mu\text{m}$ range is so small that this effect is not observed.

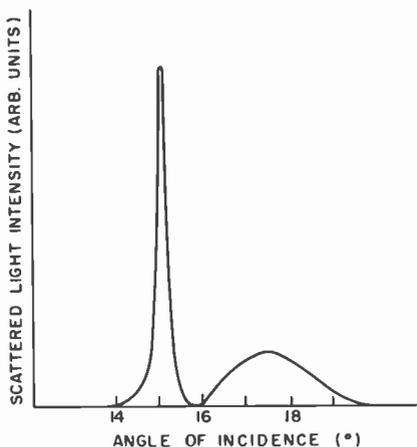


Fig. 9—Variation in scattered light intensity as a function of angle of incidence of readout beam. Scattering was caused by a beam at 15° to surface, corresponding to the center of the main scattering peak.

One way to overcome the scattering problem in doped crystals is by optical erasure. When the sample is rotated to a new angle, the incident beam tends to erase the scattering that occurs at other angles. This approach is suitable for readout of a number of fixed holograms in lightly doped crystals. The angular rotations required for access of the holograms could prevent any overall scattering problem from building up.

6. Diffusion of Transition Metals into LiNbO_3 and LiTaO_3

In addition to doping during growth, LiNbO_3 can be successfully doped by diffusion. This also is the case for LiTaO_3 , which we will also discuss in this section. Diffusion doping has been demonstrated for Fe and Cu in both LiNbO_3 and LiTaO_3 , and for Ni in LiTaO_3 . The

method used was as follows. Thin layers (1000-2000 Å) of the appropriate metal were evaporated onto crystal wafers, which were then annealed in argon for 48 hours. The temperatures used were 1200-1230°C for Fe and Ni, and 1000°C for Cu. Although the crystals became depoled when their Curie temperature was exceeded, single-domain regions could be found that were suitable for testing without the necessity for repoling the entire crystal. The penetration depth of Fe in LiNbO_3 at the end of this time is estimated to be 50-100 μm .

The absorption spectra of diffused crystals were comparable to those doped by ordinary means. The colored regions of the crystal corresponded to the shape of the evaporated film. The coloration boundary was quite sharp for Fe in LiNbO_3 but less so for Fe in LiTaO_3 , indicating that diffusion proceeds more easily in the latter. Copper was found to diffuse more readily than iron in both hosts.

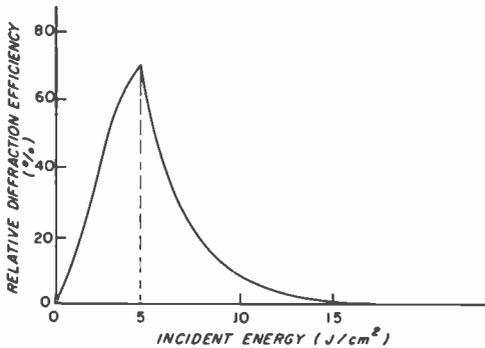


Fig. 10—Writing and erasure curves for a LiNbO_3 plate doped with iron by diffusion.

The writing and erasure curves for an iron-diffused LiNbO_3 sample are shown in Fig. 10. The writing sensitivity and diffraction efficiency are seen to be comparable to material doped by conventional means. The notable feature of Fig. 10 is that the erasure sensitivity is nearly equal to the writing sensitivity, a situation that cannot be duplicated in material doped during growth. This is a very useful feature for a number of applications, such as read-write memories. The high erase sensitivity of Fe-diffused LiNbO_3 is duplicated in LiTaO_3 . Copper-diffused crystals, on the other hand, behave in a manner similar to material doped in the conventional way. The explanations for this behavior may lie in the fact that there is a high local concentration of Fe near the surface of the diffused crystals, greatly exceed-

ing that which can be obtained in conventionally doped crystals. Since copper diffuses more rapidly than iron, it is not subject to as high surface concentrations and so does not exhibit the high erasure rates.

To determine if the thermal conditions during iron diffusion were responsible for its high erasure sensitivity, a number of conventionally doped LiNbO_3 crystals were annealed at the same time the diffusions were carried out. These crystals did have increased erase sensitivity, but it was still a factor of three smaller than in the diffused crystal. Holograms stored in Fe-diffused LiNbO_3 are found to decay thermally (in the dark) over a period of a few hours. However, Fe-diffused LiTaO_3 is quite stable thermally, showing about 10% decay in 16 hours.

7. Summary and Conclusions

It has been shown that the holographic storage performance of LiNbO_3 crystals can be increased significantly by doping with transition metal oxides. The most effective dopant, iron, results in a 500-fold increase in sensitivity over undoped LiNbO_3 . Typical iron-doped crystals require an incident energy of approximately 150 mJ/cm^2 to produce a hologram with a 1% diffraction efficiency. This figure compares favorably to other reusable storage media, and is better than the sensitivity of hologram storage in most photochromic materials. However, whereas photochromic media are limited to maximum diffraction efficiencies of a few percent, the maximum diffraction efficiency in doped LiNbO_3 approaches 100%.

The erasure sensitivity of doped LiNbO_3 crystals is about a factor of 10 lower than the record sensitivity. This characteristic is very useful for applications calling for the recording of a number of holograms in the same volume, or for repeated readout of a stored hologram. However other erasure characteristics can be obtained through changes in the processing of the material. For example in LiNbO_3 doped with iron by diffusion, the erasure sensitivity is about equal to the record sensitivity. Alternately, in lightly doped crystals, fixing techniques can be applied to produce essentially permanent, high efficiency stored holograms. Thus the properties of iron-doped crystals can be tailored to fit a number of different applications.

It should be emphasized that the work reported herein only begins to explore the potential of LiNbO_3 . It can be anticipated that other dopants or combinations of dopants will be found that further enhance the sensitivity of the material, and that new dopings or processing techniques may result in even more flexibility than is exhibited by the present materials. Finally, the results presented in this paper apply

only to LiNbO_3 (and to some extent to LiTaO_3). There is evidence that dopants can play an equally important role in hologram storage in other electrooptic materials, some of which may be found to have greater potential than LiNbO_3 .

Acknowledgments

It is a pleasure to acknowledge the skillful assistance of G. Latham, L. Levin, and R. Wer in the experimental work.

References:

- ¹ F. S. Chen, J. T. LaMacchia, and D. B. Fraser, "Holographic Storage in Lithium Niobate," *Appl. Phys. Letters*, Vol. 13, p. 223, 1968.
- ² J. J. Amodei, W. Phillips, and D. L. Staebler, "Improved Electro-Optic Materials and Fixing Techniques for Holographic Recording," *Appl. Optics*, Vol. 11, p. 390, 1972.
- ³ J. J. Amodei and D. L. Staebler, "Holographic Recording in Lithium Niobate," *RCA Review*, this issue.
- ⁴ R. C. Duncan, Jr., B. W. Faughnan, and W. Phillips, "Inorganic Photochromic and Cathodochromic Recording Materials," *Appl. Optics*, Vol. 9, p. 2236, 1970.
- ⁵ B. W. Faughnan, D. L. Staebler, and Z. J. Kiss, "Inorganic Photochromic Materials," *Applied Solid State Science*, Vol. 2, Academic Press, N. Y., N. Y., 1971.
- ⁶ J. R. Carruthers, et al, "Nonstochiometry and Crystal Growth of Lithium Niobate," *J. Appl. Phys.*, Vol. 42, p. 1846, 1971.
- ⁷ D. S. McClure, *Electronic Spectra of Molecules and Ions in Crystals*, Academic Press, N. Y., N. Y. (1959), p. 129ff.
- ⁸ G. E. Peterson, A. M. Glass, and T. J. Negran, "Control of the Susceptibility of Lithium Niobate to Laser-Induced Refractive Index Changes," *Appl. Phys. Letters*, Vol. 19, p. 130, 1971.
- ⁹ W. Phillips and D. L. Staebler (to be published).
- ¹⁰ H. Kogelnik, "Coupled Wave Theory for Thick Hologram Gratings," *Bell. Syst. Tech. J.*, Vol. 48, p. 2909, 1969.
- ¹¹ D. L. Staebler and J. J. Amodei, "Coupled-Wave Analysis of Holographic Storage in LiNbO_3 ," *J. Appl. Phys.*, Vol. 43, p. 1042, 1972.
- ¹² D. L. Staebler, J. J. Amodei, and W. Phillips, "Multiple Storage of Thick Phase Holograms in LiNbO_3 ," to be presented at the 1972 Quantum Electronics Conf., Montreal, Canada, May 1972.
- ¹³ D. L. Staebler and J. J. Amodei, "Thermally Fixed Holograms in LiNbO_3 ," *Ferroelectrics*, Feb. 1972.

Phase Holograms in Dichromated Gelatin

Dietrich Meyerhofer

RCA Laboratories, Princeton, N. J.

Abstract—Dichromate gelatin is one of the most promising recording materials for phase holograms because of the high efficiency and low noise obtainable. We describe the general properties and requirements for thick phase holograms and how dichromate gelatin fulfills these requirements. Reliable processing procedures for obtaining optimum results have been evolved. The total available phaseshift allows the superposition of up to 20 efficient holograms. The variation of gelatin thickness during the processing was studied to determine the conditions for forming the best images. Interesting applications of this material are for lenses to perform multiple imaging and photocomposition. Also, high efficiency copies can be made from low-efficiency holograms formed in more sensitive materials.

Introduction

Recently, dichromated gelatin (DCG) has been developed as potentially one of the best holographic recording materials.¹ Efficiencies close to the theoretical limits have been obtained.^{1,2} The scattering noise can be very small,³ so that high signal-to-noise ratios are possible in the reconstructed images. Some of the disadvantages of this material are the low sensitivity and the difficulties in controlling the properties and development of the emulsion. Also, the mechanism of photochemistry is only poorly understood.⁴

In a previous paper,⁵ we discussed the photochemical properties of thin, unhardened layers of DCG that behave like negative photoresist material. Its application to holography is limited by its resolution (the spatial frequency response starts to decrease around 200 line-pairs/mm). The present report is concerned with hardened layers of DCG where most of the optical path length variation takes place in the "interior" of the emulsion. Since we use spectroscopic plates as starting materials,⁶ these are relatively thick holograms. We will discuss the properties and processing of these layers and the kinds of high resolution holograms and holographic optical elements that can be produced with them.

Because the properties of phase holograms are not as widely known, we will first discuss the characteristics of the various types of holographic recording materials and then describe the specific properties of DCG and the kinds of holograms that can be made with it.

Holographic Recording Materials

Holographic recording materials have generally been divided into four different categories according to the following two criteria: is the information recorded as a variation of absorption (absorption hologram) or of phase (phase hologram) and is the recording medium thin or thick? We will discuss these distinctions in more detail below. It must be emphasized however, that these are not clear-cut separations, in that most recording media have both absorption and phase effects and the thickness often has some intermediate value. In addition to these categories, a distinction is also made depending on the recording procedure, which may cause the reference and object beams to impinge on the recording medium either from the same side (transmission hologram) or from the opposite sides (reflection hologram).

Consider first the differences between *absorption* and *phase* recording. We can make some general statements without going into the details of the recording materials or processes. When modulating the absorption, there is a well-defined upper limit for the modulation and consequently for the modulation depth, namely the difference between no absorption and complete absorption. Furthermore, at reasonable modulation depths, a substantial part of the incoming light beam must be absorbed and cannot contribute to the reconstructed image beam. In contrast to this, a material that records a signal as a modulation of phase does not have any inherent limitation to the modulation depth. The amount of phase change can in principle be made as large

as desired by increasing the amount of material to be modulated. At the same time, such a material does not absorb any light; all the incident light is available to form the image beam. Thus, a phase material can be expected to produce a higher efficiency hologram (as compared to absorption materials) and to allow multiple exposures.

The two kinds of modulation also differ in the fidelity with which the image can be reproduced. Consider the basic hologram geometry shown in Fig. 1. If r and o are the wavefronts of reference and object

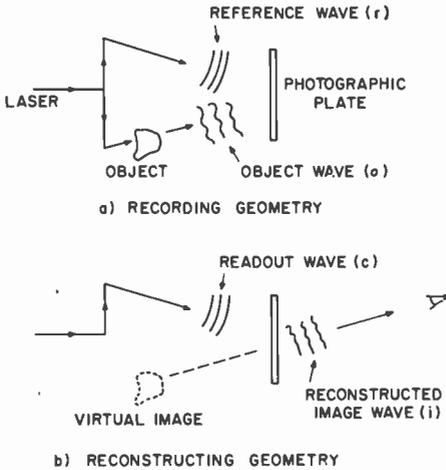


Fig. 1—Hologram geometry: (a) recording (b) readout.

beams, respectively, impinging on the light-sensitive medium (i.e., their amplitudes), then the total intensity is given by

$$\begin{aligned}
 I_{in} &= (r + o) (r + o)^* \\
 &= |r|^2 + |o|^2 + (ro^* + r^*o).
 \end{aligned}
 \tag{1}$$

The exposure r and o produces a modulation of the developed medium that varies from point to point

$$M = M(I_{in}),
 \tag{2}$$

which may be a change in absorption or phase or both. Considering *thin* holograms first, I_{in} needs to be specified only in the plane of the

hologram. When the hologram is now inserted in the setup as in Fig. 1(b) and illuminated by a readout beam c , the resulting image beam i may be written

$$i = (M) (c). \quad [3]$$

In the case where M represents *absorption*, we write it in terms of amplitude transmittance T_A

$$i = T_A c. \quad [4]$$

Consider the simplest case of a linear recording medium, i.e.,

$$T_A = t_o - t_1 (I_{in}). \quad [5]$$

Also, let the readout beam be the same as the reference beam $c = r$. Then Eq. [4] becomes

$$i = t_o r - t_1 (|r|^2 + |o|^2) r - t_1 (r)^2 o^* - t_1 |r|^2 o. \quad [6]$$

The first two terms, proportional to r , represent the attenuated readout beam (or zero order), the third term is the conjugate real image (-1 order), and the fourth is the virtual image ($+1$ order). The latter term, $i_v = \text{const} \cdot o$, shows that if $|r|^2$ is constant, the object may be reconstructed without distortion by the appropriate holographic technique. If the recording medium is not linear, then Eq. [5] only represents the first terms of a series expansion. The higher order terms will cause additions to Eq. [6] in the form $t_2 |r|^2 |o|^2 o$, etc., which must be added to i_v and which represent distortion.

When M represents a *phase* variation, the situation is very different:

$$i = \exp \{i\phi\} c. \quad [7]$$

If the modulation is linear ($\phi = f I_{in}$) and $c = r$, this becomes

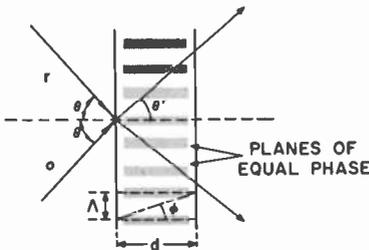
$$i = \exp \{if (|r|^2 + |o|^2)\} \exp \{if (r o^* + r^* o)\} r. \quad [8]$$

It does not produce a single term $i_v = \text{const} \cdot o$; there are always higher order terms. If I_{in} is not too large, so that the exponent can be expanded, the lowest order terms of the virtual image (the one proportional to o) becomes

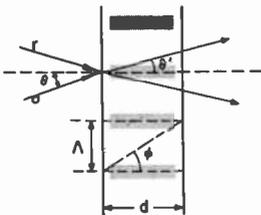
$$i_v = if|r|^2 \exp\{if|r|^2\}[1 + if|o|^2\{1 + (if/2)|r|^2\}$$

+ higher order terms in $(f|o|^2)o$. [9]

The $(f|o|^2)$ terms represent unwanted signal or noise, which can only be eliminated by using small $(f|o|^2)$ values, and therefore small efficiencies.



a) THICK HOLOGRAM: $\theta' > \phi$



b) THIN HOLOGRAM: $\theta' < \phi$

Fig. 2—Definition of thin and thick holograms (Kogelnik criterion). The angles θ and θ' will differ if the refractive indices inside and outside the emulsion differ.

These considerations apply only to thin holograms because Eqs. [4] and [7] describe a two-dimensional recording surface. *Thick* holograms consist of phase variations distributed throughout the volume of the material, usually produced by a changing index-of-refraction. (There are also thick absorption holograms but they are inefficient and will not be considered here.) In this case, the reconstruction efficiency for the first-order virtual image can approach 100%, as compared to a maximum of 30-40% for thin phase holograms. This can be shown qualitatively in the case of a simple sinusoidal grating produced by two plane waves, as sketched in Fig. 2(a). When the readout beam c is parallel to the reference beam r , the image beam i

will be parallel to the object beam o . Then the planes of constant index-of-refraction n are oriented in such a way that c is reflected into i . Since each change in n produces some reflection, and c traverses many of these index "planes," it becomes apparent that i can become very large. The other diffracted beams (-1 , $+2$ orders, etc.) do not have this mirror geometry and will be much weaker.

These considerations can be put on a quantitative basis. The two beams forming the hologram in Fig. 2(a) are plane waves making angles θ' with the y axis (the x -axis is along the emulsion surface in the plane of the figure):

$$\begin{aligned} o &= o_0 \exp[ik(x \sin \theta' + y \cos \theta')] \\ r &= r_0 \exp[ik(-x \sin \theta' + y \cos \theta')] \end{aligned} \quad [10]$$

Here, $k = 2\pi/\lambda'$, λ' is the wavelength in the medium. In this case, M of Eq. [2] represents a change in dielectric constant $\Delta\epsilon$. Inserting Eq. [10] in Eq. [1] one obtains

$$\Delta\epsilon = \Delta\epsilon_0 + \epsilon_1 \cos Kx, \quad [11]$$

where $K = 2k \sin \theta' = 2\pi/\Lambda$, Λ is the spacing of the fringes (Fig. 2). Since the modulation is always small, $\Delta\epsilon \ll \epsilon_0$, we write

$$\epsilon = \epsilon_0 + \epsilon_1 \cos Kx. \quad [12]$$

When a beam c is incident on the developed hologram, the modulation varies with distance and it is necessary to solve the wave equation in order to calculate i :

$$\epsilon k^2 E = \nabla^2 E. \quad [13]$$

From the above discussion, we know that as long as c is the same as r , one needs to consider only one reconstructed order i , so that one can write

$$E = c(y) + i(y), \quad [14]$$

with the boundary conditions

$$c(0) = c_0 \quad i(0) = 0, \quad [15]$$

Solving Eq. [13] for c and i yields the diffraction efficiency at the

output plane ($y = d$) for the special case $c = r$

$$\eta = \frac{i i^*}{c c^*} = \sin^2 \frac{\pi}{2} \frac{\epsilon_1}{\epsilon_0} \frac{d}{\lambda'} \quad [16]$$

This shows that 100% efficiency is reached for the first time when

$$d = \frac{\lambda'}{\epsilon_1 / \epsilon_0} \quad [17]$$

One can now obtain an approximate expression for the small higher order terms (i_n) and one finds⁷ that their intensities have the maximum values of

$$\frac{(i_n i_n^*)_{\max}}{c c^*} = \frac{1}{16} \left(\frac{\epsilon_1}{\epsilon_0} \frac{\Lambda^2}{\lambda'^2} \right)^2 \quad [18]$$

This permits one to define a thick hologram as one in which the value of Eq. [18] is much smaller than 1. Eliminating the modulation ϵ_1 from Eqs. [17] and [18] and expressing the result in terms of the factor Q used by Hargrove⁸ in calculating diffraction from ultrasonic acoustic waves,

$$Q = \frac{2\pi \lambda' d}{\Lambda^2},$$

one finds that for a thick hologram,

$$Q > 10. \quad [19]$$

The condition can be understood intuitively by considering Fig. 2. Using the angles $\phi = \tan^{-1} (\Lambda/d)$ and θ' , Eq. [19] is changed to

$$\theta' > \phi.$$

which evidently means that in a thick hologram c must go at least once through the entire modulation range.

When the hologram is more complicated than a simple grating, Eq. [12] becomes more complex. One can, however, expand it in a Fourier series of plane wave gratings and one then finds that a max-

imum efficiency of 100% can still be obtained in principle (even for diffuse object beams) if Eq. [19] is fulfilled. Under these conditions, there will be noise in the reconstructed image due to the $|\sigma|^2$ term, which decreases both as the reference-to-object beam ratio increases and as the hologram is made thicker.

It is the possibility of high-efficiency, low-distortion reconstruction that makes thick phase holograms so attractive compared to other kinds of holograms. To turn this possibility into reality, one requires materials with adequate or, preferably, high sensitivity, large enough modulation to allow at least a π phaseshift in the linear range of exposure, and good stability. By the last we mean that the material must have the same dimensions after development that it has during exposure. Also, the developed material should be transparent to the reconstructing radiation and not be affected by it.

Because of difficulties with one or more of these requirements, thick phase holograms have not yet come into widespread use. Bleached silver halide emulsions have been studied most thoroughly. Good results have been obtained,⁹ but the processing requirements are very severe, and effects of the reconstruction radiation are difficult to eliminate completely. Other materials that have been described make use of photochromic, electro-optic, or other photoelectronic effects.¹⁰ Because of the low values of modulation per unit thickness, crystals or other very thick materials must be used with concomitant advantages and disadvantages. They usually have low sensitivities and the quality depends strongly on difficult material development. On the other hand, very thick holograms have the advantage of high angular sensitivity, so that many holograms can be superimposed in the same space and read out separately *if* the total phase shift available is appropriately large (π phase shift for each image).

The material to be discussed here, dichromated gelatin, has many properties that promise to make it almost ideal for recording of thick phase holograms. We use Kodak 649F spectroscopic plates as starting material. Of the approximate gelatin thickness of 12 μm only about 4 μm are used in the hologram formation, as will be shown below. Inserting this value and a typical laser wavelength into Eq. [19], one finds that a thick hologram will be recorded if the smallest angle between the object and reference beam is 23° or larger.

Processing Dichromated Gelatin for Thick Phase Holograms

We now describe the procedure for sensitizing and developing dichromated gelatin plates based on Kodak 649F spectroscopic plates as

starting material.⁶ The procedure, summarized in Table 1, is based on the results of numerous trials throughout an extended period of time. This sequence of steps was found to produce optimum results with fair reproducibility. There are occasional variations in the results that appear to be caused by variations in the gelatin obtained from the spectroscopic plates.¹¹

Before we consider the various steps in the process, we summarize what is known about photochemical mechanism of DCG. Dichromated colloidal materials are among the earliest photosensitive materials known.¹² However, the details of the photochemical processes are still only poorly understood. In sensitizing the gelatin by dissolving $(\text{NH}_4)_2\text{Cr}_2\text{O}_7$, or some other dichromate, in it, the chromium ion is incorporated as Cr^{6+} . During exposure, the absorption of light changes this to some lower ionization state, possibly Cr^{3+} . The reduced ion then causes cross-linking and hardening of the gelatin. This is accompanied by small changes of optical pathlength in the exposed regions, so that weak phase recordings can be made.⁵

If the original gelatin layer was unhardened, then the exposed plate can be developed by washing away the remaining unhardened material. This technique is used to make etching resist patterns and relief phase images and holograms.⁵ In the present process, the layer is pre-hardened, so that a water wash after exposure does not remove the gelatin, but only the excess $(\text{NH}_4)_2\text{Cr}_2\text{O}_7$. The wash causes an increase in optical pathlength difference, but the recording is still rather weak.

The crucial step in producing the strong phase recording is the immersion in isopropyl alcohol following the wash. It increases the optical pathlength difference to such a large value that efficient phase holograms can be produced. It is known that part of the phase change is produced by having isopropyl alcohol molecules securely bound to reduced Cr sites where the gelatin is cross-linked.¹³ This complex has an absorption peak in the near-ultraviolet and therefore exhibits an increased dielectric constant in the blue region of the spectrum. The change can reach a value $\Delta n = 0.01-0.02$ for a fully exposed DCG layer.¹ The maximum shift in optical pathlength in a 15 μm thick layer is therefore 0.3 μm , or 0.6 λ of blue light. This is about what is required to reach 100% efficiency. However, the measured maximum pathlength difference is much larger than that, since many high efficiency holograms can be superimposed on the same 15 μm plate, or single high-efficiency holograms can be made in layers as thin as 1 μm .⁴ The latter implies a Δn of the order of 0.5, which can only be obtained by actual removal of a substantial amount of gelatin in some areas.

Table 1—Processing of Dichromate Gelatin Plates**A. Preparation of 649F Plates.**

1. 15 min. in Kodak fixer with hardener.
2. 10 min. wash in running water (20-25°C).
3. 10 min. in methyl alcohol or until plate is clear (agitate).
4. 10 min. in clean methyl alcohol.
5. Dry in vertical position.

B. Preparation of Sensitizing Solution.

1. 50g/l of Baker, Merck or equivalent ammonium dichromate fine crystals (use distilled water).
2. Filter solution before use. There is no limit as to how many plates one can sensitize with each solution, but after storing one week, a new solution should be used. The use of Kodak FotoFlow 0.2 oz/gal is optional.

C. Sensitizing Plates.

1. Soak plates 5 min. in dichromate solution in flat pan with emulsion side up.
2. Tilt plate ~10° and allow excess to run off (about 3 min.) Remove residue at edge with a paper towel.
3. Store in light tight box with same tilt as in (2) until ready to use.
4. Expose between 15 hours and 40 hours after sensitizing.

D. Development.

1. Wash 10 min. in running water at 20°C.
2. Soak 2 min. in mixture 50% isopropyl alcohol and 50% distilled water, agitated.
3. Repeat in 90% isopropyl alcohol-10% water mixture.
4. The hologram is mounted on a stainless-steel plate with gelatin layer away from the plate, and inserted vertically into fresh isopropyl alcohol with agitation for 10-20 min.
5. The plate is pulled out of the container at a rate of about 1 cm/min. At the same time, a flow of hot air is directed against the gelatin.

Shankoff⁴ confirmed that there appeared to be voids in the gelatin as observed under the microscope, and the holograms could be wiped out reversibly by immersing in a noninteracting index-matching fluid. It must therefore be assumed that this immersion in isopropyl alcohol and the consequent rapid dehydration and shrinkage of the gelatin causes it to tear apart.

We now discuss the processing outlined in Table 1. The first series of procedures is designed to produce as clean a gelatin plate as possible. The reason for starting with photographic plates rather than coating clean plates with gelatin is that this procedure leads more easily to uniform and repeatable layers. It does, however, contribute variability due to changes in gelatin quality from batch to batch.

The sensitizing solution is chosen to incorporate the highest concentration of dichromate in the gelatin without causing undesired side effects. Such effects produce noise in the final image. During the first stage of the drying process, the solution is allowed to run off the almost horizontal plates. This results in a uniform concentration except for the lowest region of the plate (about 1-cm wide). If the plates are exposed too soon after sensitizing, they tend to "crystallize" during development. By this, we mean that regions of the gelatin appear disturbed as if one of the components in the layer had crystallized. Such regions cause extra scattering and, therefore, noise in the final image. If the sensitized plates are kept too long before exposing, they start to lose sensitivity.

The real "magic" of this process is in the development step. The water wash serves two purposes. First, it softens the gelatin for the following drying step. Secondly, it removes all the unreacted dichromate. This is essential to preventing crystallization and the introduction of scattering centers upon drying. For 4×5 inch plates, 10 minutes at room temperature seems to be adequate. Lin⁶ suggested some additional procedures to improve the cleaning, but we find that they add to the difficulties of making the process reproducible and reduce the sensitivity; further, they are not necessary if the washing is done carefully.

The washed plate is immersed wet into the isopropyl alcohol. There seems to be some advantage to making the first step a mixture of alcohol and water, but we have not been able to document this conclusively. In any case, at least two steps must be used so that the last bath remains as free of water as possible. We have found that dehydration is complete after two minutes for 4×5 -inch plates. However, for larger plates (8×10 inches), the process takes a much longer time, as it appears that the reaction proceeds from the edges of the plate to the interior in a diffusion-controlled mechanism. This can be observed visually, but we cannot offer an explanation for it. It is known¹ that using the water and alcohol at elevated temperature increases the sensitivity, but we have observed that it also greatly increases the scattering of the layer. In fact, the gelatin can take on a milky appearance.

The final step of removing the alcohol is more critical than had first been suspected. When plates are removed from the alcohol and dried in air, the results are very sensitive to the humidity in the air. Low humidity encourages crystallization, while high humidity reduces the hologram efficiency. The problem was overcome by pulling the plates slowly out of the bath and drying them with a stream of

warm air directed at the liquid interface. When the plate is well cleaned, the interface forms a uniform straight line, and the dried plate has a uniform appearance and efficiency. After the drying is complete, the plates are no longer affected by relative humidity up to 90%. We have found hologram efficiencies to be unchanged after 2 years storage in room atmosphere and ambient light.

One of the characteristics of these thick DCG plates is that the development process can be repeated if the first attempt is unsatisfactory, e.g., due to crystallization. As long as the damage is not too great, the second hologram, if successful, will have the same efficiency and low scattering as the first would have had. Repeated reprocessings, however, usually result in a drop in efficiency.

Experimental Parameters of DCG Holograms in 649F Plates

Phaseshift Available

Since the available phase shift in a phase-recording material may be many times 2π , it allows the storage of a number of high efficiency holograms superimposed. To determine the maximum phase shift, one needs to know how much material is exposed, how many photosensitive centers there are, and how much phase shift each exposed center produces after development.

Sensitized, unexposed 649F plates are about $12\ \mu\text{m}$ thick and transmit 2 to 5% of 441-nm light (all exposures were made at this wavelength with He-Cd laser). This means that the absorption coefficient $\alpha = (2.5\text{--}3.3) \times 10^3\text{cm}^{-1}$ and that absorption is nonuniform. Only a layer about $4\ \mu\text{m}$ thick will be exposed and form the holograms.

We previously measured the absorption cross-section of the dichromate complex to be $\sigma = 1.5 \times 10^{-18}\text{cm}^2$ for 441-nm light.⁵ Combining α and σ leads to a density of $2 \times 10^{21}\text{cm}^{-3}$, or about 30% by weight of dichromate in the dried gelatin layer. The refractive index change produced by exposure of one Cr ion was determined on thin, unhardened layers⁵ to be $4 \times 10^{-24}\text{cm}^3$, so that the maximum possible change in index for the thick layers is 0.008. In a $4\text{-}\mu\text{m}$ thick layer, this implies a change in phase of $\Delta\phi_{\text{max}} = 0.45$ at 441 nm and 0.32 at 628 nm.

To check these deductions, the phase change was measured on the 649F plates. The experiment is shown in Fig. 3. It consists of measuring the diffraction efficiency of a simple grating during exposure. The He-Ne beam is not absorbed by the Cr and does not affect the exposure. A typical result is shown in Fig. 4. The efficiency is

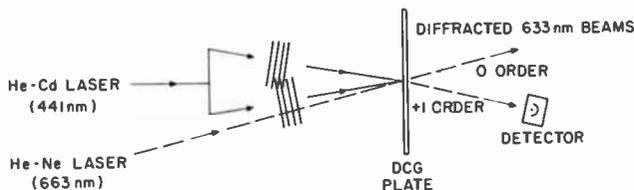


Fig. 3—Experimental arrangement for measuring the phase change due to exposure alone.

independent of grating spacing to the smallest experimental value used, $\Lambda = 0.5 \mu\text{m}$. From the diffraction curve, the phase shift is deduced and values of $\Delta\phi_{\text{max}}$ between 0.2 and 0.3 obtained (Ref. [5], Eq. [15]). This is considered to be adequate agreement with the expected value, because the variation of exposure with thickness was not taken into account explicitly. To reach these values of phase shift requires an exposure of 0.5-1.0 J/cm².

We now consider the changes that take place during the development process. As was mentioned earlier, the water-alcohol development process increases the phase shift by a large amount. The amplification can readily be determined at low exposures. To reach a phase shift of 0.2 in a developed plate, an exposure of about 2 mJ/cm² is required for a typical development (cf. Fig. 7). Thus, the procedure produces a magnification of about 300 \times in the phase shift per unit

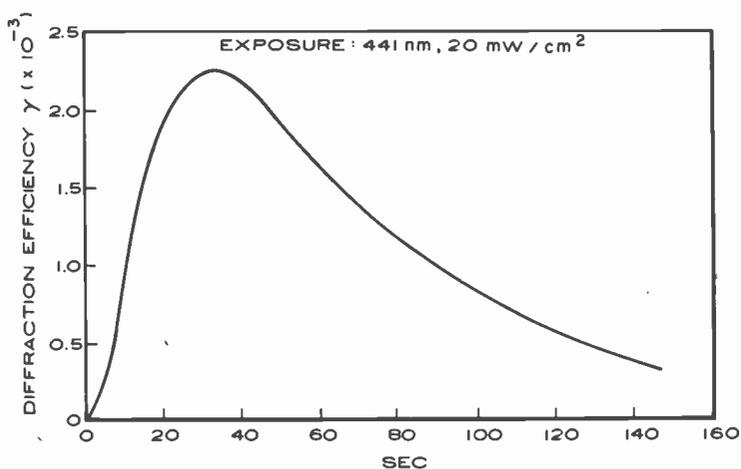


Fig. 4—Grating efficiency during hologram exposure, measured with He-Ne laser.

exposure, or per exposed Cr ion. To measure any variations at longer exposures, we made multiple holograms in which the collimated reference beam was kept constant, but the diverging spherical object beam was moved slightly for each exposure (Fig. 5). The n superimposed holograms were read out with the single reference beam. As expected from the previous discussion, total exposures over 0.5 J/cm^2 tended to saturate and wash out the holograms (all Cr ions are being exposed). Below that level, the light image was evenly divided into

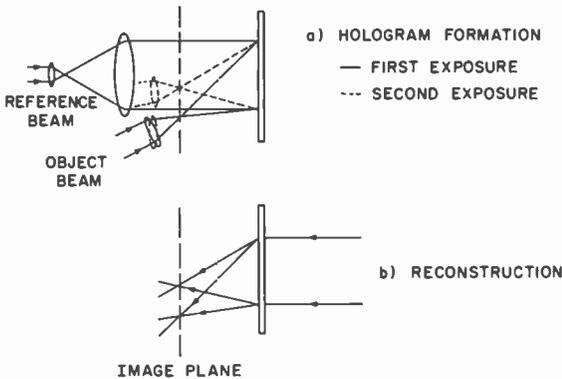


Fig. 5—Recording of a multiple hologram that focuses a collimated beam into multiple points.

all the beams, and the total quantity did not vary appreciably whether there were one or twenty exposures on the plate. Furthermore, the images were as sharp as the image for a single hologram (close to the diffraction limit), so that there was no obvious distortion introduced due to the heavy exposures.

These results of constant amplification and low distortion are difficult to reconcile with the model of the gelatin cracking apart during dehydration.⁴

Emulsion Thickness

One of the problems of thick holograms in photographic media is that the emulsions change thickness during development. This may be due to varying quantities of liquid in the gelatin or due to chemical effects such as hardening. It causes planes of constant index change in the emulsion to change their angle of tilt. For a simple holographic grating, therefore, both the angle of the readout beam and the image

beam must be changed to obtain maximum diffraction. For a picture hologram, which can be thought of as a superposition of many gratings, it is clear that each grating will have a different angular change resulting in reduced diffraction efficiency and/or distortion of the image beam. A useful holographic material must therefore have the thickness of the fully developed emulsion approximately equal to that during the exposure.

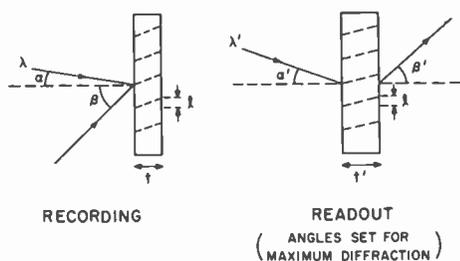


Fig. 6—Arrangement for measuring the relative gelatin thickness. The angles α and β should differ by a large amount for maximum sensitivity.

We studied the thickness of the gelatin during the various stages of processing by measuring the angular variation of a simple grating. The geometry is shown in Fig. 6. The relative thickness is given by¹⁴

$$\frac{t'}{t} = \frac{\tan[(\alpha_i - \beta_i)/2]}{\tan[(\alpha'_i - \beta'_i)/2]} \quad [20]$$

The results are listed in Table 2. The thickness of the emulsion during the first exposure is normalized to 1.0. After the first complete development, the layer thickness has increased by 30 to 50% in agreement with previous measurements.¹¹ This value is larger than the 6% observed by Lin⁶ for reflection holograms, probably due to the different recording geometry. To check the permanence of these changes, we sensitized the plate for the second time and found the thickness at the exposure time to be the same as the first time. The same was true after another development. It is known that if the developed plate is immersed in water and then dried, most of the hologram disappears. At the same time, the thickness returns to the value after sensitizing. This demonstrates that it is the water wash, followed by drying, that determines the original thickness, not the dichromate (this, despite

the high concentration of dichromate (30% by weight)). The Table also shows that the thickness of the gelatin during immersion in water increases by a factor of 3 to 4.

Table 2—Relative Thickness of the Gelatin Layer at Various Stages of the Processing of DCG.

During Exposure	1.0
After Developing	1.3-1.4
During Resensitizing	3.0
After Resensitizing (dry)	0.98
After Renewed Developing	1.3
In the Wash	4.1
Dry	1.01
Redevelop	1.3
After Heating	1.0-1.3

It should be noted that the diffraction efficiencies varied greatly during these measurements, in agreement with previous discussions. It was always possible however, to measure the angles of maximum diffraction. The efficiency could be fully recovered by going through the complete development process.

The last entry in Table 2 shows a very interesting result. If the developed layer is baked at temperatures of about 150°C, the thickness gradually reduces to the relative value 1, or even further. Often this reduction takes place without any change in efficiency. This means that the large variations in refraction index and/or density of the layer must remain, even though the average density is increased by the shrinkage. We note here that ambient humidities up to 70 or 80% do not affect the holograms in any way, unless the baking has reduced the thickness below the value of 1. Then, the layer tends to expand in high humidity back to that value.

Properties of Diffuse-Object Holograms

Previous work^{9,11} has shown that simple grating holograms behave as expected. They show a \sin^2 dependence of diffraction efficiency on exposure with peak values close to 1. We have measured similar values for simple gratings and lenses. We have also investigated this dependence for diffuse-object holograms. Fig. 7 shows the result for an object consisting of a diffuser surrounded by an opaque frame.

The reference beam to object beam ratio was 9. The resulting curve does not show a good \sin^2 dependence and reaches a maximum of only a little over 70%. This is due partly to a somewhat nonuniform exposure and partly to the thickness change during development (no heat treatment shrinkage was attempted with these plates). The latter effect also caused noise in the reconstructed image at high efficiencies.

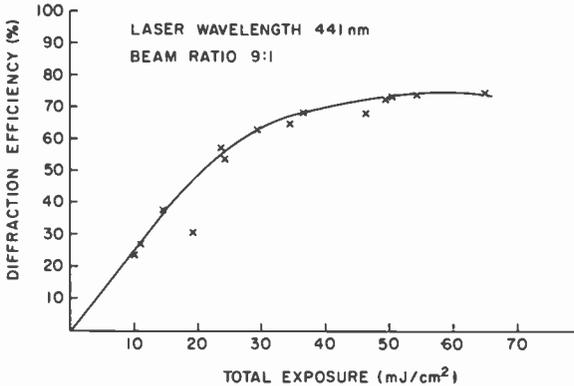


Fig. 7—Diffraction efficiency into the first order image as function of exposure. The object consisted of a random diffuser against a black background.

The plates used in Fig. 7 were sensitized and developed according to our standard recipe. Variation of these procedures can change the absolute values considerably. For example, by using hot water and isopropyl alcohol, the sensitivity is increased considerably, but at the cost of increasing noise. The curve represents our best compromise.

Holographic Optical Elements and Applications

One of the objects of developing dichromated gelatin as a holographic recording material was to make use of its high efficiencies for holographic optical elements. Efficiency is particularly important in this use, because one usually wants to work with the lowest possible exposures and because getting rid of the undiffracted beam without increasing the background illumination is generally a problem. We have studied a number of different optical elements for possible applications.

The simplest holographic optical element is the grating, which is the analog of the prism. It has the advantage that by changing the

angles of the plane waves, a thickness change in the emulsion can be compensated exactly. A somewhat more complicated element is the holographic lens, which combines a change in wavefront curvature with a deflection. It can be formed by a plane and spherical beam or by two spherical beams. When it is used in reconstruction at or near the recording directions, it is capable of forming high-resolution images, but only if the emulsion thickness does not vary from exposure to readout. This is why the final baking step for DCG gelatin was developed. It allowed lenses to be formed with over 90% efficiency and high quality imaging at the same time.

DCG has particular advantages for recording multiple elements because of its large phase change. Multiple lenses have been of interest,¹⁵ but they have been difficult to produce because of the complicated object beam required. If each lens is recorded separately, that problem can be alleviated considerably. We have shown in a previous section that we can record up to 20 such lenses. To demonstrate the multiple imaging of one object, we fabricated a multiple lens with the optical arrangement of Fig. 5, except that the collimated beam is replaced by a converging beam. On readout, the object is placed at the focus of the converging beam. An example of a small transparency imaged with a six-fold lens is shown in Figure 8. In this case, the last two lenses were under-exposed, leading to the variation in intensity. The distortion of the images is due to geometrical effects that have been eliminated in a more carefully designed setup.

Since one usually wants more than 10 or 20 lenses,¹⁵ thicker layers of DCG are needed to obtain sufficient phase shift for this kind of exposure. Sufficient phase shift can be obtained by reducing the absorption coefficient of the sensitized layer by the use of longer-wavelength light, such as 448 or 517 nm Argon laser light. As long as most of the light is absorbed in the layer, the sensitivity will not be changed.

Holographic lenses of the type described are simple and are suited for imaging from the region near one fixed point to the region near the other fixed point and at the specified wavelength. One application where this restriction is not a handicap is a proposed photoprinting scheme shown in Fig. 9. It makes use of a previously discussed holographic character generation system¹⁶ with deflection of the character onto the line of print by a series of hololenses, each for one or a few positions. This scheme is simpler than the previously used electric-pickup-and-readout arrangement. Because of the large optical aperture required by the hologram and because lenses must be moved rapidly, conventional lenses can not be used. Hololenses, on the other hand, work at very high apertures. The field of view of one character is

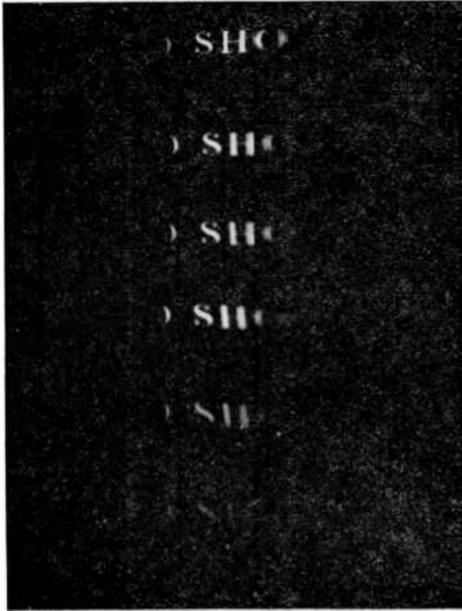


Fig. 8—Result of imaging a small transparency by a six-fold lens.

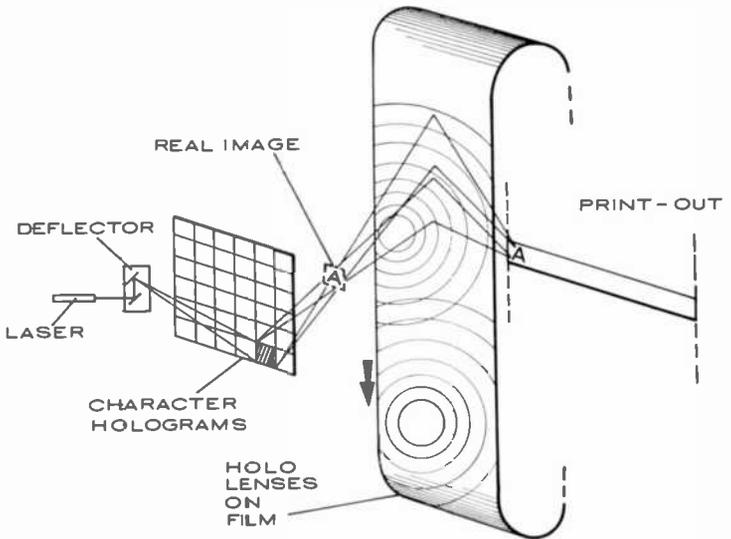


Fig. 9—Schematic view of a photocomposition system with holographic character storage and moving hololens readout.

small, so good freedom from aberration can be obtained. Furthermore, the holographic lenses can be made on very light substrates to allow the rapid movement.

To test these ideas, we fabricated lenses on film. After experimenting with various emulsions, we found that Kodak 649 films gave the best behavior, and we were able to make lenses with good imaging properties in preliminary experiments.

Duplicate Holograms in DCG

DCG holograms combine the advantage of high efficiency and low noise with the disadvantage of the low sensitivity. This is particularly important for diffuse objects, where most of the incident light goes into illumination of the object. We have therefore investigated a scheme for overcoming this problem. First, diffuse object holograms were formed in bleached 649F plates, which have very much higher sensitivity than DCG. The diffraction efficiency of these holograms was deliberately kept at 10% or lower. The holograms were then contact-printed onto DCG. The bleached hologram was located in exactly the same position as during recording and was illuminated with the reference beam only. Because of the low diffraction efficiency of the original holograms, the undiffracted "new" reference beam for the DCG exposure was adequately uniform. The diffracted image beam became the "new" object beam. The procedure produced excellent copies with the high efficiencies characteristic of DCG.

The low sensitivity of DCG is much less of a problem for copies than in making originals, first, because all the light is used to expose the DCG and, second, because much longer exposures can be tolerated in a contact printing system without encountering any coherence problems due to vibration or thermal variations. The procedure is quite similar to the hybrid technique of Pennington¹⁷ who essentially combines the silver and dichromate solutions in the same emulsion.

Acknowledgments

The competent assistance of E. J. Gavalchin throughout this work is gratefully acknowledged. Our early work benefitted greatly from the exchange of ideas with N. F. Maxemchuk and D. L. Matthies.

References:

- 1 T. A. Shankoff, "Phase Holograms in Dichromated Gelatin," *Appl. Optics*, Vol. 7, p. 2101 (1968); R. G. Brandes, E. E. Francois, and T. A. Shankoff, "Preparation of Dichromated Gelatin Films for Holography," *Appl. Optics*, Vol. 8, p. 2346 (1969).

- ² T. A. Shankoff and R. K. Curran, "Efficient High-Resolution, Phase Diffraction Gratings," **Appl. Phys. Lett.**, Vol. 13, p. 239 (1968).
- ³ D. H. R. Vilkomerson, private communication.
- ⁴ R. K. Curran and T. A. Shankoff, "The Mechanism of Hologram Formation in Dichromated Gelatin," **Appl. Optics**, Vol. 9, p. 1651 (1970).
- ⁵ D. Meyerhofer, **Appl. Optics**, Vol. 10, p. 416 (1971).
- ⁶ L. H. Lin, "Hologram Formation in Hardened Dichromated Gelatin Films," **Appl. Optics**, Vol. 8, p. 963 (1969).
- ⁷ M. Born and E. Wolf, **Principles of Optics**, Ch. 12, 2nd ed., Macmillan, New York, 1964.
- ⁸ L. E. Hargrove, "Optical Effects of Ultrasonic Waves Producing Phase and Amplitude Modulation," **J. Acoust. Soc. Amer.**, Vol. 34, p. 1547 (1962).
- ⁹ K. S. Pennington and J. S. Harper, "Techniques for Producing Low-Noise, Improved Efficiency Holograms," **Appl. Optics**, Vol. 9, p. 1643 (1970).
- ¹⁰ J. J. Amodel, W. Philips, and D. L. Staebler, this issue.
- ¹¹ K. S. Pennington, private communication.
- ¹² J. Kosar, **Light-Sensitive Systems**, Ch. 2, John Wiley and Sons, New York (1965).
- ¹³ D. L. Matthies, private communication.
- ¹⁴ D. H. R. Vilkomerson and D. Bostwick, "Some Effects of Emulsion Shrinkage on a Hologram's Image Space," **Appl. Optics**, Vol. 6, p. 1270 (1967).
- ¹⁵ Sun Lu, "Generating Multiple Images for Integrated Circuits by Fourier-Transform Holograms," **Proc. IEEE**, Vol. 56, p. 116 (1968); G. Groh, "Multiple Imaging by Means of Point Holograms," **Appl. Optics**, Vol. 7, p. 1643 (1968).
- ¹⁶ D. Meyerhofer, **Laser Focus**, Feb. 1970.
- ¹⁷ K. S. Pennington, J. S. Harper, and F. P. Laming, "New Phototechnology Suitable for Recording Phase Holograms and Similar Information in Hardened Gelatin," **Appl. Phys. Lett.**, Vol. 18, p. 80 (1971).

Redundant Holograms

A. H. Firester, E. C. Fox, T. Gayeski, W. J. Hannan, and M. Lurie

RCA Laboratories, Princeton, N. J.

Abstract—This paper reviews progress made in the development of a holographic prerecorded television system and examines the fundamental bases for the elimination of coherent-light noise. In particular, it points out that when the system noise is large, there is a maximum redundancy per unit hologram area that can be gainfully employed without introducing coarse structures within the image that are attributable to the redundancy itself.

This article discusses how high quality images of prerecorded video information can be holographically stored with a minimum of recording medium area in a manner suited to a low cost consumer product.¹ Substantial research effort has been devoted to achieving high quality within these constraints and some of what has been learned and accomplished will be reported here. This article is organized into three sections: (1) an introduction to the holographic system and its associated problems; (2) the concept of redundant information storage and a review of prior progress toward achieving highly redundant holograms; and (3) a description of the results of this theory as applied to our application.

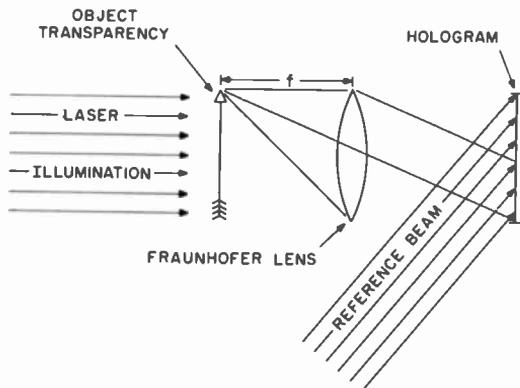
1. Introduction

Our application¹ has used Fraunhofer holograms; we will therefore restrict our discussion to this form of holography, although the extension of these results to other types of holograms will be evident to those familiar with the field. The Fraunhofer hologram records an image in such a way that, upon playback, the image remains immobile despite any translation of the hologram. This image immobilization enables us to make a holographic movie as a sequence of individual Fraunhofer holograms and to reconstruct this movie by a continuous motion of these holograms, thereby avoiding the mechanical complexities required by conventional movie projectors. Fraunhofer holograms are made by converting the light emanating from each point of the object into a collimated beam and recording the interference pattern of the ensemble of those beams with a collimated reference beam. The recording and playback systems that accomplish this are illustrated in Figs. 1(a) and 1(b) respectively.

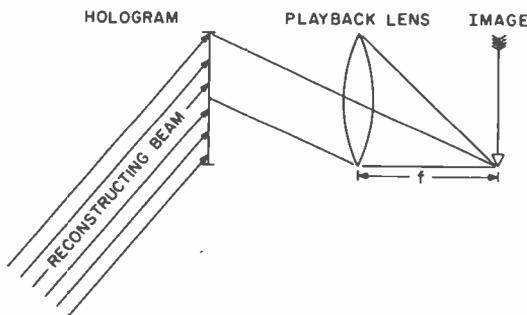
Because highly coherent light must be used for both recording and playback of holograms, small defects in the system (e.g. dust, scratches or dimples on the lenses or recording tape) will have an effect in direct proportion to the electric field amplitude that they intercept and diffract. Each of the optical elements in the system has surfaces and bulk material in which such defects may occur. When used with coherent light, a seemingly perfect optical system will produce an output obscured by diffraction rings, fringes, and speckle all of which are caused by the combined effects of dust, minute scratches, scattering particles, and reflections. This type of noise is usually indiscernible with incoherent illumination. Fortunately, in any optical system, the paths of the rays going through the system can be changed, without changing the location of the reconstructed object, by changing the location of the source of object illumination. Therefore, if many sources illuminate the object, the light from each source experiences different degradations due to the optical imperfections of the total system, and the effect of the imperfections on the image is decreased.

2. Redundancy

The ability to direct light from a particular object point through many different regions of an optical system is our most powerful tool against noise in coherent-light imaging systems. Redundancy is the number of different regions in the system traversed by the light from a particular object point. If the imaging system is used to produce



(a) RECORDING SYSTEM



(b) PLAYBACK SYSTEM

Fig. 1—Holographic optical system for (a) object recording and (b) image playback.

a hologram, the redundancy of the hologram is the number of different areas on the hologram in which light from a particular object point is recorded. Thus if one portion of a highly redundant hologram is mutilated or even obliterated, other parts of the hologram still reconstruct the entire image. Similarly, since the light that forms each portion of the hologram has experienced different degradations upon passing through the optical components, the noise introduced by optical defects in the record/playback system will be uncorrelated and these effects will also be diminished by the redundancy.

Early research in holography² revealed that holograms made from diffusely illuminated or diffusely reflecting objects have unique noise

suppression properties. These holograms have a very high degree of redundancy. Information from every point on the object is spread all over the hologram and, conversely, every point on the hologram contains information about the entire object. As a result, such holograms can be scratched, spotted with dirt, or even broken into pieces without serious loss of information, and they show little noise due to optical imperfections in the record/playback system. However, an unwanted by-product of diffuse illumination, whether or not holography is involved, is speckle noise, the noise becoming more severe as the system aperture is reduced. Assuming that the entire optical system is solely diffraction limited by an aperture of area A_h , it has been shown³ that the signal-to-speckle-noise ratio of a diffuse coherent imaging system is given by

$$(\text{SNR})_{\text{diffuse}} = I / \sqrt{\bar{I}^2} = \sqrt{A_h/A_{\text{min}}} \quad [1]$$

where $I / \sqrt{\bar{I}^2}$ is the ratio of average intensity to rms fluctuation of light intensity in the image and A_{min} is the area of the smallest aperture size at which image resolution just matches the resolution of the display. The SNR of actual diffuse holograms is probably somewhat less than Eq. [1] would indicate since optical defects in the system are not considered in the derivation of this equation.

We define the SNR, which Eq. [1] expresses as the ratio of average light intensity to rms fluctuation in light intensity as $20 \log (I / \sqrt{\bar{I}^2})$ dB. In our system, the output of the photodetector is a current that is proportional to the light intensity and, therefore, our definition is consistent with the definition $\text{SNR} = 20 \log (i_s/i_n)$, where i_s and i_n are the signal and noise current derived from a camera. In accordance with this definition we will term a picture with an $I / \sqrt{\bar{I}^2}$ ratio of 100 as a 40 dB picture.

Fig. 2 shows photographs of a transparency illuminated by diffuse coherent light with $\sqrt{A_h/A_{\text{min}}}$ ratios of 2, 4, and 8. Clearly, SNR improves as the aperture size is increased; unfortunately, the size needed for acceptable SNR is prohibitively large for most applications, particularly for high density holographic storage. Speckle can be eliminated by simply *not* using diffuse illumination. But, then, there is no redundancy and, as illustrated in Fig. 3, images are plagued by the effects of dirt, scratches, and cosmetic defects in the optical system. In other words, the high redundancy in diffuse holo-

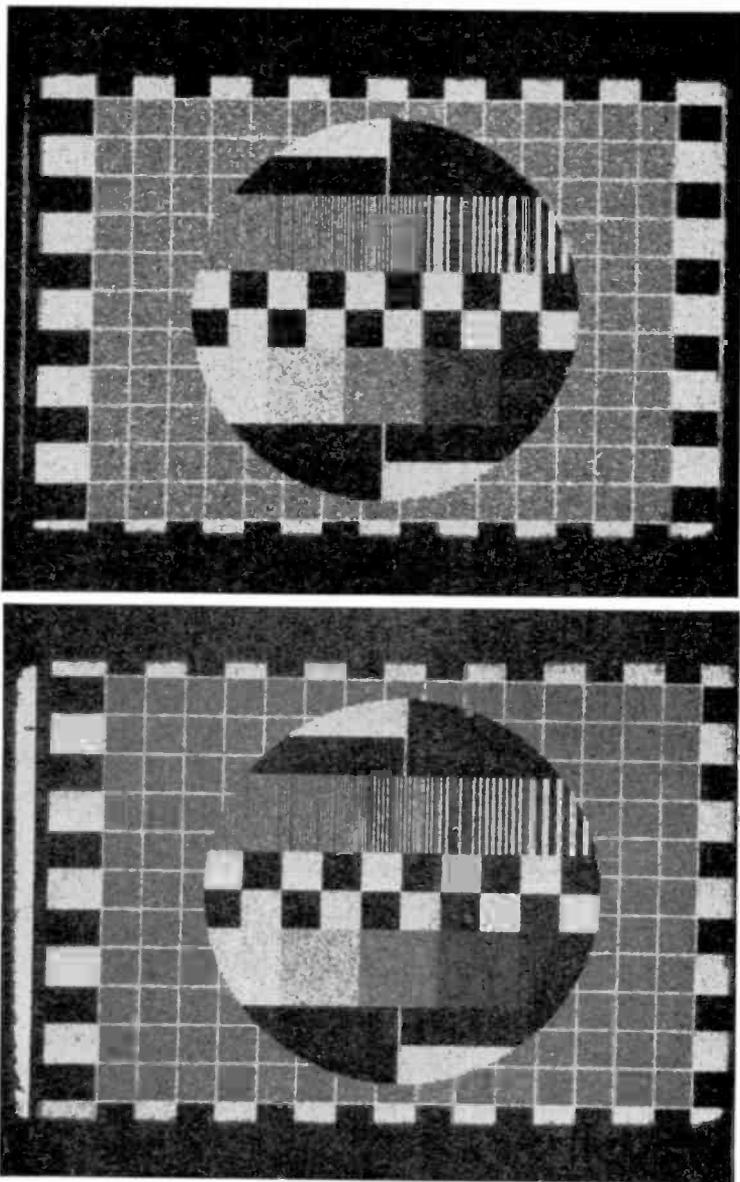


Fig. 2(a), (c)—Photographs of a transparency illuminated by diffuse coherent light. A transparency, located at the front focal plane of a 50 mm lens, was illuminated by *diffuse* coherent light (441.6 nm). The photographs in (a), (b), and (c) (see next page) were made with a camera focused at infinity and correspond to 1-mm, 2-mm, and 4-mm stops located in the back focal plane of the 50 mm lens. The system is approximately diffraction limited with a 0.5-mm stop.

grams diminishes noise due to imperfections in the optical system, but substitutes a new noise—speckle. Since diffuse holograms that yield acceptable SNR are too large for most applications and some degree of redundancy is needed in coherent imaging systems, an alternate form of object illumination is required.

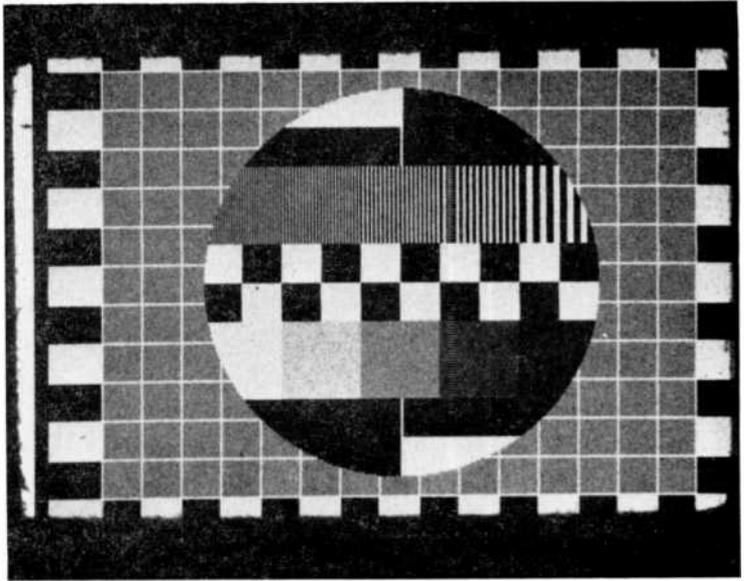


Fig. 2(c)—Photograph corresponding to 4-nm stop located at back focal plane of 50-nm lens.

A basic approach to recording redundant, speckle-free holograms involves illuminating the object transparency with a multiplicity of beams, the angular spacing between beams being sufficient to ensure that the spatial frequencies of all periodic interference patterns at the image are beyond the resolution capability of the ultimate imaging system. Each of the illumination beams produces what we will refer to as a subhologram. Accordingly, hologram redundancy is equal to the number of subholograms, or the number of object illumination beams. We will refer to the general class of holograms that are recorded with a multiplicity of object beams as “multiple-beam” holograms. Likewise, we will refer to holograms recorded with diffuse illumination as “diffuse” holograms.

A simple derivation clarifies the advantages of multiple-beam holo-

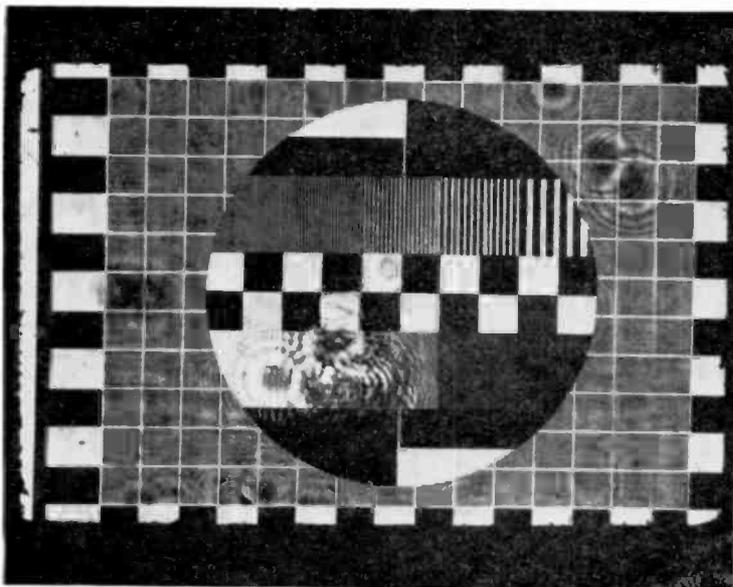


Fig. 3(a)—Photograph of a transparency illuminated by collimated coherent light. Bulls-eye patterns are due to dust specks while the “squiggles” are due to scratches. To simulate phase degradations typical of a poor optical system, a transparency, located at the front focal plane of a 50 mm lens, was illuminated by collimated coherent light (441.6 nm). A strip of 1.5-mil-thick clear vinyl was placed 2 cm in front of the back focal plane of the 50-mm lens, and the photograph was made with a camera focused at infinity. The highest spatial frequency on this test chart corresponds to 27 cycles/mm.

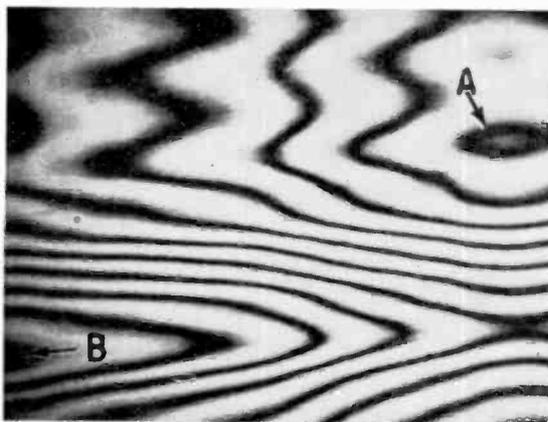


Fig. 3(b)—An interferogram of a 17.8×20.3 mm sample of the vinyl used in Fig. 3a. Each fringe represents 316.4 nm change in thickness. A is the thinnest region in this sample. The thickness increases monotonically to B, which is 3164-nm thicker than A.

grams over equal-area diffuse holograms. Later in this paper, we show that when the system noise is greatest, the maximum redundancy R that can be gainfully employed in multiple-beam holograms is given by

$$R = A_h/A_{\min} . \quad [2]$$

If the noise is uncorrelated from one subhologram to another, which is the usual case, SNR is given by

$$(\text{SNR})_{\text{multiple-beam}} = \rho\sqrt{R} = \rho\sqrt{A_h/A_{\min}} , \quad [3]$$

where $\rho = I/\sqrt{\bar{I}^2}$ is the SNR for an individual subhologram. In highly redundant multiple-beam holograms, the noise seen in reconstructed images has the appearance of speckle; therefore, by combining Eqs. [1] and [3], we find that

$$(\text{SNR})_{\text{multiple-beam}} = \rho (\text{SNR})_{\text{diffuse}} , \quad [4]$$

which indicates that the relative SNR advantage of a multiple-beam hologram over a diffuse hologram is ρ .

The subjective signal-to-noise ratio of a subhologram or a non-redundant hologram depends upon the nature of the recorded object. Highest subjective signal-to-noise ratio occurs when the object is a small white spot on a black background; lowest occurs when the object is an all white field. We shall see later that each Fourier component of any object—single spot or all white field—suffers the degradation of the white field. Under certain circumstances the degradations of Fourier components do not attract attention and the image appears to have a high SNR. For example, if one views a busy scene consisting of bare November foliage, one is unlikely to be bothered by a few twigs that are out of place or a little darker or lighter than they are supposed to be. However if one is holographically storing a line drawing from which some machine is to cut a part with a high degree of accuracy, then distortions will be significant.

Regardless of the nature of the object, it is clear that multiple-beam illumination will enhance SNR, because it spreads redundant information over a wide area on the hologram and because it tends to suppress "hot spots", or intensity peaks, that run the exposure beyond the linear range of the recording materials. Diversity and wideband modulation produce analogous effects in communication systems. The use of multiple-beam recording to record redundant holograms is

analogous to the use of frequency-division multiplexing to transmit redundant information over a communication channel.

Our first attempt to record multiple-beam holograms involved illuminating the object transparency with a beam that had passed through a symmetrical 2-D sinusoidal phase grating.³ This type grating has a transmission $T(x,y)$ given by

$$T(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} J_n(\phi) \exp\{j2n\pi x/\Delta\} J_m(\phi) \exp\{j2m\pi y/\Delta\}, \quad [5]$$

where J_n is the n^{th} order Bessel function of the first kind, Δ is the spatial period of the grating, and ϕ is the peak phase shift introduced by the grating. By constructing a grating with $\phi = 1.43$ radians, we were able to produce nine equally intense beams and thus achieve ninefold redundancy in the holograms. However, although the reconstructed images were speckle-free and had reasonably good dirt and scratch resistance, it was apparent that further improvement could be realized by employing more than nine illuminating beams, provided larger holograms could be tolerated.

To give a concrete example of the redundancy that can be obtained from a reasonably sized hologram, we will consider a microholographic storage and retrieval system in which a standard television system having a bandwidth of 3 MHz is used to read out reconstructed images. For an image width of 10 mm, the smallest horizontal spatial frequency p_w that must be resolved is $p_w = 15.9$ cycles/mm. And, if a Fraunhofer hologram is recorded with a He-Cd laser (441.6 nm) and a recording lens having a focal length of $f = 50$ mm, then the minimum hologram width is $w = 2\lambda p_w f = 0.71$ mm. Correspondingly, if the smallest vertical spatial frequency p_l that must be resolved to prevent beats with the television raster structure is $p_l = 31.8$ cycles/mm, then under similar recording conditions the minimum hologram length l is 1.42 mm. Therefore a hologram measuring, say, 6×11 mm² could be recorded with a redundancy of $(6 \times 11)/(0.71 \times 1.42) = 66$. Achieving an SNR of 100 or 40 dB from this type hologram would call for the SNR of the individual subholograms to be $I/\sqrt{I^2} = 12$, or about 22 dB. In an actual system, however, practical problems usually arise (e.g., nonuniform illumination, vignetting, etc.) that prevent the theoretical redundancy from being realized. Later we will show that there are certain cir-

cumstances under which redundancies much higher than 66 can be obtained and used. Nonetheless, it is clear that more than nine-fold redundancy can be used to advantage in most systems. In this regard, we investigated a few different techniques for generating more than nine equally intense beams.

Our first attempt involved trying to extend the performance of 2-D phase gratings. It is possible to fabricate 2-D phase gratings that produced more than nine equally intense beams by exposing a phase recording medium through a grating (absorption or phase) having high harmonic content. For successive exposures, all beams are masked except, successively, the two first orders, the two second orders, the two third orders, and so on, with exposure times adjusted to yield equal amplitudes in the harmonic components of the ultimate phase grating. Although this approach is theoretically possible, practical difficulties associated with the nonlinear characteristics of phase-recording materials, cosmetic imperfections in optical components, etc., have thus far prevented fabrication of suitable gratings.

Our second attempt involved using an array of mirrors to generate a multiplicity of beams. We achieved a modest improvement in SNR with an array of 12 mirrors, but only after spending tedious hours aligning them. Proper alignment requires that the spatial frequency of the interference pattern produced by each pair of mirrors can differ from the spatial frequency of any other similar pair of mirrors by, at most, one half cycle across the object field. This degree of alignment is required because the vidicon response, as well as that of most other light detectors, is a nonlinear function of the light intensity. This nonlinearity generates coarse intensity beats *between* the fine interference patterns produced by mirror pairs. Maintaining proper alignment to prevent these beats proved to be such a formidable problem that this approach was finally abandoned.

A third approach to multiple-beam holography, which we have explored, enabled us to achieve redundancies higher than those of Eq. [2] by using the Fresnel images⁴ of an array of pinholes whose periodicity is coarser than the sampling periodicity required by the object bandwidth. If this relatively coarse array of pinholes is illuminated by coherent light, exact images of the array are formed at finite distances from the array; at distances intermediate between these exact images, additional images called Fresnel images are formed. Some of these intermediate images are arrays of points whose periodicity is much finer than the original coarse array and can equal and even exceed the sampling requirements of the object bandwidth. For example, if an original square array of 0.1 mm periodicity, is illum-

inated with 441.6 nm light, a square array of 0.02 periodicity will be formed in space a distance 9 mm from the original. By locating the object at one of these Fresnel images, we have made holograms with over sixteen times the redundancy of Eq. [2]. Such holograms, however, are particularly sensitive to system noise and defects. The effect of such defects is to generate in the background of the image the coarse periodicity that was characteristic of the original array. An example of the failure mode is illustrated in Fig. [4].

3. Theory

All multiple-beam holograms can be represented by an illumination source consisting of a number of point sources. Our crossed gratings and mirror arrays can be modeled by a finite array of point sources at infinity, while diffuse illumination can be modeled by an infinite number of randomly phased point sources, also at infinity. We will show that, beyond a limited number of point sources in the illumination, noise in the form of spurious patterns or speckle tends to occur.

Illumination of the object by multiple coherent sources implies that there will exist at the object plane interference patterns whose coarseness will increase as the number of sources per unit area (and, therefore, the redundancy) increases. It is possible, as Gabor⁵ has pointed out, to arrange these multiple sources with any density in such a manner that the coarse patterns at the object plane are eliminated. We will show however, that the success of this strategy depends upon the quality with which the hologram and optics reconstruct the wavefronts resulting from each source. When the optics are poor, it becomes more difficult to suppress the coarse patterns associated with high redundancy. It appears that redundancy is like bank credit; the more one needs it, the worse are the terms on which it comes. In what follows, we shall put forth a quantitative theory of the behavior of multiple-object-source systems. We shall also try to make predictions as to the behavior of such systems in the presence of system constraints and degradation.

Two aspects of the redundancy device must be closely controlled. The first is the illumination pattern of the object. It can have no structures coarse enough to be visible or of such frequency that they cause objectionable moire beats with structures in the recorded object. The second aspect is the failure behavior of this illumination upon degradation by system noise and imperfections. After degradation, the illumination should not revert to any coarse patterns. This feature we term failsafe, and failure to meet this criteria is why diffuse holo-

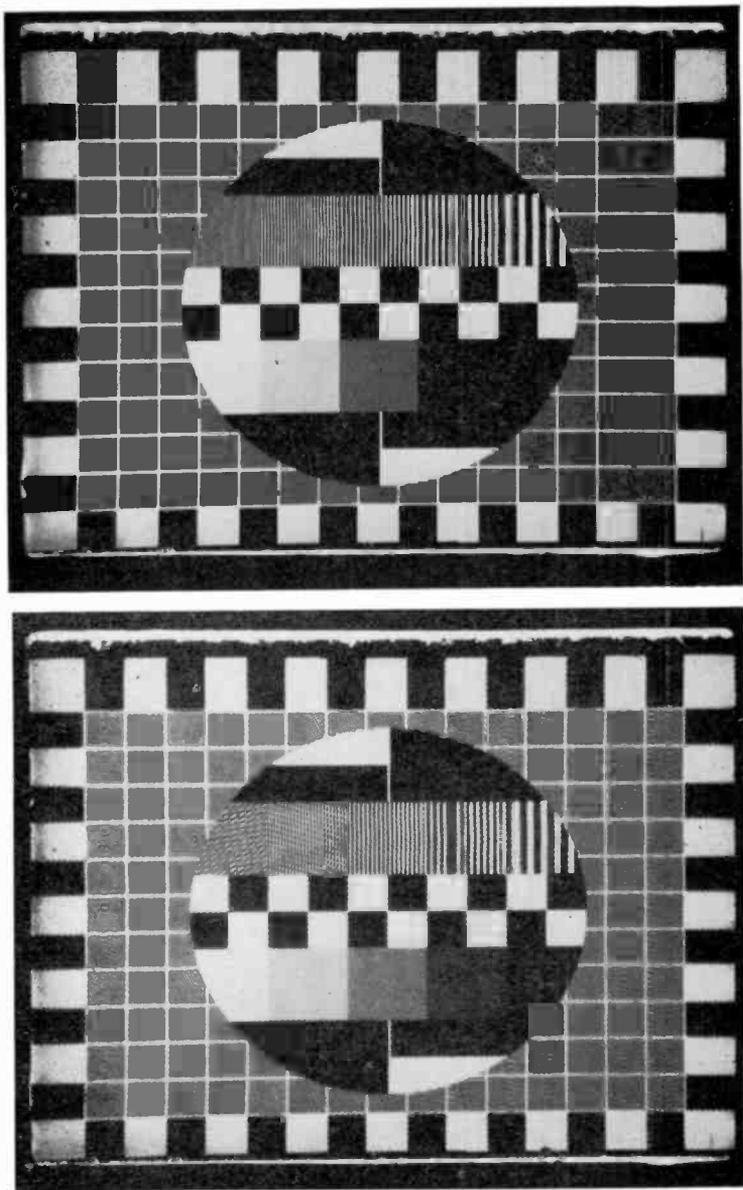


Fig. 4—Photographs of a transparency illuminated by the multiple beams derived from a square array of $4.5\text{-}\mu\text{m}$ pinholes on $36\text{-}\mu\text{m}$ centers. The transparency was located 50 mm from the pinhole array and was illuminated by the light diffracted by the array. Photograph (a) shows the transparency when its illumination is one of the high periodicity Fresnel images of the original array. Photograph (b) shows the same transparency when system degradation causes this type of illumination to display the original coarse periodicity.

grams are not satisfactory. The small aperture size of an otherwise perfect optical system causes them to produce coarse speckle. We will first discuss how this first requirement limits the redundancy device; we will then show theoretically how system noises and imperfections prevent some schemes that successfully satisfy the first requirement from satisfying the failsafe requirement.

We model the redundancy device by any array of perfect point sources located a distance z from the object plane; each point source completely illuminates the object and none are vignetted anywhere in the system. This model is useful both experimentally and theoretically. In experimentation, an actual array of point-sources can be reliably generated with behavior approaching the theoretical behaviour through the use of spatial filtering with pinholes; in a theoretical analysis, an array of point sources can be used to represent any redundancy device. For example, if the light from any of the point sources fails to cover the entire object, or is vignetted anywhere in the system, the system noise functions can be altered to take this into account.

Having chosen as a model a physical array of pinhole sources such that each illuminates the entire object and then passes through the system unobstructed, we are left to make an optimum decision, if possible, on the phase, amplitude, and location of these sources. The first requirement on our point-source array is that when it illuminates the object plane, this illumination shall contain no spatial frequencies so low that their beat products with the object fall within the spatial passband of the object. The fact that each pair of point-sources illuminating the object plane forms a sinusoidal intensity grating whose spatial frequency is proportional to the separation of the point-source pair suggests a lower bound on the separation of any pair in the array (we will show later that there are "non-failsafe" arrays that can have closer spacing). Since each unvignetted point in the array will reconstruct the entire object with a different (largely uncorrelated) noise, we expect the SNR of the total reconstruction to increase as the square root of the number of points in the array. Because we want as many points as possible in the array and a lower bound on separation, a regular array having all points equally intense is called for rather than a random array.

Fig. 5 shows the schematic layout of a generalized multiple-object-beam record/playback system for Fraunhofer holograms. For the sake of simplicity, we assume that playback occurs with a reconstructing beam identical in wavelength and angle to the recording reference beam. The redundancy device, a regular array of equally intense

point sources with arbitrary phase, lies in a plane with coordinate system p, q and located at a distance z behind the object plane. Since the light from each point in the redundancy device is first diffracted by the object, the object itself can be replaced by modifying each point in the redundancy device so that, at a distance z along the optic axis, the diffraction pattern of the device is identical to the actual illuminated object. This modification of the redundancy device to incorporate

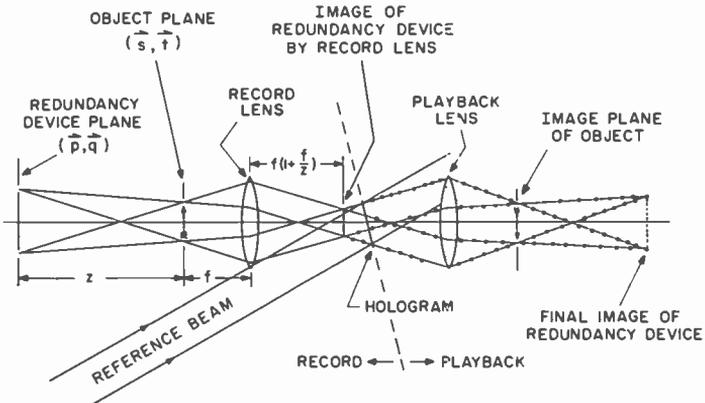


Fig. 5—Multiple object beam record/playback system for Fraunhofer holograms.

the object consists of Fresnel transforms⁶ of the object located at each point of the redundancy device. It is convenient to discuss the object in terms of its spatial Fourier transform, which is z independent, rather than its Fresnel transform, which does depend upon z . Recall, however, that for infinite z , the two transforms are equal, and for finite z they are interrelated. Thus for individual Fourier components of the object, the object-modified redundancy device is the one actually present, but translated according to the vector frequency of the Fourier component and with a relative amplitude and time phase corresponding to the relative amplitude and spatial phase of the object Fourier component modified by the Fresnel phase factor.

Light beams from each point of the object-modified redundancy device progress through the system and each is further diffracted by all optical apertures and inhomogeneities in the system. The optical inhomogeneities include those that arise from slight amplitude and phase differences between the reconstructing beam and the recording reference beam, as well as surface and bulk optical defects of the system components. Every Fourier component of the object will there-

fore be modulated by the pattern its corresponding redundancy device and set of fictitious system noise, amplitude, and phase filters impose on the object plane. These patterns will be statistically similar for all Fourier components of the object. Thus the key to understanding of noise in multiple-object-beam holograms lies in calculating the pattern modulating the zero-frequency component of the reconstructed object. Data relevant to this method of modeling the sources of holographic noise can be obtained by measurements of the amplitude and phase noise of holographic reconstructions of white-field objects.

Our treatment will be to consider the holographic system with a white-field object (zero spatial frequency component of the object) as reproducing a distorted image of the redundancy device. We will then determine the intensity pattern on the reproduced object field due to the distorted image of the redundancy device. A model of any degradation in the image of the redundancy device can be made by the use of two system filters: (1) by convolution of the perfect redundancy device with one complex filter and (2) by multiplication by a second complex filter. The object itself is a filter of the first type, while a field stop for the redundancy device is an example of the second type. We will term these two filters N_C and N_M and, for convenience, will locate the convolutional filter N_C at the plane of the object or its image and the multiplicative filter N_M at the plane of the redundancy device or its image. We will consider only the effects of the multiplicative filter N_M . The effect of the convolutional filter N_C is trivial; the object intensity is modulated by $|N_C|^2$, since we assume it to be located at the plane of the object or its image.

The redundancy device consists of an array of point sources located on a lattice whose basic vectors are \mathbf{p} and \mathbf{q} in a plane perpendicular to the optic axis \mathbf{z} as illustrated in Fig. 6(a). The field at \mathbf{r}_1 in the plane of the redundancy device, $z = 0$, is thus

$$U(\mathbf{r}_1; 0) = \sum_{m,n} U_{mn} \delta(\mathbf{r}_1 - \mathbf{r}_{mn}). \quad [6]$$

where

$$\mathbf{r}_{mn} \equiv m\mathbf{p} + n\mathbf{q}; \quad [7]$$

m, n are integers, and U_{mn} is the complex field amplitude of the point source at \mathbf{r}_{mn} as modified by multiplicative system filter N_M .

Within the usual parabolic approximation to the Fresnel-Kirchhoff equation, the field at point (\mathbf{r}_0, z) within the object plane, produced

by this modified array of point sources, is described by

$$U(\mathbf{r}_0; z) = \exp\{jkz\} \sum_{m,n} U_{mn} \exp\left\{\left(\frac{jk}{2z}\right) |\mathbf{r}_0 - \mathbf{r}_{mn}|^2\right\}, \quad [8]$$

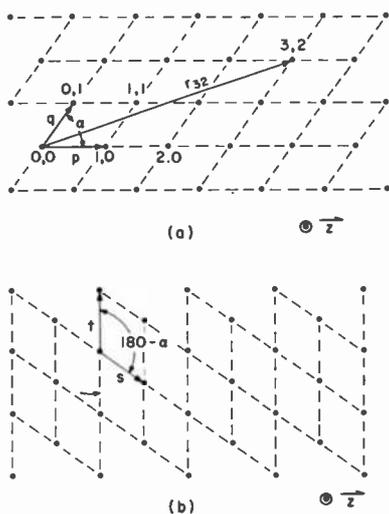


Fig. 6—Coordinate System of (a) the redundancy device space and (b) the reciprocal lattice space.

where

$$k = \frac{2\pi}{\lambda}$$

To describe the location of points within the object plane it is convenient to define a new coordinate system in terms of the reciprocal lattice of \mathbf{p} and \mathbf{q} .⁷ This new coordinate system, illustrated in Fig. 6(b) has basis vectors \mathbf{s} and \mathbf{t} defined by

$$\mathbf{s} = \frac{\mathbf{q} \times \mathbf{z}}{\mathbf{p} \cdot \mathbf{q} \times \mathbf{z}} \lambda z \quad \text{and} \quad \mathbf{t} = \frac{\mathbf{z} \times \mathbf{p}}{\mathbf{p} \cdot \mathbf{q} \times \mathbf{z}} \lambda z. \quad [9]$$

These new basis vectors have the property that

$$\begin{aligned} \mathbf{s} \cdot \mathbf{p} &= \lambda z & \mathbf{s} \cdot \mathbf{q} &= 0 \\ \mathbf{t} \cdot \mathbf{q} &= \lambda z & \mathbf{t} \cdot \mathbf{p} &= 0 \end{aligned} \quad [10]$$

Expressing \mathbf{r}_o in terms of these basis vectors as $\mathbf{r}_o = \alpha \mathbf{s} + \beta \mathbf{t}$, Eq. [8] can be written

$$U(\alpha, \beta; z) = \exp\{jkz\} \sum_{m,n} U_{mn} \exp \left\{ \left(\frac{jk}{2z} \right) (r_o^2 + r_{mn}^2) \right\} \exp\{-2\pi j(\alpha m + \beta n)\}. \quad [11]$$

Our interest, however is not in the field at the object plane but rather the intensity at that plane. The intensity $I(\alpha, \beta; z)$ is the squared magnitude of the field, i.e.,

$$\begin{aligned} I(\alpha, \beta; z) &= U(\alpha, \beta; z) U^*(\alpha, \beta; z) \\ &= \sum_{m,n} \sum_{m',n'} U_{mn} U_{m'n'}^* \exp \left\{ \left(\frac{jk}{2z} \right) (r_{mn}^2 - r_{m'n'}^2) \right\} \\ &\quad \exp\{2\pi j[\alpha(m' - m) + \beta(n' - n)]\} \end{aligned} \quad [12]$$

The intensity pattern in the object plane, $I(\alpha, \beta; z)$, can be rewritten

$$I(\alpha, \beta; z) = \sum_{M,N} I_{MN} \exp\{2\pi j(\alpha M + \beta N)\} \quad [13]$$

where

$$\begin{aligned} I_{MN} &\equiv \exp\{(-jk/2z)(M^2 p^2 + 2MN \mathbf{p} \cdot \mathbf{q} + N^2 q^2)\} \\ &\quad \sum_{m,n} U_{mn} U_{M+m, N+n}^* \exp\{-jk/z\} \\ &\quad [Mmp^2 + (Mn + Nm) \mathbf{p} \cdot \mathbf{q} + Nnq^2] \end{aligned} \quad [14]$$

Eq. [13] expresses the intensity at the object plane as a Fourier series with the basic periodicity of the \mathbf{s}, \mathbf{t} lattice, *never* coarser. No matter how the multiplicative filter modifies either the magnitude or the phase of the point sources in the redundancy array, the object-plane pattern will never have patterns coarser than the basic periodicity of the \mathbf{s}, \mathbf{t} lattice. If the modified redundancy array maintains equal intensity and cophasarity of all points, however, intensity patterns *finer* than the \mathbf{s}, \mathbf{t} periodicity can be achieved. For example for the particularly simple case of a rectangular array, i.e., $\mathbf{p} \cdot \mathbf{q} = 0$, it can be shown that a distance z can be chosen such that $(kp^2/z) = 2\pi(H+1/J)$ where H and J are integers. It should be noted that for a given J , a set of distances z exist that correspond to all the integers H

that satisfy this equation. If all U_{mn} are equally intense and cophasor and z satisfies the above relation, the only nonzero $I_{M,N}$ are those for which $M = 0, J, 2J, 3J, \dots$. These distances are the object-plane locations where the illumination periodicity is J times the periodicity, which is expected if the array points are not cophasor or equally intense. These distances are *not* the planes of the J^{th} harmonic of the array itself as described in Ref. [4]. Indeed multiplication of the basic s, t spatial periodicity is not necessary for our application. To eliminate the fundamental s, t periodicity, for example, it is only necessary to prevent beats with the object frequencies from falling within the system pass-band; the elimination of higher harmonics is not necessarily required. However, with equally intense, cophasor point sources, a distance z can be found for which the basic spatial-pattern repeat frequency in both the s and t directions can be multiplied by an integer.

Although the optic system can be designed so that the multiplicative modifications of the point sources still permit the utilization of these finer spatial patterns, the noise component of the multiplicative filter will generally destroy the amplitude and phase relationship among the modified point sources and will cause the intensity pattern at the object plane to revert to the basic s, t periodicity. It is only when the noise component of the multiplicative system filter can be controlled, that it is possible to use schemes that result in intensity patterns which are spatial harmonics of the basic s, t periodicity.

Needless to say, the higher the sought-for harmonic of the s, t periodicity, the more difficult it is to achieve. By way of example the J^{th} harmonic in the s direction imposes amplitude and phase requirements over groups of J point sources in the modified redundancy array.

Before we consider a specific application of this theory, let us derive the maximum redundancy that we can achieve assuming equally intense but randomly phased point sources in the modified redundancy device. Under these circumstances, the intensity pattern on the object plane will have the basic s, t periodicity. Referring back to Fig. 5, we see that the redundancy device is imaged in holospace. If our hologram is of finite size, the number of array points whose light can pass through the system without vignetting will have an upper bound set by the number of points in the holospace image of the array that fit within the hologram area A_H . The redundancy R is the hologram area A_H divided by the area of each subhologram. The area of each subhologram is the area of unit cell in the p, q lattice times the square of the linear magnification $(f/z)^2$, where f is the focal length of the Fraunhofer lens. Thus,

$$R = \frac{A_H}{\mathbf{p} \cdot \mathbf{q} \times (\mathbf{z}/z)} (z/f)^2. \quad [15]$$

However, the unit cell area of the \mathbf{p}, \mathbf{q} lattice is related to the unit cell area of the reciprocal lattice A_s , by

$$(\mathbf{p} \cdot \mathbf{q} \times \mathbf{z}/z)^{-1} = (\mathbf{s} \cdot \mathbf{t} \times \mathbf{z}/z)/(\lambda z)^2 = A_s/(\lambda z)^2. \quad [16]$$

Thus, in terms of the area of a sample on the object A_s , the redundancy is

$$R = \frac{A_H A_s}{\lambda^2 f^2}. \quad [17]$$

Any increase in redundancy beyond this value will involve dependence upon multiplication of the effective sampling rate as discussed previously. The success of sampling-rate multiplication will depend not only upon both the amplitude and phase noise in the optical system and the hologram itself, but also on the location of the hologram with respect to the holospace image of the redundancy device. If the hologram is located at or near the holospace image of the redundancy device, both hologram amplitude and phase noise will be uncorrelated from one redundancy array point to another. Since, for example, 2×2 multiplication of the failsafe sampling rate imposes amplitude and phase requirements upon 2×2 groups of array points, the effect of hologram noise will be modulation of each Fourier component of the object by the coarse, objectionable patterns that are characteristic of the unmultiplied \mathbf{s}, \mathbf{t} periodicity. If, however, the hologram is not located at the image plane of the redundancy device, the hologram phase and amplitude defects will not only tend to bring out these coarse patterns but will also degrade the image with defects that could have even lower spatial frequencies. The degradation caused by hologram defects, when the hologram is not in the plane of the holospace image of the redundancy device, is similar to the degradation caused by the hypothetical system noise filter N_c . Both can introduce image noise with no lower bound on spatial frequency.

4. Experimental Results

Applying the foregoing results to a practical system calls for designing a pinhole array and selecting a recording geometry that will provide the desired sampling rate at the plane of the object transparency. Once

the desired sampling structure is established, an appropriate pinhole array can be designed such that the fundamental s,t periodicity provides the desired sampling rate. One can then go on to design geometrically similar arrays of finer structure that will yield higher redundancy. But, as mentioned previously, the finer structures can be used effectively only if the optical fidelity of the system is high enough to ensure that the fundamental s,t components will be adequately suppressed.

It is theoretically possible to calculate the increase of redundancy that can be achieved through multiplication, given sufficient amplitude and phase statistics of the overall optical system, but an experimental approach has proved to be the shortest route to determining maximum redundancy.

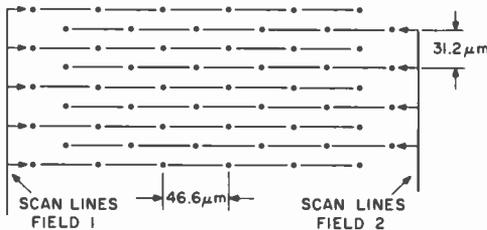


Fig. 7—Sampling structure for television playback system.

Let us consider, as an example, a typical television playback system with these parameters:

Object frame size.....	12.2 × 7.8 mm
Focal length of Fraunhofer lens.....	50 mm
Recording wavelength.....	441.6 nm
Hologram size.....	6.4 × 11.1 mm
Television bandwidth.....	5 MHz

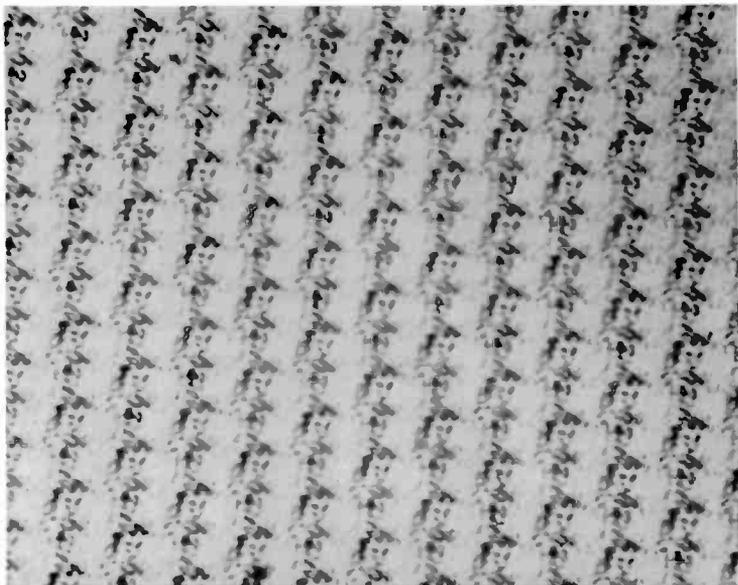
As a starting point we will consider the SNR that can be achieved without relying on multiplication of the basic s,t periodicity. Accordingly, the sampling structure shown in Fig. 7 represents a good design compromise between achieving high redundancy and introducing serious beat problems. For this structure $A_s = 7.15 \times 10^{-4} \text{ mm}^2$, providing 104-fold redundancy on the $6.4 \times 11.1 \text{ mm}^2$ hologram. Experimental results show that such holograms replicated on commercial-grade vinyl

yield a low frequency SNR of 36 dB. Since this SNR is barely acceptable, it is desirable to invoke multiplication of the effective sampling rate to achieve higher redundancy. Doubling, for example, would yield 416-fold redundancy, which in turn would increase SNR to 42 dB.

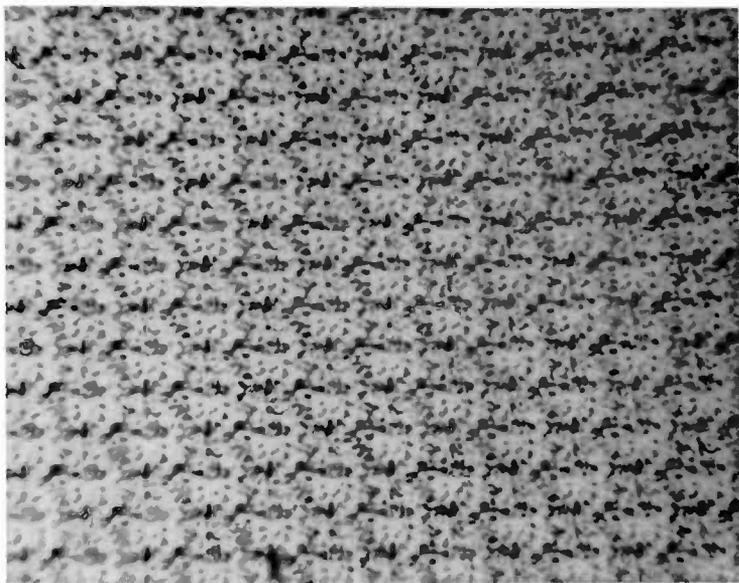
Fortunately, it appears that the doubling strategy is at least partially successful. In Fig. 8(a) we see a photograph of the fine structure at the object plane resulting from an array with $p = 1.90$ mm, $q = 2.13$ mm, and $\alpha = 55.9^\circ$, operating at $z = 223$ mm. This photograph was taken by placing an unexposed film in the object plane, exposing it to the object plane illumination, and taking a microphotograph of the result. It can be shown, however, that I_{01} , I_{03} , $I_{05} \dots$ should be nearly zero at this distance, corresponding to the doubling of the pattern repeat in the s direction seen in Fig. 8(a). Fig. 8(b) is the same as Fig. 8(a) except that a low quality sheet of 1-mil-thick vinyl of the type used in replicating holograms has been placed over the array. We can see that the coarser structure has been brought back. This is an exaggerated simulation of phase errors that might occur in the record-playback process. In practice the phase degradations are not this severe.

5. Conclusions

We have shown that although SNR increases with increasing redundancy, for a given hologram area and object resolution ever increasing redundancy does not generally result in higher and higher signal-to-noise ratios. Great care must be exercised in the application of large amounts of redundancy. Under the worst conditions, when the need for SNR improvement is greatest, redundancy must be limited so that coarse spurious structures will not be produced by the redundancy itself. If on the other hand, system fidelity is sufficient, amounts of redundancy larger than the "failsafe" limit can be safely employed. That the amount of failsafe redundancy is limited has not been previously recognized and is indeed the crux of the matter. The key lies in recognizing that the amount of failsafe redundancy is limited, and that this limit should be exceeded only if the system can tolerate the requirements of higher fidelity. Within this context, one can understand speckle in a diffuse hologram as an attempt to exceed the failsafe requirements. One can also see that illumination by Fresnel images is dangerous, since such illuminations will result in a coarse



(a)



(b)

Fig. 8—Microphotographs of the object plane when illuminated by an array of $4\text{-}\mu\text{m}$ pinholes ($p = 1.90\text{ mm}$, $q = 2.13\text{ mm}$, $\alpha = 55.9^\circ$ and $z = 223\text{ mm}$).

self-image of the point-source array in the presence of system noise. With this understanding of the properties and limitations of redundancy, one can proceed to maximize both the redundancy and the system SNR in a systematic manner.

References:

- ¹ R. Bartolini, W. Hannan, D. Karlsons, and M. Lurie, "Embossed Hologram Motion Pictures for Television Playback," *Appl. Opt.*, Vol. 9, p. 2283, 1970.
- ² E. N. Leith and J. Upatnieks, "Wavefront Reconstruction with Diffused Illumination and Three-Dimensional Objects," *J. Opt. Soc. Am.*, Vol. 54, p. 1295, 1964.
- ³ H. J. Gerritsen, W. J. Hannan, and E. G. Ramberg, "Elimination of Speckle Noise in Holograms with Redundancy," *Appl. Opt.*, Vol. 7, p. 2301, 1968.
- ⁴ J. T. Winthrop and C. R. Worthington, "Theory of Fresnel Images: I. Plane Periodic Objects in Monochromatic Light," *J. Opt. Soc. Am.*, Vol. 55, p. 373, 1965.
- ⁵ D. Gabor, *IBM J. Res. Develop.*, Vol. 14, p. 509, 1970.
- ⁶ A. Papoulis, *Systems and Transforms with Applications in Optics*, McGraw Hill Book Co., N. Y. (1968), Chap. 9, p. 322.
- ⁷ C. Kittel, *Introduction to Solid State Physics*, Chap. 2, p. 52, John Wiley and Sons, Inc., N. Y. (1953).

Wavelength Dependent Distortion in Fraunhofer Holograms and Applications to RCA Holotape®

R. A. Bartolini, D. Karlsons, and M. Lurie

RCA Laboratories, Princeton, N. J.

Abstract—Distortion, the only aberration introduced by reading out a Fraunhofer hologram with light of a wavelength different from that with which it was recorded, can be changed by changing several system parameters. For example the record and readout field angles can be changed as well as the angles between the hologram plane, the reference-beam axis, and the object-beam axis. In the single-frame Holotape system, optimization of these system parameters can reduce distortion from 18.2% to 4.0%. In the two-frame color Holotape system, misregistration is the most serious result of distortion. Maximum misregistration can be reduced from 3.3% to 1.4% by proper choice of system parameters.

1. Introduction

It is sometimes necessary to record and playback holograms at different wavelengths,¹ most often because many useful recording materials are UV or blue sensitive, whereas the low-cost, highly visible, red HeNe laser is the preferred reconstructing source. The change in wavelength causes aberrations. In this paper we consider these aberrations in detail for Fraunhofer holograms recorded as in Fig. 1. A plane

® Registered trademark of the RCA Corporation.

object transparency in the first focal plane of a lens produces the Fraunhofer diffraction pattern on the other side of the lens. That pattern interferes with an off-axis plane-wave reference beam to form the hologram. For this type of hologram, the only wavelength-dependent aberration is distortion.

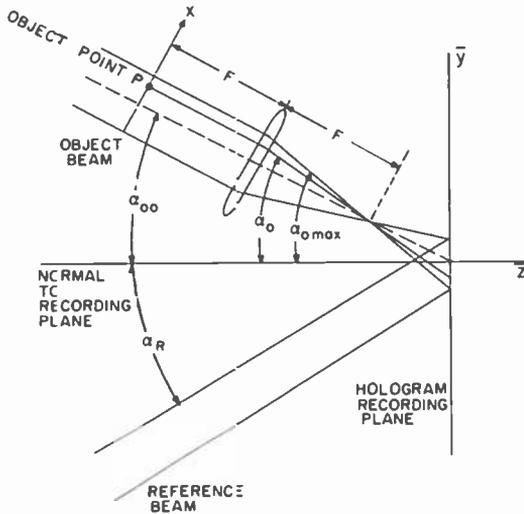


Fig. 1—Fraunhofer hologram recording arrangement.

The plane object transparency in Fig. 1 can be thought of as composed of many points each producing a wavefront that leaves the lens as a plane wave. Therefore, the hologram of each object point is a diffraction grating with uniform period. When this hologram is illuminated with a plane wave, each of the component gratings produces a plane wave that can be focused by a lens to form an image point. The position of the image point is determined by the period and orientation of the component grating and by the wavelength of the illumination. The planarity of the reconstructed wave is not dependent on wavelength. Thus, since only the location and not the shape of the reconstructed point is wavelength dependent, the only aberration present is distortion. More general discussions of holographic aberrations can be found in the literature.²⁻⁶

2. Theory of Distortion

Light of wavelength λ_1 from a point p on the object, after passing through the recording optics, reaches the hologram recording plane

as a plane wave whose direction is described by the angle β_o that it makes with the y - z plane, and the angle α_o that its projection on the y - z plane makes with the z axis as in Fig. 2. The reference beam direction can be described by similar angles β_R and α_R . When recon-

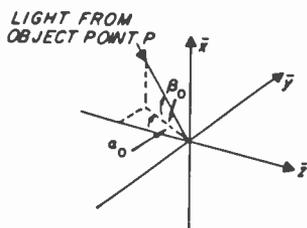


Fig. 2—Coordinate geometry for a point in the object plane.

structed with a plane wave of wavelength λ_2 described by angles β_c , α_c , the following relationships between image and object points result:

$$\sin\alpha'_o \cos\beta'_o = \frac{\lambda_2}{\lambda_1} [\sin\alpha_o \cos\beta_o + \sin\alpha_R \cos\beta_R] - \sin\alpha_c \cos\beta_c \quad [1]$$

$$\sin\beta'_o = \frac{\lambda_2}{\lambda_1} (\sin\beta_o - \sin\beta_R) + \sin\beta_c \quad [2]$$

where β'_o , α'_o are the angles describing the reconstructed plane wave direction. If we assume that the reference and reconstructing beams lie in the y - z plane, i.e., $\beta_R = \beta_c = 0$, then Eqs. [1] and [2] become

$$\sin\alpha'_o \cos\beta'_o = \frac{\lambda_2}{\lambda_1} [\sin\alpha_o \cos\beta_o + \sin\alpha_R] - \sin\alpha_c \quad [3]$$

$$\sin\beta'_o = \frac{\lambda_2}{\lambda_1} \sin\beta_o. \quad [4]$$

From Fig. 1, we can relate the location of an object point to the resulting plane-wave direction α_o , β_o . Let the center of the object be in the y - z plane, at angle α_{o0} with the z axis as in the figure. Then, for a distortion-free lens (e.g. a pinhole or wide-angle small-aperture lens as opposed to a lens satisfying the sine condition),

$$\alpha_o = \alpha_{oo} + \tan^{-1} \frac{x}{F}, \quad [5]$$

and

$$\beta_o = \tan^{-1} \frac{y}{F} \quad [6]$$

where x and y are the object coordinates measured from the center of the object, F is the focal length of the recording lens, and α_{oo} is the direction of the optical axis of the object system. On readout, the corresponding image point is

$$\alpha'_o = \alpha'_{oo} + \tan^{-1} \frac{x'}{F'} \quad [7]$$

and

$$\beta'_o = \tan^{-1} \frac{y'}{F'}, \quad [8]$$

where α'_{oo} is the image system optic axis angle, x' and y' are the image coordinates measured from the center of the image, and F' is the readout lens focal length. Substituting Eqs. [5] to [8] into Eqs. [3] and [4] gives the following relationships between the object and image coordinates and the system angles:

$$\sin \left[\alpha'_{oo} + \tan^{-1} \frac{x'}{F'} \right] \cos \left[\tan^{-1} \frac{y'}{F'} \right] = \frac{\lambda_2}{\lambda_1} \quad [9a]$$

$$\left[\sin \left(\alpha_{oo} + \tan^{-1} \frac{x}{F} \right) \cos \left(\tan^{-1} \frac{y}{F} \right) + \sin \alpha_R \right] - \sin \alpha_C$$

$$\sin \left[\tan^{-1} \frac{y'}{F'} \right] = \frac{\lambda_2}{\lambda_1} \sin \left[\tan^{-1} \frac{y}{F} \right]. \quad [9b]$$

Obviously, x' is not a linear function of x , and y' is not a linear function of y . Thus, the magnification across the image varies, i.e., there is distortion. It is apparent from these equations that magnification as usually defined, $\Delta x'/\Delta x$ and $\Delta y'/\Delta y$ referred to the optic axis, is an unwieldy expression. Expressed in this manner, magnification is best obtained by numerical evaluation of x' and y' using Eqs. [9a] and [9b]. However, the paraxial magnification, dx'/dx and dy'/dy can be expressed more simply. Because it will be important in the example given later, we define a more general *local* magnification as

$$m_x = \frac{dx'}{dx}, \quad [10a]$$

and

$$m_y = \frac{dy'}{dy}, \quad [10b]$$

for any x or y . Differentiating Eqs. [9a] and [9b] gives

$$m_x = \frac{\lambda_2 F' \cos \left[\alpha_{00} + \tan^{-1} \frac{x}{F} \right] \left[\bar{1} + \left(\frac{x'}{F'} \right)^2 \right] \cos \left[\tan^{-1} \frac{y}{F} \right]}{\lambda_1 F \cos \left[\alpha'_{00} + \tan^{-1} \frac{x'}{F'} \right] \left[1 + \left(\frac{x}{F} \right)^2 \right] \cos \left[\tan^{-1} \frac{y'}{F'} \right]} \quad [11a]$$

and

$$m_y = \frac{\lambda_2 F' \cos \left[\tan^{-1} \frac{y}{F} \right] \left[1 + \left(\frac{y'}{F'} \right)^2 \right]}{\lambda_1 F \cos \left[\tan^{-1} \frac{y'}{F'} \right] \left[1 + \left(\frac{y}{F} \right)^2 \right]}, \quad [11b]$$

which reduce to the paraxial magnifications for $x = y = 0$. m_x is a function of x and y , but m_y is a function of y only. Determining m_x and m_y over the field still requires numerical computation because Eqs. [9] must be used to find x' and y' for each value of x and y . But for $x = y = 0$, we also have $x' = y' = 0$, so Eqs. [11] reduce to

$$m_{ox} = \frac{\lambda_2 F' \cos \alpha_{oo}}{\lambda_1 F \cos \alpha'_{oo}} \quad [12a]$$

$$m_{oy} = \frac{\lambda_2 F'}{\lambda_1 F} \quad [12b]$$

for the paraxial case.

The paraxial aspect ratio is then

$$\frac{m_{ox}}{m_{oy}} = \frac{\cos \alpha_{oo}}{\cos \alpha'_{oo}}, \quad [13]$$

which is usually not equal to one.

Generally, the distortion depends on many hologram parameters, as well as any lens used. The aspect ratio can be controlled by the choice of α_{oo} and α'_{oo} , but these alter the other forms of distortion. Furthermore, α_{oo} , with α_R , affects the hologram carrier frequency, and hence the required resolution of the recording material. If we were free to choose the field of view, i.e., the limits on x and y , and the lens focal lengths, we could set $F \gg x$ and y , and $F' \gg x'$ and y' . In that case, there would be no distortion, and the magnification would be everywhere equal to the paraxial magnification in Eq. [12]. In practice, these parameters are restricted by the size and resolution of the hologram and object, by the required image size, and by available lenses.

Thus, in spite of the many parameters involved, the number of changes that can be considered to minimize distortion is small. Usually the reference-to-object beam angle $\alpha_{oo} + \alpha_R$ is kept constant, while α_{oo} is varied. Eqs. [11] and [13] are used to determine the distortion for the various choices of α_{oo} .

3. Calculations of Distortion in the RCA Holotape System

We will use the RCA Holotape television storage system¹ as an example of the application of these results. A Holotape is a long series of Fraunhofer holograms replicated on a flexible tape, each hologram being a recording of a frame of a motion picture. The images are played back sequentially and projected into a vidicon camera to reproduce television program material. The holograms are recorded on a photoresist requiring blue light using a He-Cd laser at 442 nm. For economy and reliability, they are played back with a HeNe laser

at 633 nm. The required object and image formats and the available recording media restrict many system parameters. Each hologram records one 16-mm motion picture frame, which is 7.5×10 mm. The resulting image is approximately 9.5×12.5 mm for detection with a one-inch vidicon. Then the required magnification is $m_{ox} = m_{oy} = 1.25$. With $F = 50$ mm, $\lambda_1 = 442$ nm, $\lambda_2 = 633$ nm, Eq. [12a] gives for $m_{ox} = 1.25$.

$$F' = \frac{43.7 \cos \alpha'_{oo}}{\cos \alpha_{oo}} \quad [14]$$

To limit the period of the hologram to no less than $0.75 \mu\text{m}$ we set an upper limit of 36° on the reference-to-object beam angle. The object beam subtends a half angle of approximately 6° , so the maximum reference-to-object axis angle, $\alpha_{oo} + \alpha_R$, is 30° . The relative sizes of α_{oo} and α_R can be varied by rotating the hologram recording plane, as one can see from Fig. 1. The only other constraint is that the readout geometry be symmetrical, $\alpha_C = \alpha'_{oo}$, to minimize the effect of twisting of the holograms,¹ (this is important because they are on tape). Table 1 gives the values of α'_{oo} and α_C , and F' required for magnification $m_{ox} = 1.25$, for several choice of α_{oo} , rounded to the nearest integer.

Table 1—Requirements for Magnification $m_{ox} = 1.25$.

Object* Recording Angle α_{oo}	Image Readout Angle $\alpha'_{oo} = \alpha_C$	F'
0°	21°	41 mm
10°	22°	41 mm
22°	22°	44 mm
30°	21°	47 mm

* Reference angle $\alpha_R = 30^\circ - \alpha_{oo}$

For any choice of α_{oo} , we can calculate the magnification over the field using Eq. [11]. In the Holotape system we have been describing, the local distortion is critical because of the technique that is used to record the chrominance information in the color television picture. The chrominance is encoded by modulating one or more spatial frequency carriers (vertical or tilted bars in the image). Changes in local magnification, m_x and m_y , change the spatial frequency, which, in turn, can affect the demodulation and hence the color image (to which the eye is very sensitive). The local magnification over the field for various values of α_{oo} is plotted in Figs. 3 and 4. In both figures, F' was chosen to give $m_{ox} \sim 1.25$.

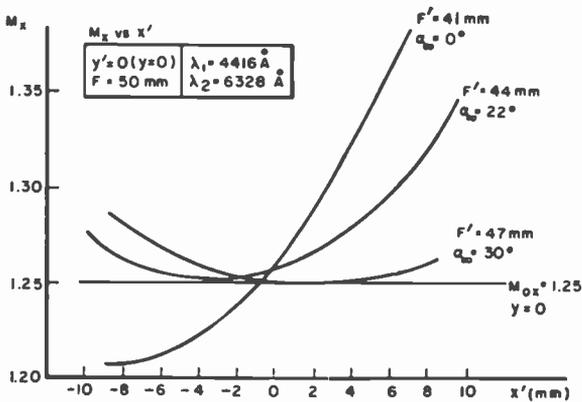


Fig. 3a— m_x vs x' for $y' = 0$ ($y = 0$) for several values of α_{00} .

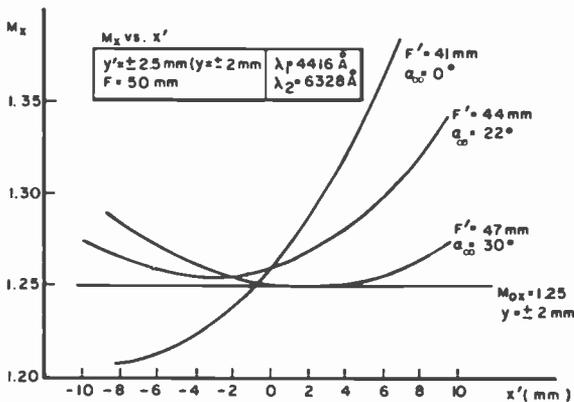


Fig. 3b— m_x vs x' for $y' = \pm 2.5$ mm ($y = \pm 2$ mm) for several values of α_{00} .

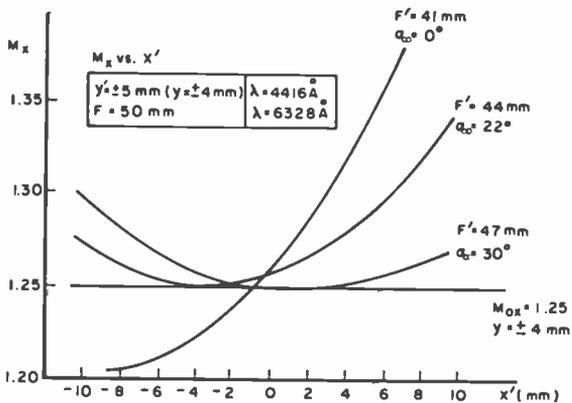


Fig. 3c— m_x vs x' for $y' = \pm 5$ mm ($y = \pm 4$ mm) for several values of α_{00} .

The figures show that minimum local distortion is achieved with $\alpha_{00} = 30^\circ$ (the reference beam perpendicular to the recording plane), but that by minimizing the distortion, the aspect ratio has been altered. The aspect ratio is shown in Table 2.

Table 2—Aspect Ratio for Different Values of α_{00}

α_{00}	m_{0x}	m_{0y}	m_{0x}/m_{0y}
0°	1.25	1.17	1.07
10°	1.25	1.17	1.07
22°	1.25	1.26	.99
30°	1.25	1.35	.93

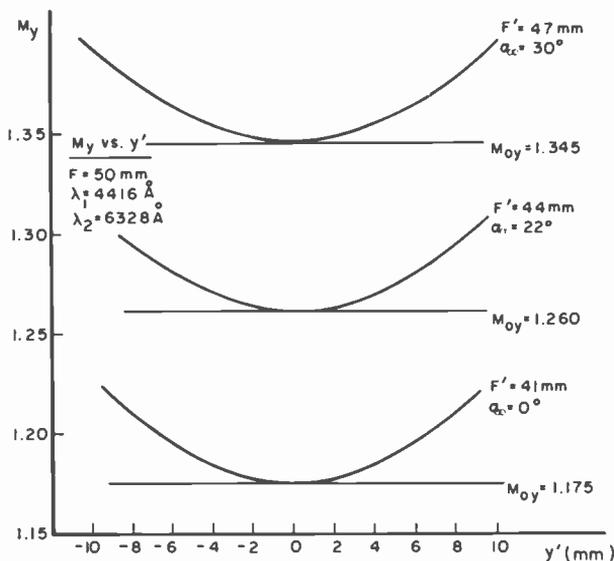


Fig. 4— m_y vs y' for several values of α_{00} .

By choosing $\alpha_{00} = 22^\circ$, the range of local magnification for the given field is limited to 1.25 to 1.30, giving a local distortion of 4.0% (nearly minimum) with an aspect ratio very close to 1.0. That is the chosen compromise for this system.

In a different version of the Holotape system, the modulated spatial frequency carrier for the chrominance signal is recorded in a separate motion picture frame from the luminance image, as shown in Fig. 5. Changes in local magnification have the same effect as in the previous example, except that the larger field angles must be considered. But

variation of the conventional magnification, as opposed to the differential or paraxial magnification, will now result in distortion that results in pairs of points in the separate chrominance and luminance images not being superimposed. Thus, portions of the field will show the color misregistered with the luminance.

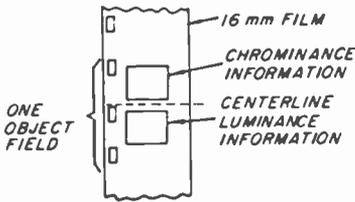


Fig. 5—Two-frame color-encoded transparency.

To study this problem it was more convenient to divide each frame into a matrix of 7×7 points and calculate image position directly from Eq. [9] than to derive an analytic expression for magnification (as explained in Section 2). Misregistration was defined as a fraction of picture width or height:

$$\text{horizontal misregistration at point K} = \frac{(x_K - x_R)_{\text{lum}} - (x_K - x_R)_{\text{chrom}}}{\text{image width}} \quad [15]$$

where x_K is any one of the 49 points in the luminance frame or the corresponding 49 points in the chrominance frame and x_R is the center of the image. Similarly,

$$\text{vertical misregistration at point K} = \frac{(y_K - y_R)_{\text{lum}} - (y_K - y_R)_{\text{chrom}}}{\text{image height}} \quad [16]$$

where y_K is any one of the 7 points on the luminance y axis or the corresponding 7 points on the chrominance y axis (since y is independent of x , only 7 points are necessary) and y_R is the center of the image.

The distance from the center of the luminance frame to the center of the chrominance frame must be correct to give proper registration at the center of the color picture. The proper distances in the Holotape system require $m_{oy} = 1.38$. To produce the correct spatial frequency for the color carrier, we also require $m_{ox} = 1.38$. Using Eqs. [9] and

Table 3—Values of Various Parameters for $m_{os} = 1.38$

α_{oo}	$\alpha'_{oo} = \alpha_c$	F'	m_{os}	m_{oy}
0°	21°	45 mm	1.38	1.29
10°	22°	45 mm	1.38	1.29
22°	22°	48 mm	1.38	1.37
30°	21°	52 mm	1.38	1.49

[12], various combinations of parameters resulting in $m_{ox} = 1.38$ can be derived as shown in Table 3.

As in the system described previously, $\alpha_{oo} = 22^\circ$ produces an aspect ratio of 1.0, giving the required value of m_{oy} . Using Eqs. [15] and [16], Table 4 gives the percent vertical and horizontal misregistrations for the four different values of α_{oo} . From Tables 3 and 4 it is apparent that the least misregistration occurs at $\alpha_{oo} = 30^\circ$ and that the minimum difference between m_{ox} and m_{oy} occurs at $\alpha_{oo} = 22^\circ$.

Table 4—Percent Vertical and Horizontal Misregistration With Values of $\alpha_{oo} = 0^\circ, 10^\circ, 22^\circ,$ and 30° . Values Are for Points Within 7×7 Matrix (One Frame) (See Eq. [15] and [16]).

$\alpha_{oo} = 0^\circ$

Vertical Misregistration

0.62
0.28
0.07
0.00
-0.07
-0.28
-0.62

Horizontal Misregistration

-1.95	-2.12	-2.30	-2.50	-2.73	-3.00	-3.32
-1.30	-1.41	-1.53	-1.66	-1.82	-1.99	-2.20
-0.65	-0.70	-0.76	-0.83	-0.91	-1.00	-1.10
0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.65	0.70	0.76	0.83	0.91	1.00	1.10
1.30	1.41	1.63	1.66	1.82	1.99	2.20
1.95	2.12	2.30	2.50	2.73	3.00	3.32

$\alpha_{oo} = 10^\circ$

Vertical Misregistration

0.62
0.28
0.07
0.00
-0.07
-0.28
-0.62

Table 4 (continued)

Horizontal Misregistration						
-1.23	-1.39	-1.57	-1.77	-1.98	-2.22	-2.50
-0.82	-0.93	-1.04	-1.17	-1.32	-1.48	-1.66
-0.41	-0.46	-0.52	-0.59	-0.66	-0.74	-0.83
0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.41	0.46	0.52	0.59	0.66	0.74	0.83
0.82	0.93	1.04	1.17	1.32	1.48	1.66
1.23	1.39	1.57	1.77	1.98	2.22	2.50

 $\alpha_{00} = 22^\circ$

Vertical Misregistration

0.62
0.28
0.07
0.00
-0.07
-0.28
-0.62

Horizontal Misregistration

-0.30	-0.46	-0.64	-0.82	-1.00	-1.21	-1.43
-0.20	-0.31	-0.42	-0.54	-0.67	-0.80	-0.95
-0.10	-0.15	-0.21	-0.27	-0.33	-0.40	-0.47
0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.10	0.15	0.21	0.27	0.33	0.40	0.47
0.20	0.31	0.42	0.54	0.67	0.80	0.95
0.30	0.46	0.64	0.82	1.00	1.21	1.43

 $\alpha_{00} = 30^\circ$

Vertical Misregistration

0.62
0.28
0.07
0.00
-0.07
-0.28
-0.62

Horizontal Misregistration

0.42	0.25	0.08	-0.09	-0.26	-0.44	-0.62
0.28	0.17	0.05	-0.06	-0.17	-0.29	-0.41
0.14	0.08	0.03	-0.03	-0.09	-0.14	-0.21
0.00	0.00	0.00	0.00	0.00	0.00	0.00
-0.14	-0.08	-0.03	0.03	0.09	0.14	0.21
-0.28	-0.17	-0.05	0.06	0.17	0.29	0.41
-0.42	-0.25	-0.08	0.09	0.26	0.44	0.62

4. Measurement of Distortion in the RCA Holotape System

To verify these results, Fraunhofer holograms were recorded with various values of α_{oo} using a He-Cd laser 422 nm as the light source. They were readout with 633 nm light from a HeNe laser. The recording lens focal length (F) was 50 mm.

For the distortion and aspect ratio measurements, the test pattern shown in Fig. 6 was used as the object transparency. Figs. 6(a) and 6(b) are photographs of images from holograms recorded with $\alpha_{oo} = 30^\circ$ and $\alpha_{oo} = 22^\circ$, respectively. The photographs verify the changes in aspect ratio in Table 2. Magnification at different points can also be determined from these photographs. Horizontal magnifications m_x are shown in Table 5(a) and 5(b) for 9 points on the image—the center, the 4 corners, top and bottom center (3 mm away from center), and left and right center (4 mm away from center). The calculated results are in parenthesis.

Table 5—Comparison of Measured and Calculated (in Parenthesis) Values of m_x

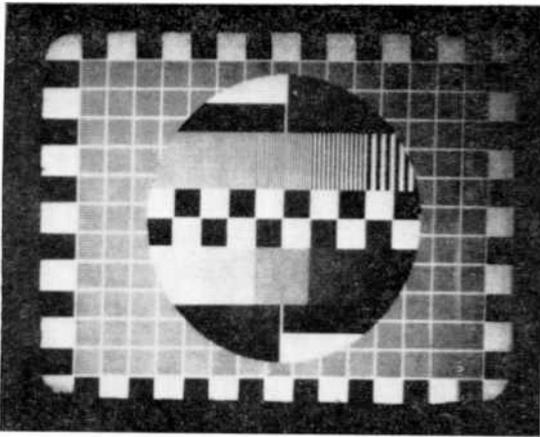
(a)— m_x With $\alpha_{oo} = 30^\circ$

1.29	1.27	1.27
(1.27)	(1.25)	(1.26)
1.29	1.27	1.27
(1.26)	(1.25)	(1.25)
1.29	1.27	1.27
(1.27)	(1.25)	(1.26)

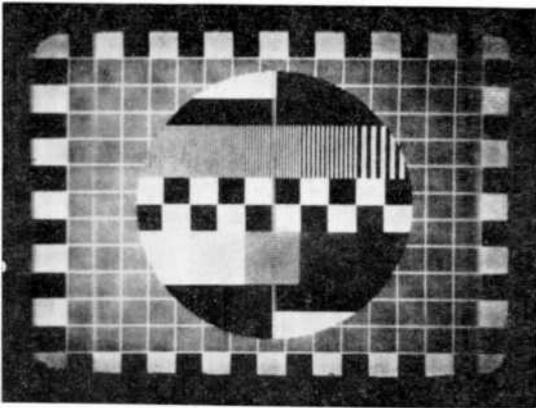
(b)— m_x With $\alpha_{oo} = 22^\circ$

1.25	1.28	1.30
(1.26)	(1.26)	(1.29)
1.25	1.25	1.30
(1.26)	(1.25)	(1.29)
1.25	1.28	1.30
(1.26)	(1.26)	(1.29)

For the misregistration measurements, holograms were recorded using the 2-frame format with a uniform crosshatch pattern as the object in both frames. Figs. 7(a) and 7(b) are photographs of images from these holograms recorded with $\alpha_{oo} = 30^\circ$ and $\alpha_{oo} = 22^\circ$, respectively. These photographs were enlarged to 8×10 inches and one image laid on top of the other with the centers aligned to measure misregistration. The enlarging lens was checked with a test reference frame to insure that it did not introduce any additional distortion into the system. Tables 6(a) and 6(b) show the horizontal misregistration for 6 points on the image (4 corners and top and bottom center). The calculated results are in parenthesis.



(a)



(b)

Fig. 6—Test pattern image from Hologram (a) with $\alpha_{oo} = 30^\circ$ and (b) with $\alpha_{oo} = 22^\circ$.

Table 6—Horizontal Misregistration as a Percent of Picture Width

$\alpha_{oo} = 30^\circ$			
+ .45%	unmeasurable	-.74%	
(+.42%)	(-.09%)	(-.62%)	
- .45%	unmeasurable	+ .74%	
(-.42%)	(+.09%)	(+.62%)	
$\alpha_{oo} = 22^\circ$			
- .35%	-.86%	-1.72%	
(-.30%)	(-.82%)	(-1.43%)	
+ .35%	+ .86%	+1.72%	
(+.30%)	(+.82%)	(+1.43%)	

5. Conclusions

Distortion is the only aberration introduced by reading out a Fraunhofer hologram with light of a wavelength different from that with which it was recorded. Magnification due to the change in wavelength varies across the field, and with direction. Distortion can be changed by changing several system parameters, such as record and readout field angles and the angles between the hologram plane, the reference beam axis, and the object beam axis. Measurements confirm the calculated distortions for two examples.

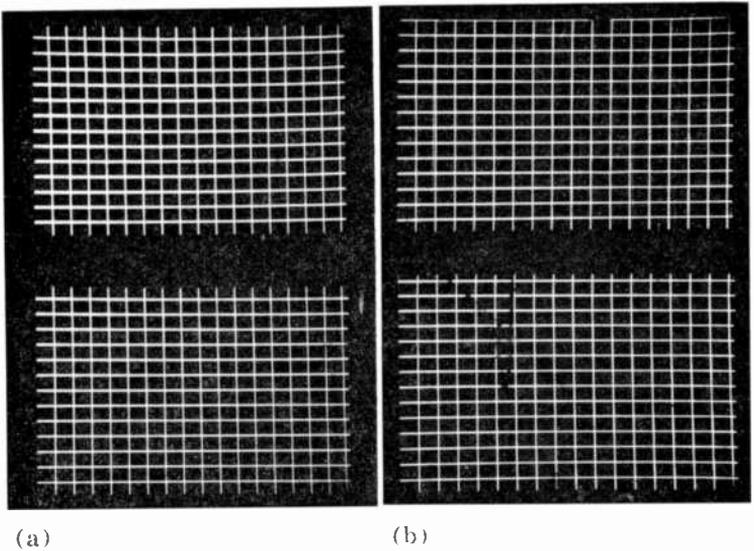


Fig. 7—Uniform crosshatch image from Hologram (a) with $\alpha_{oo} = 30^\circ$ and (b) with $\alpha_{oo} = 22^\circ$.

In the Holotape system, two forms of distortion are important, the usual variation of magnification over the field, measured with respect to the center, and the variation of local magnification measured in any small region of the field.

In the single-frame Holotape system, optimized parameters produce a local magnification in the plane of the reference and object beam axes that varies from $m_x = 1.25$ to 1.30 (Table 5(b)), or 4.0% distortion, with an aspect ratio $m_{ox}/m_{oy} = 1.0$. Arbitrary choice of these parameters could produce m_x values varying from 1.21 to 1.43, or 18.2% distortion, or an aspect ratio up to 1.07.

In the two-frame color Holotape system, misregistration is the most serious result of distortion. It is possible to reduce maximum registration from 3.3% to 1.4%, while maintaining an aspect ratio of 1.0, by proper choice of parameters, particularly reference- and object-beam angles with respect to the hologram plane. With a different choice of angles, the misregistration can be reduced to half that number, but the aspect ratio becomes 0.93. Errors of this magnitude can be detected in a television receiver, but are tolerable to most observers. In the Holotape system, further correction can be obtained in a straightforward way by predistorting the object.

References:

- ¹ R. A. Bartolini, W. J. Hannan, D. Karlsons, and M. Lurie, "Embossed Hologram Motion Pictures for Television Playback," *Appl. Opt.*, Vol. 9, p. 2283, Oct. 1970.
- ² R. W. Meier, "Magnification and Third-Order Aberrations in Holography," *J. Opt. Soc. Amer.*, Vol. 55, p. 987, Aug. 1965.
- ³ E. N. Leith, J. Upatnieks, and K. A. Haines, "Microscopy by Wavefront Reconstruction," *J. Opt. Soc. Amer.*, Vol. 55, p. 981, Aug. 1965.
- ⁴ E. B. Champagne, "Nonparaxial Imaging, Magnification, and Aberration Properties in Holography," *J. Opt. Soc. Amer.*, Vol. 57, p. 51, Jan. 1967.
- ⁵ D. I. Bostwick, D.N.R. Vilkomersen, and R. S. Mezrich, "Techniques for the Removal of Distortion in Hologram Images Caused by a Change of Playback Wavelength," *J. Opt. Soc. Amer.*, Vol. 58, p. 730, May 1968.
- ⁶ J. N. Latta, "Computer-Based Analysis of Hologram Imagery and Aberrations. 1. Hologram Types and Their Nonchromatic Aberrations," *Appl. Opt.*, Vol. 10, p. 599, March 1971.

Recording Considerations for RCA Holotape®

Robert A. Bartolini, Joseph Bordogna,† and Dainis Karlsons

RCA Laboratories, Princeton, N. J.

Abstract—The differences among Fresnel, Fraunhofer, and lensless Fourier transform holograms are reviewed. Paraxial and nonparaxial analyses are included. Practical arrangements for recording small, thin, relief phase holograms of photographic transparencies of these types are discussed.

1. Introduction

In any given application of holography, there is always a group of fundamental factors to be considered. These include type of hologram, hologram dimensions, redundancy, replication, playback requirements, recording materials, and recording techniques. We address our attention here to recording techniques for off-axis, thin, relief phase holograms of photographic transparencies. The interested reader can find related information on the other factors in other papers of this issue and elsewhere in the literature.^{1,2}

The primary motivation for this study of recording techniques resulted from a requirement to produce a type of hologram that could

® Registered trade name of the RCA Corporation.

† Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia.

be replicated quickly and inexpensively and whose reconstructed wavefront yielded an image unperturbed by hologram motion. At the outset it became clear that the terms used to describe various types of holograms were loosely defined and implied different meanings to different workers in the field. To help end this confusion we include careful consideration of the reasons leading to the present terminology.

Our discussion begins with a description of the arrangements currently considered most useful in hologram recording. All these are variations of the off-axis type of hologram, in which spatial separation of a hologram's primary and conjugate images³ is possible.

2. The Basic Recording Arrangements

A hologram is a recording of the interference pattern produced by two light beams that are coherent to each other.^{4,5} In recording a hologram of a photographic transparency, one of the beams, the object beam, passes through the transparency prior to mixing with the other beam, the reference beam. The angular orientation of these two beams with respect to each other classifies the resulting hologram as either an "on-axis" hologram* (object and reference beams parallel, coincident, and normal to hologram plane) or an "off-axis" hologram† (object and reference beams meet at a nonzero angle at the hologram plane). Most of the time, the "off-axis" designation refers specifically to the axis of the reference beam because, in practice, the axis of the object beam is usually oriented perpendicular to the hologram plane. However, there do exist important practical reasons in certain cases (described below) for orienting *both* beams at nonzero angles with respect to the normal to the hologram recording plane; therefore, any general analysis should allow for this situation.

Further thought about the reference beam reveals that its wavefront shape is arbitrary as long as it is coherent with the object beam. Thus, in general, a hologram may be recorded with a diverging, parallel, or converging reference beam. There are important reasons, however, for the selection of a specific wavefront shape; it is a source of confusion that these reasons are usually not explicitly stated when classifying holograms.

Holograms are further classified with respect to the distance between the object and the hologram recording plane. On this basis, there are two major classifications of hologram recording arrange-

* Also referred to frequently as a Gabor hologram.

† Also referred to frequently as a Leith-Upatnieks hologram.

ments—Fresnel and Fraunhofer. In analogy with the concepts of Fresnel (near-field) and Fraunhofer (far-field) diffraction, a Fraunhofer hologram is formed when the object is at a great distance* from the recording medium and a Fresnel hologram is formed when this distance is small. Note that these definitions say nothing about the shape of the reference beam; indeed it is not necessary that they do so in order to satisfy clearly the analogy with diffraction pattern terminology. However, since most holographers use collimated reference beams when recording holograms, this plane wavefront shape is generally implied in the definition. While this implication is of little consequence for a Fresnel hologram, it is of considerable importance in recording a Fraunhofer hologram, because it establishes the property of image immobility on readout. Thus, the term Fraunhofer hologram has come to imply image immobility which, in turn, sets a restriction on the use of this term for describing holograms of Fraunhofer diffraction patterns recorded with collimated reference beams.

When recording holograms of plane object transparencies, a Fraunhofer hologram can be formed at a reasonable distance from the object by placing the object at the front focal plane of a lens. Illumination of the object then produces its Fraunhofer diffraction pattern (where the Fraunhofer approximation is made with respect to hologram size) at any plane on the back side of the lens and, thus, a Fraunhofer hologram at any such plane when mixed with a *collimated* reference beam. The lens of course produces in its back focal plane an exact Fourier transform of the complex field amplitude produced at its front focal plane by the object transparency.⁶ Thus, a Fraunhofer hologram recorded at the back focal plane of a lens is often referred to as a Fourier transform hologram. This particular terminology has created another source of confusion because when a holographer speaks of a Fourier transform hologram, he usually is not implying the preceding definition. Rather, he is speaking of the so-called lensless Fourier transform (LFT) hologram.⁷ This designation is somewhat of a misnomer, because such a hologram is actually a simulated Fraunhofer hologram; that is to say, it is an approximation to the interference pattern produced when a Fraunhofer diffraction pattern mixes with a collimated reference beam. The important of this hologram lies in the fact that it yields image immobility without the use of a lens, as discussed in the next section.

* Theoretically, an infinite distance, but in practice, a distance compared to object *or* hologram size that yields linear phase change across the wavefront at the recording plane.

3. Mathematical Fundamentals

Since we are interested here in thin holograms of photographic transparencies, the mathematical discussion simplifies to two dimensions in the hologram recording plane. Thus, consider the generalized arrangement^{8,9} for recording an off-axis, thin hologram of an object transparency as shown in Fig. 1. The reference beam emanates from

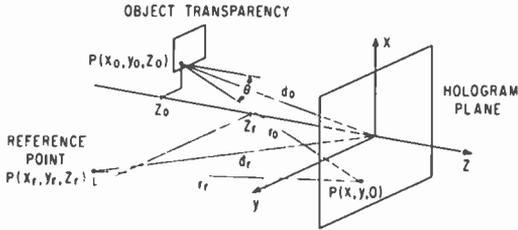


Fig. 1—Generalized arrangement for recording thin, off-axis holograms.

the point (x_r, y_r, z_r) which is, therefore, the origin of a spherical wave. (For the case of a plane reference wave the implication is that $z_r \rightarrow -\infty$.) The location of the object transparency is defined by a plane surface generally not including the point (x_r, y_r, z_r) * and also is oriented such that no point of its surface lies on a line joining the reference point to any point on the hologram plane.† Further, since the polarization of the reference and object beams must be identical before they can fully interfere, scalar theory can be used for analysis.

A typical point (x_o, y_o, z_o) of the object transparency gives rise to a spherical wave when illuminated with monochromatic light. Hence, the complex field amplitude (CFA) at the hologram plane caused by the object can be expressed by the Fresnel-Kirchhoff diffraction formula,¹⁰

$$u_o(x, y, z) = \frac{jk_o}{4\pi} \iint_{S_o} u_o(x_o, y_o, z_o) (1 + \cos \theta) \frac{\exp \{-jk_o r_o\}}{r_o} dS_o. \quad [1]$$

Here the exponential expresses the change in phase of the wave propa-

* In the special case of a lensless Fourier transform hologram the reference point does indeed lie in the plane defined by the transparency.

† When the reference point lies on a line joining any point of the object to any point on the hologram plane, an on-axis (or Gabor) type hologram results.

gating from (x_o, y_o, z_o) to (x, y, z) ; $1/r_o$ expresses the reduction in amplitude of u_o caused by spreading of the spherical wave; $k_o = 2\pi/\lambda_o$ is the magnitude of the wave vector; θ is the angle between the normal to the surface at (x_o, y_o, z_o) and the line joining (x_o, y_o, z_o) with (x, y, z) ; λ_o is the wavelength of the monochromatic light illuminating the transparency; and the integration is carried out over the surface of the object transparency. If $dx_o(x, y, z)$ and $dy_o(x, y, z)$ are defined as incremental line elements within the plane of the object transparency, then $dS_o = dx_o dy_o$.

As mentioned previously, we can classify holograms according to the magnitude of the distance between object and hologram recording plane, or more specifically, the distance r_o between the object point $P(x_o, y_o, z_o)$ and the hologram point $P(x, y, z)$:

$$r_o = [(x - x_o)^2 + (y - y_o)^2 + (z - z_o)^2]^{1/2}. \quad [2]$$

Since we are interested here in thin holograms, we can set $z = 0$ in Eq. [2] and consider the CFA only in the plane defined by $z = 0$ in Fig. 1. With the resulting exact expression for r_o , the integration in Eq. [1] cannot usually be performed explicitly and, in any case, is difficult. Thus, we search for an appropriate expansion for Eq. [2] that will permit a reasonable approximation for r_o . Looking at Fig. 1, we realize that there are two distances about which we can make useful expansions of Eq. [2] to achieve an approximation for r_o , namely, z_o and d_o . To investigate the implications of these two different, yet equivalent, expansions we rewrite Eq. [2] with $z = 0$, in the following forms:

$$r_o = z_o \left[\frac{(x - x_o)^2 + (y - y_o)^2}{z_o^2} + 1 \right]^{1/2} \quad [3a]$$

or, noting that $x_o^2 + y_o^2 + z_o^2 = d_o^2$ in Fig. 1,

$$r_o = d_o \left[\frac{x^2 + y^2}{d_o^2} - 2 \frac{(xx_o + yy_o)}{d_o^2} + 1 \right]^{1/2}. \quad [3b]$$

Expansion of Eq. [3a] in a Maclaurin series about z_o leads to an analytic description of hologram construction and reconstruction which, in most cases, requires a paraxial ray restriction about the z axis. Most analyses in the literature follow this route. Expansion of Eq. [3b] about d_o allows a nonparaxial analysis, which is important when

rays from object points make large angles with respect to the normal to the hologram recording plane; such an analysis is required, for example, for understanding the image reconstruction properties of the lensless Fourier transform hologram. Champagne¹¹ has shown that the two expansions appear to differ only when the lower terms in each are retained, but that the expansions are actually identical if all terms are considered.

We consider first the approximation of r_o by retaining the first few terms of an expansion of Eq. [3a] in a Maclaurin series about z_o .* Although in many practical applications of holography the first term in the brackets of Eq. [3a] may appear to be greater than unity (thereby causing the common binomial series expansion of Eq. [3a] to be nonconvergent), the use of appropriate optics or geometrical configuration can force $z_o^2 > [(x - x_o)^2 + (y - y_o)^2]$. Then, Eq. [3a] may be expanded in a Maclaurian series about the point $z_o = 0$:

$$r_o = z_o \left\{ 1 + \frac{x^2 + y^2 + x_o^2 + y_o^2 - 2xx_o - 2yy_o}{2z_o^2} - \frac{1}{8} \left[\frac{(x - x_o)^2 + (y - y_o)^2}{z_o^2} \right]^2 + \dots \right\}.$$

With respect to this Maclaurin series expansion for r_o , the question arises as to which terms are significant: in other words, what are the practical conditions for which the higher order terms can be neglected? To answer this question, we recall that to form the expansion in the first place we had to assume that $[(x - x_o)^2 + (y - y_o)^2]/z_o^2 < 1$ for convergence. This assumption restricts rays from object points to make small (paraxial) angles with the z axis† and allows us further to eliminate the term $[1/8]/z_o^4 [(x - x_o)^2 + (y - y_o)^2]^2$ as well as succeeding terms to arrive at a reasonable approximation to r_o , namely

$$r_o \approx z_o \left(1 + \frac{x_o^2 + y_o^2 - 2xx_o - 2yy_o}{2z_o^2} + \frac{x^2 + y^2}{2z_o^2} \right). \quad [4]$$

This particular approximation is known as the Fresnel approximation

* The approximation of r_o in the form of Eq. [3b] is considered below when discussing the lensless Fourier transform hologram.

† In other words, the object transparency must lie close to the normal to the hologram recording plane so that rays from it make small angles with the z axis. Note that this in no way restricts the orientation of the reference beam with respect to the z axis.

and defines the conditions for a Fresnel hologram. If the additional approximation, $(x^2 + y^2)/2z_o^2 \ll 1$, is imposed on Eq. [4], the conditions for a Fraunhofer hologram pertain as will be explained below following Eq. [34].

If we look carefully at the implications of the Fraunhofer hologram approximation on the Fresnel-Kirchhoff formula, we see that we are requiring

$$k_o z_o \left(\frac{x^2 + y^2}{2z_o^2} \right) \ll 2\pi$$

in Eq. [1], i.e., there must be little phase change in $u_o(x,y,z)$ across the recording plane with respect to all points in the object plane. At optical wavelengths (say, $\lambda_o = 441.6$ nm) and for holograms of, say, the size of a 16-mm movie frame (10×7.5 mm), the preceding phase condition demands that

$$z_o \gg \frac{x^2 + y^2}{2\lambda_o} = \frac{10^{-4} + 0.56 \cdot 10^{-4}}{0.884 \cdot 10^{-6}} \approx 177 \text{ m.}$$

Although this value of z_o may at first appear unreasonable for practical application, it is quite a simple matter to simulate even larger values of z_o by placing the object transparency in the front focal plane of a lens.

Along with the description of the object beam at the hologram plane in Eq. [1] and the location of the object beam defined by Eq. [2] or Eq. [4], we need to describe the reference beam and its orientation with respect to the hologram plane in order to continue our analysis. The descriptions for finite z_r (a spherically-shaped reference beam) and for infinite z_r (a collimated reference beam) are different.

For a spherically-shaped reference beam emanating from the reference point (x_r, y_r, z_r) in Fig. 1, the geometry dictates that the phase difference between the reference beam CFA at the origin of the hologram plane and its CFA at the point (x,y,z) is $\phi_r = \mathbf{k}_r \cdot \mathbf{r}_r - \mathbf{k}_r \cdot \mathbf{d}_r$ so that, with $z = 0$ as before,

$$u_r(x,y,0) = A_r \frac{\exp \{j\mathbf{k}_r \cdot (\mathbf{r}_r - \mathbf{d}_r)\}}{r_r} \quad [5]$$

with

$$r_r = [(x - x_r)^2 + (y - y_r)^2 + z_r^2]^{1/2}, \quad [6]$$

where \mathbf{k}_r is the wave vector of the reference beam and A_r is the amplitude of the reference wave at the reference point.

For a plane reference beam, $z_r \rightarrow -\infty$, thus causing r_r to become infinite and, therefore, untenable for analysis. This is no problem in Eq. [5], however, because in the first place, since the plane wave does not spread as it propagates, there is no reduction in amplitude of u_r , and the factor $1/r_r$ can be eliminated from its description. Also, the phase change for a plane wave can be derived directly from the exponent of Eq. [5] by describing \mathbf{k}_r in terms of the angle β_r it makes with its projection in the y - z plane and the angle α_r this projection makes with the z axis. Thus, since

$$\mathbf{k}_r = \frac{2\pi}{\lambda_o} (\mathbf{a}_x \sin \beta_r + \mathbf{a}_y \cos \beta_r \sin \alpha_r + \mathbf{a}_z \cos \beta_r \cos \alpha_r) \quad [7a]$$

and

$$\mathbf{r}_r - \mathbf{d}_r = \mathbf{a}_x x + \mathbf{a}_y y,$$

then

$$\mathbf{k}_r \cdot (\mathbf{r}_r - \mathbf{d}_r) = \frac{2\pi}{\lambda_o} (x \sin \beta_r + y \cos \beta_r \sin \alpha_r) \quad [7b]$$

where \mathbf{a}_x , \mathbf{a}_y , \mathbf{a}_z are unit vectors of the cartesian coordinate system and $\lambda_r = \lambda_o$. If we further consider the reference beam to lie in the y - z plane (a typical situation), then $\beta_r = 0$, and

$$\mathbf{k}_r \cdot (\mathbf{r}_r - \mathbf{d}_r) = \frac{2\pi}{\lambda_o} y \sin \alpha_r.$$

The CFA of the reference beam at the recording plane then becomes

$$u_r(x, y, 0) = A_r \exp \left\{ j \frac{2\pi}{\lambda_o} y \sin \alpha_r \right\}. \quad [8]$$

We define

$$u_o(x, y, 0) = u_o(x, y) = |u_o(x, y)| \exp \{ j\phi_o(x, y) \} \quad [9]$$

$$u_r(x, y, 0) = u_r(x, y) = |u_r(x, y)| \exp \{ j\phi_r(x, y) \}, \quad [10]$$

where $|u_o(x,y)|$ and $|u_r(x,y)|$ are real and $\phi_o(x,y)$ and $\phi_r(x,y)$ describe the spatial phase variation (relative to the phase at the origin of the hologram plane) of the object and reference beams, respectively. The total CFA at the hologram plane then is

$$u_h(x,y) = u_o(x,y) + u_r(x,y), \quad [11]$$

which produces an irradiance H at the hologram plane described by

$$H(x,y) = u_h u_h^* = |u_h(x,y)|^2 = |u_o(x,y) + u_r(x,y)|^2 \quad [12]$$

with units of power per unit area (for example, W/cm²). If a photoresist coated film is exposed to this irradiance, its exposure (i.e., energy per unit area) is given by

$$E(x,y) = H(x,y) T, \quad [13]$$

where T is the time interval of exposure. This exposure results in a change in thickness of the photoresist given, as found in our laboratory, by

$$\Delta d \approx cE, \quad [14]$$

where Δd is the change in photoresist layer thickness caused by exposure and c is an exposure constant.

Here the nonlinearities of the photoresist have been neglected, a valid approximation achieved in practice by proper exposure and development. Since the processed photoresist film has (ideally) a transmittance of unity, the hologram will alter only the phase of incident coherent illumination during reconstruction and is thus referred to as a *phase hologram*. Specifically, the wave transmitted through the hologram will experience a phase change (or modulation) of the form $k_R(n-1)\Delta d$, where $k_R = 2\pi/\lambda_R$ (the subscript R pertains to parameters of the reconstruction or readout beam) and n is the index of refraction of the photoresist. This phase modulation yields diffracted orders, the two first orders producing the desired reconstructions of primary and conjugate images.

When compared to that of an ideal absorption hologram, the transmittance function $t(x,y)$ of an ideal phase hologram is purely imaginary, while that of an ideal absorption hologram is purely real. Thus, for a phase hologram, object information is recorded in the phase factor of a unity-magnitude transmission function. Therefore, assum-

ing that the recording medium changes its thickness in proportion to the exposing radiation as in Eq. [14], the phase variation $\phi(x,y)$ imposed by the hologram on a readout beam can be described using Eq. [13] as

$$\begin{aligned}\phi(x,y) &= k_R(n-1)\Delta d = k_R(n-1)cE(x,y) \\ &= k_R(n-1)cTH(x,y).\end{aligned}\quad [15]$$

The transmittance function of the hologram becomes

$$\begin{aligned}t(x,y) &= \exp [j\phi(x,y)] = \exp [jk_R(n-1)cTH(x,y)] \\ &= \exp [jk_RKH(x,y)],\end{aligned}\quad [16]$$

where K is defined as $(n-1)cT$ and is a constant of the hologram. From Eqs. [12], [9], and [10], Eq. [16] becomes

$$\begin{aligned}t(x,y) &= \exp \{jk_RK|u_o(x,y) + u_r(x,y)|^2\} \\ &= \exp \{jk_RK|u_o|^2\} \exp \{jk_RK|u_r|^2\} \exp \{j2k_RK|u_o||u_r| \\ &\quad [\cos(\phi_o - \phi_r)]\}.\end{aligned}\quad [17]$$

Note that the phase (i.e., the interference term) in the third factor of Eq. [17] is *linear* in the CFA coming from the point object. Thus, if $u_o(x,y)$ is given by the superposition of several points of the plane object transparency, e.g.,

$$u_o(x,y) = \sum_i u_{oi}(x,y) = \sum_i |u_{oi}(x,y)| \exp \{j\phi_{oi}(x,y)\},$$

then the interference term will be given by the superposition of terms of the form (neglecting small object-to-object intermodulation terms)

$$(\text{constant}) \sum_i 2k_RK|u_r||u_{oi}| [\cos(\phi_{oi} - \phi_r)].$$

This implies that the point-object analysis we are developing can be applied directly to the analysis of a complete plane object transparency. That is, we decompose the object transparency into a superposition of point objects, describe how the CFA from each point affects the hologram formation, and then add the results to describe the total effect from the actual object.

To simplify Eq. [17] somewhat, we define

$$A_i \hat{=} \exp \{jk_RK|u_o|^2\} \exp \{jk_RK|u_r|^2\}$$

which is a constant for a single point object or, generally, if u_o is sufficiently small. We write furthermore that

$$\begin{aligned}
 a &\hat{=} 2k_R K |u_o| |u_r| \\
 \rho &\hat{=} \phi_o - \phi_r.
 \end{aligned}
 \tag{18}$$

Example: In a situation where a hologram is recorded on vinyl tape, typical parameters are $n = 1.5$, $\Delta d = 0.1 \mu\text{m}$, $\lambda_R = 0.4416 \mu\text{m}$, $|u_r| = 10 |u_o|$, $H \approx |10u_o|^2$. Thus, a typical value for a in Eq. [18] is

$$\begin{aligned}
 a &= 2k_R K |u_o| |u_r| = 2 \left(\frac{2\pi}{\lambda_R} \right) (n-1) cT |u_o| |u_r| \\
 &= \frac{4\pi}{\lambda_R} (n-1) \frac{cE}{H} |u_o| |u_r| = \frac{4\pi}{\lambda_R} (n-1) \frac{\Delta d}{H} |u_o| |u_r| \\
 &= \frac{4\pi (1.5-1) (10^{-7}) (10 |u_o|^2)}{(4.416 \cdot 10^{-7}) (10 |u_o|)^2} = 0.14.
 \end{aligned}$$

Using Eq. [18] in Eq. [17],

$$t(x,y) = A_t [\cos (acos \rho) + j \sin (acos \rho)], \tag{19}$$

and using the Bessel function expansions for $\cos (acos \rho)$ and $\sin (acos \rho)$,

$$\begin{aligned}
 t(x,y) &= A_t \left[J_0(a) + 2 \sum_{m=1}^{\infty} (-1)^m J_{2m}(a) \cos(2m\rho) \right. \\
 &\quad \left. + j2 \sum_{m=0}^{\infty} (-1)^m J_{2m+1}(a) \cos(2m+1)\rho \right], \tag{20}
 \end{aligned}$$

where the $J_i(a)$ are Bessel functions of the first kind. Each $J_i(a)$ in Eq. [20] is the amplitude of the i th diffracted order produced by the grating structure of the phase hologram. For $m = 0$, we obtain the first order term which, upon readout, yields two holographic images of practical interest.* Thus, from Eq. [20]

* Of course, the remaining infinite number of orders also yield images. These are attenuated (because of the decreasing magnitude of higher order Bessel functions in Eq. [20]), are spatially separated from the first orders and from each other, and are of little practical importance here. Also, Eq. [20] is applicable only to individual image points; otherwise ρ is not a linear function of distance and there are no grating orders.

$$t(x,y)|_{m=0} = A_t J_0(a) + j A_t J_1(a) [\exp \{j(\phi_o - \phi_r)\} + \exp \{-j(\phi_o - \phi_r)\}]. \quad [21]$$

When the hologram is illuminated for readout, the wave produced by the term containing the phase factor $+j(\phi_o - \phi_r)$ is defined as the "primary" image wave while the wave produced by the term containing the phase factor $-j(\phi_o - \phi_r)$ is called the conjugate image. We distinguish these terms as $t_p(x,y)$ and $t_c(x,y)$, respectively. Thus,

$$t(x,y)|_{m=0} = A_t J_0(a) + t_p(x,y) + t_c(x,y), \quad [22]$$

where

$$t_p(x,y) = \exp \left\{ j \left(k_R K |u_o|^2 + k_R K |u_r|^2 + \frac{\pi}{2} \right) \right\} J_1(2k_R K |u_o| |u_r|) \exp \{j(\phi_o - \phi_r)\} \quad [23]$$

$$t_c(x,y) = \exp \left\{ j \left(k_R K |u_o|^2 + k_R K |u_r|^2 + \frac{\pi}{2} \right) \right\} J_1(2k_R K |u_o| |u_r|) \exp \{-j(\phi_o - \phi_r)\} \quad [24]$$

Illuminating the phase hologram with a readout beam $u_R(x,y)$ thus produces primary and conjugate images $u_p(x,y)$ and $u_c(x,y)$, given by

$$\begin{aligned} u_p(x,y) &= u_R(x,y) t_p(x,y) \\ &= u_R(x,y) \exp \left\{ j \left(k_R K |u_o|^2 + k_R K |u_r|^2 + \frac{\pi}{2} \right) \right\} \\ &J_1(2k_R K |u_o| |u_r|) \exp \{j(\phi_o - \phi_r)\} \end{aligned} \quad [25]$$

$$\begin{aligned} u_c(x,y) &= u_R(x,y) t_c(x,y) \\ &= u_R(x,y) \exp \left\{ j \left(k_R K |u_o|^2 + k_R K |u_r|^2 + \frac{\pi}{2} \right) \right\} \\ &J_1(2k_R K |u_o| |u_r|) \exp \{-j(\phi_o - \phi_r)\}. \end{aligned} \quad [26]$$

In general, the readout beam,

$$u_R(x,y) = |u_R(x,y)| \exp \{j \phi_R(x,y)\}, \quad [27]$$

can be arbitrary (i.e., plane or spherical), in which case Eqs. [25] and [26] become

$$u_p(x,y) = |u_R(x,y)| \exp \left\{ j \left(k_R K |u_o|^2 + k_R K |u_r|^2 + \frac{\pi}{2} \right) \right\} \\ J_1(2k_R K |u_o| |u_r| \exp \{j(\phi_o - \phi_r + \phi_R)\}) \quad [28]$$

$$u_c(x,y) = |u_R(x,y)| \exp \left\{ j \left(k_R K |u_o|^2 + k_R K |u_r|^2 + \frac{\pi}{2} \right) \right\} \\ J_1(2k_R K |u_o| |u_r| \exp \{-j(\phi_o - \phi_r - \phi_R)\}). \quad [29]$$

With these expressions, the position and form of the primary and conjugate images resulting from readout of a hologram with an arbitrary readout (i.e., reconstruction) beam can be determined. Since the phase angles in Eqs. [28] and [29] are functions of beam wavelength and orientation, differences between ϕ_r and ϕ_R will produce images that are aberrated (i.e., magnified or demagnified nonuniformly). Thus, the situation for fidelity is improved if the wavelengths and orientations of the reference and readout beams are identical; for example, when $\phi_r = \phi_R$, Eq. [28] becomes

$$u_p(x,y) = |u_r| \exp \left\{ j \left((k_R K |u_o|^2 + k_R K |u_r|^2 + \frac{\pi}{2}) \right) \right\} \\ J_1(2k_R K |u_o| |u_r|) \exp \{j\phi_o\}. \quad [30]$$

Since*

$$J_1(a) = \frac{1}{2} a - \frac{\left(\frac{1}{2} a\right)^3}{1^2 \cdot 2} + \frac{\left(\frac{1}{2} a\right)^5}{1^2 \cdot 2^2 \cdot 3} - \dots \\ \approx \frac{1}{2} a, \text{ for small } a, \quad [31]$$

Eq. [30] becomes

* Recall from the example associated with Eq. [18] that in practice a can easily be made to satisfy the criterion for the validity of Eq. [31].

$$\begin{aligned}
 u_p &\approx \left\{ k_R K |u_r| |u_r| \exp \left\{ j \left(k_R K |u_o|^2 + k_R K |u_r|^2 + \frac{\pi}{2} \right) \right\} \right\} |u_o| \exp \{j\phi_o\} \\
 &\approx \text{constant } |u_o| \exp \{j\phi_o\}
 \end{aligned} \tag{32}$$

which is the original object transparency CFA (Eq. [9]) multiplied by a constant.

Eq. [32] includes the effects of the intrinsic linearization of Eq. [31], as well as the recording material linearization of Eq. [14]. These linearizations can be accomplished in practice and must be kept in mind when interpreting experimental results.

4. Distinctions Among Fresnel, Fraunhofer, and Lensless Fourier Transform (LFT) Holograms

The different kinds of phase-hologram recording arrangements are distinguished by the mathematical structure of ϕ_o in Eq. [32]. In the case of a single-point transparency, for example, $\phi_o(x,y)$ is simply the phase of the object wave CFA in the hologram plane at $(x,y,0)$ relative to that of its CFA at the origin of the hologram plane. Thus, from the geometry of Fig. 1,

$$\begin{aligned}
 u_o(x,y) &= |u_o(x,y)| \exp \{j\phi_o(x,y)\} \\
 &= A_o \frac{\exp \{j\mathbf{k}_o \cdot (\mathbf{r}_o - \mathbf{d}_o)\}}{r_o},
 \end{aligned} \tag{33}$$

where A_o is the amplitude of the object wave as it leaves the object point and (using r_o from Eq. [4])

$$\phi_o(x,y) = k_o r_o - k_o d_o = \frac{2\pi}{\lambda_o} (r_o - d_o) \approx \frac{2\pi}{\lambda_o} \left(\frac{x^2 + y^2}{2z_o} - \frac{xx_o}{z_o} - \frac{yy_o}{z_o} \right). \tag{34}$$

Recalling the discussion associated with Eq. [4], ϕ_o in Eq. [34] represents the phase expression for a Fresnel hologram of a single-point object transparency, i.e., a hologram for which the second-order terms in x and y cannot be neglected. If these second-order terms can be neglected, ϕ_o reduces to the phase expression for a Fraunhofer hologram:

$$\phi_o(x,y) \approx \frac{2\pi}{\lambda_o} \left(-\frac{xx_o}{z_o} - \frac{yy_o}{z_o} \right). \tag{35}$$

Let us look carefully at some implications of the Fraunhofer hologram. To start with, since the hologram plane is a very large distance from the object Eq. [1] becomes

$$u_o(x,y,z) \approx \frac{jk}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_o(x_o, y_o) \frac{\exp\{-jk_o r_o\}}{r_o} dx_o dy_o. \quad [36]$$

In addition, with z_o large (i.e., $z_o^2 \gg (x - x_o)^2 + (y - y_o)^2$), Eq. [4] reduces to

$$r_o \approx z_o + \frac{x_o^2 + y_o^2}{2z_o} - \frac{xx_o}{z_o} - \frac{yy_o}{z_o}. \quad [37]$$

In the denominator of Eq. [36], r_o can safely be approximated by z_o and taken outside the integration; however, the full expression for r_o given by Eq. [37] must be used for r_o in the exponential term in Eq. [36], since the integrand is sensitive to small deviations in its phase. Thus, Eq. [36] can be written

$$u_o(x,y) \approx \frac{jk_o}{2\pi z_o} \exp\left\{-jk_o z_o\right\} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_o(x_o, y_o) \exp\left\{-jk_o \frac{x_o^2 + y_o^2}{2z_o}\right\} \exp\left\{\frac{jk_o}{z_o} (xx_o + yy_o)\right\} dx_o dy_o. \quad [38]$$

Eq. [38] is, of course, truly valid (i.e., truly the Fraunhofer diffraction pattern of the object) only in the limiting case $z_o \rightarrow -\infty$. However, in practice we can place this Fraunhofer diffraction pattern at a finite distance from the object by placing the object transparency in the front focal plane of a lens. In this way, light emanating from any point in the transparency becomes plane at the lens output, thus satisfying the condition that the object be at an infinite distance. This condition holds for any plane (and, therefore, for any value of z_o) on the back side of the lens. At the back focal plane, a further condition exists—the Fraunhofer diffraction pattern becomes the *exact* Fourier transform of the CFA in the front focal plane. This is the well-known

Fourier transforming property of a lens.⁶ By taking into account the lens phase factor $\exp\{-jk(x^2 + y^2)/2F\}$ where F is the lens focal length and integrating with respect to $dx'dy'$ over the object, $u_o(x,y)$ becomes instead of the expression given by Eq. [38],

$$u_o(x,y) \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_o(x_o, y_o) \exp \left\{ -j \frac{2\pi}{\lambda_o F} (x_o x + y_o y) \right\} dx_o dy_o. \quad [39]$$

Letting $f_x = x/\lambda_o F$, $f_y = y/\lambda_o F$, Eq. [39] becomes

$$u_o(x,y) = u_o(f_x, f_y) \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_o(x_o, y_o) \exp \left\{ -j2\pi(x_o f_x + y_o f_y) \right\} dx_o dy_o, \quad [40]$$

which says that $u_o(x,y)$, the CFA in the hologram recording plane, is the Fourier transform of the object CFA, $u_o(x_o, y_o)$.

In review, a *Fresnel phase hologram* is constructed by recording the intensity of the interference pattern produced by a Fresnel diffraction pattern of the object (the object beam) mixed with a coherent reference beam as an optical path difference. A *Fraunhofer phase hologram* is constructed by recording the intensity of the interference pattern produced by a Fraunhofer diffraction pattern of the object (the object beam) mixed with a coherent reference beam as an optical path difference. When the recording of a Fraunhofer hologram occurs in the back focal plane of a lens with the object transparency sitting at the front focal plane, the recorded interference pattern consists of the Fourier transform of the object CFA mixed with a coherent reference beam; we call this interference pattern a *Fourier transform phase hologram*.

Usually a Fraunhofer hologram is recorded using a *collimated* reference beam, and when one speaks of a Fraunhofer hologram he generally implies this situation. This particular recording arrangement amounts to the mixing of plane wavefronts producing essentially a superposition of plane diffraction gratings, each grating corresponding to one point of the object transparency. Images produced on read-out by such a hologram are immobile with any linear translation of the hologram because of the property of a plane diffraction grating that it always diffracts incident light in the same direction no matter how the grating is translated in its x - y directions. Mathematically,

this property derives from a consideration of Eq. [39]. During reconstruction with a lens, the CFA of the reconstructed image (say, $u_i(x_i, y_i)$) is related to the CFA in the hologram plane (say, $u_h(x_h, y_h)$) as

$$u_i(x_i, y_i) \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_h(x_h, y_h) \exp \left\{ -j \frac{2\pi}{\lambda_o F} (x_i x_h + y_i y_h) \right\} dx_h dy_h.$$

Suppose now that the hologram is translated in the x_h direction a distance h ; then the magnitude of $u_i(x_i, y_i)$ becomes

$$|u_i(x_i, y_i)| \propto \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_h(x_h + h, y_h) \exp \left\{ -j \frac{2\pi}{\lambda_o F} [(x_h) x_i + y_h y_i] \right\} d(x_h) dy_h \right|.$$

Letting $\xi = x_h + h$,

$$|u_i(x_i, y_i)| \propto \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_h(\xi, y_h) \exp \left\{ -j \frac{2\pi}{\lambda_o F} (\xi x_i + y_h y_i) \right\} d\xi dy_h \right|,$$

or, since ξ is a dummy variable of integration,

$$|u_i(x_i, y_i)| \propto \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_h(x_h, y_h) \exp \left\{ -j \frac{2\pi}{\lambda_o F} (x_h x_i + y_h y_i) \right\} dx_h \cdot dy_h \right|,$$

which indicates *no* change in the image due to hologram translation. Since the "Fourier transform" hologram is a special case of the Fraunhofer hologram, it too yields an immobile image on readout.

It is interesting to note that an approximation to the Fraunhofer hologram can be recorded *without* the use of a lens.⁷ This may seem odd at first, since an essential ingredient for practical Fraunhofer holography is to use a lens to place the Fraunhofer diffrac-

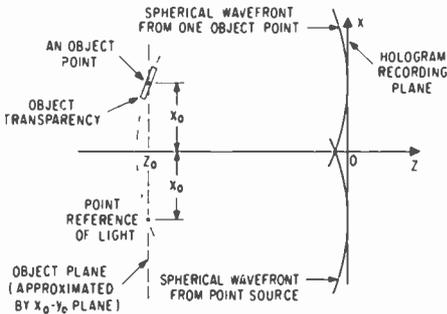


Fig. 2—Lensless recording of a Fourier transform hologram. The two spherical wavefronts shown (one from the reference point source of light and the other from one point on the object) have the same curvature at the hologram plane since they originate at an equal distance from it.

tion pattern of the object at a finite distance. However, if the object and reference source points are equidistant from the hologram plane and *no* lenses are used (as shown in Fig. 2), the sphericities of the two spherical waves at the hologram plane (one from an object point, the other from the reference point) tend to cancel. This action produces a spatial frequency grating for each object point approximately the same as that for each object point in a Fraunhofer hologram where two *plane* waves beat. The difference between the constant single-object-point spatial frequency grating of a Fraunhofer hologram and the *approximately* constant spatial frequency grating recorded by lensless Fourier transform holography is illustrated in Fig. 3. A mathematical explanation of this “trick” of simulating a Fraunhofer hologram can be found in Ref. [14].

5. Relationship Between Image and Object Coordinates

As indicated by Eq. [22], transmission readout of a phase hologram yields three major fields of light that can be readily observed—the

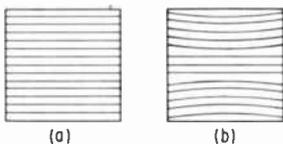


Fig. 3—Lines of maximum intensity for (a) Fraunhofer (b) lensless Fourier transform holograms of a point object.

zero order (unmodulated) beam, which follows along the direction of the readout beam and two first-order (modulated) beams, which produce images of the original object transparency. Since all the image information is contained (for a thin relief phase hologram) in the phase factors of the transmittance function of the hologram, the positions of the images can be determined by describing the phase factors in detail for the simple case of recording the hologram of a point object transparency.

5.1 Fresnel Hologram

For the Fresnel hologram, we can distinguish two cases depending on whether the object beam is normal to the recording plane or not.

5.1a. Object and Reference Beams Both Off Normal to Recording Plane (Small-Angle Analysis)

In a manner similar to that used to derive the phase factor ϕ_o in Eq. [34] for a Fresnel hologram of a point object, the phase factors of the reference and readout beams are given, respectively, as

$$\phi_r(x,y) \approx \frac{2\pi}{\lambda_o} \left(\frac{x^2 + y^2}{2z_r} - \frac{xx_r}{z_r} - \frac{yy_r}{z_r} \right) \quad [41]$$

$$\phi_R(x,y) \approx \frac{2\pi}{\lambda_R} \left(\frac{x^2 + y^2}{2z_R} - \frac{xx_R}{z_R} - \frac{yy_R}{z_R} \right) \quad [42]$$

where we must remember that the approximations leading to these descriptions presuppose a paraxial analysis; i.e., although off the z axis, the beams are permitted to make only small angles (< 6 degrees) with respect to it.

From Eqs. [28] and [29], the phase factor ϕ_p of the primary image and the phase factor ϕ_c of the conjugate image are given, respectively, by

$$\phi_p = \phi_o - \phi_r + \phi_R \quad [43]$$

$$\phi_c = -(\phi_o - \phi_r - \phi_R). \quad [44]$$

To avoid the confusion of aberrations generated by wavelength shift on readout¹² we assume at this point that $\lambda_R = \lambda_r = \lambda_o$. With this

convenience, substituting Eqs. [34], [41] and [42] into Eqs. [43] and [44] yields

$$\phi_p \approx \frac{2\pi}{\lambda_0} \left(\frac{x^2 + y^2}{2z_p} - \frac{xx_p}{z_p} - \frac{yy_p}{z_p} \right) \quad [45]$$

$$\phi_c \approx \frac{2\pi}{\lambda_0} \left(\frac{x^2 + y^2}{2z_c} - \frac{xx_c}{z_c} - \frac{yy_c}{z_c} \right) \quad [46]$$

where

$$z_p = \frac{z_0 z_r z_R}{z_r z_R - z_0 z_R + z_0 z_r} \quad [47]$$

$$x_p = \frac{x_0 z_r z_R - x_r z_0 z_R + x_R z_0 z_r}{z_r z_R - z_0 z_R + z_0 z_r} \quad [48]$$

$$y_p = \frac{y_0 z_r z_R - y_r z_0 z_R + y_R z_0 z_r}{z_r z_R - z_0 z_R + z_0 z_r} \quad [49]$$

$$z_c = \frac{z_0 z_r z_R}{z_0 z_r - z_r z_R + z_R z_0} \quad [50]$$

$$x_c = \frac{x_R z_0 z_r - x_0 z_r z_R + x_r z_R z_0}{z_0 z_r - z_r z_R + z_R z_0} \quad [51]$$

$$y_c = \frac{y_R z_0 z_r - y_0 z_r z_R + y_r z_R z_0}{z_0 z_r - z_r z_R + z_R z_0} \quad [52]$$

Note that Eqs. [45] and [46] have the same form as, for example, Eq. [34]; thus, we interpret x_p , y_p , z_p as the coordinates of the primary image of the original point object. Similarly, x_c , y_c , z_c describe the conjugate image. [Also, just as in going from Eq. [34] to [35], Eq. [45] and [46] become ϕ_p and ϕ_c for a Fraunhofer hologram (with object, reference, and readout beams making small angles with the z axis) when the terms $(x^2 + y^2)/2z_p$ and $(x^2 + y^2)/2z_c$ are deleted.]

For the Fresnel hologram constructed with a plane reference beam

and reconstructed with a plane readout beam, $z_r = z_R \rightarrow -\infty$. Usually, these sources also lie in the y - z plane of Fig. 1 so that $x_r = x_R = 0$. In this case the preceding equations reduce to

$$\begin{aligned} z_p &= z_o & z_c &= -z_o \\ x_p &= x_o & x_c &= x_o \\ y_p &= y_o - z_o \frac{y_r}{z_r} + z_o \frac{y_R}{z_R} & y_c &= y_o - z_o \frac{y_r}{z_r} - z_o \frac{y_R}{z_R}. \end{aligned}$$

Defining

$$\alpha_r = \tan^{-1} \frac{y_r}{z_r} \text{ and } \alpha_R = \tan^{-1} \frac{y_R}{z_R},$$

the equations of the image coordinates can be written

$$\begin{aligned} z_p &= z_o & z_c &= -z_o \\ x_p &= x_o & x_c &= x_o \\ y_p &= y_o - z_o \tan \alpha_r + z_o \tan \alpha_R & y_c &= y_o - z_o \tan \alpha_r - z_o \tan \alpha_R. \end{aligned}$$

For identically oriented reference and readout beams, $\alpha_R = \alpha_r$, and the image coordinate equations reduce further to

$$\begin{aligned} z_p &= z_o & z_c &= -z_o \\ x_p &= x_o & x_c &= x_o \\ y_p &= y_o & y_c &= y_o - 2z_o \tan \alpha_R. \end{aligned}$$

Thus, we see that there exist both real ($z_p = z_o$) and virtual ($z_c = -z_o$) images on readout and that, in addition, the y -coordinate of the virtual image is displaced by the term $-2z_o \tan \alpha_r$.

It is interesting to note that if $\alpha_R = -\alpha_r$, the image coordinate equations yield

$$\begin{aligned} z_p &= z_o & z_c &= -z_o \\ x_p &= x_o & x_c &= x_o \\ y_p &= y_o - 2z_o \tan \alpha_R & y_c &= y_o. \end{aligned}$$

Thus, reading out with a beam lying on the opposite side of the z axis produces an undisplaced virtual image and a displaced real image.

5.1b. Object Beam Normal to Recording Plane

An important fact in the practical application of the preceding Fresnel hologram image coordinate equations is that they apply only for the condition of *all* beams making *small angles* with respect to the z axis. As discussed in the preceding section, however, the axis of the object beam is usually oriented normal (i.e., along the z axis in Fig. 1) to the hologram recording plane, and the shape of the reference and readout beam wavefront are usually plane. These facts allow a simplification of the discussion of the preceding section and actually provide more practically useful mathematical descriptions of the relationships among image and object coordinates.

With the object beam along the z axis, the small angle approximation leading to Eq. [34] is valid, and for a collimated reference beam oriented at *any angle* α_r with respect to the z axis and lying in the y - z plane, Eq. [8] yields a reference beam CFA phase angle

$$\phi_r = k_o y s \sin \alpha_r = \frac{2\pi}{\lambda_o} y s \sin \alpha_r. \quad [53]$$

Similarly, the phase angle of the readout beam CFA (for a beam lying in the y - z plane) for $\lambda_R = \lambda_r = \lambda_o$ becomes

$$\phi_R = k_o y s \sin \alpha_R = \frac{2\pi}{\lambda_o} y s \sin \alpha_R. \quad [54]$$

Thus, substituting Eqs. [34], [53] and [54] into Eqs. [43] and [44] yields

$$\phi_p \approx \frac{2\pi}{\lambda_o} \left[\frac{x^2 + y^2}{2z_o} - \frac{xx_o}{z_o} - \frac{yy_o}{z_o} \left(1 + \frac{z_o}{y_o} \sin \alpha_r - \frac{z_o}{y_o} \sin \alpha_R \right) \right]$$

$$\phi_c \approx \frac{2\pi}{\lambda_o} \left[\frac{x^2 + y^2}{2z_o} - \frac{xx_o}{z_o} - \frac{yy_o}{z_o} \left(1 + \frac{z_o}{y_o} \sin \alpha_r + \frac{z_o}{y_o} \sin \alpha_R \right) \right].$$

For identical reference and readout beams, $\alpha_R = \alpha_r$, we identify image coordinate from these equations, as

$$\begin{aligned} z_p &= z_o & z_c &= -z_o \\ x_p &= x_o & x_c &= x_o \\ y_p &= y_o & y_c &= y_o - 2z_o \sin \alpha_R \end{aligned}$$

We see that both real ($z_p = z_o$) and virtual ($z_c = -z_o$) images are produced on readout and that, in addition, the y -coordinate of the virtual image is displaced by the term, $-2z_o \sin\alpha_R$.

For $\alpha_R = -\alpha_r$,

$$\begin{array}{ll} z_p = z_o & z_c = -z_o \\ x_p = x_o & x_o = x_o \\ y_p = y_o - 2z_o \sin\alpha_R & y_o = y_o \end{array}$$

Thus, reading out with a beam lying on the opposite side of the z axis produces an undisplaced virtual image and a displaced real image.

In using the preceding Fresnel-hologram image-coordinate expressions, we must keep in mind that they apply for the condition of axis of object beam *normal* to recording plane and plane reference beam oriented at *any angle* with respect to object beam. Equality of this situation with that where both object and reference beams make small angles with respect to the normal to the hologram recording plane requires that $|\alpha_R|$ (and thus $|\alpha_r|$) are such that $\sin\alpha_r \approx \tan\alpha_r$.

5.2 Fraunhofer Hologram

For the Fraunhofer hologram, in addition to $z_r = z_R \rightarrow -\infty$ (as for the Fresnel hologram recorded with plane reference and readout beams), $z_o \rightarrow -\infty$. Thus, following similar reasoning as in the several paragraphs leading to the descriptions of a plane reference beam given by Eqs. [7] and [8], ϕ_o in Eq. [33] becomes

$$\phi_o = \mathbf{k}_o \cdot (\mathbf{r}_o - \mathbf{d}_o) \quad [55]$$

with

$$k_o = \frac{2\pi}{\lambda_o} (\mathbf{a}_x \sin\beta_o + \mathbf{a}_y \cos\beta_o \sin\alpha_o + \mathbf{a}_z \cos\beta_o \cos\alpha_o) \quad [56]$$

$$\mathbf{r}_o - \mathbf{d}_o = \mathbf{a}_x x + \mathbf{a}_y y, \quad [57]$$

where β_o is the angle k_o makes with its projection in the y - z plane of Fig. 1, and α_o is the angle between this projection and the z axis. Substituting Eqs. [56] and [57] in Eq. [55] gives

$$\phi_o = \frac{2\pi}{\lambda_o} (x \sin\beta_o + y \cos\beta_o \sin\alpha_o). \quad [58]$$

(Note that this expression for ϕ_o could also have been obtained by letting $z_o \rightarrow -\infty$ in Eq. [35]).

Assuming both reference and readout beams are plane, of identical wavelength, and lie in the y - z plane, we substitute Eqs. [53], [54] and [58] into Eqs. [43] and [44], giving for a Fraunhofer hologram

$$\phi_p \approx \frac{2\pi}{\lambda_o} \left(x \sin\beta_o + y \cos\beta_o \sin\alpha_o - y \sin\alpha_r + y \sin\alpha_R \right) \quad [59]$$

$$\phi_c \approx \frac{2\pi}{\lambda_o} \left(-x \sin\beta_o - y \cos\beta_o \sin\alpha_o + y \sin\alpha_r + y \sin\alpha_R \right). \quad [60]$$

If it so happens that the point object lies in the y - z plane then $\beta_o = 0$ and Eqs. [59] and [60] become

$$\phi_p \approx \frac{2\pi}{\lambda_o} y \left(\sin\alpha_o - \sin\alpha_r + \sin\alpha_R \right) \quad [61]$$

$$\phi_c \approx \frac{2\pi}{\lambda_o} y \left(-\sin\alpha_o + \sin\alpha_r + \sin\alpha_R \right). \quad [62]$$

These equations tell us the directions in which the Fraunhofer hologram reconstructed images lie for an object point located in the y - z plane at $-\infty$; i.e., for $x_o = 0$ and $z_o \rightarrow -\infty$. If we define α_p and α_c , respectively, as the angles between the positive direction of the z axis in Fig. 1 and the direction in which the primary and conjugate images lie, we can identify

$$\sin\alpha_p = \sin\alpha_o - \sin\alpha_r + \sin\alpha_R \quad [63]$$

$$\sin\alpha_c = -\sin\alpha_o + \sin\alpha_r + \sin\alpha_R \quad [64]$$

from Eqs. [61] and [62]. For identical reference and readout beams, $\alpha_r = \alpha_R$, these become

$$\sin\alpha_p = \sin\alpha_o \quad [65]$$

$$\sin\alpha_c = -\sin\alpha_o + 2\sin\alpha_r. \quad [66]$$

When using these formulas we must keep in mind our convention that all α -angles lie in the y - z plane and are considered positive when referenced counterclockwise from the positive direction of the z axis.

Example: For a point object located at $z_o \rightarrow -\infty$ in the second quadrant of the y - z plane at 12° from the negative z axis, $\alpha_o = -12^\circ$. For a reference beam in the third quadrant coming into the origin of the y - z plane at an angle of 18° with respect to the negative z axis, $\alpha_r = 18^\circ$. Substituting these values into Eqs. [65] and [66], we find that on playback with $\alpha_R = \alpha_r$, the primary and conjugate images lie, respectively, in the fourth quadrant at $\alpha_p = -12^\circ$ and in the first quadrant at $\alpha_c \approx 56^\circ$.

If $\alpha_R = -\alpha_r$, Eqs. [63] and [64] become

$$\sin\alpha_p = \sin\alpha_o - 2\sin\alpha_r \quad [67]$$

$$\sin\alpha_c = -\sin\alpha_o \quad [68]$$

Example: For the same hologram recording geometry as in the previous example, but with $\alpha_R = -\alpha_r$, the primary and conjugate images lie, respectively, in the fourth quadrant at $\alpha_p = -56^\circ$ and in the first quadrant at $\alpha_c = 12^\circ$.

These examples reveal a striking playback difference between Fresnel and Fraunhofer holograms. The Fresnel hologram reconstructed images lie on *opposite* sides of the hologram plane (i.e., one is virtual and one is real); reconstruction from a Fraunhofer hologram produces primary and conjugate images on the *same* side of the hologram plane.

5.3 Lensless Fourier Transform Hologram

As introduced in Section 2 in the discussion associated with Eq. [2], the way in which r_o is expanded in the exponent of Eq. [1] in order to make mathematical analysis of the hologram process tenable can result in the restrictions of paraxial analysis. Eq. [3a] does imply such a restriction but is certainly of practical use in some cases; e.g., recording Fresnel holograms with collimated reference beam and object beam normal (and therefore satisfying the paraxial restriction) to the hologram plane. For a Fraunhofer hologram, the question of expanding r_o does not arise since both object and reference beams are

collimated; however, in the case of lensless Fourier transform holography the proper expansion of r_o is of great importance, since both object and reference points are at finite distances from and usually far off the perpendicular to the hologram plane. Thus, we use Eq. [3b] and expand r_o in a Maclaurin series about d_o rather than z_o . Such an expansion is convergent for $d_o^2 > x^2 + y^2 - 2xx_o - 2yy_o$. This convergence requirement is less severe than that of the paraxial approximation associated with Eq. [4], because it requires only that light rays emanating from the object transparency make small angles with respect to d_o rather than z_o . Thus, the dimensions (x_o, y_o) of the object transparency and those of the hologram (x, y) must be small relative to d_o , but the object may be located at any angle relative to the z axis (see Fig. 1).

Expanding Eq. [3b] about d_o and retaining terms only through the second order yields

$$r_o \approx d_o + \frac{x^2 + y^2}{2d_o} - \frac{xx_o}{d_o} - \frac{yy_o}{d_o} - \frac{1}{8d_o^3} \left[(x^2 + y^2)^2 - 4(x^2 + y^2)(xx_o + yy_o) + 4(xx_o + yy_o)^2 \right] \quad [69]$$

or

$$r_o \approx d_o + \frac{x^2 + y^2}{2d_o} - \frac{xx_o}{d_o} - \frac{yy_o}{d_o}. \quad [70]$$

Similarly, expansions of r_r and r_R about the distances d_r and d_R , respectively, of the reference and readout point sources from the origin yield

$$r_r \approx d_r + \frac{x^2 + y^2}{2d_r} - \frac{xx_r}{d_r} - \frac{yy_r}{d_r} \quad [71]$$

$$r_R \approx d_R + \frac{x^2 + y^2}{2d_R} - \frac{xx_R}{d_r} - \frac{yy_R}{d_R}. \quad [72]$$

Using these relationships, the phase factors ϕ_o , ϕ_r , and ϕ_R become

$$\phi_o = \frac{2\pi}{\lambda_o} (r_o - d_o) \approx \frac{2\pi}{\lambda_o} \left(\frac{x^2 + y^2}{2d_o} - \frac{xx_o}{d_o} - \frac{yy_o}{d_o} \right) \quad [73]$$

$$\phi_r = \frac{2\pi}{\lambda_o} (r_r - d_r) \approx \frac{2\pi}{\lambda_o} \left(\frac{x^2 + y^2}{2d_r} - \frac{xx_r}{d_r} - \frac{yy_r}{d_r} \right) \quad [74]$$

$$\phi_R = \frac{2\pi}{\lambda_R} (r_R - d_R) \approx \frac{2\pi}{\lambda_R} \left(\frac{x^2 + y^2}{2d_R} - \frac{xx_R}{d_R} - \frac{yy_R}{d_R} \right). \quad [75]$$

As shown in Fig. 2, $d_r \approx d_o$ for all object points; also, $\lambda_R = \lambda_o$ and $d_R = d_r$ usually. Using these equalities and substituting Eqs. [73], [74], and [75] into Eqs. [43] and [44] yields

$$\begin{aligned} \phi_p &= \phi_o - \phi_r + \phi_R \\ &= \frac{2\pi}{\lambda_o} \left(\frac{x^2 + y^2}{2d_o} - \frac{xx_o}{d_o} - \frac{yy_o}{d_o} \right) \end{aligned} \quad [76]$$

$$\begin{aligned} \phi_c &= -(\phi_o - \phi_r - \phi_R) \\ &= \frac{2\pi}{\lambda_o} \left(\frac{x^2 + y^2}{2d_o} - \frac{x(2x_r - x_o)}{d_o} - \frac{y(2y_r - y_o)}{d_o} \right). \end{aligned} \quad [77]$$

Comparing these with ϕ_o in Eq. [73], we see that the primary and conjugate image coordinates are

$$\begin{aligned} d_p &= d_o & d_c &= d_o \\ x_p &= x_o & x_c &= 2x_r - x_o \\ y_p &= y_o & y_c &= 2y_r - y_o. \end{aligned} \quad [78]$$

To interpret this result, we note first that the angular direction of d_o is determined from the signs of the object coordinates x_o, y_o, z_o . Thus, Eq. [78] indicates that if the object point is to the left of the hologram plane (as illustrated in Fig. 1), so are the primary and conjugate images. We must keep in mind, of course, that this is true because we assumed the reference point source was also located to the left of the hologram plane in our derivation of Eq. [78]. On readout we have an alternative choice for the location of the *readout* point source. It could be located either to the left *or* right of the hologram plane. In other words, since the readout beam must be spherically shaped it

either emanates from a point to the left of the hologram plane ($z_r = z_o$, z_o negative) or converges to a point to the right of the hologram plane ($z_r = -z_o$).

For $z_r = z_o$ and $x_r = -x_o$, Eq. [78] tells us there exist two *virtual* images—a primary aberration-free image at the same position ($x_p \equiv x_o$) as the original object and a conjugate image located at $x_c = 2(-x_o) - x_o = -3x_o$, or a distance $|2x_o|$ below the position of the reference point source in Fig. 2. Of course, this position is specifically that for the object point located exactly at a distance $|x_o| = |x_r|$ from the z axis. Since there is only one point reference source but many object points, the general relationship is $x_c \approx 2x_r - x_o = -3x_o$.

For $z_r = -z_o$ and $x_r = -x_o$, the readout beam converges to a point to the right of the hologram plane, i.e., a mirror image (with respect to the x - y plane) of the original reference point source location. In this case, Eq. [78] indicates that two *real* images are produced that are inverted with respect to the previous readout condition. That is, the change of signs of x_R in Eq. [75] with respect to x_r in Eq. [74] causes Eq. [78] to yield an aberration-free conjugate image.

Comparing these lensless Fourier transform hologram results with those of the Fresnel and Fraunhofer holograms, we note the unique distinction that in the LFT case either virtual or real images are produced on readout but *never* a combination.

6. Recording Arrangements and Practical Considerations

For the recording of sequences of holograms with the intention of producing motion picture images on playback, the discussions above reveal that CW readout is possible with Fraunhofer and lensless Fourier transform holograms, while Fresnel holograms require a pulsed readout beam to preserve image stability. Thus, for such an application the Fraunhofer and lensless Fourier transform holograms permit construction of a more simple playback mechanism, e.g., since CW readout is possible, frame-to-frame synchronization is not required. The lensless Fourier transform hologram has the obvious advantage of requiring no lens, but also the slight disadvantage (because of the approximate Fraunhofer nature of the lensless Fourier transfer hologram) of yielding not quite as good image immobility as a Fraunhofer hologram. The emphasis on phase holograms in this paper is based, of course, on the desire to record easily-replicable holograms. Thus, a pragmatic thread running through the presentation of recording arrangements that follows is the necessity to record these holograms

in the form of relief patterns to permit mass production of inexpensive vinyl tapes, as discussed in detail elsewhere in the literature.²

Recording arrangements for the different types of holograms discussed above are illustrated in Figs. 4, 5, and 6. The entire recording-table arrangement is shown in Fig. 4 for the Fresnel hologram; Figs. 5 and 6 illustrate just the portions of the recording arrangement that differ from those in Fig. 4 for the Fraunhofer hologram and the lensless Fourier transform hologram.

In the selection of components for the various recording arrangements, consideration must first be given to stability. Since optical holography demands the recording of fringe spacings on the order of a micrometer, the recording table must be isolated from building vibrations. Isolation can be accomplished by a variety of techniques, the most common being suspension of a massive slab of granite or rigid metal (typically aluminum or steel) or a wood platform on (1) air-inflated rubber tubes, (2) an interfaced series of thick rubber sheets and sandboxes, or (3) a set of air-piston legs. If cost is not a factor, the massive granite slab supported by air-piston legs seems to be the optimum configuration; however, quite satisfactory "stable tables" of honeycombed metal and wood structures supported by air-piston legs are also available at much lower costs. (Some manufacturers also offer magnetic-based optical bench components as a convenience for use on steel-topped tables, while other manufacturers usually provide drilled and $\frac{1}{4}$ -20 tapped holes spaced on 2-4 inch centers to permit the fastening of components to the table top.) In any case, it is advisable to encase the table-top within an enclosure in order to minimize air currents which tend to move (refract) the object and reference beams during recording. The usual material used for this purpose is plastic, which is opaque at the wavelength of the laser light and therefore isolates it from the rest of the laboratory for personnel safety.

The laser must be chosen to satisfy the irradiance and wavelength-sensitivity requirements of the hologram recording material, and its coherence length must be of sufficient magnitude to permit flexible arrangements on the recording table. At the present state of the art of hologram recording on polyester (e.g., Cronar) tape coated with Shipley AZ-1350 photoresist,² an argon or helium-cadmium laser is used. For the Ar laser, $\lambda = 457.9$ nm, output power = 250 mW, and coherence length = 6 cm; for the HeCd laser, $\lambda = 441.6$ nm, output power = 100 mW, and coherence length = 25 cm for isotope Cd or 6 cm for natural Cd. For uniform irradiance across the hologram recording plane, these lasers are operated in the TEM₀₀ mode. In this mode,

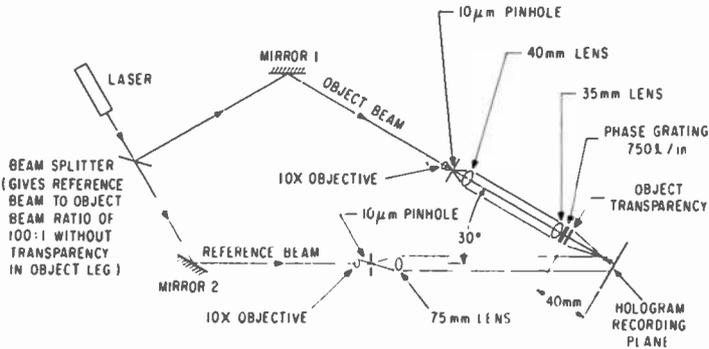


Fig. 4—Typical Fresnel hologram recording arrangement.

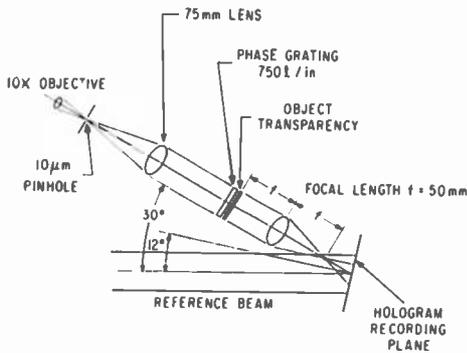


Fig. 5—Typical Fraunhofer hologram recording arrangement. Remainder of recording arrangement to left same as in Fig. 4.

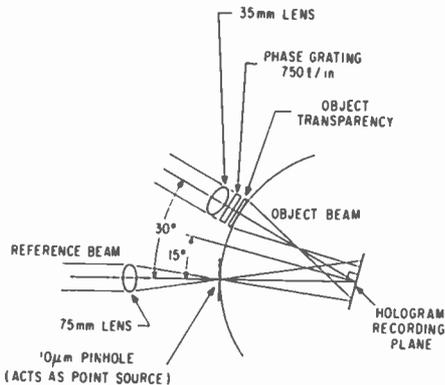


Fig. 6—Typical lensless Fourier transform hologram recording arrangement. Remainder of recording arrangement to left same as in Fig. 4.

the laser output is essentially a plane wavefront with gaussian-distributed energy cross section—the so-called “gaussian energy profile”.¹⁵

Since the laser generates considerable heat and therefore is likely to create air currents, it should not be included with the other recording components within the table-top plastic enclosure. Further if it is mechanically cooled in any way (e.g., by blower or water), it should be moved to a separate table to isolate its vibrations. In any case, care should still be given to enclosing the output beam within a tube of some sort along its entire length.

As shown in Fig. 4, the coherent light from the laser is separated into object and reference beams by a variable beam splitter. The “variability” property of this beam splitter is quite important, because it permits the careful control of the radiance of its reflected and transmitted beams for optimum reference-beam/object-beam radiance ratio and, therefore, balancing linearity of hologram recording against holographic diffraction efficiencies. Also, from another point of view, the use of a variable beam splitter permits maintaining a desired object-to-reference beam ratio for a variety of object transparencies.

The mirrors following the beam splitter in Fig. 4 enable proper orientation of the object and reference beams to achieve a desired angle of interference at the hologram recording plane. To minimize radiance loss, the coatings on these mirrors should be selected for optimum reflectance at the desired wavelength and angle. To maintain spatial coherence after reflection, the coating must be uniform across the mirror surface.

The beam expanders in both the reference- and object-beam legs of the recording arrangements must be selected with several constraints in mind. Basically, the beam expanders are used to provide a uniform irradiance across the hologram plane and to allow insertion of a pinhole spatial filter at their common internal focal point. Since the cross-sectional output radiance distribution of the laser beam is gaussian shaped,¹⁵ the beam-expander parameters are chosen to spread the beam sufficiently so that the central beam area covers the hologram area at the recording plane. The formula for this design aspect is $d_2/d_1 = f_2/f_1$, where d_2 is the desired output beam diameter, d_1 the input beam diameter, f_2 the focal length of the output lens, and f_1 is the focal length of the input lens. The lenses should be AR-coated and wavefront-corrected to $\lambda/20$ at least; obviously no chromatic corrections are needed, however, because of the single-color recording light. The pinhole spatial filters “clean” the light in the object and reference beams, essentially by filtering out high-spatial-frequency light components and passing primarily the fundamental. The diameter of the

pinhole is chosen to provide a diffraction limited spot at the $1/e^2$ points of the gaussian-shaped beam at the output of the beam expander in accordance with the formula $D = 4\lambda f_1/\pi d_1$. In practice, the pinhole diameter is usually chosen equal to $2D$.

The angular orientation of the two beams shown in Fig. 4 is typical of the present state of the art. The selection of this angle derives from a consideration of several factors. Foremost among these is the resolution capability of the recording material—a photoresist in the case of embossable phase holograms.³ The spatial resolution capability of a given photoresist is usually expressed in terms of the highest spatial frequency f it can reliably record. Knowing this frequency, the maximum angles between the reference and object beams and the normal to the recording plane can be found from the equation³

$$f \cong f_c + f_i = \frac{\sin\theta_r + \sin\theta_o}{\lambda},$$

where f = maximum spatial frequency resolution capability of recording material

f_c = carrier spatial frequency of the hologram

$2f_i$ = information spatial frequency bandwidth of the object transparency

λ = recording laser wavelength

θ_r = maximum angle between reference beam and normal to recording plane

θ_o = maximum angle between object beam and normal to recording plane.

While this equation yields maximum reference- and object-beam angles based on maximum resolving power of the photoresist, we must also be concerned about minimum angles in order to avoid hologram spectrum overlap. To avoid spectrum overlap, $f_c \cong 3f_i$. As an example, $f \approx 1500$ line pairs/mm for Shipley AZ-1350.³ For a typical motion-picture-frame object transparency, $2f_i = 600$ line pairs/mm or $f_i = 300$ line pairs/mm. Since $f_c \cong 3f_i$, f_c can be chosen as 1200 line pairs/mm and, from the above equation, the angle between reference and object beams for $\lambda = 441.6$ nm is 30° .

In the recording of hologram motion pictures, another factor influencing angular orientation is the physical limitation placed on the

mechanical design of the object-film and recording-tape transports by the desired hologram size. If very small holograms are to be recorded a requirement for a particular angle between the object and reference beams may have to be modified in order to accommodate the constraints imposed by the transport structures.

The phase grating in the object-beam leg in Figs. 4, 5 and 6 is not a necessary component of a hologram recording arrangement; however, it offers certain advantages. Without the phase grating, a nonredundant hologram recording results; i.e., the object information is recorded uniquely across the hologram plane. Insertion of a phase grating (of which there are several varieties) into the object beam in the position shown in Fig. 4 spreads information from each object point over several portions of the recording plane, resulting in a redundant hologram. In addition to making the hologram immune to mutilation, redundant recording causes the reconstructed images to be relatively free of defects caused by all the optical elements in the system. These include optical surface scratches, dirt particles, and similar defects that plague coherent optical systems. The improvement effected by this use of a grating is vividly demonstrated in Ref. [1] and [2] and elsewhere in this issue of RCA Review.*

The final lens in the object-beam leg of Fig. 4 is used to control hologram size at the recording plane; in particular, its focal length (e.g., 35 mm in Fig. 4) is chosen as a tradeoff among desired hologram size, required signal-to-noise ratio, acceptable resolution, and resolution capability of the recording material.¹ For the system under consideration in this paper, this tradeoff allows a minimum Fresnel hologram size of approximately 1 mm². For the Fraunhofer hologram setup of Fig. 5, the focal length of the lens (50 mm in Fig. 5) is again chosen on the basis of a tradeoff involving hologram size, signal-to-noise ratio, and resolution, but notice that it follows rather than precedes the object transparency.² The object transparency is positioned at the back focal plane of the lens in order to produce the Fourier transform of the object transparency at positions beyond the lens. The exact Fourier transform is formed of course, at the front focal plane, but because of the limited dynamic range of present recording materials, the recording is made to either side of the front focal plane.

Notice in Fig. 6 that lenses are placed in both the object and reference beams in order to provide spherical wavefronts of equal sphericity and thus to form a lensless Fourier transform as explained

* A. H. Firester, E. C. Fox, T. E. J. Gayeski, W. J. Hannan, and M. J. Lurie, "Redundant Holograms," this issue, p. 131.

in Sec. 5.3. In accordance with the requirements of forming a lensless Fourier transform hologram, the final reference-beam lens is positioned so that its front focal point is located at the same distance from the recording plane as the object transparency. A pinhole spatial filter is placed at this focal point so as to closely approximate an ideal point source. The focal length of this lens is chosen to provide a reference-beam size at the recording plane that overlaps the object-beam size determined by the final lens in the object-beam leg.

Note the different orientations of the recording plane in Figs. 4, 5, and 6. In Fig. 4, the recording plane is oriented normal to the object-beam axis in order to satisfy the Fresnel hologram analysis above, while in Figs. 5 and 6 the recording plane lies oriented symmetrically with respect to both the object- and reference-beam axis. For Fig. 6, the LTF theory requires that the recording plane lie normal to the bisector of the angle between the object- and reference-beam legs. For the Fraunhofer arrangement in Fig. 5, the hologram recording plane lies normal to the "approximate" bisector of the angle between the object- and reference-beam legs. This arrangement offers a greater degree of image immobility on transmission readout when the hologram is subjected to an incremental rotation about the axis in the plane of the hologram.³ According to the theory of Ref. [2], the hologram plane should lie normal to the "exact" bisector of the angle between object- and reference-beam legs; in practice, however, this would cause the reference beam to reflect into the image on recording. Therefore, the hologram is positioned slightly off the normal (e.g., 12° instead of 15° in Fig. 5) on recording but is read out symmetrically. The image is of course, somewhat more sensitive to holotape twist due to this change but, fortunately, the resulting image defects are quite tolerable.*

Playback arrangements for the three hologram types are shown in Fig. 7, where only the positions of the aberration-free images are indicated. One obvious advantage of the Fresnel hologram evident from this figure is that no lens is needed on readout. However, a significant disadvantage is that such a hologram does not yield immobile images with respect to hologram motion; thus, for a motion picture application, either a pulsed laser or frame shuttering with appropriate synchronization would be required to prevent image smear.

Since both the Fraunhofer and LFT holograms do yield immobile images on readout, a simple cw laser (e.g., HeNe) can serve as the coherent source. Both require lenses but the LFT has a more stringent

* R. A. Bartolini, M. Lurie, and D. Karlsons, "Wavelength Dependent Distortion in Fraunhofer Holograms—Applications to RCA Holotape", this issue, p. 154.

lens requirement—a precision duplication of the reference beam is necessary to provide acceptable wavefront reconstruction. In either case, one demand on the playback transport mechanism is to keep the tape as flat as possible, which is a requirement for proper Fourier transformation. In this regard, to help minimize image motion caused

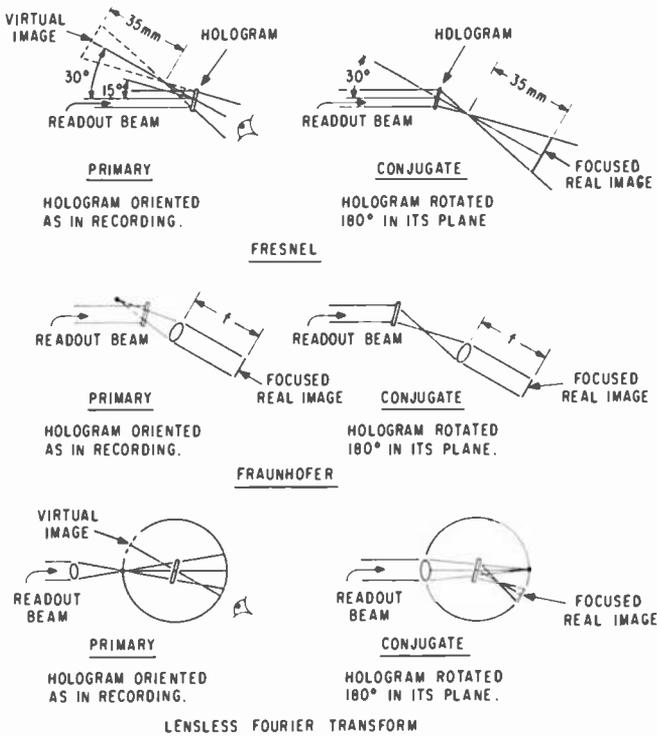


Fig. 7—Transmission playback arrangements for reading out *undistorted* primary and conjugate images from Fresnel, Fraunhofer, and LFT holograms.

by unavoidable time-varying tape twist as the tape moves through the readout beam, a readout angle equal to the diffraction angle is chosen (both angles referred to the normal to the hologram plane). It has been shown² that this “symmetrical” angle readout technique offers the greatest degree of tolerance to hologram misalignment, consistent with image-resolution requirements. In addition, use of Fraunhofer holograms permits simple readout alignment, since the resulting image is in focus at infinity, i.e., the position of the hologram with respect to the readout lens is noncritical within the field of view of the readout lens.

Summary

The theory, rules of thumb, and practical aspects of recording replicable holograms on vinyl tape for use in RCA's Holotape Television playback system have been presented. The requirements of replication dictate that phase holograms be recorded, and the use of Fraunhofer holography along with symmetrical angle readout provides motion-picture immobility with no need for a pulsed laser or a shutter. The recording configurations for comparing the several types of hologram recording were shown in detail and the selection of each component was discussed. The theory underlying these configurations was presented and the distinctions between the paraxial and nonparaxial approaches to analysis were delineated.

References:

- ¹ H. J. Gerritson, W. J. Hannan, and E. G. Ramberg, "Elimination of Speckle Noise In Holograms with Redundancy," *Appl. Opt.*, Vol. 7, p. 2301, Nov. 1968.
- ² R. A. Bartolini, W. J. Hannan, D. Karlsons, and M. J. Lurie, "Embossed Hologram Motion Pictures for Television Playback," *Appl. Opt.* Vol. 9, No. 10, p. 2283, Oct. 1970.
- ³ E. N. Leith and J. Upatnieks, "Reconstructed Wavefronts and Communication Theory," *J. Opt. Soc. Am.*, Vol. 52, p. 1123, Oct. 1962.
- ⁴ E. G. Ramberg, "The Hologram—Properties and Application," *RCA Review*, Vol. XXVII, p. 467, Dec. 1966.
- ⁵ M. Françon, *Holographie*, Masson et Cie, Paris, 1969.
- ⁶ J. W. Goodman, *Introduction to Fourier Optics*, pp. 86-87, McGraw-Hill Book Co., New York, 1968.
- ⁷ G. W. Stroke, "Lensless Fourier-Transform Method for Optical Holography," *Appl. Phys. Ltrs.*, Vol. 6, p. 201, May 15, 1965.
- ⁸ H. J. Smith, *Principles of Holography*, John Wiley & Sons, New York, 1969.
- ⁹ R. W. Meier, "Magnification and Third-Order Aberrations in Holography," *J. Opt. Soc. Am.*, Vol. 55, p. 987, Aug. 1965.
- ¹⁰ M. Born and E. Wolf, *Principles of Optics*, 3rd (revised) edition, p. 380, Pergamon Press, New York, 1965.
- ¹¹ E. B. Champagne, *A Qualitative and Quantitative Study of Holographic Imaging*, Ph.D. Dissertation, The Ohio State University, Columbus, Ohio 1967.
- ¹² J. N. Latta, "Computer-Based Analysis of Hologram Imagery and Aberrations II: Aberrations Induced by a Wavelength Shift," *Appl. Opt.* Vol. 10, p. 609, March 1971.
- ¹³ E. B. Champagne, "Nonparaxial Imaging, Magnification and Aberration Properties in Holography," *J. Opt. Soc. Am.*, Vol. 57, p. 51, Jan. 1967.
- ¹⁴ G. W. Stroke and D. G. Falconer, "Theoretical and Experimental Foundations of Wavefront-Reconstruction Imaging," in J. T. Tippett, et al (editors), *Optical and Electro-Optical Information Processing*, 2nd printing, The M.I.T. Press, Cambridge, Mass., 1968.
- ¹⁵ A. L. Bloom, *Gas Lasers*, John Wiley & Sons, New York, 1968.

Thermoplastic Media for Holographic Recording

T. L. Credelle and F. W. Spong

RCA Laboratories, Princeton, N. J.

Abstract—Thermoplastic-photoconductor media for holographic recording combine the advantages of high sensitivity and resolution, dry and nearly instantaneous *in situ* development, erasability, and high readout efficiency. This article reviews the structure, fabrication, and techniques for holographic recording in this medium. Several new results are reported including holographic grating efficiencies of 40%, image reconstruction efficiencies of 35% with little distortion, resolution in excess of 4100 cycles/mm, insensitivity to high ambient light levels during recording, and a post development step to increase the diffraction efficiency of weak holograms.

Introduction

There has been an interest in deformable media for recording and display devices since the early 1940's¹⁻⁷ and an interest in deformable films for holography since 1966.^{8,9} As a recording material, the deformable film has several advantages: all processing is dry, development is essentially instantaneous and can be done *in situ*, information can be erased and the medium reused, dark handling is eliminated in most cases, resolution in excess of 4000 cycles per millimeter can be achieved, and either light or electrical charges can be used for exposure.

As a holographic recording material, thermoplastics—one class of deformable films—have the following desirable properties: good sensitivity (comparable with very high resolution photographic emulsions, e.g., Kodak 649), high efficiency (since thermoplastics are a thin phase media), erasability, development *in situ*, and bandpass spatial frequency response, which we have shown reduces intermodulation distortion.

In this article we review holographic recording in thermoplastic-photoconductor media including sample fabrication and recording techniques. Examples of holographic gratings and Fraunhofer holograms are presented. We also report several new results in thermoplastic holographic recording: holographic interference gratings with 40% diffraction efficiency, image reconstruction efficiencies of 35% with little intermodulation distortion, resolution in excess of 4100 cycles/mm, demonstration of insensitivity to high ambient light levels while recording, holograms of incandescent objects, and a post development step to increase the diffraction efficiency of weak holograms.

Thermoplastic Holography

The thermoplastic hologram is recorded as a thickness variation corresponding to the light intensity pattern of a holographic fringe pattern. Since the thermoplastic is usually light-insensitive, it must be combined with a photoconductor, either in a "sandwich" structure or in a single layer where the photoconductor is dissolved or dispersed in the thermoplastic.^{3,4} The sandwich structure, i.e., a thermoplastic layer overcoated on a photoconductor layer, has been used throughout these experiments because higher photosensitivity is obtainable for the thermoplastic thicknesses used to record at high resolution (e.g., 0.4 μm at 1200 cycles/mm). For this reason the discussion will be limited to a description of the sandwich structure only.

The recording medium consists of this sandwich structure on a conductively-coated substrate. Shown in Fig. 1 is the basic structure; there are three elements of interest coated on a smooth, optical quality substrate, e.g. glass or Cronar (an optical quality polyester film available from E. I. DuPont Co. Inc.). The conductive coating is a transparent or semi-transparent layer either sputtered or evaporated to obtain the necessary surface resistance. Next an organic photoconductor is coated from a solvent solution to a thickness of 1-3 μm for these experiments. Finally a layer of thermoplastic is overcoated onto the photoconductor—the thickness is determined by the spatial frequency response desired.¹⁰

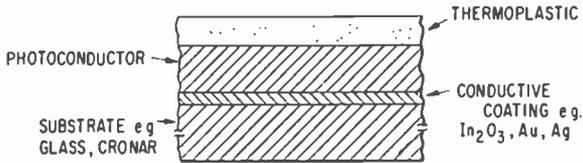


Fig. 1—Basic recording medium—"sandwich" structure.

The usual technique of recording can be described by referring to Fig. 2. The first step is to establish a uniform charge on the free surface of the thermoplastic, e.g., with a corona-charging device. The voltage, typically several hundred volts, is capacitively divided between the photoconductor and thermoplastic layers. The second step is exposure to illumination; the photoconductor conducts in the illuminated areas and acts to discharge the voltage across it. However, the surface charge density on the free thermoplastic surface is unchanged; hence, the electric field at the surface is uniform and, therefore, the electrostatic force is constant along the surface. To establish a force variation on the thermoplastic, the surface is recharged to a uniform potential, thus adding charge to the thermoplastic-photoconductor sandwich where illuminated. Finally, the thermoplastic is heated to its softening temperature, typically 50°-100°C, and the electrostatic forces deform the surface as shown. Deformation proceeds until the electrostatic forces are balanced by surface tension restoring forces, unless limited by viscous forces. When the thermoplastic cools, the

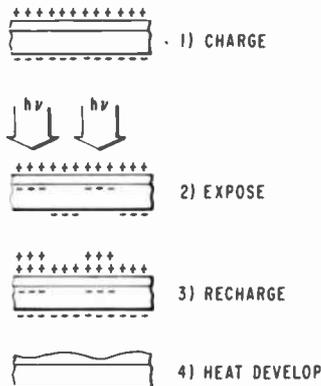


Fig. 2—Sequential recording technique. For simplicity, only the thermoplastic (top layer) and photoconductor are shown.

information is "frozen in" as a surface relief pattern. To erase the information the thermoplastic is heated so that surface tension smooths out the thickness variation. Any residual charge pattern will be discharged by the increased electrical conductivity of the thermoplastic and photoconductor.

We have used two other techniques with success in the recording of thermoplastic holograms. These techniques involve "simultaneous" rather than "sequential" charging and exposure; that is, exposure to incident radiation is made during corona charging. The advantage of the simultaneous technique is that a larger charge-density modulation can be created on the thermoplastic surface; it also reduces recording to a two-step process.

Simultaneous recording can be done in one of two ways: (1) charging and exposure before heating and (2) charging and exposure immediately after or during heating. In the first case the charge pattern (corresponding to the holographic fringe pattern) is created while the film is at ambient temperature; a developing heat pulse is then applied. In the second case, the charge pattern is formed as the thermoplastic cools. Since charging and exposure continues while the thermoplastic surface is deforming, there exists an increasing electrostatic force; i.e., if the potential on the thermoplastic surface is maintained (by continuous charging, for example) as deformation occurs, the charge density in the grooves will be increased. Thus the electrostatic forces will also be increased in that region. In this mode of operation, the deformation will continue until either the surface tension and viscous forces balance the increased electrostatic force or until the thermoplastic film is torn apart along the bottom of the grooves. We have seen this tearing phenomenon in scanning electron micrographs of high efficiency thermoplastic holograms.

Sample Fabrication

When choosing materials to be used in thermoplastic recording, certain requirements on the substrate, the photoconductor, and the thermoplastic must be considered. The substrate should have a good, optically flat surface and should be at least semi-transparent for viewing in transmission. If the conductive coating is to serve as a resistive heater, it must withstand 1-10 J/cm² heat pulses. In addition it should be semi-transparent, but not necessarily anti-reflecting, as reflection fringes are not recorded in thermoplastic holograms (see experimental results and discussion). If an external heat source is used (e.g., hot air or radiant heating), the conductive coating can have a surface resist-

ance as high as 10^5 ohms/square and still provide a ground for charging. The photoconductor may be any organic or inorganic photoconductor that is sensitive to the incident radiation used and that has a low enough dark current to prevent rapid dark decay. The main requirements on the thermoplastic are that it have a high enough softening temperature to prevent cold flow at room temperature, that it have a high enough resistivity so that the charge pattern placed on the free thermoplastic surface does not decay before the pattern is heat developed, and that the solvent used for coating the thermoplastic layer does not attack the photoconductor layer.

The substrate materials used in most of these experiments were commercial indium oxide coated glass,* evaporated chromium-gold on glass, and evaporated chromium-gold on 0.004-inch Cronar. Hologram size was typically 10×10 mm. As substrate heating was used in this study, the surface resistance of the conductive coating was from 10 to 50 ohms/square.

An organic photoconductor, polyvinyl carbazole (PVK),† sensitized with 2,4,7 trinitro-9-fluorenone (TNF),† was chosen. It has nearly panchromatic photoresponse,¹¹ a low dark current, and can be coated easily from a solvent solution. PVK/TNF with a molar ratio of 8:1, was dissolved in a 1:1 mixture of 1,4 dioxane and dichloromethane. Concentration (approximately 6%) of solids was adjusted to yield a $2.5 \mu\text{m}$ coating when "pulled" at approximately 13 cm/min. Samples were "pulled" in a dry environment (relative humidity < 25%) to prevent fogging of the photoconductor layer. We have used other solvents, including tetrahydrofuran and 1,1,2 trichloroethane; the latter solvent may be used with humidities up to 65%.

Photosensitivity at 441.6 nm of this photoconductor was measured with a Monroe feedback electrometer; typical results are shown in Fig. 3. The photoconductor layer was first charged to its breakdown limit. The photosensitivity curves were obtained by exposing the charged layer to short pulses of light and monitoring the change in surface potential. The data shown in Fig. 3 is for a $2.9\text{-}\mu\text{m}$ -thick photoconductor layer with a gold electrode. The breakdown-field strength was 1.1×10^6 V/cm. Sensitivity at 441.6 nm was 22 nC/ μJ for a charge contrast of 110 nC/cm². The average quantum efficiency was 0.11 electrons/absorbed photon. PVK/TNF has a field dependent quantum efficiency;¹¹ the value of 0.11 is an average from the breakdown field to half that value.

* Available from Pittsburgh Plate Glass

† Available from Polysciences, Inc.

The thermoplastic used in most of these experiments was Staybelite Ester 10,* a derivative of a natural tree resin, dissolved in petroleum ether or hexane. A 0.5- μm coating was obtained when the sample was pulled from a 20% solution at a rate of about 13 cm/min. All layer thicknesses were measured with a Watson Barnet interferometric microscope objective. The "write" temperature of this thermoplastic is in the 40-50°C range; the "erase" temperature is approximately 70°C. The complete thermoplastic-photoconductor film was baked for 1 hour at 60°C.

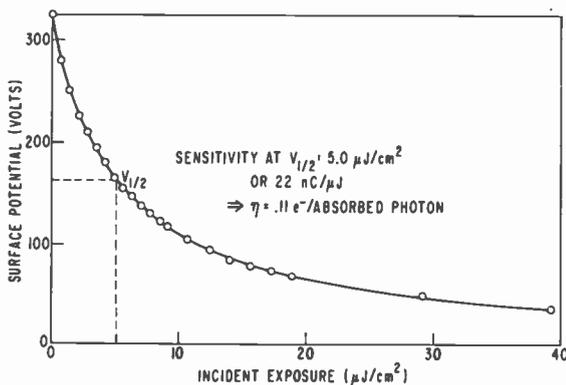


Fig. 3—Photosensitivity of 2.9 μm layer 8:1 molar ratio PVK/TNF. Conductive substrate consists of evaporated Au-Cr on glass.

Experimental Apparatus

The techniques used to investigate the properties of thermoplastic holograms involved making both two-beam holographic gratings and Fraunhofer holograms. The basic experimental apparatus, shown in Fig. 4, consists of a standard holographic table where the object and reference beams intersect at an angle of 30°. A He-Cd laser provides approximately 100 mW of power at 441.6 nm. The corona power supply used to charge the thermoplastic-photoconductor film is a 0-10 kV supply and can be operated at either positive or negative polarity. Typically, positive voltage of 6-10 kV were used in our experiments. The heater pulse supply consists of a pulse generator, a mercury relay, and a dc power supply for resistively heating the conductive layer on glass slides; to heat the flexible substrates, the pulse generator is used

* Available from Hercules, Inc.

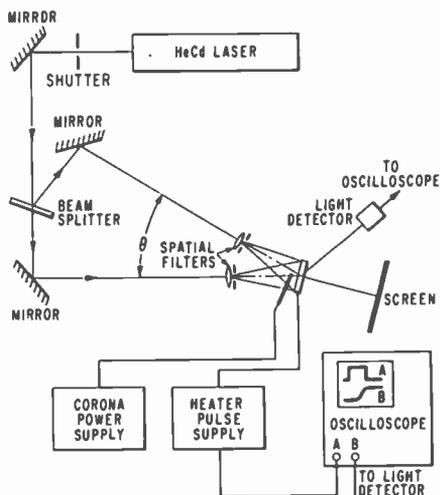


Fig. 4—Experimental apparatus for two-beam interference-grating exposure.

in conjunction with an rf transmitter (described below). Readout of these holograms is immediate, as the development is done *in situ*. The projected interference pattern between the object and reconstructed wavefronts can be viewed on a screen, or the power diffracted into the first order can be monitored with a Si photodiode. Use of this detector with an oscilloscope, as shown in Fig. 4, allows accurate determination of the heat pulse required for full development.

The rf heating technique is used with the flexible Cronar substrates.³ Heating is still through Joule losses in the resistive layer, but contact is made capacitively through the Cronar. Fig. 5 is a diagram of the heating apparatus and equivalent circuit. Two brass

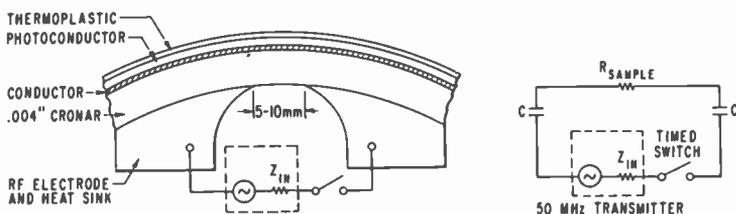


Fig. 5(a)—RF heating apparatus and (b) simplified equivalent circuit. Capacitance is determined by thickness of Cronar and area of tape in contact with rf electrode.

“shoes” serve as rf electrodes and as heat sinks to localize heating to the gap. Rf power, approximately 50-100 watts, is capacitively coupled through the Cronar; rf current then flows through the resistive layer, heating the thermoplastic. Capacitive reactance of the 16-mm Cronar at 50 MHz is approximately 25 ohms for a 3-cm electrode. Using this technique, series of holograms have been made on tape without erasing adjacent information.

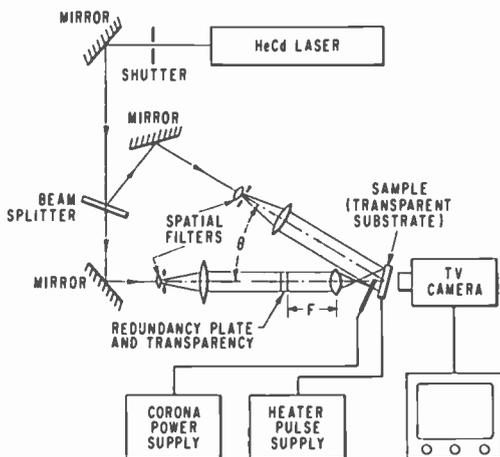


Fig. 6—Fraunhofer recording and readout apparatus.

The Fraunhofer recording and readout apparatus is shown in Fig. 6. It consists of a collimated reference beam and an object path including a redundancy plate, a film transparency, and a Fraunhofer lens that is a focal distance from the object.¹² A television camera is used to monitor the “object” image and the reconstructed image after the hologram is formed. The charging and heating techniques are the same as used for the grating experiments.

Experimental Results and Discussion

The response to the simplest holographic signal, a sinusoidal charge pattern, is a periodic rippling of the thermoplastic surface. In the absence of any signal charge pattern, i.e., with a uniform surface charge on the thermoplastic, a “frost” pattern can be developed.³ This frosting is a random deformation of the thermoplastic and it has been observed that this deformation has a dominant spatial

frequency dependent primarily on the thermoplastic layer thickness (the dominant spatial wavelength is approximately twice the thickness).¹⁰ Associated with this dominant or "resonant" spatial frequency is a bandpass spatial-frequency response. A typical example of frost deformation is shown in Fig. 7. The frost was formed by uniformly charging a thermoplastic layer and then heat developing. Thermoplastic thickness was $0.4 \mu\text{m}$ and dominant wavelength was approximately $1 \mu\text{m}$.

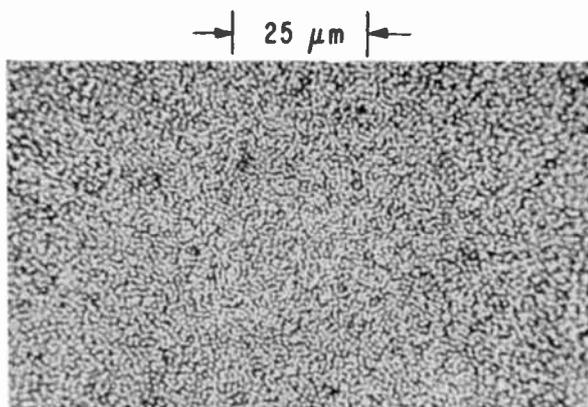


Fig. 7—Example of quasi-random frost deformation (thermoplastic thickness is $0.4 \mu\text{m}$ and dominant wavelength is approximately $1 \mu\text{m}$).

The thickness is adjusted for the particular holographic recording geometry to center the bandpass spatial-frequency response on the information spatial frequencies. Holographic "signal" information then replaces the quasi-random frost.¹³ The information bandwidth should be less than the thermoplastic bandwidth to prevent shading or vignetting in the reconstructed image, since there is, in our Fraunhofer recording setup, a one-to-one correspondence between locations in the object plane and spatial frequencies in the hologram. Using the sequential-recording technique, we have demonstrated holographic recording with very little shading at a center frequency of 1200 cycles/mm and an object bandwidth of 500 cycles/mm for a thermoplastic thickness of approximately $0.4 \mu\text{m}$.

Simultaneous Recording

We have recorded high-efficiency, low-noise, low-distortion thermoplastic holograms using one of the simultaneous recording techniques (i.e.,

heat, then charge and expose simultaneously). In this mode the electrostatic driving forces are stronger and one need not record at the resonant spatial frequency; in fact, much thinner thermoplastic layers can be used than predicted by the factor of two rule relating dominant wavelength and thickness. Fraunhofer holograms, with a center frequency of 1170 cycles/mm and a bandwidth of 500 cycles/mm, have been recorded in thermoplastics with thicknesses of 0.10-0.20 μm where the predicted resonant spatial frequency is 5,000-2,500 cycles/mm. Little or no shading was seen in the reconstructed images.

If the application permits, the simultaneous mode with thin thermoplastic layers is preferred to the sequential mode because virtually frost-free, highly efficient holograms can be recorded. The superior signal-to-frost ratio of the simultaneous mode probably can be explained in terms of the following considerations: (a) a threshold electric field for frost exists,¹⁴ whereas an applied signal does not have such a threshold response; (b) any small deformation of the surface will be enhanced by electrostatic forces, which increase as deformation proceeds. Thus, an applied signal can form a small deformation that is driven by an increasing field as corona charging proceeds. By the time the frost threshold is reached, the signal deformation is strong enough to suppress frost.

The strong electrostatic forces associated with simultaneous recording permit the achievement of both higher efficiency and higher resolution than obtainable with sequential recording. We have achieved first-order diffraction efficiencies of up to 40%, which exceeds the theoretical diffraction efficiency for sinusoidal thin phase gratings (34%). The gratings produced have a symmetrical profile or wave-shape. If the gratings are tilted from normal incidence, however, up to 60% of the transmitted light can be diffracted into one of the first orders. This "blaze effect" is a result of refracted light coinciding in direction with one of the first-order diffracted beams and is a symmetrical effect.

The high efficiency gratings were recorded in a sample of 1.2- μm photoconductor and $\approx 0.2 \mu\text{m}$ thermoplastic at a spatial frequency of 1170 cycles/mm. (30° between beams at 441.6 nm). Efficiency was measured at 441.6 nm. Scanning electron micrographs (SEM) of these high efficiency gratings show a nonsinusoidal waveshape as expected; Fig. 8 is a typical example of the grating, where the SEM photograph was taken at an angle of 60° from the normal to the sample. It can be seen that, in this case, the electrostatic force during deformation was strong enough to force all of the thermoplastic material into the ridges, leaving the underlying photoconductor

layer exposed. The rough edges suggest that a tearing apart of the thermoplastic material in the valleys has occurred. The shape of these gratings approximates that of a half-wave rectified sine wave.

We have found the limiting resolution of these films to be in excess of 4,100 cycles/mm (130° between beams at 441.6 nm). Diffraction efficiency at 4,100 cycles/mm was approximately 1%. At 1850 cycles/mm, it was greater than 10%. Perhaps these efficiencies might be increased by optimizing thermoplastic and photoconductor thicknesses.

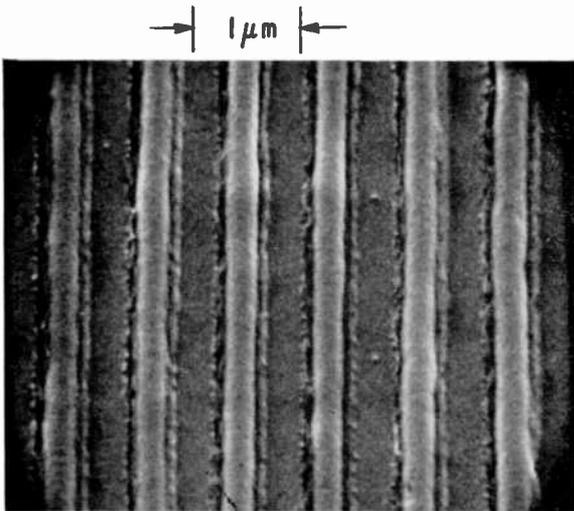


Fig. 8—Scanning electron micrograph of high efficiency, low noise, 1200 cycles/mm holographic grating recorded in simultaneous mode. Grating profile approximates a half-wave rectified sine wave.

Sequential Recording

Although the simultaneous mode with thin thermoplastic layers has excellent properties with respect to efficiency and signal-to-noise ratio, it is desirable in several applications to separate exposure from development. For example, when the use of flexible substrates is required, substrate deformation during heating seriously degrades the holographic fringe contrast. This temporary deformation causes no problem in the sequential mode of recording. Furthermore, in application to real-time holographic interferometry, the "zero fringe" condition between object and reconstructed wave fronts can be obtained only with the sequential mode of recording. Moreover, in the sequential mode the zero fringe condition is readily achieved even with flexible substrates, e.g., Cronar.

To record in the sequential mode, the thermoplastic thickness is adjusted to record at the resonant spatial frequency. For example, a thermoplastic thickness of $0.45\ \mu\text{m}$ is used for holographic recording at 1170 cycles/mm. Thinner layers used at this frequency are "off resonance"; the signal-to-noise ratio is drastically reduced, and the reconstructed images are shaded. We have recorded holograms in this mode with frequencies up to 1850 cycles/mm, with corresponding thermoplastic thicknesses, down to approximately $0.25\ \mu\text{m}$. The effi-

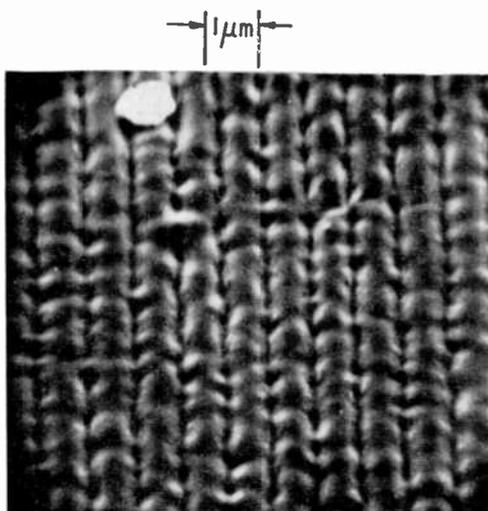


Fig. 9—Scanning electron micrograph of 1170 cycles/mm holographic grating recorded in sequential mode. Frost noise is seen as a quasi-random modulation of the fringes.

ciency is low, 2-5%, at these high frequencies. This is not surprising, since charge contrast is limited by breakdown in the thin layers, and surface-tension restoring forces at a given corrugation depth are stronger for high spatial frequencies. Fast heat pulses (2-3 msec) were necessary to develop these holograms.

Sequential recording at 1170 cycles/mm has been used to record holograms with diffraction efficiencies of up to 22% (exposure $\sim 160\ \mu\text{J}/\text{cm}^2$ at 441.6 nm). Heat development is not critical and heat pulses, typically 100 msec in duration, could be applied several minutes after exposure.

An example of a holographic grating made in a sample having $2.5\ \mu\text{m}$ photoconductor and $0.45\ \mu\text{m}$ thermoplastic is shown in Fig. 9. This SEM photograph, taken at an angle of 45° from the normal, is a

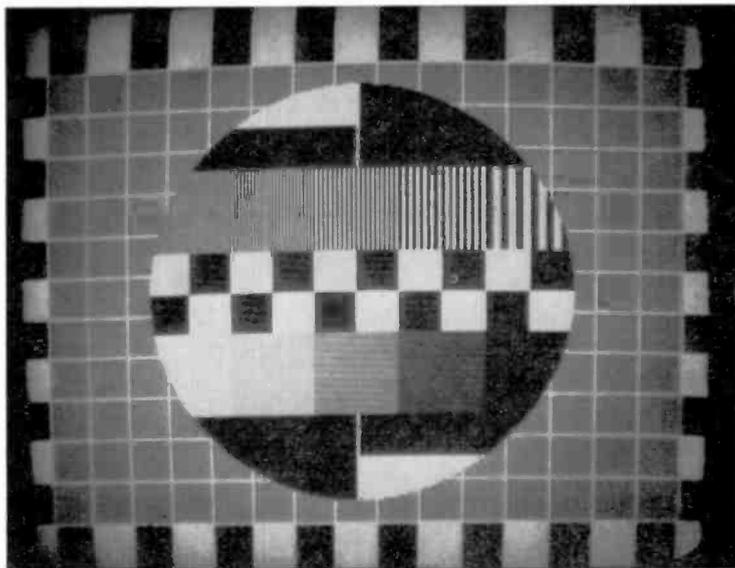
"noisy" section of the grating; the frost noise is seen as a quasi-random modulation of the ripples. Diffraction efficiency was approximately 10% at 441.6 nm.

Fraunhofer Holograms

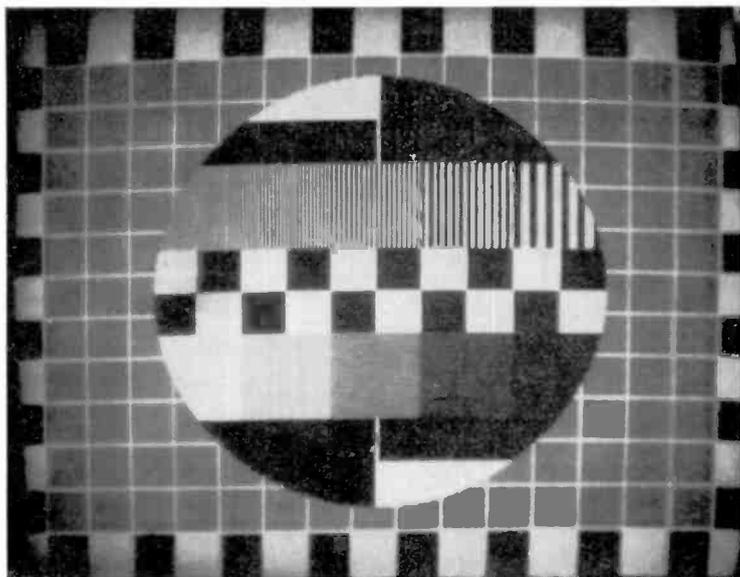
Using the experimental apparatus shown in Fig. 6, Fraunhofer holograms have been recorded in thermoplastics on both glass and Cronar substrates. This experimental arrangement allowed monitoring of the object transparency and viewing of the reconstructed image, using the reference beam for readout. Recording angle was 30° and the 441.6 nm He-Cd laser line was used. Information bandwidth was approximately 500 cycles/mm centered at 1170 cycles/mm.

Fraunhofer holograms have been recorded using both the sequential and simultaneous recording techniques with diffraction efficiencies up to 35% at 441.6 nm for the simultaneous method. An example of a simultaneous mode thermoplastic hologram is shown in Fig. 10; the object transparency, a standard test pattern, and the reconstructed image were displayed on a television monitor and show the full 5 MHz resolution. Thermoplastic layer thickness was approximately $0.2 \mu\text{m}$; diffraction efficiency was 21% at 441.6 nm. The main source of noise in these holograms was the random frost deformation. Signal-to-noise ratio for a standard test pattern was measured to be 27 dB over the 0-5 MHz frequency range (peak-to-peak signal to rms noise). There is also a slight shading from right to left which corresponds to a roll off in the spatial frequency response at low frequencies. With sequential recording, the thermoplastic thickness was adjusted so that shading in the hologram image was minimized. By varying the thermoplastic thickness $\pm 10\%$ from a nominal $0.45 \mu\text{m}$, the shading could be shifted from the high spatial frequency side of the image to the low frequency side. With good control on thermoplastic thickness, the shading is practically eliminated; this is an indication that the spatial frequency bandwidth is at least 500 cycles/mm for sequential recording. Fig. 11 is an illustration of the information storage application of thermoplastics; a redundant, Fraunhofer hologram, 3×3 mm, of a standard $8\frac{1}{2} \times 11$ -inch page of text is shown.

Fraunhofer holograms on 16-mm 0.004-inch Cronar were made using sequential recording with rf heating. Holograms with little frost have been recorded on this flexible substrate with diffraction efficiencies of 17% at 441.6 nm.



(a)



(b)

Fig. 10—(a) Coherently illuminated object transparency viewed directly with TV camera and (b) reconstructed image. Diffraction efficiency is 21% at 441.6 nm.

Suppression of Intermodulation Distortion

The achievement of highly efficient, good quality, Fraunhofer holograms demonstrates an important property of thermoplastic holograms—the intermodulation distortion, intrinsic to most thin phase

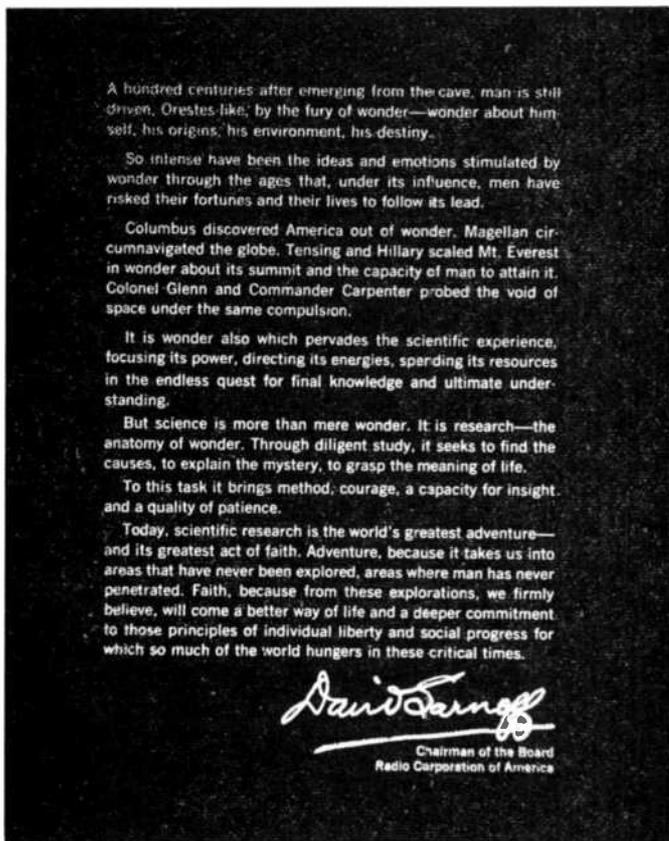


Fig. 11—Reconstructed image of standard $8\frac{1}{2} \times 11$ page of text; hologram size is 3×3 mm.

holograms,¹⁵ is greatly reduced due to the bandpass spatial-frequency response. The intermodulation distortion, which is manifested as ghost images in the reconstructed image, is due to beats between the object spectrum and low-frequency autocorrelation terms; since low frequency response is greatly attenuated with thermoplastics, intermodulation is reduced. To demonstrate this suppression, we recorded a Fraunhofer hologram of a test pattern with an efficiency of 35% at 441.6 nm (see

Fig. 12). To the best of our knowledge, this is the first time a thin, phase hologram (other than a simple grating) has been reported with such high efficiency without appreciable intermodulation distortion. These highly efficient holograms were made in the simultaneous mode with thin (approximately $0.2 \mu\text{m}$) thermoplastic layers.

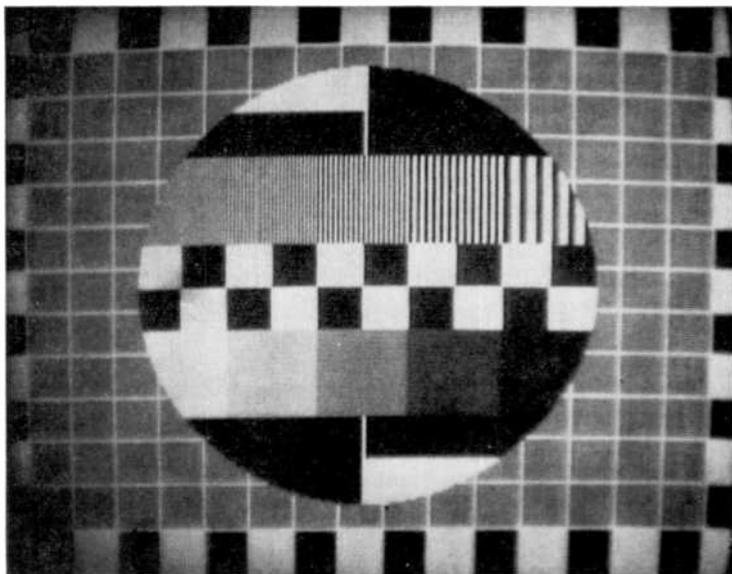


Fig. 12—Reconstructed image. Diffraction efficiency is 35% at 441.6 nm. Intermodulation distortion is barely discernible.

Immunity from Reflection Fringes and Ambient Illumination

Another interesting property of thermoplastic-photoconductor media is their relative insensitivity to illumination other than the holographic fringe pattern, even though the photoconductor is nearly panchromatic. This phenomenon results both from the photoelectronic (rather than photochemical) nature of the process and from the bandpass response of the thermoplastic.

As an example of how this works to our advantage, consider the problem of reflections from a substrate. In addition to the holographic fringes that one desires to record, reflections give rise to other sets of spurious interference fringes. Photochemical media, such as silver halide emulsions and photoresist record all such fringes. The photo-

conductor-thermoplastic structure responds strongly only to those fringes that are parallel to the electric field, i.e., normal to the surface. Fringes whose planes lie at an angle or parallel to the surface produce a nearly uniform distribution of charge as the photogenerated carriers are swept across the photoconductor by the electric field. Moreover, since the response of the thermoplastic is "tuned" to the spatial frequencies of the desired holographic fringes, the spurious, weak, "ac" terms are further attenuated because they are off resonance, i.e., they have low spatial frequencies. The only effect of the "dc" and low-frequency terms is a moderate reduction of charge contrast or signal strength of the desired information. This filtering effect permits the use of any conductive substrate, no matter how reflective. We have made good holograms on chrome-gold and silver substrates that were good mirrors.

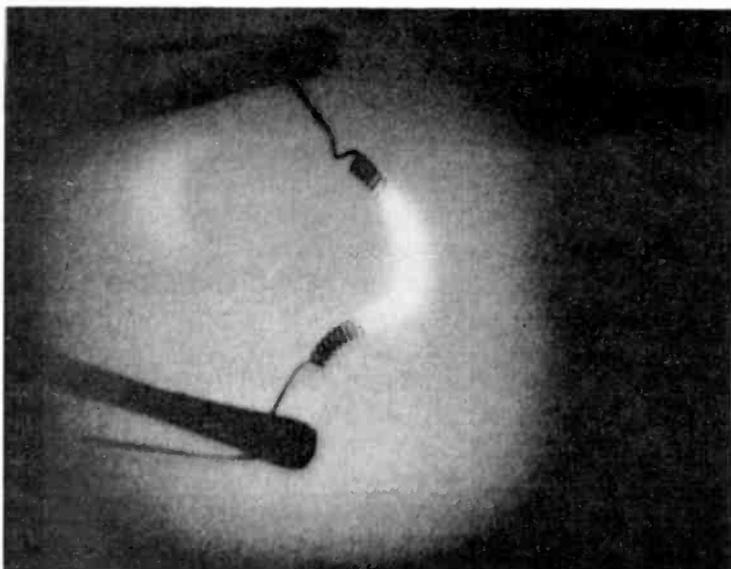
Similarly, the lack of response to uniform or low spatial-frequency illumination allows holographic recording at high ambient light levels (even though the medium is panchromatic) and permits the recording of holograms even of incandescent objects.

We have recorded good holograms in the usual manner, with coherent 441.6 nm light, but with the additional flood illumination of a 60-watt light bulb 10 cm from the recording medium during exposure! This recording in high ambient light levels is most effective when done with continuous charging during exposure. Thermal development, however, can be done either simultaneously or sequentially.

Real-Time Interferometry

The immunity from ambient light, together with *in situ* development, suggests the application of thermoplastic holography to interferometry in such fields as nondestructive testing.^{16,17} *In situ* development is well suited to "real time" holographic interferometry,¹⁸ a technique in which the object is interferometrically compared with the reconstructed image of the object. Small changes in the object are observed as fringes, which can be monitored in real time. The *in situ* development of thermoplastic holograms permits the achievement of perfect match, or "zero fringe" condition, of reconstructed and object wavefronts. The immunity from incoherent ambient light extends the class of objects that can be examined by this technique to brightly illuminated or even incandescent objects. Previously, such objects could be holographed only with Q-switched lasers and gated exposures.¹⁹

To illustrate this capability we made a thermoplastic hologram of an incandescent light bulb filament. Fig. 13(a) is the "zero fringe"



(a)



(b)

Fig. 13—Real time interferometry with *in situ* developed hologram of an incandescent object: (a) “zero fringe” interference of reconstructed image with the direct image of an incandescent light bulb filament; (b) interference pattern after the light bulb has cooled.

interference of the reconstructed image with the hot object. Fig. 13(b) shows the interference after the bulb has cooled. Real-time monitoring was done with a television camera.

Charge-Assisted Post Development

In the usual sequential mode, i.e., charge, expose, recharge, and heat develop, the resulting diffraction efficiency may be less than desired. We have discovered that a post-development step consisting of recharging and reheating (either simultaneously or sequentially) considerably enhances the diffraction efficiency. The recharging establishes a uniform potential on the deformed hologram surface; however, the charge distribution is nonuniform, with higher charge density in the "valleys". Thus there will be electrostatic forces, proportional to the square of surface-charge density, tending to deepen the corrugations and hence increase the diffraction efficiency of the hologram. We have used charge-assisted post development to enhance the efficiency of a Fraunhofer hologram from less than 0.1% to more than 8% with little loss in signal-to-noise ratio. This gain mechanism is, however, ultimately limited by an increase in frost noise.

Summary

The thermoplastic-photoconductor medium for holographic recording combines a number of important advantages:

- (1) it has high sensitivity ($5\text{-}100 \mu\text{J}/\text{cm}^2$) comparable to high-resolution silver halide emulsions (Kodak 649),
- (2) it is grainless,
- (3) resolution is greater than 4000 cycles/mm,
- (4) bandpass spatial frequency response reduces intermodulation distortion and permits the achievement of undistorted high efficiency reconstructions,
- (5) development is dry, nearly instantaneous, and can be done *in situ*,
- (6) readout is efficient and nondestructive,
- (7) the medium is erasable and reusable,
- (8) the holograms can be replicated since the information is recorded as a surface relief pattern.

Factors that currently limit more widespread use of the medium are the complexity of the recording apparatus; problems of maintaining

uniformity in thickness, charging, and heating; random frost noise; and vulnerability to erasure by elevated ambient temperatures.

In this article we have reported the achievement of several new results and techniques. A "simultaneous" technique for holographic recording has been used to produce holograms that have high resolution and high efficiency. Two beam-interference gratings have demonstrated resolutions exceeding 4,100 cycles/mm, and efficiencies up to 40%. Reconstructed image efficiencies up to 35% have been achieved with little intermodulation distortion or frost noise. We have observed a remarkable immunity from ambient illumination even though the medium is panchromatic. This property makes possible the holography of illuminated or even incandescent objects without Q -switched lasers and gated exposures. A similar effect is the immunity from spurious fringes due to reflections from the substrate. Finally a charge-assisted post development technique for enhancing the efficiency of weak holograms has been used to achieve efficiency gains of more than 100.

Acknowledgments

The authors wish to acknowledge the technical assistance of A. Matzelle. We also wish to acknowledge the support and encouragement of I. Gorog and W. J. Hannan.

References:

- ¹ E. I. Sponable, "Eidophor System for Theater Television," *J.SMPTE*, Vol. 60, p. 337 (1953).
- ² W. E. Glenn, "Thermoplastic Recording," *Appl. Phys.*, Vol. 30, p. 1870 (1959).
- ³ R. W. Gundlach and C. J. Claus, "A Cyclic Xerographic Method Based on Frost Deformation," *Photograph. Sci. Eng.*, Vol. 7, p. 14 (1963).
- ⁴ J. Gaynor and S. Aftergut, "Photoplastic Recording," *Photograph. Sci. Eng.*, Vol. 7, p. 209 (1963).
- ⁵ N. E. Wolff, "A Photoconductive Thermoplastic Recording System," *RCA Review*, Vol. 25, p. 200 (1964).
- ⁶ E. C. Giaimo, "Thermoplastic Organic Photoconductive Recording Media—Electrophotographic Characteristics and Processing Techniques," *RCA Review*, Vol. 25, p. 692 (1964).
- ⁷ J. T. Bickmore and C. J. Claus, "Charge Transfer Frost Xerography," *Photograph. Sci. Eng.*, Vol. 9, p. 283 (1965).
- ⁸ J. C. Urbach and R. W. Meier, "Thermoplastic Xerographic Holography," *Appl. Opt.*, Vol. 5, p. 666 (1966).
- ⁹ L. H. Lin and H. L. Beauchamp, "Write-Read-Erase in Situ Optical Memory Using Thermoplastic Holograms," *Appl. Opt.*, Vol. 9, p. 2088 (1970).
- ¹⁰ H. F. Budd, "Dynamical Theory of Thermoplastic Deformation," *J. Appl. Phys.*, Vol. 36, p. 1613 (1965).

- ¹¹ R. M. Schaffert, "A New High-Sensitivity Organic Photoconductor for Electrophotography," **IBM J. Res. Dev.**, Vol. 15, p. 75 (1971).
- ¹² R. A. Bartolini, W. J. Hannan, D. Karlsons, and M. J. Lurle, "Embossed Hologram Motion Pictures for Television Playback," **Appl. Opt.**, Vol. 9, p. 2283 (1970).
- ¹³ J. C. Urbach, "The Role of Screening in Thermoplastic Xerography," **Photograph. Sci. Eng.**, Vol. 10, p. 287 (1966).
- ¹⁴ P. J. Cressman, "New Type of Thermoplastic Deformation," **J. Appl. Phys.**, Vol. 34, p. 2327 (1963).
- ¹⁵ J. C. Urbach and R. W. Meier, "Properties and Limitations of Hologram Recording Materials," **Appl. Opt.**, Vol. 8, p. 2269 (1969).
- ¹⁶ J. C. Bellamy, O. B. Ostrowsky, M. Poindron, and E. Spitz, "In Situ Double Exposure Interferometry Using Photoconductive Thermoplastic Film," **Appl. Opt.**, Vol. 10, p. 1458 (1971).
- ¹⁷ J. M. Holeman, "Use of Photoplastic Film in Holographic Interferometry," paper presented at S.P.S.E. Seminar on Novel Audio-visual Imaging Systems, New York, Sept. 24, 1971.
- ¹⁸ R. J. Collier, E. T. Doherty, and K. S. Pennington, "Application of Moire Techniques to Holography," **Appl. Phys. Lett.**, Vol. 7, p. 223 (1965).
- ¹⁹ L. O. Heflinger, R. F. Wuerker, and R. E. Brooks, "Holographic Interferometry," **J. Appl. Phys.**, Vol. 37, p. 642 (1966).

Recyclable Holographic Storage Media

Joseph Bordogna,* Scott A. Keneman, and Juan J. Amodei

RCA Laboratories, Princeton, N. J.

Abstract—Performance parameters of state-of-the-art recyclable holographic recording media are compared to develop tradeoffs for their use in holographic storage and imaging applications. Included are record energy, record time, erase time, diffraction efficiency, linearity, resolution, cycle lifetime, natural decay time, and a figure of merit. Materials and devices discussed include: magneto-optic, electro-optic, and photochromic materials; metal films; ferroelectric-photoconductor, thermoplastic deformation, elastomer, liquid-crystal-photoconductor, and photoconductor-Pockel's effect devices.

Introduction

The advent of the laser as a reliable commercial product has made possible the use of optical holography for information storage, processing, display, and image conversion systems.¹⁻³ This paper discusses briefly and compares the materials⁴ and devices that researchers hope will store the holographic data and perform the real-time holographic imaging in such systems. In particular, an overview is given of the physical basis, fabrication, and operating parameters of available recyclable holographic storage media.

* Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia.

Recyclable media are characterized by their ability to be erased and reused for sequentially storing different information patterns. Referring to such media as "holographic" at least implies that they are capable of supporting micrometer resolution. "Storage" implies an ability of the medium to keep the information intact after the hologram is recorded. In view of the embryonic state of development of light-sensitive media for information applications, this paper is not limited in scope to those media that satisfy all the above criteria; materials and devices that satisfy some, but not all, of the requirements of ideal recyclable holographic storage media are also discussed. Some possess insufficient resolution; others are subject to decay. Some make no attempt at permanence but are useful for holographic imaging. Others fatigue after numerous operating cycles. We include materials subject to these deficiencies, because it is likely that as technology advances and understanding of the various media increases, some of these deficiencies may be removed and the realization of higher-performance recyclable holographic storage media achieved.

The media discussed may be characterized as *materials* or *devices* depending on their fabrication and use. The former possess a single constituent that is directly light sensitive. The latter are composite structures containing a light-sensitive element, but requiring more than light alone to operate properly. The *materials* discussed include magneto-optic films, electro-optic crystals, photochromic materials, and metal films. The *devices* include ferroelectric-photoconductor devices, thermoplastic-deformation devices, elastomer devices,* liquid-crystal-photoconductor devices, and photoconductor-Pockel's-effect devices.

The media are compared on a number of operating parameters: record or write energy (quantity of energy per unit area required for recording—frequently referred to as "sensitivity"); record time (time required to record a hologram); erase time (time to clear storage medium for next hologram); diffraction efficiency (percentage of light energy in reconstructed image compare to incident energy of readout light); linearity (measure of attainable contrast or "gray scale"); resolution capability (number of resolvable lines per linear distance); cycle lifetime (number of times medium can be recorded and erased without deterioration of performance); and natural decay time (length of time hologram can be stored with no sustaining power). A comparison summary is presented in Table 1. Unfortunately, this table does not tell the complete story. Factors such as difficulty of fabrication, reproducibility of parameters, and projected eventual cost are missing. Further research on these and other factors is in progress.

* The elastomer devices, although not light sensitive, are included for generality, since they may be useful for holographic imaging.

Magneto-Optic Recording Materials

Curie-point writing on manganese-bismuth films⁵⁻¹⁰ can be used for direct storage of digital information or for holographic storage of digital or pictorial information. Writing is achieved by switching the magnetization of a semi-transparent film (~ 300 to 700 nm thick) in response to the heating caused by local light absorption. Readout can be accomplished through the magneto-optic Faraday or Kerr effect, which gives rise to rotation of the polarization of the light transmitted or reflected. Because the writing mechanism relies on selectively heating the film above the Curie temperature in the regions of higher light intensity, while leaving the regions of low intensity below the Curie temperature, this medium requires that writing be accomplished in times that are short compared to the thermal time constant of the film; this time is normally of the order of 100 nanoseconds. As a result high-power pulsed lasers or focused cw lasers are required in order for high-resolution writing to supply the energy necessary to reach the Curie temperature in a sufficiently short time. Erasure of the pattern is accomplished by uniformly remagnetizing the film by means of an external magnetic field. The film-deposition technique and other details of the operation and recording process are discussed at length by Mezrich and Cohen elsewhere in this issue.*

The most common magneto-optic recording material is manganese bismuth (MnBi). Its pertinent performance parameters are listed in Table 1. Other important parameters include: Faraday rotation $> 5 \times 10^5$ degrees/cm, Kerr rotation $\approx 3^\circ$, absorption coefficient $= 3.5 \times 10^5$ cm^{-1} ; additional physical parameters are given by Mezrich and Cohen.* The disadvantages of MnBi are the need to write in a very short time, the low diffraction efficiency of MnBi Holograms, and the fact that performance deteriorates under humidity conditions greater than 50% ; however, this effect is not yet completely understood.¹¹

Other materials being investigated include gadolinium-iron-garnet (GdIG)^{12,*} and europium oxide (EuO).^{13,*} EuO has a major disadvantage in that it operates only at liquid helium temperatures. All things considered, MnBi appears to be the best candidate for magneto-optic holographic storage at the present state of the art.

Electro-Optic Recording Materials

High-efficiency volume phase holograms can be recorded in certain transparent electro-optic crystals. Examples are lithium niobate (LiNbO_3), lithium niobate co-doped with iron ($\text{LiNbO}_3:\text{Fe}$), lithium

* R. S. Mezrich and R. Cohen, "Material for Magneto-Optic Memories," *RCA Review*, this issue, p. 54.

Table 1—Performance Parameters for Recyclable Holographic Storage Media.

Storage Medium	Material (or principal material in device)	Reference	Recording and Erasure				
			Recording Process	Write Energy ($\text{mJ} \cdot \text{cm}^{-2}$)	Record Time	Erase Process	Erase Time
Magneto-Optic Materials	MnBi	5-13	Heating above Curie point	10	10 ps \rightarrow 50 ns	Same as recording	$\sim \mu\text{s}$
Electro-Optic Materials	LiNbO ₃	14-19	Optically induced charge separation	$10^3 \rightarrow 5 \cdot 10^5$	10 s \rightarrow min	Optically or thermally induced charge redistribution	Same as record
Photochromic Materials	Many types (see text)	20-29	Optically induced charge transfer	50	\sim ns	Same as recording	Same as record
Metal Films	Bi	30	Evaporation by heating	50	5 \rightarrow 20 ns	Redistribution by contained evaporation	\sim 1s
Ferroelectric Photoconductor Devices	Bi ₄ Ti ₃ O ₁₂	38-43	Light-modulated ferroelectric switching	1	\sim ms	Same as recording	Same as record
	PLZT	34, 35		10	\sim s		Same or less than record
Thermoplastic Deformation Devices	Staybelite	47-55	Electric field produced deformation of heated material	10^{-3} (theoretical)	1 ms	Heat	\sim 1s
Elastomer Devices	Siloxane (γ -ruticon)	56	Electric field produced deformation	10^{-1}	\sim ms	Removal of electric field	\sim 10
Liquid Crystal Photoconductor Devices	Mixed liquid crystal; ZnS, CdS photoconductor	59, 60	Light-modulated dynamic scattering	$\sim 5 \cdot 10^{-3}$	$<$ 0.1 ms	Application of ac electric field	20-500
Photoconductor-Pockel's Effect Devices	Bi ₁₂ SiO ₂₀	61-63	Light-modulated electro-optic effect	$\sim 10^{-4}$	$\sim \mu\text{s}$	Light at appropriate wavelength	\sim μs

Type of hologram	Readout			Cycle Lifetime		Decay Time (Dark Storage)	Figure of Merit
	Readout Efficiency (%)	Linearity	Resolution ($\mu\text{p} \cdot \text{mm}^{-1}$)	Write—Erase	Read		
in phase	10^{-2} (Faraday) 10^{-1} (Kerr)	Fair	~ 1000	Indefinite	∞	Indefinite—limited by humidity	100
in phase	80	Good	~ 1500	Indefinite	Reading at same λ as writing destroys information	Weeks to months	24
amplitude	1.2 \rightarrow 3.7	Very good	$\sim 10,000$	Indefinite	Bleaching occurs	Minutes to months	740
amplitude	6	Good	~ 1000	~ 10 cycles	∞	Indefinite	120
in phase	10^{-2}	Fair	~ 800	$> 10^5$ cycles	Indefinite	Indefinite	8
	Unknown	Good	~ 50	Limited by fatigue (see text)	Indefinite	Indefinite	Readout efficiency unknown
in phase	15	Good	~ 1000	100 cycles	∞	Indefinite	$7 \cdot 10^6$ (theoretical)
in phase	15	Good	~ 1000	$> 10^4$ cycles	Depends on device structure	Not a permanent storage device	$1.5 \cdot 10^5$
in phase	Unknown	Good	~ 10	Unknown	Unknown	Unknown	Readout efficiency unknown
in phase	Unknown	Good	~ 1000	Unknown	Unknown	\sim hours	Readout efficiency unknown

tantalate (LiTaO_3), strontium-barium niobate [$(\text{Sr,Ba})\text{Nb}_2\text{O}_6$, SBN, or $\text{Ba}_{0.75}\text{Sr}_{0.25}\text{Nb}_2\text{O}_6$], and doped barium-sodium niobate ($\text{Ba}_2\text{NaNb}_5\text{O}_{15}$).¹⁴⁻¹⁸ The recording process relies on the optical excitation and subsequent drift or diffusion of electrons that originate from localized centers in the crystal. This creates fields that modulate the index of refraction of the material; the result is a phase hologram whose efficiency could theoretically be as high as 100%. Proper recording procedure requires that the crystal optic axis be perpendicular to the fringe gratings of the hologram. High-efficiency readout can be accomplished only with light polarized in the plane containing the optic axis.

Holographic diffraction efficiencies as high as 80% have been achieved in $\text{LiNbO}_3:\text{Fe}$ crystals about 0.2 cm thick. The sensitivity of this material (which is relatively low compared to other recording materials) varies between 1 and $500 \text{ J} \cdot \text{cm}^{-2}$ (at $\lambda = 488 \text{ nm}$) depending on the doping or other treatment to which the material was subjected. Storage times of many days are possible at ordinary room illumination levels, and erasure can be achieved either optically or by heating the crystal to 300°C . The holograms stored in the material can be "fixed" utilizing a simple thermal process that causes the patterns to be optically nonerasable.¹⁹ Application of an electric field has been shown to improve the performance of $(\text{Ba,Sr})\text{Nb}_2\text{O}_6$ as an electro-optic recording material,¹⁵ but the published diffraction efficiencies (2%) are considerably lower than in LiNbO_3 and LiTaO_3 . Details on LiNbO_3 and doped LiNbO_3 are given elsewhere in this issue.*

In addition to the materials discussed above, several other electro-optic hosts have been tried but with only marginal success. These include barium titanate (BaTiO_3), bismuth silicon oxide ($\text{Bi}_{12}\text{SiO}_{20}$), and bismuth germanate ($\text{Bi}_4\text{Ge}_3\text{O}_{12}$).^{17,18}

Photochromic Materials

Color centers in transparent crystals are caused by the presence of impurities and imperfections that give rise to localized states that trap electrons or holes within the forbidden energy gap of the material.²¹ Such states may show absorption at visible wavelengths where the pure crystal is transparent. In particular, the photochromic materials

* J. J. Amodi and D. L. Staebler, "Holographic Recording in Lithium Niobate: Model, Theory and Experiments," *RCA Review* this issue, p. 71, and W. Phillips, J. J. Amodi and D. L. Staebler, "Transition Metal Doping of Lithium Niobate," *RCA Review*, this issue, p. 94.

have the ability to switch colors under the influence of optical irradiation. This change is normally caused by the transfer of an optically excited electron from one type of color center to another, with the absorption properties of both centers being changed accordingly. Light of one wavelength produces a given change in coloration, and the crystal returns to its original state when exposed to light of another wavelength.

The performance of inorganic photochromics can be compared to other media by referring to Table 1. They are characterized by the maximum absorption change induced by light, the energy required to produce a given change, the background absorption, and the storage time of the material. The inherent resolution of the materials is practically unlimited, as illustrated by the fact that reflection holograms with fringe spacing of less than 100 nm have been successfully recorded by Amodei in $\text{SrTiO}_3:\text{Fe,Mo}$.

The sensitivity of inorganic photochromic materials varies considerably from hundreds of $\text{J} \cdot \text{cm}^{-2}$ to $\text{mJ} \cdot \text{cm}^{-2}$, depending on the quantum efficiency of the process. The maximum storage time of the crystals, which is usually determined by the thermal activation energy of the centers, varies from minutes to months.

Examples of *inorganic* photochromic materials include silver halide photochromic glasses^{21,23} and inorganic photochromic crystals (e.g., $\text{SrTiO}_3:\text{Fe,Mo}$; $\text{BaF}_2:\text{La}$; Sodalite:Cl; electron beam colored NaCl; and fluorescein-boric acid glass organophosphor).^{24,26}

Successful cyclic recording and erasure of holographic information on thin *organic* photochromic films has been reported by Mikaeliane and his associates in the Soviet Union²⁷ and M. Lescinsky and M. Miller in Czechoslovakia.²⁸ Performance data for these materials (essentially photochromic spiropyran dissolved in styrene polymer) are rather sketchy at this time, however. In the United States, Ross²⁹ has investigated the use of thioindigo dye. The efficiency of this material is temperature dependent, but its optical densities are higher than those of the inorganic photochromic materials.

In summary, a clearly distinct advantage of photochromic materials is that there is no inherent resolution limit, since absorption takes place on an atomic or molecular scale. Disadvantages include: sensitivity limited to narrow spectral regions; relatively thick samples (1-10 mm) required to obtain useful optical densities; image deterioration due to both thermal decay and optical bleaching during readout; and the fact that two wavelengths must be used for writing and erasing.

Metal Films

Concentrated laser energy may be used to raise the temperature of a material to its vaporization point. This principle has been used to store evaporation holograms in thin metal films.³⁰ Films may initially be deposited on substrates using conventional technology. Then, during hologram recording, locations of constructive interference rise to higher temperatures than do points of destructive interference.

Holographic recording by vaporization relies on the fact that the energy absorbed by a thin film (7.5-20 nm) during a very short laser pulse does not have time to diffuse away. Exposure to such light pulses, therefore, causes the material to vaporize from each point of the surface in amounts that are nearly proportional to the integrated intensity of the light absorbed at that point. When a holographic intensity pattern is applied to a film in this manner, it leaves a film-thickness profile that closely resembles the spatial intensity variations of the recording light. Such a pattern is considered an amplitude hologram because, while the absorption is substantial, the film is not sufficiently thick to significantly affect the phase of the light traversing it. The hologram thus stored can be read out by reflection or transmission with comparable efficiencies. In tests with 7.5-20 nm thick bismuth (Bi) films vacuum deposited on glass substrates,³⁰ both transmission and reflection holograms had efficiencies of 6%, which is very close to the theoretical maximum for amplitude holograms.

The sensitivity of this direct holographic recording material depends to some extent on film thickness; for films approximately 10 nm thick, the energy required to record was experimentally³⁰ found to be less than 50 mJ/cm², which compares favorably with other materials such as photochromics and magneto-optic thin films (see Table 1).

Another feature of the thin metal films is that they may be operated in a linear region, because the energy required to reach boiling temperature is much lower than the vaporization energy. Also, since the recording technique is heat- rather than light-dependent, lasers of any wavelength can be used for recording. A chief disadvantage, of course, is the difficulty in recycling. It is possible to build a contained cell where the material is evaporated between two surfaces in very close proximity and is thus redistributed during the recording cycle. Such a device could then, in principle, be erased by exposure to a spatially uniform light pulse that would redistribute the film evenly over the surface. Experiments by Amodei have shown that this approach is feasible, at least for a limited number of cycles (5 to 10),

and it is conceivable that with proper design, cycle lifetime could be extended considerably.

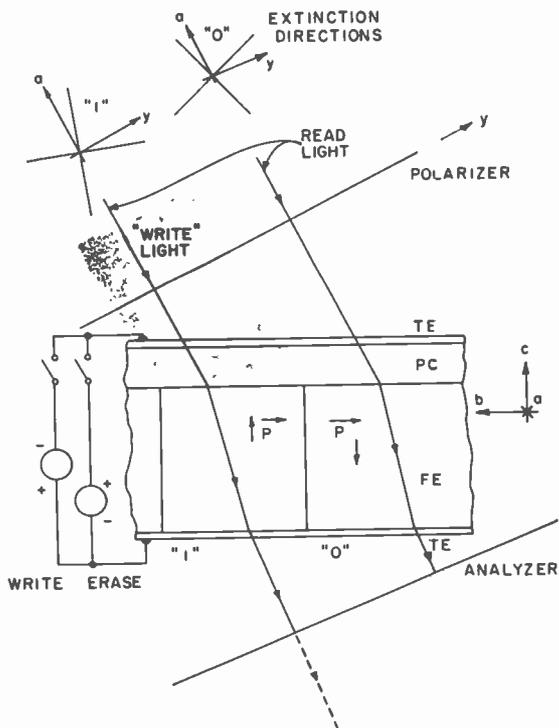
Ferroelectric-Photoconductor Devices

The use of a ferroelectric-photoconductor sandwich device for information storage is not a particularly new idea.³¹ However, it is only recently that the suggestion was made to operate such a device with optical readout.³² Early devices incorporated an optically controlled record operation, but readout was electrical.

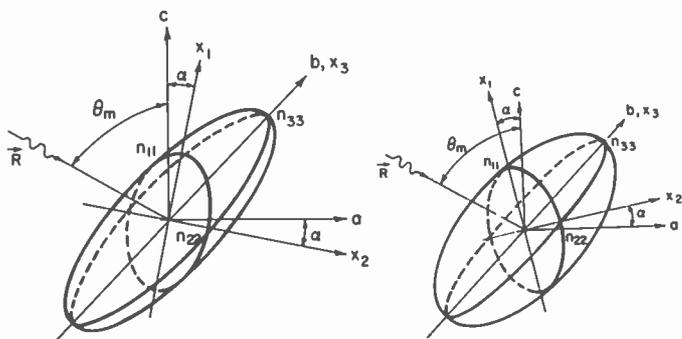
In many ferroelectric materials, the two stable remanent polarization (P_R) states may not be distinguished optically. Recently, two material configurations not possessing this disadvantage have been suggested: (1) use of ferroelectric bismuth titanate, $\text{Bi}_4\text{Ti}_3\text{O}_{12}$, in which non-180° switching of P_R may be performed,³³ and (2) "strain-biasing" of ceramic PLZT (lanthanum-doped lead zirconate-lead titanate) ferroelectrics to allow effectively non-180° switching of P_R .^{34,35} With these ferroelectric materials, ferroelectric-photoconductor (FE-PC) devices capable of optical (with applied electric field) record and optical readout are possible.

The FE-PC device geometry is shown schematically in Fig. 1(a). The ferroelectric (FE) slab (either single crystal, ceramic, or thin film) is covered with a photoconductive (PC) layer, and the combination is sandwiched between transparent electrodes. External batteries or pulse generators are used to apply switching fields to the device. The operational characteristics of the FE-PC device are determined by the specific materials used. Two typical state-of-the-art materials are $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ and strain-biased PLZT.

Bismuth titanate, a member of crystal class m ,³³ is prepared in single crystal platelets by flux growth.³⁶ It possesses four stable remanent polarization states (having both c -axis and a -axis components) in a detwinned crystal. If P_{Ra} is first poled³⁷ (in the plane of the crystal) uniformly, then switching P_{Rc} between its two states leads to the non-180° switching described earlier. As shown in Fig. 1(b), the orientations of the optical indicatrix for the two states differ by a rotation of 2α (roughly 50°) around the crystalline $b(x_3)$ -axis. Since $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ is biaxial, $n_{11} \neq n_{22} \neq n_{33}$. Two techniques for observing the change in optical indicatrix orientation (or equivalently, reading out a stored pattern in an FE-PC device) are possible: (1) tilting the crystal about the a -axis to achieve a difference in extinction directions for the two states^{32,38-40} (illustrated in Fig. 1a) and (2) tilting the crystal about the b -axis to achieve a phase difference for light polarized



(a)



(b)

Fig. 1—FE-PC (a) $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ device geometry showing WRITE and READ processes and (b) orientation of optical indicatrix for $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ for the two states of c -axis remanent polarization P_{re} (with fixed P_{na}).

perpendicular to b .^{39,41} The reader may easily convince himself that light normal to the ab plane (i.e., parallel to the c -axis) will not be adequate to distinguish the two states. Both techniques have been used to reconstruct holograms stored in FE-PC devices.^{38,41}

PLZT (lanthanum-doped lead zirconate-lead titanate ceramic material) platelets,³⁴ coated with a photoconductor layer and transparent electrodes, are held in tension or compression, causing ferroelectric domains to line up parallel and anti-parallel to the strain axis (in the platelet) and resulting in zero net remanent polarization. Switching the device with fields normal to the platelet produces a normal component of remanent polarization. The platelet birefringence is related to this remanent polarization, allowing readout of a stored pattern.

Both bismuth titanate and PLZT FE-PC devices are included among the comparisons in Table 1. To date, holographic storage has been reported only in the former. Both devices possess their own unique advantages and disadvantages. Both have an inherent gain mechanism, similar to that in photographic film. Incident light for recording does not directly cause the storage; rather, photoelectrons in the photoconductor are "created" by the light, causing an increase in conductivity and allowing the battery to switch the ferroelectric.

The bismuth titanate device involves difficult fabrication and processing techniques because of its single-crystal nature. Large-area devices are difficult to achieve. Furthermore, an optical degradation has been observed after numerous switching cycles with metal electrodes directly on the $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ surfaces.⁴² However, such degradation has not yet been observed in FE-PC devices with photoconductive layers.⁴³

The PLZT device is simpler to fabricate because of its ceramic form. However, ferroelectric ceramics are notorious for their electrical and optical fatigue,⁴⁴ which would restrict their useful lifetime.

Perhaps the ultimate FE-PC for holographic applications would utilize thin films of $\text{Bi}_4\text{Ti}_3\text{O}_{12}$. Such films have recently been produced,^{45,46} and they might represent the optimum compromise between the difficulties of single-crystal technologies and long-term ceramic fatigue.

Thermoplastic Deformation Devices

The technique of thermoplastic recording was described by Glenn in 1959.^{47,48} Prior to its application to holography, this recording technique used an electron beam to scan a thermoplastic film (e.g., polystyrene, chlorinated polyphenyl, or alphas-methylstyrene) and record

video signals in the form of a pictorial image; the image was visible on the film immediately following recording. The technique offered very high resolution and recording bandwidth with the option of erasure if desired.

The making of phase holograms on thermoplastic film was demonstrated by Urbach and Meier in 1966.⁴⁹ As most recently described,⁵⁰ a phase hologram is recorded in a transparent thermoplastic film as a spatial variation of film thickness corresponding to the light-intensity variation of a holographic fringe pattern. Since the thermoplastic is usually a light-insensitive resin, it is combined with a photoconductor in a film structure that can respond to light. Although manufacturing a film structure with the photoconductor dispersed through the thermoplastic is possible, the usual structure consists of a layer of thermoplastic coated over a layer of photoconductor, because it yields higher photosensitivity.⁵⁰ In practice, the complete structure consists of a glass substrate covered by a grounded transparent conductive film, on top of which lie the photoconductor and thermoplastic films. The sequence of a complete hologram record-erase cycle is described elsewhere in this issue by Credelle and Spong,* whose extensive experiments with the material staybelite have resulted in improved readout efficiency and resolution.

It is important to note that, unlike most other hologram recording media, thermoplastic material responds only to a band of spatial frequencies.^{50,51} Thus, there is an inherent tradeoff between efficiency and resolution that depends on the angle between reference and object beams in system design. Also, thermoplastic devices have a relatively poor cycle lifetime, although significant improvement in this parameter has been reported recently.⁵²

In 1968, Chang⁵³ investigated the physical parameters required of thermoplastic film for a system in which an electron beam was used for both recording and readout. More recently, Doyle and Glenn^{54,55} have described a real time optical processing system that records with an electron beam and reads out with visible light. Their device, named "lumatron", makes use of a fixed, reusable transparent thermoplastic surface mounted in a glass envelope. Surface modulation of the thermoplastic is accomplished by a high-resolution ($\sim 10\mu\text{m}$) electrostatically focused (at 6.5 kV) and deflected electron beam. Readout is performed with a single-lens schlieren optical system external to the vacuum but with a light path coaxial with the recording electron beam and normal to the thermoplastic coated faceplate, as shown in Fig. 2.

* T. L. Credelle and F. W. Spong, "Thermoplastic Media for Holographic Recording," *RCA Review* this issue, p. 206.

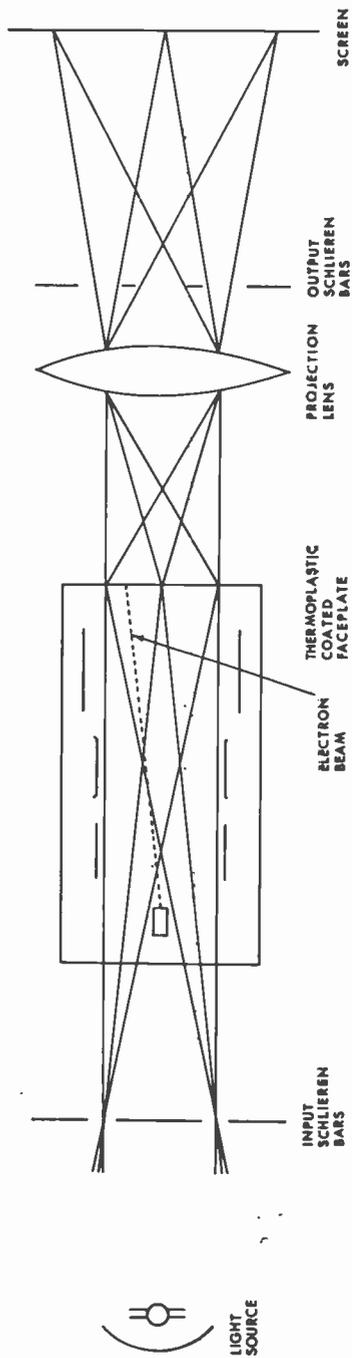


Fig. 2—Lumatron and Schlieren projector (after Doyle and Glenn⁸⁹).

The thermoplastic is coated onto a transparent conductor on the faceplate which serves both as the final faceplate reference potential and as a heater. Erasure is accomplished in less than a second by supplying a pulse of heater current to the transparent conductor to thermally remove all deformations stored in the thermoplastic surface. In real-time operation, the thermoplastic is kept at a temperature between the deformation and erase temperatures, and recording is continuous with the scanning electron beam. In this mode of operation, the mechanical time constant of the thermoplastic determines the rate at which the recorded pattern can be changed. According to Doyle and Glenn,⁵⁴ cycling can be accomplished at 1/30 second and gray scale is greater than 8 shades. Thus, the lumatron certainly has potential for use in real time incoherent-to-coherent holographic image conversion systems that need such a device for their successful implementation.³

Elastomer Devices

If the thermoplastic material in thermoplastic-deformation devices discussed in the preceding section is replaced by an elastomer, another useful device for holographic imaging (but not permanent storage) is created. Such a device, the "ruticon",* has been investigated by Sheridan.⁵⁶

The ruticon device has a layered structure consisting of a conductive transparent substrate, a thin photoconductor, a thin deformable elastomer layer, and a deformable electrode. The deformable electrode can take three forms: (1) a conductive liquid, (2) a conductive gas, or (3) a thin optically opaque flexible metal; the separate devices resulting from these three different deformable electrodes are identified, respectively, as α ruticon, β ruticon, and γ ruticon. The most interesting of these is the γ ruticon. In operation, a dc electric field (< 600 V) between the thin flexible metallic electrode and the transparent electrode establishes the initial uniform charge. Holographic information impinging on the device through the transparent substrate causes changes in the electric fields across the photoconductor and, hence, across the elastomer. These fields create mechanical forces that cause the elastomer, and consequently the thin metal layer, to deform and form a phase hologram.

In readout, light is reflected with high efficiency from the metal

* From the Greek "ruti" for wrinkle and "icon" for image.

layer to reconstruct the original object wavefront. If the metal layer is non-transparent, readout light will not affect the conductivity modulation of the photoconductor layer. In some forms, the permanence of ruticon storage may be enhanced by utilizing an image locking mode. Erasure is accomplished by removing the dc field. This is a distinct advantage over the lumatron, where heating is required for erasure.

Liquid-Crystal-Photoconductor Devices

The ability to store optical images in liquid crystals⁵⁷ has prompted their use in both electrically- and photoconductively-controllable real-time imaging devices. Kiemle and Wolff⁵⁸ have reported recyclable holographic storage in an electrically-controllable device consisting of mixed (90% nematic, 10% cholesteric) liquid-crystal sandwich cells having matrix-shaped electrodes. Holographic recording was achieved as a result of dynamic scattering induced by application of a dc or low-frequency ac electric field across the electrodes. While initially the nematic ordering is intact and the nematic-cholesteric mixture is clear, the dynamic scattering caused by the applied field emulsifies the cholesteric material. Because this colloidal-like suspension can influence the coherence of an incident laser-light wavefront, holographic recording is feasible. Recent experiments⁵⁸ indicate that this optical scattering state can persist for weeks without a sustaining field or can be erased by an ac field in 20-500 ms. A typical restoration field strength is $5 \text{ V} \cdot \mu\text{m}^{-1}$ (rms) at 3 kHz.

The matrix-shaped electrode structure required for the electrically-controllable device is complicated to construct and, worse, limits resolution capability. The implementation of the photoconductively-controlled liquid-crystal device overcomes, in principle, both of these problems. Such a device is achieved simply by replacing the electrode matrix structure by a photoconductive layer. Of the photoconductive materials tested, zinc sulfide⁵⁹ (ultraviolet recording) and cadmium sulfide⁶⁰ (visible light recording) have given the best results thus far. Holographic recording has not yet been reported, however, because resolution of the device is on the order of only a few line-pairs per millimeter. In structure, the device consists of a layer of liquid crystal and a layer of photoconductor sandwiched between two electrode layers of conductive material (e.g., tin oxide). The photoconductive layer is designed to have a "nonilluminated" resistance that is higher than that of the liquid-crystal layer but an "illuminated" resistance that is lower. In the nonilluminated condition, an applied dc voltage across

the electrodes of the device lies almost entirely across the photoconductive layer. Thus, the low voltage across the liquid-crystal layer is not sufficient to produce scattering effects, and the liquid crystal remains transparent. However, when a wavefront illuminates the photoconductor, its resistance is lowered, causing more voltage to appear across the liquid crystal. This higher voltage produces sufficient current to produce dynamic scattering and allow image storage.

The liquid-crystal-photoconductor devices permit image recording and readout at a rate controllable by the magnitude of the applied voltage and composition of the materials used. The sensitivity appears to be better than that of other recyclable materials, being several orders of magnitude greater than that of photochromic films.

Photoconductor-Pockel's Effect Devices

This device consists of a layer of photoconductive, electro-optic, single-crystal material covered by insulating dielectric layers on one or both faces and sandwiched between transparent electrodes (or one electrode and a simple ohmic contact). Experiments using highly resistive photoconductive ZnS and $\text{Bi}_{12}\text{SiO}_{20}$ as the crystal materials have been reported in the literature.⁶¹⁻⁶³ In operation, a dc voltage applied across the material is modulated by an incident optical wavefront in accordance with the photoconductive properties of the crystal, thereby storing the information in the wavefront structure. Since the material is simultaneously electro-optic, readout of the stored electrostatic pattern can be accomplished by the Pockel's effect, i.e., the pattern is reconverted to an optical wavefront by local phase retardations resulting from the transmission of uniformly polarized light through the material. The combination of both photoconductive and electro-optic functions in a single material has an advantage over a device using separate laminated photoconductive and electro-optic films because of the fabrication problems inherent in interfacing a multilayered structure of this kind. On the other hand, the complexities involved in fabricating a uniform single material which *optimally* exhibits both functions is actually more difficult than the interfacing problem at the present state of the art.

The ZnS and $\text{Bi}_{12}\text{SiO}_{20}$ materials used in the reported devices are films on the order of 30- μm and 150- μm thickness, respectively. At the present state of the art, they appear to offer comparable resolution capability: 10- μm and approximately 12 μm . Images have been recorded and read out in both devices but there have been no reported experiments of holographic storage.

Summary

A comparison of performance parameters for recyclable holographic storage media is given in Table 1. In most cases we have listed the "best" state-of-the-art material in the storage-medium column as an example, although there may be potentially better materials being investigated. Detailed information on these may be found in the list of references accompanying this paper.

Although quantitative information for recyclable holographic storage media is in a rapid state of flux, we have attempted to offer a quantitative comparison in Table 1 by defining a "Figure of Merit". In deciding on a reasonable definition for this relative "tradeoff" value, we have omitted record- and erase-time parameters, because they differ by such great orders of magnitude (e.g., picoseconds versus minutes). These particular parameters can be considered independently in any specific design application depending on the desired recycling speed. The remaining quantitative parameters (resolution, readout efficiency, and write energy) listed in Table 1 thus yield the following definition:

$$\text{Figure of merit} = \frac{\text{Resolution}(lp \cdot mm^{-1}) \times \text{Diffraction efficiency}(\%)}{\text{Write energy}(mJ \cdot cm^{-2})}$$

The superior figure-of-merit values in Table 1 for the thermoplastic and elastomer devices are due primarily to their low write-energy values. A look at the record-time column, however, indicates that their record-time values are several orders of magnitude higher than many media that have lower figure-of-merit values. Also, the thermoplastic and elastomer devices require more than light alone to operate properly, a fact which moderates the meaning of their high figure of merit in certain applications. In addition, the thermoplastic device has poor cycle lifetime at present and the elastomer device is not useful for permanent storage.

Because of their low write-energy values, the liquid-crystal-photoconductor device and the photoconductor-Pockel's effect device probably have figure-of-merit values equivalent to those of the thermoplastic and elastomer devices. However, since their diffraction efficiency values are not available, we cannot compute a specific figure of merit for them at present. Comparing these devices further we note that the photoconductor-Pockel's effect device has excellent sensitivity relative to the other recyclable media, while the liquid-crystal-photo-

conductor device has poor resolution at present (although holographic recording has been reported in the more complicated electrically-controlled device⁵⁸). On the other hand, the liquid-crystal-photoconductor device is simple in structure, has low power consumption, and requires only low control voltages.

Generally speaking, the thermoplastic, liquid-crystal-photoconductor, and photoconductor-Pockel's effect devices can find significant application as incoherent-to-coherent image converters.³ Most of the other materials in Table 1 find use in applications in which frequent readout and infrequent write-in are required.

Reviewing these other materials we see that magneto-optic materials are quite useful for two-dimensional storage and have the advantage of relative ease of fabrication compared to crystal devices and the disadvantages of requiring accurate temperature control and pulsed operation. Electro-optic materials are useful for volume storage and, therefore, have high storage capacity which, in turn, implies that selective erasure is difficult. Also, being single-crystal materials, they are relatively expensive compared to film media. One of their important advantages, however, is in nonoptically erasable applications because holographic information can be "fixed" in the material by a simple thermal process. In this respect, electro-optic materials are superior to photochromic materials, which, while three-dimensional storage materials in principle, cannot be used as such because information stored in them cannot be fixed. Also, photochromic materials require two wavelengths for operation but have inherently unlimited resolution capability.

In the photoconductor devices, although the photoconductive material is not the key material for storage, it does affect sensitivity both in terms of amount of light required and the wavelength that can be used; it also affects readout efficiency, since light is absorbed if reading out with the same wavelength as writing in. In this regard, the real beauty of the FE-PC devices is their ability to retain information on readout when light is absorbed; in comparison, information in the photoconductor-Pockel's effect device decays on readout.

In summary, Table 1 presents both quantitative and qualitative information on the state of the art of recyclable holographic recording media. The calculated figures of merit are intended only as an indication of possible potential performance. Since the state of the art in this area is so fluid at present, a very careful and perhaps partly intuitive tradeoff must be made among all performance parameters for any projected application.

References:

- ¹ J. A. Rajchman, "Promise of Optical Memories," *Appl. Phys.*, Vol. 41, p. 1376, March 1970.
- ² R. A. Bartolini, W. J. Hannan, D. Karlsons, and M. J. Lurie, "Embossed Hologram Motion Pictures for Television Playback," *Appl. Optics*, Vol. 9, p. 2283, Oct. 1970.
- ³ N. H. Farhat and W. R. Guard, "Holographic Imaging at 70 GHz," *Proc. IEEE*, Vol. 58, p. 1955, Dec. 1970.
- ⁴ J. J. Amodei, "Holographic Materials," in R. J. Pressley (editor), *CRC Handbook of Lasers with Selected Data on Optical Technology*, The Chemical Rubber Co., Cleveland, Ohio, 1971.
- ⁵ L. Mayer, "Curie-Point Writing on Magnetic Films," *J. Appl. Physics*, Vol. 29, p. 1003, June 1958.
- ⁶ L. Mayer, "Magnetic Writing with an Electron Beam," *J. Appl. Physics*, Vol. 29, p. 1454, Oct. 1958.
- ⁷ D. Chen, J. F. Ready and E. Bernal, "MnBi Thin Films: Physical Properties and Memory Applications," *J. Appl. Physics*, Vol. 39, p. 3916, July 1968.
- ⁸ R. S. Mezrich, "Curie-Point Writing of Magnetic Holograms on MnBi," *Appl. Phys. Letters*, Vol. 14, p. 132, Feb. 15, 1969.
- ⁹ G. Y. Fan, K. Pennington, and J. H. Greiner, "Magneto-Optic Hologram," *J. Appl. Phys.*, Vol. 40, p. 974, March 1, 1969.
- ¹⁰ M. J. Freiser, "A Survey of Magneto-Optical Effects," *IEEE Trans. Magnetism*, Vol. MAG-4, p. 152, June 1968.
- ¹¹ R. S. Mezrich, Private Communication.
- ¹² N. Goldberg, "A High-Density, Magneto-Optic Memory," *IEEE Trans. Magnetism*, Vol. MAG-3, p. 605, Dec. 1967.
- ¹³ D. I. Tehernev and G. W. Lewicki, "Extremely High-Density Magnetic Information Storage," *IEEE Trans. Magnetism*, Vol. MAG-4, p. 75, March 1968.
- ¹⁴ F. S. Chen, T. T. LaMacchia, and D. B. Fraser, "Holographic Storage in Lithium Niobate," *Appl. Phys. Ltrs.*, Vol. 13, p. 223, Oct. 1, 1968.
- ¹⁵ J. B. Thaxter, "Electrical Control of Holographic Storage in Strontium-Barium Niobate," *Appl. Phys. Ltrs.*, Vol. 15, p. 210, Oct. 1969.
- ¹⁶ J. J. Amodei, D. L. Staebler, and A. W. Stephens, "Holographic Storage in Doped Barium Sodium Niobate ($\text{Ba}_2\text{NaNb}_5\text{O}_{15}$)," *Appl. Phys. Ltrs.*, Vol. 18, p. 507, June 1, 1971.
- ¹⁷ F. S. Chen, "Optically Induced Change of Refractive Indices in LiNbO_3 and LiTaO_3 ," *J. Appl. Phys.*, Vol. 40, p. 3389, July 1969.
- ¹⁸ J. J. Amodei, W. Phillips, and D. L. Staebler, "Improved Electro-Optic Materials for Holographic Storage Applications," *IEEE J. Quant. Elec.*, Vol. QE-7, p. 321, June 1971.
- ¹⁹ J. J. Amodei and D. L. Staebler, "Holographic Pattern Fixing in Electro-Optic Crystals," *Appl. Phys. Ltrs.*, Vol. 18, p. 540, 15 June 1971.
- ²⁰ G. Jackson, "The Properties of Photochromic Materials," *Optica Acta*, Vol. 16, p. 1, Jan.-Feb. 1969.
- ²¹ S. Herman, "Properties of Photochromic Memories," *Proc. Symp. Modern Optics.*, p. 743, John Wiley and Sons, New York, 1967.
- ²² J. P. Kirk, "Hologram on Photochromic Glass," *Appl. Optics*, Vol. 5, p. 1684, Oct. 1966.
- ²³ R. J. Araujo, "Properties and Applications of Photochromic Glasses," *Recent Advances in Display Media*, NASA Symposium Proceedings, Cambridge, Mass., p. 63, Sept. 19-20, 1967.
- ²⁴ D. R. Bosomworth and H. J. Gerritsen, "Thick Holograms in Photochromic Materials," *Appl. Opt.*, Vol. 7, p. 95, Jan. 1968.
- ²⁵ A. S. Mackin, "Holographic Recording on Electron Beam Colored Sodium Chloride Crystals," *Appl. Opt.*, Vol. 9, p. 1658, July 1970.
- ²⁶ T. A. Shankoff, "Recording Holograms in Luminescent Materials," *Appl. Optics*, Vol. 8, p. 2282, Nov. 1969.
- ²⁷ A. L. Mikaeliane, A. P. Axenchikov, V. I. Bobrinev, E. H. Gulianiane, and V. V. Shatun, "Holograms on Photochromic Films," *IEEE J. Quantum Elec.*, Vol. EQ-4, p. 757, Nov. 1968; or *Soviet Physics-Doklady*, Vol. 13, p. 810, 1968.
- ²⁸ M. Lescinsky and M. Miller, "Holographic Record of Information onto Organic Photochromic Film," *Optics Communications*, Vol. 1, p. 417, April 1970.

- ²⁹ D. L. Ross, "Photochromic Indigoids, III: A Photochromic Element Based on the cis-trans Photoisomerization of a Thioindigo Dye," *Appl. Optics*, Vol. 10, p. 571, March 1971.
- ³⁰ J. J. Amodei and R. S. Mezrich, "Holograms in Thin Bismuth Films," *Appl. Phys. Lett.*, Vol. 15, p. 45, July 15, 1969.
- ³¹ R. M. Schaffert, "Photoelectric Recording Apparatus," U.S. Patent 3,148,354, issued September 8, 1964.
- ³² S. E. Cummins, "A New Optically Read Ferroelectric Memory," *Proc. IEEE*, Vol. 55, p. 1536, Aug. 1967.
- ³³ S. E. Cummins and L. E. Cross, "Electrical and Optical Properties of Ferroelectric $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ Single Crystals," *J. Appl. Phys.*, Vol. 39, p. 2268, April 1968.
- ³⁴ J. R. Maldonado and A. L. Meitzler, "Strain-Biased Ferroelectric-Photoconductor Image Storage and Display Devices," *Proc. IEEE*, Vol. 59, p. 368, March 1971.
- ³⁵ J. R. Maldonado and L. K. Anderson, "Strain-Biased Ferroelectric-Photoconductor Image Storage and Display Devices Operated in a Reflection Mode," *IEEE Trans. Elec. Dev.*, Vol. ED-18, p. 774, Sept. 1971.
- ³⁶ A. D. Morrison, F. A. Lewis, and A. Miller, "The Growth of Large Single-Crystal $\text{Bi}_4\text{Ti}_3\text{O}_{12}$," *Ferroelectrics*, Vol. 1, p. 75, May 1970.
- ³⁷ M. M. Hopkins and A. Miller, "Preparation of Poled, Twin-Free Crystals of Ferroelectric Bismuth Titanate, $\text{Bi}_4\text{Ti}_3\text{O}_{12}$," *Ferroelectrics*, Vol. 1, p. 37, March 1970.
- ³⁸ S. A. Keneman, G. W. Taylor, A. Miller, and W. H. Fonger, "Storage of Holograms in a Ferroelectric-Photoconductor Device," *Appl. Phys. Letters*, Vol. 17, p. 173, 15 Aug. 1970.
- ³⁹ S. E. Cummins and T. E. Luke, "A New Method of Optically Reading Domains in Bismuth Titanate for Display and Memory Applications," *IEEE Trans. Elec. Dev.*, Vol. ED-18, p. 761, Sept. 1971.
- ⁴⁰ S. A. Keneman, G. W. Taylor, and A. Miller, "Ferroelectric-Photoconductor Optical Storage Medium Utilizing Bismuth Titanate," *Ferroelectrics*, Vol. 1, p. 227, Oct. 1970.
- ⁴¹ S. A. Keneman, A. Miller, and G. W. Taylor, "Phase Holograms in a Ferroelectric-Photoconductor Device," *Appl. Optics*, Vol. 9, p. 2279, Oct. 1970.
- ⁴² G. W. Taylor, S. A. Keneman, and A. Miller, "Depicting of Single-Domain Bismuth Titanate," *Ferroelectrics*, Vol. 2, p. 11, Jan. 1971.
- ⁴³ S. A. Keneman, A. Miller, and G. W. Taylor, "The Bismuth Titanate FE-PC Optical Storage Medium—Further Developments," *Ferroelectrics*, Vol. 3, p. 131, Feb. 1972.
- ⁴⁴ 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.
- ⁴⁵ W. J. Takei, N. P. Formigoni, and M. H. Francombe, "Preparation and Properties of Epitaxial Films of Ferroelectric $\text{Bi}_4\text{Ti}_3\text{O}_{12}$," *Appl. Phys. Letters*, Vol. 15, p. 256, 15 Oct. 1969.
- ⁴⁶ S. Y. Wu, W. J. Takei, M. H. Francombe, and S. E. Cummins, "Domain Structure and Polarization Reversal in Thin Films of $\text{Bi}_4\text{Ti}_3\text{O}_{12}$," *Ferroelectrics*, Vol. 3, Dec. 1971.
- ⁴⁷ W. E. Glenn, "Thermoplastic Recording," *J. Appl. Phys.*, Vol. 30, p. 1870, Dec. 1959.
- ⁴⁸ W. E. Glenn, "Thermoplastic Recording: A Progress Report," *J. SMPTE*, Vol. 74, p. 663, Aug. 1965.
- ⁴⁹ J. C. Urbach and R. W. Meier, "Thermoplastic Xerographic Holography," *Appl. Optics*, Vol. 5, p. 666, April 1966.
- ⁵⁰ L. H. Lin and H. L. Beauchamp, "Write-Read-Erase in Situ Optical Memory Using Thermoplastic Holograms," *Appl. Optics*, Vol. 9, p. 2088, Sept. 1970.
- ⁵¹ J. Urbach and R. W. Meier, "Properties and Limitations of Hologram Recording Materials," *Appl. Optics*, Vol. 8, p. 2269, Nov. 1969.
- ⁵² H. R. Anderson, Jr., E. A. Bartkus, and J. A. Reynolds, "Molecular Engineering in the Development of Materials for Thermoplastic Recording," *IBM J. Res. Develop.*, p. 140, March 1971.
- ⁵³ L. S. Chang, "The Physical Parameters of a Thermoplastic Polymer Film in an Electron Beam Read/Write System," *Photographic Science and Engineering*, Vol. 12, p. 238, Sept.-Oct. 1968.
- ⁵⁴ R. J. Doyle and W. E. Glenn, "Lumatron, A High Resolution Storage and Projection Display Device," *IEEE Conf. Record 1970 IEEE Conf. on Display Devices (70 C. 55-ED)*, p. 88, Dec. 2-3, 1970.
- ⁵⁵ R. J. Doyle and W. E. Glenn, "Lumatron: A High-Resolution Storage and Projection Display Device," *IEEE Trans. Electron Devices*, Vol. ED-18, p. 739, Sept. 1971.
- ⁵⁶ N. K. Sheridan, "A New Optical Recording Device," *IEEE International Electron Devices Meeting*, Washington, D.C., October 28, 1970.

⁵⁷ G. H. Hellmeier and J. E. Goldmacher, "A New Electric Field Controlled Reflective Optical Storage Effect in Mixed Liquid Crystal Systems," *Proc. IEEE*, Vol. 57, p. 34, Jan. 1969.

⁵⁸ H. Kiemle and U. Wolff, "Electrically Controllable Holograms Using Nematic Liquid Crystals," 1970 Annual Meeting of Optical Society of America, 28 Sept.-2 Oct. 1970.

⁵⁹ J. D. Margerum, J. Nimoy, and S. Y. Wong, "Reversible Ultraviolet Imaging with Liquid Crystals," *Appl. Phys. Ltrs.*, Vol. 17, p. 51, 15 July 1970.

⁶⁰ G. Assouline, M. Hareng, and E. Leiba, "Liquid Crystal and Photoconductor Image Converter," *Proc. IEEE*, Vol. 59, p. 1355, Sept. 1971.

⁶¹ D. S. Oliver, P. Vohl, R. E. Aldrich, M. E. Behrndt, W. R. Buchan, R. C. Ellis, J. E. Genthe, J. R. Goff, S. L. Hou, and G. McDaniel, "Image Storage and Optical Readout in a ZnS Device," *Appl. Phys. Ltrs.*, Vol. 17, p. 416, 15 Nov. 1970.

⁶² D. S. Oliver and W. R. Buchan, "An Optical Image Storage and Processing Device Using Electro-Optic ZnS," *IEEE Trans. Elec. Dev.*, Vol. ED-18, Sept. 1971.

⁶³ S. L. Hou and D. S. Oliver, "Pockel's Readout Optical Memory Using $\text{Bi}_{12}\text{SiO}_{20}$," *Appl. Phys. Ltrs.*, Vol. 18, p. 325, 15 April 1971.

Erase-Mode Recording Characteristics of Photochromic CaF_2 , SrTiO_3 , and CaTiO_3 Crystals*

R. C. Duncan, Jr.

RCA Laboratories, Princeton, N. J.

Abstract—Erase-mode optical recording characteristics of photochromic crystal wafers of $\text{CaF}_2:\text{La,Na}$; $\text{CaF}_2:\text{Ce,Na}$; $\text{SrTiO}_3:\text{Ni,Mo,Al}$; and $\text{CaTiO}_3:\text{Ni,Mo}$ have been measured. An argon laser operating at 5145 \AA was used for both optical recording and optical readout. Sensitometric curves of optical-density change versus logarithm of exposure are shown for a number of erase-beam intensities between 0.2 mW/cm^2 and 2 W/cm^2 . In this range, time-intensity reciprocity holds for the CaF_2 materials but fails for the titanates, particularly at low intensities. The dependences of sensitivity, gamma, and maximum transmission contrast ratio on wafer thickness and material are discussed. Wafers of SrTiO_3 , CaTiO_3 , and CaF_2 exhibiting approximately equal maximum contrast ratios have relative sensitivities approximately in the ratio 5:2:1, respectively, at an erase intensity of 1 W/cm^2 . In the regime studied, the ratio of gamma to maximum optical-density change for each material is found empirically to be independent of wafer thickness.

1. Introduction

Photochromic materials are materials that change color reversibly under appropriate optical illumination. The effect has been observed and studied in a variety of materials, both organic¹⁻³ and inorganic.³⁻⁵ In

* The research reported in this paper was jointly sponsored by the National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland under contract No. NAS 5-10335; and RCA Laboratories, Princeton, N. J.

particular, a number of recent papers have discussed the photochromic behavior of CaF_2 doped with certain rare earths ($\text{CaF}_2:\text{RE}$)⁶⁻¹³ and of SrTiO_3 (and CaTiO_3) doped with transition metals ($\text{SrTiO}_3:\text{TM}$ and $\text{CaTiO}_3:\text{TM}$).¹⁴⁻¹⁷ At the present time, these are the inorganic photochromic materials that, in the form of single-crystals wafers or slabs, offer the greatest promise for applications requiring all-optical recording, readout, and erasure. The recent studies of these materials have focused primarily on (a) identifying the active photochromic color centers, (b) understanding the fundamental photochromic mechanisms, and (c) maximizing the photochromically induced changes in absorption.

Far less attention has been paid, thus far, to detailed applications-oriented characterization of these materials. It is the purpose of this paper, therefore, to report quantitative and comparative studies of the photochromic "erase-mode" recording characteristics of crystal-wafer samples of several of these state-of-the-art materials.

"Erase-mode" operation refers to the use of visible light to optically record on photochromic materials which have previously been uniformly "switched" (or darkened) by violet or ultraviolet light. This mode of use is of particular practical interest because only laser sources operating at visible wavelengths have sufficient power output to accomplish the fast optical recording usually required. In the work reported here, the argon-laser 5145 Å line was used both to record and to "read" the resulting changes in absorption. This line lies very close to the peak of the visible readout and erase absorption band of each of the materials chosen for study. "Write-mode" sensitometric measurements now in progress will be reported in another paper.

2. Materials

2.1 General Properties

Samples of four different photochromic host-dopant combinations were included in this study. The host crystals, dopants, and nominal dopant concentrations are listed in Table 1.

Of the photochromic $\text{CaF}_2:\text{RE}$ systems, $\text{CaF}_2:\text{La,Na}$ exhibits the strongest photochromic effects.^{9,13} Charge compensation with Na inhibits the growth of a frequently observed nonphotochromic coloration and thereby permits greater photochromic coloration and correspondingly greater photochromic switching.⁹ The La, Ce, and Na concentrations indicated in Table 1 lie in the ranges found to give the maximum photochromic absorption changes.⁹ The CaF_2 crystals were grown at RCA Laboratories by H. Temple using a gradient freeze technique

described elsewhere.^{9,13} Photochromic color centers were generated in the optically polished wafer samples by appropriate additive coloration.⁹ The optical quality of carefully prepared wafers is generally excellent.

Of the photochromic SrTiO₃:TM systems, SrTiO₃:Ni,Mo is one of the most extensively studied and exhibits among the strongest photochromic switching and the longest switched-state thermal lifetimes.^{5,8,17} With Mo omitted, the SrTiO₃:Ni system shows both weaker switching and a shorter lifetime. The Al impurity indicated in Table 1 does not

Table 1—Materials

<i>Host</i>	<i>Dopants and Nominal Concentrations</i>
CaF ₂	LaF ₃ 0.05 mol %
	NaF 0.1 mol %
CaF ₂	CeF ₃ 0.05 mol %
	NaF 0.1 mol %
SrTiO ₃	NiO 0.35 mol %
	MoO ₃ 0.34 mol %
	Al ₂ O ₃ 0.18 mol %
CaTiO ₃	NiMoO ₄ 0.19 mol %

affect the photochromic behavior but has sometimes been included to provide additional charge compensation. CaTiO₃:TM materials typically exhibit somewhat greater photochromic color changes and somewhat longer switched-state thermal lifetimes than do the corresponding SrTiO₃:TM materials.¹⁷ The Ni and Mo concentrations indicated in Table 1 lie in the ranges found to produce maximum photochromic effects.⁸ Single crystals of SrTiO₃:Ni,Mo,Al and CaTiO₃:Ni,Mo grown by flame fusion were obtained from National Lead Co. (Titanium Division). Wafer samples cut from these crystals were optically polished, then subjected to mild reduction treatments^{15,17} to reduce to a minimum their unswitched-state visible absorption. The optical quality of the SrTiO₃:Ni,Mo,Al wafers was generally good, although some signs of strain were frequently evident in crossed-polarizer examination. The optical quality of the CaTiO₃:Ni,Mo wafers was significantly poorer, with extensive lamellar twinning evident in all samples in spite of the extraordinary crystal growth and annealing procedures used.¹⁹

2.2 Photochromic Absorption Spectra

Photochromic absorption spectra of wafer samples of each material are shown in Figs. 1 through 4. The transmission optical density (*D*),

the logarithm of the ratio of the incident (I_0) and transmitted (I) light intensities, is plotted as a function of the wavelength of the light. The various features of these spectra have been discussed in detail elsewhere.^{5,18,17} For present purposes, it will be sufficient to briefly describe the photochromic behavior indicated and to mention several practical considerations.

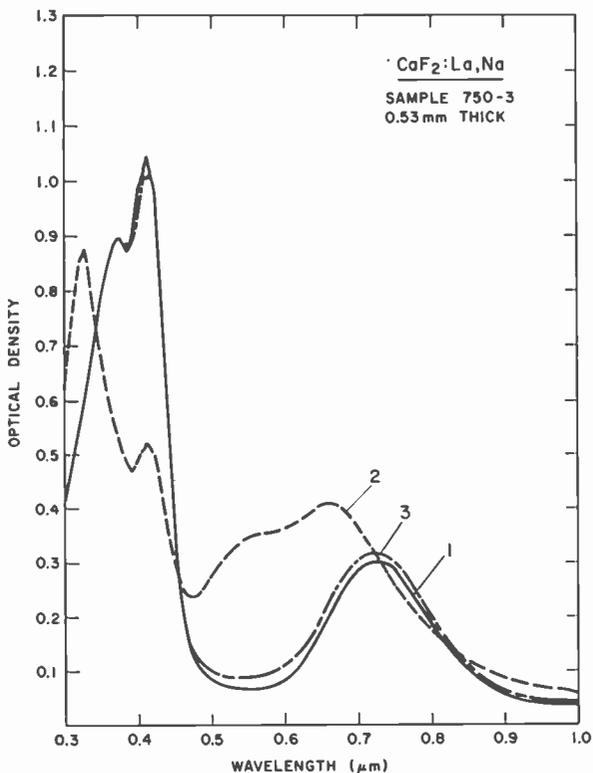


Fig. 1—Photochromic absorption spectra of CaF₂:La,Na in the unswitched (curve 1), switched (curve 2), and optically erased (curve 3) states.

The spectra of Figs. 1 through 4 were recorded on a Cary 14 Spectrophotometer. Both switching and optical erasure were carried out using suitably filtered and focused radiation from a 500-Watt Hg arc lamp. The filters used for switching the CaF₂ wafers, about a one centimeter thickness of saturated CuSO₄ solution and a Corning 7-59, 0-51 filter combination, produced switching radiation in the range $3800 \text{ \AA} < \tilde{\lambda} < 4600 \text{ \AA}$. For switching the titanate wafers, the CuSO₄ solution and a Corning 7-54 filter produced switching radiation in the

range $3300 \text{ \AA} \lesssim \lambda \lesssim 3900 \text{ \AA}$. Optical erasure of all samples was carried out using a Corning 3-71 filter and about one centimeter of water to limit the erasing light to the range $4800 \text{ \AA} \lesssim \lambda \lesssim 9500 \text{ \AA}$.

The spectra of the thermally stable (unswitched) states are indicated by the solid curves (curves 1 in Figs. 1 and 2). These were recorded after the wafers had been heated to about 150°C for about two minutes and cooled to room temperature in the dark. In this state, these materials are all nearly (though not absolutely) nonabsorbing at visible wavelengths between 5000 and 6000 \AA .

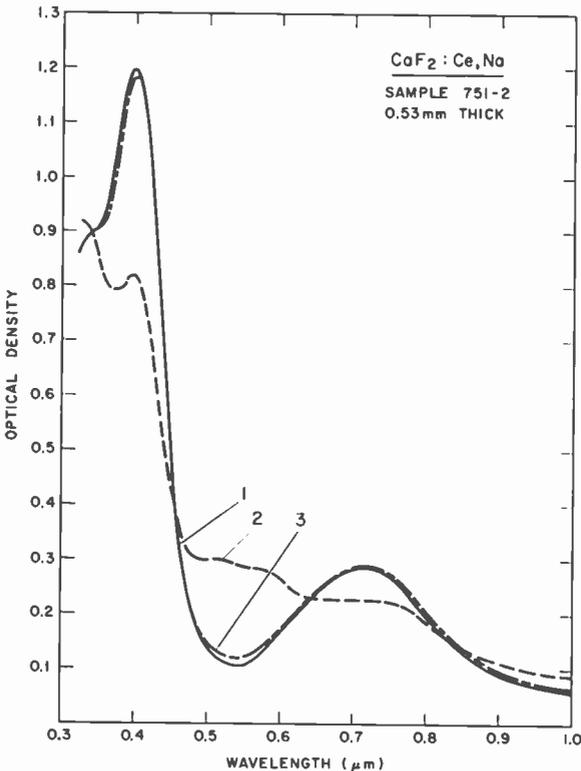


Fig. 2—Photochromic absorption spectra of $\text{CaF}_2:\text{Ce,Na}$ in the unswitched (curve 1), switched (curve 2), and optically erased (curve 3) states.

The switched-state spectra, indicated by the broken curves (curves 2 in Figs. 1 and 2), were then recorded after the wafers had been exposed to the appropriate strong violet or ultraviolet switching radiation for 2 minutes, which is long enough to saturate the photochromic changes in absorption. In the titanate materials, effective switching is

produced by radiation in and beyond (shorter wavelengths) the sharp absorption edge at about 4000 Å. The visible switched-state absorption of these wafers exhibited appreciable thermal decay while these spectra were being recorded. The spectra shown here, the broken curves of Figs. 3 and 4, have been corrected for this decay.

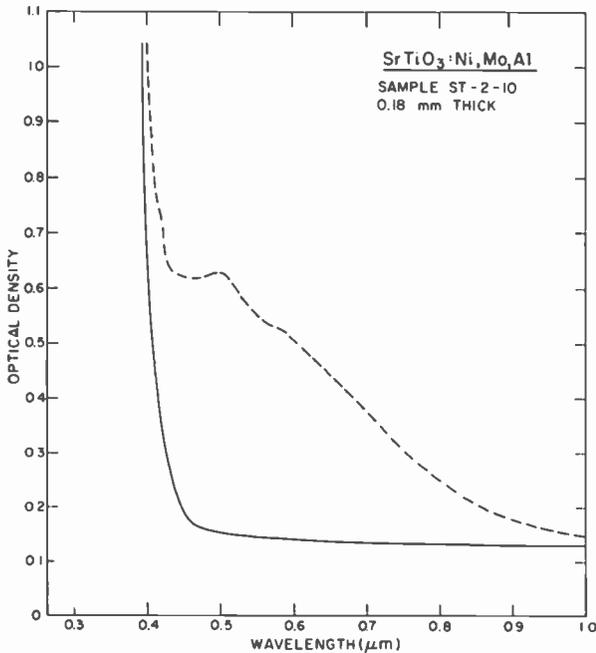


Fig. 3—Photochromic absorption spectra of $\text{SrTiO}_3:\text{Ni,Mo,Al}$ in the unswitched (solid curve) and switched (broken curve) states.

In the case of the CaF_2 materials, the switching radiation is absorbed primarily by the strong unswitched-state absorption band near 4000 Å. Absorption by the accompanying band near 7000 Å is ineffective in switching. Radiation at wavelengths shorter than about 3800 Å can produce up to about a 50% greater photochromic change. This additional switching is not optically reversible, however, and thermal erasure is required.

Finally, optical erasure was carried out by exposing the wafers to light lying within the visible switched-state absorption bands. The resulting absorption spectra of the CaF_2 materials are shown by long-short dashed curves (curves 3 in Figs. 1 and 2). The residual switched-state absorption indicated by these curves typically amounts to about 10% of the initial switched-state absorption and is *not* cumulative (<1%) with successive photochromic cycling. It can be erased therm-

ally. For the titanates, the spectra produced by optical erasure are indistinguishable from those of the initial thermally stable states.

Because of this erasing effect of visible light, it is clear that switching radiation at wavelengths longer than about 4600 Å for the CaF_2 materials or 4300 Å for the titanates is not only ineffective but actually becomes detrimental as switching proceeds, limiting both the rate of change and the ultimate saturated level of photochromic absorption.¹⁶

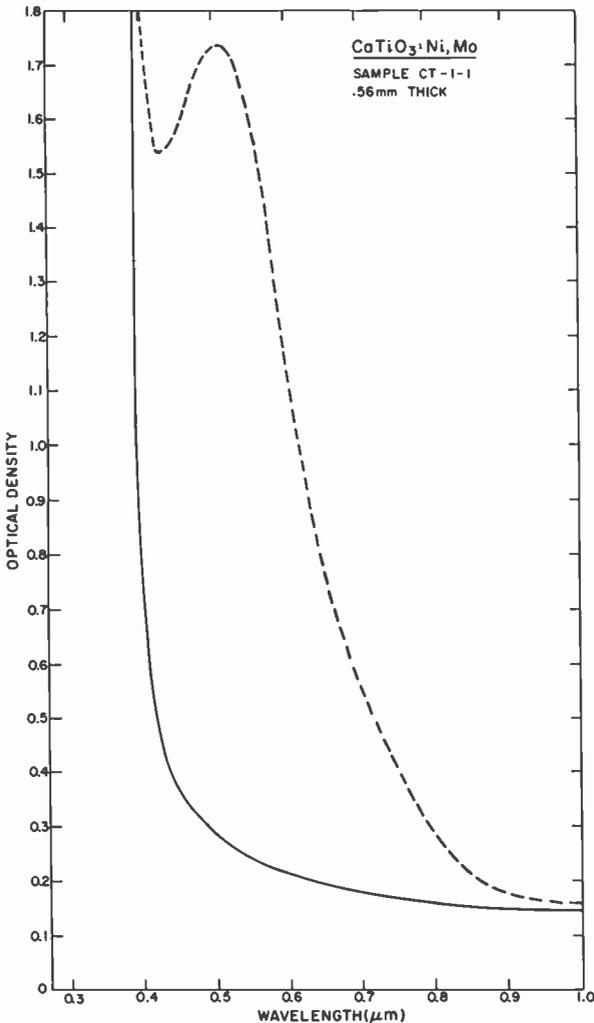


Fig. 4—Photochromic absorption spectra of $\text{CaTiO}_3:\text{Ni,Mo}$ in the unswitched (solid curve) and switched (broken curve) states.

2.3 Optical Recording Modes

The optical reversibility of these materials clearly makes them subject to two modes of optical recording. In the write mode, appropriate violet or ultraviolet radiation (as described above) is used to switch material initially in its thermally stable or unswitched state. In the erase mode, appropriate visible "erase" light is used to record on material previously darkened by uniform exposure to switching radiation. In either case, readout of the information thus recorded is obtained by monitoring the absorption in the region of the visible switched-state bands shown in Figs. 1 through 4. With CaF_2 wafers, readout (of the opposite sign) is also possible at wavelengths lying in the strong photochromic absorption bands near 4000 Å.

The photochromic sensitivity of these materials is from two to ten times greater for write-mode recording than for erase-mode recording.^{5,8,13} On the other hand, the only available laser sources with sufficient power output to accomplish the fast addressable optical recording required for many applications are those operating at visible (erase-mode) wavelengths.

In the work reported here, therefore, the 5145 Å argon laser line was used both for erase-mode optical recording and for optical readout. From Figs. 1 through 4 it is clear that 5145 Å corresponds very closely to the peaks of the visible switched-state absorption bands of all of the materials studied. The poorest match occurs for $\text{CaF}_2:\text{La,Na}$. Even here, however, the photochromic absorption change at 5145 Å is still about 80% of that at 5750 Å where the maximum change does occur. This fortunate correspondence means that we have near-maximum utilization (absorption) of the optical-recording radiation from a high intensity, coherent, and addressable source; in addition, readout with coherent light is possible (e.g., for optical processing applications) at a near-optimum readout wavelength.

2.4 Optical Readout at 5145 Å

It is of interest to examine quantitatively and comparatively the ranges of optical density change available for 5145 Å optical readout. Photochromic absorption spectra similar to those of Figs. 1 through 4 were determined for sample wafers of several thicknesses of each of the four materials studied. The wafer thicknesses, d , and the corresponding maximum photochromic optical density changes at 5145 Å, ΔD , are given in the first two columns of Table 2. The corresponding maximum obtainable transmission contrast ratios, $R = 10^{\Delta D}$, and aver-

age bulk absorption coefficient changes, $\overline{\Delta\alpha} = 2.3 \Delta D/d(\text{cm})$, are also indicated.

a. Thickness Effects

The CaF_2 materials exhibit saturation optical density changes that are nearly directly proportional to wafer thickness or, alternatively, maximum changes in average bulk absorption coefficient that are independent of wafer thickness. This is an indication that, at saturation, the photochromic coloration is essentially uniformly distributed throughout the bulk of these materials.

For the $\text{SrTiO}_3:\text{Ni,Mo,Al}$, on the other hand, while wafer thickness varies by a factor of six, the maximum ΔD varies by less than a factor of two, and the maximum $\overline{\Delta\alpha}$ varies by a factor of almost three. These wafers absorb strongly at wavelengths shorted than 4000 Å (see Fig. 3), and photochromic coloration is confined to the relatively limited penetration depth of the switching radiation in this wavelength range. The effective penetration depth is about 0.1 to 0.2 mm, comparable to the thickness of the thinnest wafer listed.

Although generally more similar to $\text{SrTiO}_3:\text{Ni,Mo,Al}$ in other respects, the $\text{CaTiO}_3:\text{Ni,Mo}$ behaves more like the CaF_2 materials in the dependence of its maximum ΔD and $\overline{\Delta\alpha}$ on wafer thickness. This is a surprising result, and perhaps not a completely general one. Its explanation in the present case is probably twofold. First, the CaTiO_3 absorption edge occurs about 200 to 300 Å further toward the ultraviolet than that of SrTiO_3 ,¹⁷ so that some of the switching radiation used ($3300 < \lambda < 3900$ Å) can penetrate more deeply into the bulk material. Second, the longer (by a factor of more than 10) switched-state lifetime of the $\text{CaTiO}_3:\text{Ni,Mo}$ permits this penetrating radiation, even if fairly weak, to induce relatively large saturated photochromic switching.

b. Contrast Ratio

The thickest CaF_2 wafers included in this study exhibited maximum transmission contrast ratios at 5145 Å of only about 2:1. Because of the depth of photochromic coloration in these materials, larger contrast ratios can be obtained with thicker wafers. In general, however, increased wafer thickness means reduced image resolution and reduced storage capacity. In cases where coherent readout light is not necessary, a better readout wavelength for these materials is near 4000 Å (see Figs. 1 and 2). Here the maximum values of ΔD are somewhat more

than twice those at 5145 Å, and the maximum values of R are approximately the squares of those at 5145 Å.

Markedly greater transmission contrast ratios are attainable with the titanate wafers, as shown in Table 2. Because of the penetration limitations discussed above, SrTiO₃:Ni,Mo,Al wafers thicker than about 1 mm provide little increase in R . The values of R available with CaTiO₃:Ni,Mo wafers even under 0.5 mm in thickness are probably adequate for most applications. Unfortunately, the usefulness of CaTiO₃ is limited by its relatively poor optical quality.

3. Erase-Mode Sensitivity

3.1 Experimental Procedures

The photochromic wafer samples were switched to saturation as described earlier and, within two seconds, shifted into the beam of the argon 5145 Å laser. The beam intensity transmitted through the wafer was monitored by a linear PIN diode and displayed as a function of time on either an image storage oscilloscope or an x - y recorder, depending on the speed at which erasure proceeded. Thus, readout and erasure were accomplished with a single source and a single beam.

The maximum power output of the laser in the 5145 Å line was about one watt. The beam intensity at a 1/16-inch sample aperture could be varied from less than 0.2 mW/cm² to more than 2 W/cm², a range of more than four decades, and was measured directly with a recently calibrated Eppley thermopile. Wafer sample transmission as a function of erase exposure time was recorded for a number of beam intensities throughout this range.

Room-temperature thermal (dark) decay characteristics were determined by exposing the wafers to the lowest detectable beam intensity for only very brief periodic measurements.

3.2 Sensitivity Curves

Families of erase-mode sensitivity curves for wafer samples of each of the four materials studied are shown in Figs. 5 through 8. The photochromic change in optical density at 5145 Å is plotted as a function of time of exposure to the erasing beam. Optical density change (ΔD) plotted here is given by $\Delta D = \log(T_u/T_s)$, where T_u and T_s represent the transmission of the wafer in its unswitched state and its (partially or completely) switched state, respectively. The labels on the individual curves of each family indicate the erase beam intensities (in mW/cm²) to which they correspond.

It is clear that for each family of curves (each material) the slowest optical-density decay corresponds to the room-temperature thermal-erase process. The CaF_2 materials exhibit the longest thermal switched-state lifetimes, the optical density dropping by about 20% to 30% in 1000 sec for $\text{CaF}_2:\text{La,Na}$ (Fig. 5) and about 10-15% in

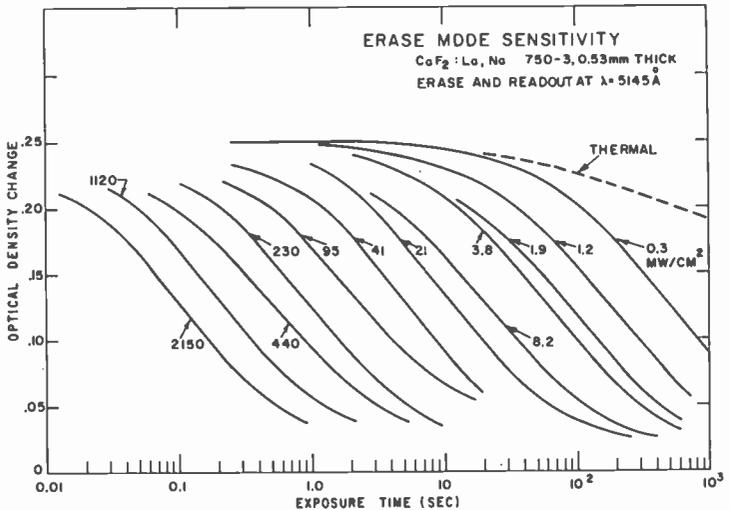


Fig. 5—Erasure of a $\text{CaF}_2:\text{La,Na}$ wafer as a function of erase exposure time for various erase-beam intensities.

1000 sec for $\text{CaF}_2:\text{Ce,Na}$ (Fig. 6). The $\text{SrTiO}_3:\text{Ni,Mo,Al}$ (Fig. 7), on the other hand, exhibits the shortest switched-state lifetime, the optical density falling by a factor of two in 15 to 40 seconds for various wafer samples examined. The corresponding decay times for $\text{CaTiO}_3:\text{Ni,Mo}$ (Fig. 8) are of the order of 20 minutes, almost two orders of magnitude longer. These room-temperature thermal-decay rates are in substantial agreement with those previously reported for these photochromic materials.^{8,13}

As the erase-beam intensity is increased, in each case, the switched-state optical density falls away from the thermal-decay curve at progressively shorter erase-exposure times. For low beam intensities, thermal decay continues to influence the erasure, and the slope of the sensitivity curve varies with beam intensity. At higher beam intensities, however, the erase curves for each material approach a nearly fixed shape with a central region of constant slope or gamma.

The existence of this linear central region of these sensitivity curves is important for many potential applications and enables us to

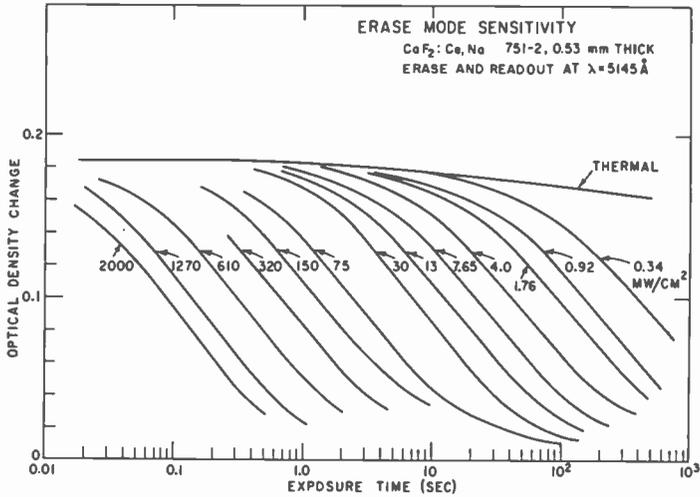


Fig. 6—Erasure of a $\text{CaF}_2:\text{Ce,Na}$ wafer as a function of erasure exposure time for various erasure-beam intensities.

apply the concept of a gamma to these photochromic materials. In each of the Figs. 5 through 8, these regions of constant slope include between 60% and 80% of the total optical density change.

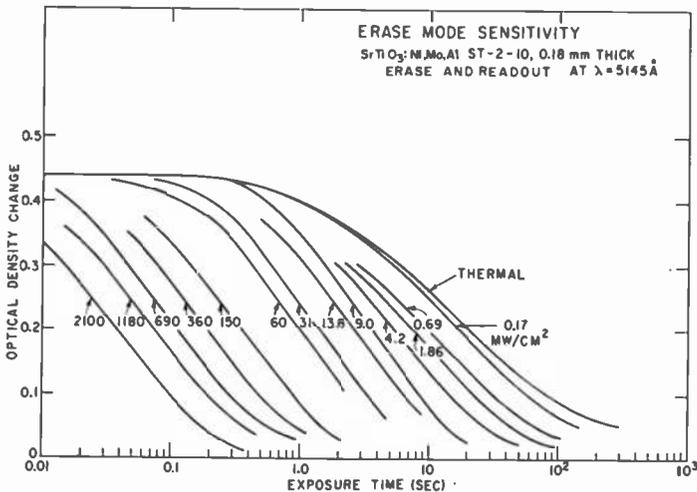


Fig. 7—Erasure of a $\text{SrTiO}_3:\text{Ni,Mo,Al}$ wafer as a function of erasure exposure time for various erasure-beam intensities.

3.3 Sensitometric Characteristics

If time/intensity reciprocity holds for the erase-mode operation of these photochromic wafers, then the entire family of sensitivity curves for each wafer should coalesce into a single sensitometric characteristic when the same optical-density changes are replotted as a function of erase-exposure energy density (i.e., the product of erase-beam intensity and exposure time) rather than of exposure time alone. Sensitometric characteristics calculated in this way from the data of Figs. 5 through 8 are shown in Figs. 9 through 12, respectively.

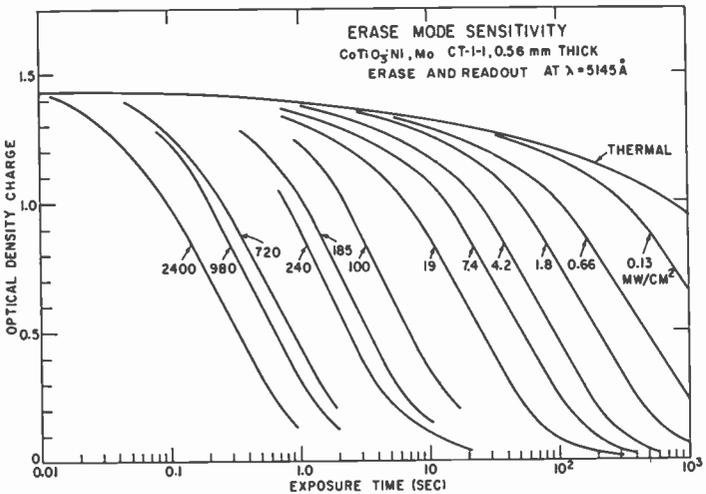


Fig. 8—Erasure of a CaTiO₃:Ni,Mo wafer as a function of erase exposure time for various erase-beam intensities.

3.3.1 CaF₂:La,Na and CaF₂:Ce,Na

The solid points in Figs. 9 and 10 indicate the erasures induced in the two CaF₂ wafers by exposure for various times to erase beams of selected *fixed intensities*. The abscissae are the corresponding erase-exposure energy densities. Thus, these points constitute a straight-forward replotting of selected sensitivity curves from Figs. 5 and 6, respectively. The selected beam intensities span approximately three decades.

In each case, the several separate curves have indeed coalesced sufficiently to be well represented by single sensitometric characteristics. With few exceptions, the plotted points along the central linear

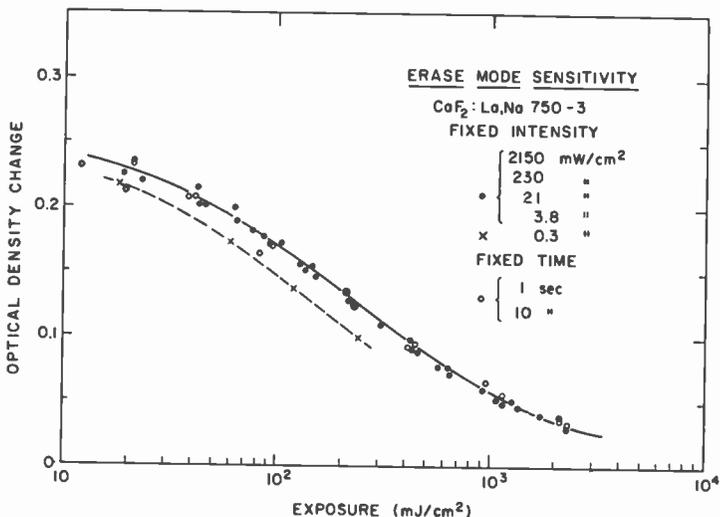


Fig. 9—Erase-mode sensitometric characteristic of a CaF_2 :La,Na wafer, showing extensive validity of time/intensity reciprocity.

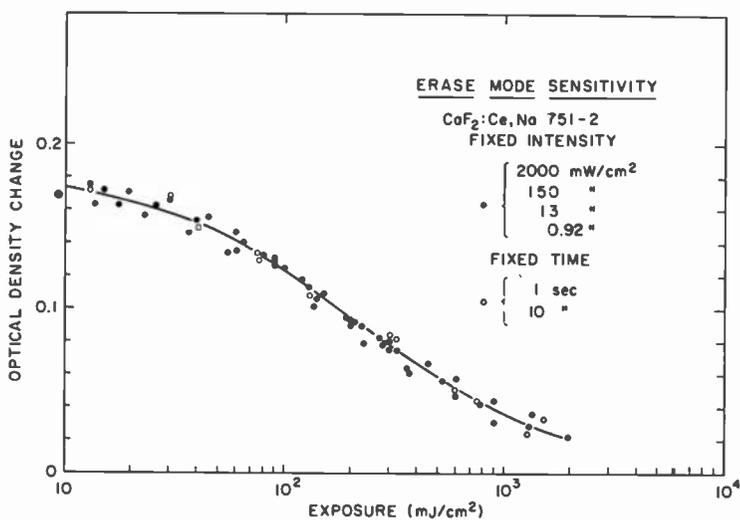


Fig. 10—Erase-mode sensitometric characteristic of a CaF_2 :Ce,Na wafer, showing extensive validity of time/intensity reciprocity.

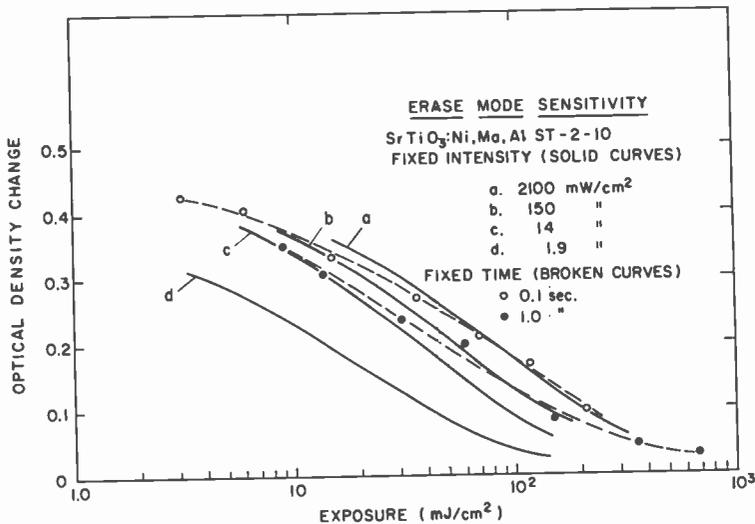


Fig. 11—Erase-mode sensitometric characteristics of a SrTiO₃:Ni,Mo,Al wafer, showing failure of time/intensity reciprocity.

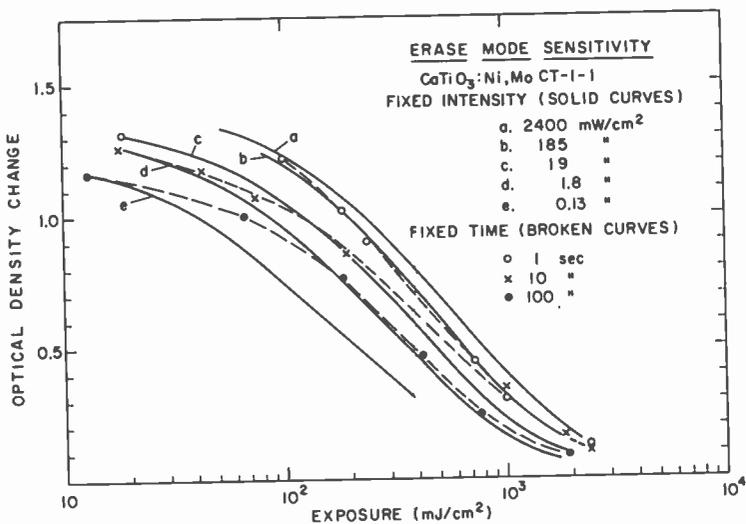


Fig. 12—Erase-mode sensitometric characteristics of a CaTiO₃:Ni, Mo wafer, showing failure of time/intensity reciprocity.

regions of these characteristics deviate from the curves drawn by less than $\pm 10\%$ in exposure or $\pm 5\%$ in optical density. Thus, in this intensity range, time/intensity reciprocity does hold, and these single characteristic curves provide condensed displays of the sensitivity data contained in the families of curves in Figs. 5 and 6.

The open-circle points in Figs. 9 and 10 indicate the erasures induced in selected *fixed times* by erase beams of various intensities. These points, also directly transposed from Figs. 5 and 6, sample ten or more different intensities in each case. Their close agreement with the previously determined sensitometric characteristics is a further demonstration of the validity of time/intensity reciprocity for these wafers in this intensity range.

Because of the effects of room temperature thermal erasure, time/intensity reciprocity cannot be expected to hold for arbitrarily low erase-beam intensities. The primary effect is that, as thermal processes begin to contribute, a given integrated erase-energy density appears to induce greater erasure. The x-points in Fig. 9 were determined in the same manner as were the solid points (which span three decades in beam intensity), but for an erase intensity smaller by another factor of ten, 0.3 mW/cm². It is clear from Fig. 5 that the time required for appreciable erasure at this beam intensity is also that required for thermal erasure to become significant. The accompanying reciprocity failure is evident in the x-point curve in Fig. 9. Time/intensity reciprocity in the erase-mode operation of these materials is discussed further in a later section.

3.3.2 $\text{SrTiO}_3:\text{Ni},\text{Mo},\text{Al}$ and $\text{CaTiO}_3:\text{Ni},\text{Mo}$

The solid curves in Figs. 11 and 12 indicate the erasures induced in the two titanate wafers by exposure for varying times to erase beams of selected *fixed intensities*. The abscissae are the corresponding erase-exposure energy densities. Thus, selected sensitivity curves of Figs. 7 and 8 are simply replotted with a new abscissa scale. The selected beam intensities span more than three decades (more than four in Fig. 12).

Unlike the similarly derived curves for the CaF_2 materials (Figs. 9 and 10), these curves have not coalesced to single sensitometric characteristics for each wafer. The closest pairs of curves differ by more than 30% in exposure for the SrTiO_3 and by about 20% for the CaTiO_3 . Clearly, then, for the erase-mode operation of these titanate-based wafers, time/intensity reciprocity does not hold over any extended region in this range of erase intensities. As a result, the sensitivity

data contained in the families of curves in Figs. 7 and 8 cannot be condensed into any single characteristic curves.

The broken curves of Figs. 11 and 12 give further support to this conclusion. These curves, also derived from Figs. 7 and 8, indicate the erasures induced in selected *fixed times* by erase beams of various intensities. Each of these curves samples six or more different intensities. These curves, too, are distinctly separate.

As we have seen (Figs. 5 through 8), the titanate-based photochromics exhibit much more rapid thermal decay than do the CaF_2 -based materials. It is not surprising, then, that they suffer a more severe reciprocity failure as well. The subject of time/intensity reciprocity in these materials and the relationship between reciprocity failure and thermal decay are examined more carefully in the following section.

3.4 Time/Intensity Reciprocity

Erase-mode time/intensity reciprocity is investigated more generally and more directly in Figs. 13 through 16. Here, the time-of-erase is plotted against the erase-beam intensity for each of the fifteen sample wafers studied (see Table 2). The data plotted for each wafer were

Table 2—Optical Readout at 5145 Å

Wafer Thickness d (mm)	Maximum Absorption Changes		
	ΔD	R	$\Delta\alpha$ (cm ⁻¹)
CaF ₂ :La,Na			
0.81	0.33	2.1	9.4
0.53	0.23	1.7	10
0.29	0.12	1.3	9.5
0.13	0.06	1.1	11
CaF ₂ :Ce,Na			
0.86	0.27	1.9	7.2
0.53	0.18	1.5	7.8
0.32	0.10	1.3	7.2
SrTiO ₃ :Ni,Mo,Al			
1.02	0.86	7.2	19
0.53	0.73	5.4	32
0.28	0.60	4.0	49
0.18	0.45	2.8	58
CaTiO ₃ :Ni,Mo			
0.81	2.20	158	63
0.56	1.44	28	59
0.30	0.80	6.3	59
0.18	0.55	3.5	70

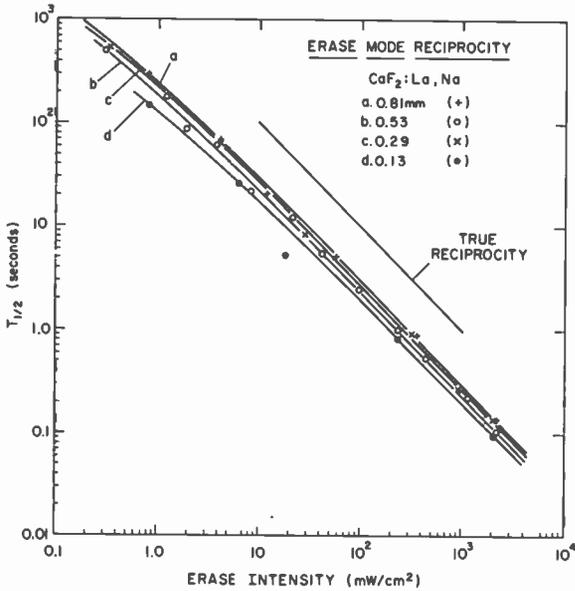


Fig. 13—Erase time $T_{1/2}$ (to $1/2 \Delta D_{max}$) versus erase-beam intensity for four wafers of $\text{CaF}_2:\text{La,Na}$.

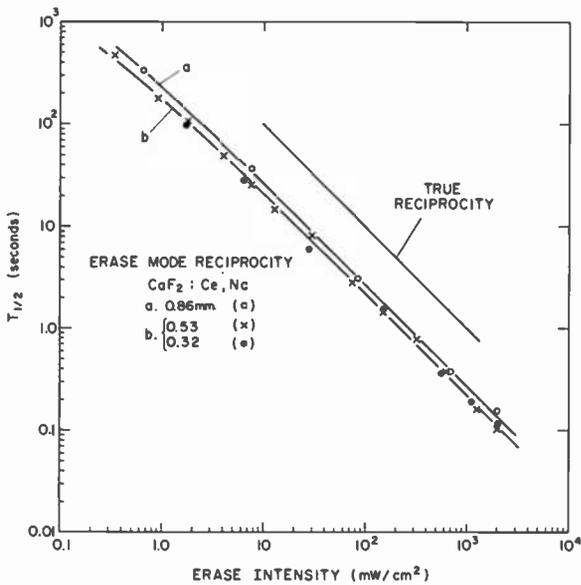


Fig. 14—Erase time $T_{1/2}$ (to $1/2 \Delta D_{max}$) versus erase-beam intensity for three wafers of $\text{CaF}_2:\text{Ce,Na}$.

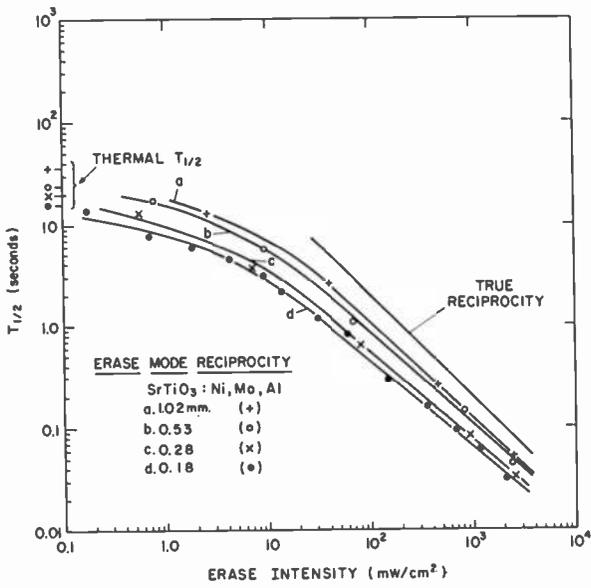


Fig. 15—Erase time $T_{1/2}$ (to $1/2 \Delta D_{\text{max}}$) versus erase-beam intensity for four wafers of $\text{SrTiO}_3:\text{Ni,Mo,Al}$.

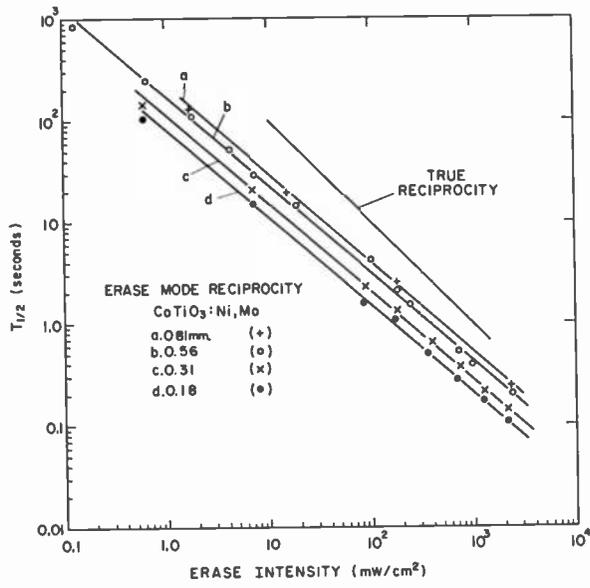


Fig. 16—Erase time $T_{1/2}$ (to $1/2 \Delta D_{\text{max}}$) versus erase-beam intensity for four wafers of $\text{CaTiO}_3:\text{Ni,Mo}$.

taken directly from families of sensitivity curves such as those shown in Figs. 1 through 4. In most cases, the range of erase intensities spans nearly four decades, in some cases more than four. The time-of-erasure used is $T_{1/2}$, the time required for the readout optical density to be reduced to half its initial (saturated) value. The "true reciprocity" straight lines in each figure have slopes of -1.0 .

3.4.1 $\text{CaF}_2:\text{La,Na}$ and $\text{CaF}_2:\text{Ce,Na}$

Figs. 13 and 14 show these reciprocity characteristics for four $\text{CaF}_2:\text{La,Na}$ wafers and three $\text{CaF}_2:\text{Ce,Na}$ wafers, respectively. For erase-beam intensities greater than about 10 mW/cm^2 , all seven wafers exhibit almost the perfect inverse relationship required for true reciprocity. Reciprocity begins to fail very gradually for smaller intensities, occurring earlier (i.e., at higher intensities) for $\text{CaF}_2:\text{La,Na}$ than for $\text{CaF}_2:\text{Ce,Na}$, and slightly earlier for the thinner wafers than for thick ones. The values of $T_{1/2}$ corresponding to the first signs of reciprocity failure are $T_{1/2} \approx 10 (\pm \times 3)$ sec for $\text{CaF}_2:\text{La,Na}$ wafers and $T_{1/2} \approx 20 (\pm \times 2)$ sec for $\text{CaF}_2:\text{Ce,Na}$. In the intensity range between 1 and 10 mW/cm^2 , i.e., for approximately the first decade of reciprocity failure, $T_{1/2}$ varies as about the inverse 0.9 power of erase intensity.

The reported values of $T_{1/2, \text{thermal}}$ corresponding to room-temperature thermal decay alone are about 2.6×10^4 sec for $\text{CaF}_2:\text{La,Na}$, about 4×10^4 sec for $\text{CaF}_2:\text{La}$, and about 4×10^5 sec for $\text{CaF}_2:\text{Ce}$.¹³ Thus, gradual reciprocity failure first begins while the ratio $T_{1/2}/T_{1/2, \text{thermal}}$ is still only about 10^{-3} or smaller. This at first seems remarkably early for thermal effects to become significant. But the thermal-erase curves for $\text{CaF}_2:\text{La,Na}$ (Fig. 5) and $\text{CaF}_2:\text{Ce,Na}$ (Fig. 6) do indeed indicate appreciable thermal decay ($\sim 5\%$) in times as short as 30 sec ($\approx 10^{-3} \times T_{1/2, \text{thermal}}$) and 40 sec ($\approx 10^{-4} \times T_{1/2, \text{thermal}}$), respectively. At least to this semi-quantitative extent, the observed reciprocity failure of the CaF_2 materials can be attributed to thermal decay processes.

3.4.2 $\text{SrTiO}_3:\text{Ni,Mo,Al}$ and $\text{CaTiO}_3:\text{Ni,Mo}$

Figs. 15 and 16 show the reciprocity characteristics for four $\text{SrTiO}_3:\text{Ni,Mo,Al}$ wafers and four $\text{CaTiO}_3:\text{Ni,Mo}$ wafers, respectively. Even at the highest erase intensities used, about 2 W/cm^2 , none of these wafers show the true time/intensity reciprocity exhibited by the CaF_2

wafers. Instead, $T_{1/2}$ varies as about the inverse 0.9 power of the erase intensity in this range. As we have already seen, however, this is not necessarily inconsistent with the CaF_2 results. There, the criterion for true reciprocity appeared to be $T_{1/2}/T_{1/2, \text{thermal}} \lesssim 10^{-3}$ to 10^{-4} . For $\text{SrTiO}_3:\text{Ni,Mo,Al}$ and $\text{CaTiO}_3:\text{Ni,Mo}$ the values of $T_{1/2, \text{thermal}}$, are much smaller, about 25 ± 10 sec and about 2×10^3 sec, respectively. Thus, even for the smallest values of $T_{1/2}$ observed, the criterion above is, at best, barely satisfied. In the case of the SrTiO_3 wafers (Fig. 15), the shortest observed optical erasure times are $T_{1/2} \approx 0.04$ sec, which corresponds to $T_{1/2}/T_{1/2, \text{thermal}} \approx 2 \times 10^{-3}$. For the CaTiO_3 wafers (Fig. 16), the smallest observed values of $T_{1/2}$ are about 0.2 sec, or $T_{1/2}/T_{1/2, \text{thermal}} \approx 10^{-4}$.

At the other extreme, reciprocity fails very badly for the SrTiO_3 wafers at erase intensities smaller than about 40 mW/cm^2 . In this range, $T_{1/2} \gtrsim 1.5$ sec and lies within about a factor of ten of its limiting value, the thermal half-life $T_{1/2, \text{thermal}}$. These limiting values are indicated for each wafer in Fig. 15. For the CaTiO_3 wafers, on the other hand, the inverse 0.9 power dependence of $T_{1/2}$ on erase intensity continues to somewhat below 10 mW/cm^2 and values of $T_{1/2} \approx 20 \pm 10$ sec. For smaller intensities, the curves of Fig. 16 also begin to flatten.

The observed reciprocity failure at the highest erase intensities is not as readily attributable to thermal-decay processes as was the onset of reciprocity failure in the CaF_2 case. The thermal-erase curves of Figs. 7 and 8 indicate that, for both the SrTiO_3 and CaTiO_3 wafers, the times required for appreciable thermal decay ($\gtrsim 5\%$) of the initial photochromic optical-density change are about an order of magnitude longer than the shortest optical-erase times observed. On the basis of these data, then, thermal decay appears to be too slow to account for the observed reciprocity failure in the corresponding range (highest decade) of erase intensities. And this, in turn, suggests the existence of some other mechanism that acts to prevent time/intensity reciprocity in this range.

However, it is just these data that are subject to the greatest experimental uncertainties. The absolute magnitude of the initial maximum-optical-density change for a given wafer (i.e., that observed in times $\lesssim 1$ sec) was frequently only reproducible to within about 5 to 10%. As a result, the detailed shapes of both the thermal- and optical-erase curves (Figs. 5 through 8) in this short time and small erasure regime also exhibited poor reproducibility.

On the basis of the present data, therefore, no firm conclusion can be drawn regarding the existence in these materials of a mechanism preventing reciprocity, other than thermal decay. Clarification of this

point could be gained by extending the measurements reported here to higher erase intensities, shorter erase times, and greater experimental precision in the intensity range in question.

One possible cause of reciprocity failure at high erase-beam intensities merits a brief comment here. While the effect of wafer heating by absorbed erase-beam energy was not specifically explored, such heating would tend to increase the thermal decay rate at high intensities and, thus, to steepen rather than flatten the reciprocity characteristics of Figs. 15 and 16.

Table 3—Erase-Mode Recording at 5145 Å

Wafer Thickness d (mm)	ΔD_{\max} at 5145 Å	γ		$E_{1/2}$ at 1 W/cm ² (mJ/cm ²)
		$\bar{\gamma}$	$\bar{\gamma}/D_{\max}$	
CaF₂:La,Na				
0.81	0.36	0.18	0.50 ± .04	290
0.53	0.25	0.12	0.48 ± .02	230
0.29	0.13	0.058	0.45 ± .03	270
0.13	0.076	0.034	0.45 ± .03	190
CaF₂:Ce,Na				
0.86	0.27	0.14	0.52 ± .03	270
0.53	0.18	0.095	0.53 ± .04	220
0.32	0.10	0.057	0.57 ± .06	220
SrTiO₃:Ni,Mo,Al				
1.02	0.90	0.57	0.63 ± .05	120
0.53	0.79	0.46	0.58 ± .04	110
0.28	0.48	0.31	0.65 ± .02	77
0.18	0.44	0.26	0.59 ± .03	65
CaTiO₃:Ni,Mo				
0.81	2.06	1.34	0.65 ± .06	530
0.56	1.43	0.97	0.68 ± .08	420
0.31	0.96	0.65	0.68 ± .04	270
0.18	0.52	0.33	0.63 ± .02	200

4. Summary and Discussion

The photochromic erase-mode recording characteristics of all fifteen wafer samples are summarized in Table 3. The maximum changes in optical density at 5145 Å, ΔD_{\max} , are those obtained from sensitivity curves similar to (and including) those of Figs. 5 through 8. In all but three cases, these values lie within 10% of the corresponding ΔD values determined from the measured photochromic absorption spectra (see Table 2). The largest difference, 27%, occurs for the thinnest CaF₂:La,Na wafer, the case subject to the greatest experimental uncertainty.

The quantity γ is defined as the slope of the nearly linear central region of a given sensitivity curve, i.e., $\gamma = \Delta D / \log(\text{exposure})$ in that region. As was pointed out earlier, the sensitivity curves for each wafer approach a relatively fixed shape with constant γ as the erase-beam intensity increases above about 1 mW/cm². The quantity $\bar{\gamma}$ in Table 3 denotes, for each wafer, the average value for only these erase intensities. It is clear that for each material the values of $\bar{\gamma}$ are approximately proportional to ΔD_{max} . The ratio $\bar{\gamma} / \Delta D_{\text{max}}$, listed in the fourth column of Table 3, appears to be a characteristic of a given photochromic material. Even the variations in $\bar{\gamma} / \Delta D_{\text{max}}$ among the four different materials are not large. In practical terms, γ can be adjusted somewhat for a particular application by varying wafer thickness, and thus ΔD_{max} . Similar adjustments in ΔD_{max} and γ can be achieved by varying the readout wavelength.

Dependence of γ on ΔD_{max} is not unexpected. At smaller values of ΔD_{max} , a smaller fraction of the incident radiation is absorbed and the erasure would be expected to proceed more slowly. An equation for the transmission of a photobleachable material as a function of bleaching exposure has been developed by Kessler²⁰ (and, independently, by Staebler²¹). The model on which the equation is based is not applicable in complete detail to the experiments reported here. The equation can be used, however, to show a dependence of γ on ΔD_{max} that, in a limited regime, might be approximated by $\gamma / \Delta D_{\text{max}} = \text{constant}$. The fact that this constant varies little among the four materials examined is indicative of the fundamental similarities in the photochromic processes.

The right-hand column of Table 3 shows the erase exposure $E_{1/2}$ required, at an erase beam intensity of 1 W/cm², to reduce the readout optical densities to half their initial (saturated) values. Clearly the reciprocal of $E_{1/2}$ is a measure of erase-mode sensitivity. The $E_{1/2}$ values tabulated are the products of the appropriate abscissae and ordinates of the various reciprocity characteristic curves in Figs. 13 through 16. (It should be noted that these are incident exposures, not absorbed exposures.) Values of $E_{1/2}$ for other beam intensities can be readily determined in similar manner. For each material, the values of $E_{1/2}$ exhibit a dependence on ΔD_{max} , although not a consistent one. For SrTiO₃:Ni,Mo,Al the two are very nearly proportional, while for the CaF₂ materials, the dependence is much weaker. Qualitatively, this dependence is again consistent with the arguments given above for the variation of γ with ΔD_{max} and with the equation derived by Kessler.²⁰

For each material, Fig. 17 shows $I/E_{1/2}$ plotted as a function of

ΔD_{\max} (See Table 3). For a particular application requiring a specific contrast ratio or ΔD_{\max} , this graph gives an indication of the relative sensitivities of the four materials. Thus, wafers of $\text{CaF}_2:\text{La,Na}$ and $\text{CaF}_2:\text{Ce,Na}$ of comparable ΔD_{\max} exhibit comparable sensitivities. For $\Delta D_{\max} = 0.4$, the $\text{SrTiO}_3:\text{Ni,Mo,Al}$ is about twice as sensitive as the $\text{CaTiO}_3:\text{Ni,Mo}$ and five times as sensitive as the CaF_2 materials.

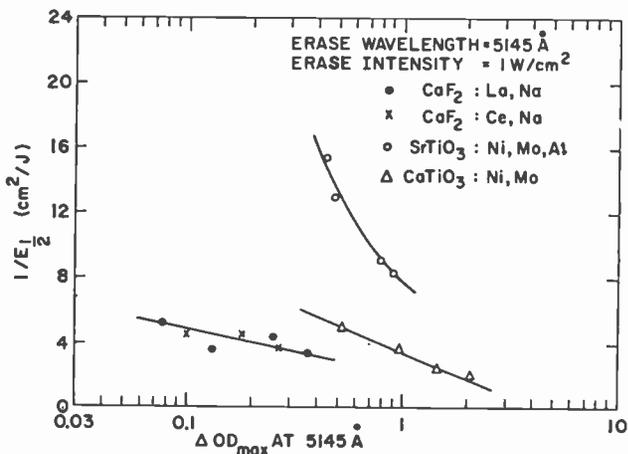


Fig. 17—Reciprocal of erase exposure $E_{1/2}$ (to $1/2 \Delta D_{\max}$) as a function of ΔD_{\max} for the several wafer samples of four materials.

Acknowledgments

The author gratefully acknowledges numerous helpful discussions with B. W. Faughnan and D. L. Staebler and the valuable experimental assistance of R. Infanti.

References:

- ¹ G. H. Brown and W. G. Shaw, "Phototropism," *Rev. Pure Appl. Chem.*, Vol. 11, p. 2 (1961).
- ² R. Exelby and R. Grinter, "Phototropy (or Photochromism)," *Chem. Rev.*, Vol. 65, p. 247 (1965).
- ³ J. J. Amodel, Chapter 20 in *Handbook of Lasers*, R. J. Pressley, ed., Chemical Rubber Company, Cleveland, 1971.
- ⁴ Z. J. Kiss, "Photochromic Materials for Quantum Electronics," *IEEE J. Quantum Electron.*, Vol. QE-5, p. 12 (1969); "Photochromics," *Phys. Today*, Vol. 23, p. 42 (1970).
- ⁵ B. W. Faughnan, D. L. Staebler, and Z. J. Kiss, "Inorganic Photochromic Materials," *Applied Solid State Science*, Vol. 2, R. Wolfe, ed., Academic Press, Inc., New York, 1971.
- ⁶ D. L. Staebler and Z. J. Kiss, "Photo-Reversible Charge Transfer in Rare-Earth-Doped CaF_2 ," *Appl. Phys. Lett.*, Vol. 14, p. 93 (1969).

- ⁷ D. L. Staebler, S. E. Schnatterly, and W. Zernik, "Linear Dichroism in Photochromic CaF_2 ," **IEEE J. Quantum Electron.**, Vol. QE-4, p. 575 (1968).
- ⁸ R. C. Duncan, Jr., B. W. Faughnan, and W. Phillips, "Inorganic Photochromic and Cathodochromic Recording Materials," **Appl. Opt.**, Vol. 9, p. 2236 (1970).
- ⁹ W. Phillips and R. C. Duncan, Jr., "Preparation of Photochromic Calcium Fluoride by Additive Coloration," **Metallurg. Trans.**, Vol. 2, p. 769 (1971).
- ¹⁰ D. L. Staebler and S. E. Schnatterly, "Optical Studies of a Photochromic Color Center in Rare-Earth-Doped CaF_2 ," **Phys. Rev. B**, Vol. 3, p. 516 (1971).
- ¹¹ C. H. Anderson and E. S. Sabisky, "EPR Studies of Photochromic CaF_2 ," **Phys. Rev. B**, Vol. 3, p. 527 (1971).
- ¹² R. C. Alig, "Theory of Photochromic Centers in CaF_2 ," **Phys. Rev. B**, Vol. 3, p. 536 (1971).
- ¹³ D. L. Staebler and R. C. Duncan, Jr. (to be published).
- ¹⁴ B. W. Faughnan and Z. J. Kiss, "Photoinduced Reversible Charge-Transfer Process in Transition-Metal-Doped Single-Crystal SrTiO_3 and TiO_2 ," **Phys. Rev. Lett.**, Vol. 21, p. 1331 (1968).
- ¹⁵ B. W. Faughnan and Z. J. Kiss, "Optical and EPR Studies of Photochromic SrTiO_3 Doped with Fe/Mo and Ni/Mo," **IEEE J. Quantum Electron.**, Vol. QE-5, p. 17 (1969).
- ¹⁶ J. J. Amodei, Ph.D. Thesis, Univ. of Pennsylvania, 1968.
- ¹⁷ B. S. Faughnan, "Photochromism in Transition Metal Doped SrTiO_3 ," **Phys. Rev. B**, Vol. 4, p. 3623 (1971).
- ¹⁸ B. W. Faughnan, private communication.
- ¹⁹ L. Merker, "Synthesis of Calcium Titanate Single Crystals by Flame Fusion Technique," **J. Am. Ceramic Soc.**, Vol. 45, p. 366 (1962); also private communication.
- ²⁰ H. C. Kessler, Jr., "The Kinetics of Light Absorption in Photobleaching Media," **J. Phys. Chem.**, Vol. 71, p. 2736 (1967).
- ²¹ D. L. Staebler, unpublished.

High Contrast, High Sensitivity Cathodochromic Sodalite for Storage and Display Applications

B. W. Faughnan and I. Shidlovsky

RCA Laboratories, Princeton, N. J.

Abstract—Recent developments in high contrast, thermal erase, cathodochromic materials are presented. The distinction between photochromism and optically reversible and thermally reversible cathodochromism is clarified. The synthesis and sensitization of hydrothermal sodalite is discussed. Experimental results in the form of contrast ratio versus electron beam flux are presented for a wide variety of bromide sodalite powders. Measurements were made in both reflection and transmission modes on both loose powders and thin settled slides. Data is presented on the sensitivity of the materials as a function of material preparation, particle size, method of measurement, and electron-beam voltage. With the most sensitive materials, a CR of 10:1 is achieved for ϕ less than $15 \mu\text{C}/\text{cm}^2$, measured in transmission using a narrow-band interference filter at the bromide absorption peak. Under identical conditions a CR in excess of 100:1 is obtained for a flux of several hundred $\mu\text{C}/\text{cm}^2$. However, the most sensitive materials also show significant room temperature decay after several minutes.

The intrinsic absorption coefficient k of a settled sodalite screen is measured by optical transmission using an index matching technique; for CR = 10:1, $k = 1000 \text{ cm}^{-1}$ and for CR = 100:1, $k = 4000 \text{ cm}^{-1}$. The effective gain g of a cathodochromic screen is defined by the equation $\text{CR} = e^{gkD}$, where D is the electron beam penetration depth, and CR is measured in transmission. g is found to be a monotonically decreasing function of CR with $g = 5.4, 4.1,$ and 2.3 for CR = 1:1, 10:1, and 100:1, respectively. The amount by which g exceeds unity represents the enhanced absorption introduced by multiple scattering inside and between the multi-layered particle screen.

1. Introduction

The interest in cathodochromic materials for storage and display tubes has increased recently¹⁻⁸, mainly because of the unusual cathodochromic and photochromic properties of sodalite. The application of materials with color centers induced by electron beam irradiation was first studied intensively during World War II.⁹ The need was for a long-persistence radar screen and the material chosen was KCl. KCl has many short-comings as a cathodochromic material, however, and interest in cathodochromic (also called "dark trace") tubes in general declined.

Since the first paper on the photochromic properties of sodalite appeared¹⁰ the material has been increasingly studied,¹¹⁻¹³ with most of the recent emphasis on the cathodochromic properties of sodalite for use in cathodochromic cathode-ray tubes.

Sodalite represents a considerable improvement over KCl in terms of sensitivity, contrast ratio achievable, and variety of light bleaching and thermal decay rates. Cathodochromic sodalite, as developed at a number of laboratories, can achieve a contrast ratio CR of approximately 2 or 3 to 1 at an electron beam flux of $\phi = 1-3 \mu\text{C}/\text{cm}^2$ using 20 to 30 kV electrons.^{2,3,*} The early material was doped with sulfur,^{3,11} since that was found to aid its photochromic properties. Fe doping was also employed. It was found independently by a number of workers,^{3,8} including ourselves, that doping is not necessary for good, optically reversible, cathodochromic properties. It is necessary however, for the photochromic effect, and can effect the light bleaching properties of optically reversible cathodochromism.⁸ In some cases, doping may degrade the cathodochromic performance. For example sulfur is known to diffuse through sodalite at elevated temperatures forming a variety of sulfur radicals.¹⁴ The "fatigue", or loss in sensitivity, always found in sulfur doped sodalite must somehow be associated with the motion of sulfur ions during irradiation, although the details of this process are not understood. It should be mentioned, however that even some undoped forms of sodalite show fatigue.⁸

* Ref. [8] reports an order of magnitude better than this figure for reflection contrast ratio on thick settled slides, i.e. CR = 2:1 at 20 KV for $\phi = 0.15 \mu\text{C}/\text{cm}^2$. However we show in Ref. [15] that our measured initial sensitivity corresponds to an F' center production efficiency of $\eta_{op} = 80 \text{ eV}/F'$ center. This is a very reasonable number since it takes at least three bandgaps of energy to create an electron hole pair ($\sim 25 \text{ eV}$) and we do not expect 100% efficiency for the trapping of electron hole pairs. η_{op} is somewhat less (20-40%) than the corresponding number found for KCl at room temperature in the fast stage of coloration (Ref. [20]). The initial efficiency in Ref. [8] would have to be 8 eV/ F' center which does not seem possible.

One form of fatigue that all sodalites manifest in the optically reversible mode of operation is the gradual build up of coloration (F band absorption) which cannot be bleached by light. We have recently shown¹⁵ that this coloration results from a different physical mechanism i.e., from an ionic displacement which creates a new vacancy, rather than merely from the trapping of an electron by an already existing vacancy. This situation is analogous to the early- and late-stage coloration in alkali halides.¹⁶ Not surprisingly, it takes more energy to create the vacancy than just to trap an electron in an existing vacancy. Nevertheless, by materials improvement, it has been possible to increase the efficiency of this process to the point where the device applications for this mode of coloration look very promising. A significant bonus for operation in this mode (called the thermal erase mode, since it can only be erased or bleached by heat) is that the density of centers obtainable is increased by more than an order of magnitude. Other advantages have already been presented,⁵ not the least of which is the absence of any fatigue in this mode of operation.

Some of the materials to be described here have already been incorporated in working devices⁵ and systems.⁸ In this paper we will describe some of their properties in detail. In addition significant improvements in material quality and understanding have been made.

2. Model for Cathodochromic Coloration in Sodalite

The nature of the cathodochromic coloration in sodalite has been studied by a variety of physical and chemical techniques. Recently, the distinction between the various types of coloration in sodalite have been clarified, and a model has been proposed for the nature of the high-contrast thermally-erasable coloration in sodalite.¹⁵

To understand cathodochromic coloration in sodalite, we have to distinguish three different modes of coloration—(1) photochromic, (2) optically reversible cathodochromic, and (3) thermally reversible cathodochromic. The three modes refer to different physical mechanisms for creating the coloration, but the actual color centers are identical in each case, i.e., an F center or an electron trapped at a halogen vacancy site.¹²

Photochromism was the first of the three mechanism's observed in sodalite.¹⁰⁻¹² It results from a light-induced transfer of charge from an electron donor to an already existing halogen vacancy. Sodalite has been doped with sulfur¹¹, iron³, and selenium and tellurium³ to supply the necessary donors. *Optically reversible cathodochromism*

also results from the trapping of an electron in a halogen vacancy site. But in this case, many electron hole pairs are created by the primary electron beam. Doping with donor impurities is not necessary in this case, i.e., sodalite can exhibit optically reversible cathodochromism without being photochromic. Finally, we have shown that *thermally reversible cathodochromism* results from the ejection of an ion from inside a normal aluminosilicate cage to a defect cage.¹⁵ A defect cage in sodalite is one which is missing either a Na ion, a halogen ion, or both. The defect cages are produced during the sensitization process that removes NaBr from the lattice.

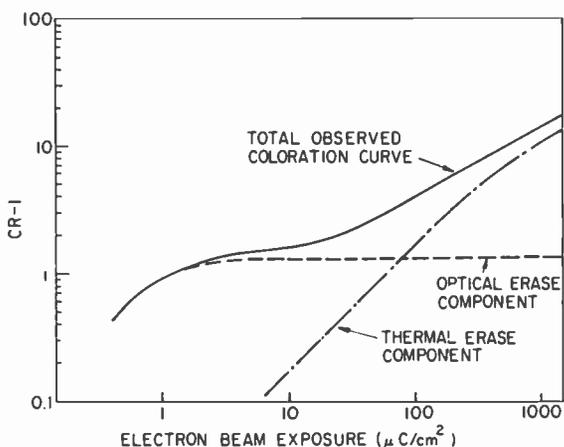


Fig. 1—Diffuse reflectance contrast ratio CR versus electron beam exposure ϕ for a bromide sodalite powder (electron-beam voltage $V = 20$ KV).

A cathodochromic coloration curve of sodalite is illustrated in Fig. 1. Contrast ratio (minus one) is plotted versus electron beam exposure. The solid line is the experimentally observable coloration curve. It consists of an optical erase component, which is sensitive initially, but tends to saturate at a contrast ratio of about 2.5:1. The contrast ratio then continues to increase, although at a slower rate, as a result of the thermal erase component. The dotted curves break up the experimental curve into its optical and thermal parts. In very sensitive material the two component nature of the coloration is no longer obvious, although its existence can still be verified by other means, such as by optical bleaching or heating, which affect the two components differently.

3. Material Preparation

Bromide sodalite ($\text{Na}_6\text{Al}_6\text{Si}_6\text{O}_{24} \cdot 2\text{NaBr}$) is one representative of the sodalite group of materials. Fig. 2 shows the cage-like structure of sodalite. A three-dimensional aluminosilicate structure encloses each Br ion and its four nearest neighbor tetrahedrally coordinated Na ions. Sodalite has body centered cubic symmetry and the cages are

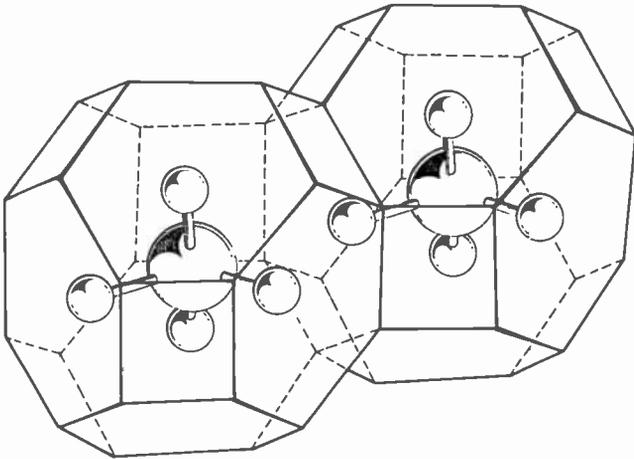


Fig. 2—Schematic diagram of the sodalite cage structure. Large circles on the center of the cages represent halogen ions. The smaller circles represent Na ions. The solid lines outline two attached aluminosilicate cages. The Al, Si, and O ions forming the cage are not shown for clarity.

closely packed. In this respect the sodalites differ from all other zeolite materials, which can have other types of cages in addition to the sodalite cage. In sodalite, the bromine can be completely or partially replaced by chlorine, iodine or OH, and partially by Group VI elements. The sodalite structure is maintained in nonstoichiometric form of the material for example, when some of the alkali and halogen atoms are missing. Maintenance of this structure is an important fact relating to the cathodochromic and photochromic properties of the sodalites.

Alkali halide sodalite, mainly in the chloride and hydroxy forms, have been previously prepared by hydrothermal methods.^{11,14,17,18} All the reported preparations have been done in basic (NaOH) solutions, and the product was always found to contain some fraction of hydroxy groups on the halogen sites. Bye and White¹⁸ have reported the growth

of large single crystals (> 5mm) of chloride sodalite, but the stoichiometry was found to be $\text{Na}_6\text{Al}_6\text{Si}_6\text{O}_{24} \cdot (1.4\text{NaCl}_{0.6}\text{NaOH})$. In the references cited above, the main emphasis was to produce sodalite in crystalline form with the typical dodecahedral shape. To this end, the alkali solution was used as a complexing media to help in transporting the material along a thermal gradient. But this was also the source for the hydroxy/contamination.

As grown, the sodalites are neither photochromic nor cathodochromic. Heat treatment is needed to sensitize the materials and that is accompanied by weight loss. The ease of sensitization is influenced by the crystallinity. Ballentine and Bye¹⁷ reported that after prolonged treatment at elevated temperatures in vacuum, they were not able to sensitize large single crystals of chloride-hydroxy sodalite. They did find, however, that it was easy to sensitize 15-25 μm crystallites, which suggests that the rate of weight loss from sodalite is influenced by the surface area and the crystallinity. The sensitization treatment is usually carried out in a hydrogen atmosphere. Chemical reduction is needed to create the electron donor centers responsible for the photochromic and, possibly, the optical erase cathodochromic properties. At the same time, halogen is removed from the sodalite cages creating vacancies which then become F' centers after trapping an electron. None of the previous work dealt with thermal erase cathodochromic sodalite. In this case it is necessary to remove an even larger quantity of sodium halide to create the defect cages necessary for the ionic-displacement mechanism. In the case of optically reversible cathodochromism, some success in limiting the thermal mode coloration has been obtained by putting sodium back into the lattice after the sensitization treatment.^{2,8}

We have attempted to develop and define the best thermal erase cathodochromic bromide sodalite. To accomplish this goal the previously used conditions of hydrothermal growth of sodalite had to be modified.

The general reaction for sodalite bromide synthesis is as follows:



Starting materials are reagent grade chemicals: NaOH, NaBr, SiO₂ and Al₂O₃. The charge for the hydrothermal reaction is prepared by mixing stoichiometric amounts of the chemicals in water, plus excess of NaOH and NaBr.

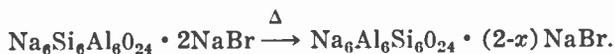
The charge is reacted in a high pressure bomb for about seven days. Typical reaction temperature and pressure are 400°C and 15,000 psi. All the reaction parameters are interrelated in their affect on

the final product, and the optimum condition must be found experimentally.

At the end of the reaction, the reaction vessel is cooled to room temperature. The powder forms loose aggregates in the container, which are separated in the washing procedure. The excess of NaOH and NaBr are washed out in water, and the sodalite, in the form of fine white powder, is dried and ready for sensitization.

When synthesizing doped or substituted sodalite bromide, a similar preparation procedure is applied. Ga and Ge are introduced as oxides; Group VI elements, such as S, Se and Te, are reproduced as the oxyanion XO_4 , which is later reduced with hydrogen to X^- .

In the sensitization process, NaBr is diffused out from the sodalite structure,



x can be as high as 0.3, which is equivalent to a 15% deficiency in NaBr. The removal of the NaBr is done without any phase transformation, as confirmed by x-ray analysis.

Careful examination of the x-ray diffraction patterns of the sodalite before and during different steps of the heat temperature shows that, during the heat treatment, the unit cell expands by 0.2%. In as-grown material, the unit cell is 8.932 ± 0.005 Å. The change in the unit cell is not linear, but occurs after a weight loss of about 1%, which corresponds to 5% of the NaBr in sodalites. This result is in agreement with fact that the coloration properties up to 1% weight loss are less than expected for the corresponding number of defects. However, an exact correspondence has not been shown and further investigation of this result is needed. One possibility is that the initial loss is from the surface, where the density of defects formed is higher than it is deeper inside the particles.

4. Contrast Ratio Measurements

Experimental Techniques

The measurement method is very important in the evaluation of cathodochromic powders. The contrast ratio versus electron-beam flux of one material can be superior to another in one method of measurement and inferior in another. Since the primary application is for targets for cathodochromic dark trace tubes, one might suppose that only the

contrast ratio of settled targets in actual tubes should be measured. This approach would be impractical and very time consuming. Also, it would not differentiate between intrinsic sensitivity of the powder, particle size, film thickness, and the effect of tube processing on materials. The approach we have followed is to use a variety of techniques that accentuate the importance of different parameters until our knowledge of the whole system is built up. Finally, one technique can be standardized when all the variables are under control.

Ideally, we would like to measure the intrinsic absorption versus the total electron beam flux k versus ϕ , the so-called "growth curve". This curve would describe the coloration property of the material completely. Then, the contrast ratio would depend on this k -versus- ϕ , curve as well as on particle size, screen weight, etc. This measurement will be discussed later. In this section only the contrast ratio measurements are considered.

The diffuse contrast ratio, is defined as $CR = R_o/R$ for reflection, or I_o/I for transmission where R_o and I_o are the initial diffuse reflection and transmission, respectively, and R and I are the final diffuse reflection and transmission, respectively, after electron-beam coloration. The contrast ratio was measured with a spot-meter ($1/2^\circ$ acceptance angle) and sometimes with a narrow band (100 Å) interference filter at the bromide sodalite absorption peak. The beam was deposited in the form of a square television raster, typically 1×1 cm. The part of the screen not colored by the electron beam was masked out so that very high contrast ratios ($> 20:1$) could be measured accurately. A schematic diagram of the experimental setups is shown in Fig. 3. Contrast ratio versus ϕ measurements are made by three different methods. In two of these methods, CR is measured continuously with the electron beam on, and in one point-by-point data is taken.

(1) *Reflectance measurement on settled slides* (method I) is the most unreliable and gives the lowest values for the contrast ratios. It is also the most time consuming, since a complete slide must be settled and mounted for each material. The electron beam strikes the back or powder side of the slide which has an evaporated Al layer, approximately 500 Å thick, that serves to draw off the current. The CR is measured from the front side through the glass or mica substrate.

This method poses certain problems. CR is a very sensitive function of slide thickness t . If $t < D$, where D is the penetration depth of electron beam, electron-beam energy is wasted. If $t \gg D$, the front surface of the slide never gets colored, yet this surface is most heavily weighted in the CR measurement. Also, CR is a sensitive function of

powder particle size which must be accounted for. Our early measurements were performed using this method, but we shall not discuss it further here.

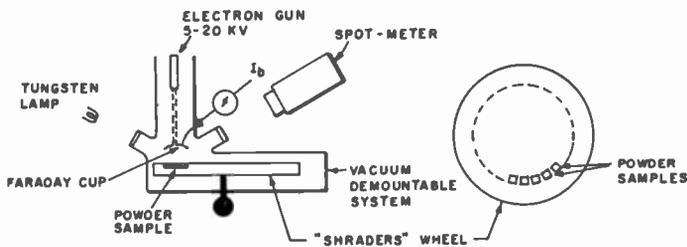
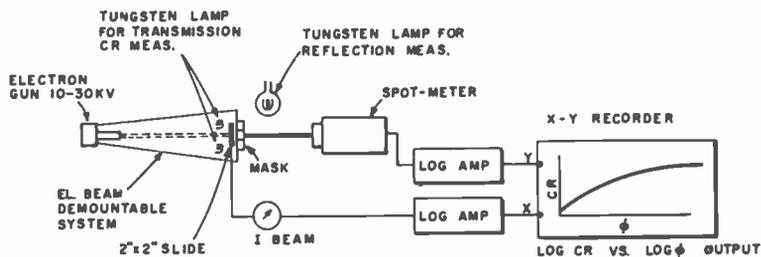


Fig. 3—Schematic diagram of experimental apparatus used for measuring contrast ratio versus electron-beam exposure of sodalite powders. The log amplifiers and x-y recorder shown for settled-slide measurements are used in method III to obtain a plot of log CR versus log ϕ directly.

(2) *Reflection measurements on loose powders* (method II) has several advantages. (1) CR is measured on the same side as the electron beam. Therefore, the best value is obtained and does not depend on the thickness of the powder. (2) As many as 30 different samples can be mounted at one time and measured during the same run under identical conditions. (3) Settled slides do not have to be made.

Because of (2) and (3) this method is very fast and convenient. It is very suitable for a rapid survey of a wide variety of differently prepared materials. The disadvantage is that the results cannot be applied directly to a cathodochromic cathode-ray-tube configuration. One reason is that loose powder may behave differently from a settled slide, and a second is that the reflection and the transmission CR are not exactly the same.

(3) *Transmission measurements on settled slides* (method III) combines advantages and disadvantages of the first two methods. It is ideally suited to the final phase, where detailed measurements on a few "best" materials are desired in a configuration close to the actual tube operation. The main points about this method are as follows. (1) High CR is obtained. There is no loss of sensitivity because of viewing from the side opposite to where the electron beam strikes. (2) The measurement is practically independent of the thickness of the slide. (3) It is found experimentally that CR is essentially independent of particle size. (4) The method is not as fast as Method II. However, if the heat erasable In_2O_3 coated mica substrate is used⁵, it is possible to thermally erase the slide. This allows one to make many runs on the same slide for detailed investigation.

To summarize, we found increasingly higher contrast ratios in going from Method I \rightarrow II \rightarrow III, even for the same material. It is interesting to note that in the published literature, Phillips² uses Method I, Bolwijn et al³ use Method II (but with a deposited thick layer), and Taylor et al⁴ use Method III.

Detailed Experimental Results

Method II

Many different materials were evaluated using this method. For clarity, only representative groups will be shown to illustrate the main conclusions.

Fig. 4 shows the total contribution to the CR, i.e., optical mode plus thermal mode. The upper group of points shows the CR-minus-one versus \emptyset curve for eight runs on six different materials that were among the most sensitive of all materials tested. All these runs have the same sensitivity within the experimental error. We concluded from this data that (1) there is no effect on cathodochromic sensitivity whether the material is undoped, Fe doped, or Se doped and (2) the main correlation with sensitivity is method and conditions of growth.

Although some dopants have no effect on material sensitivity, others have a deleterious effect. Specifically, the addition of sulfur, gallium, or calcium greatly reduces the thermal mode sensitivity, as shown by the lower curve in Fig. 4.

The effect of electron-beam voltage on reflection contrast ratio is shown in Fig. 5. As expected, higher beam voltages produce higher contrast ratios. This effect occurs not primarily because of the increased energy carried by the electron, but rather because of the

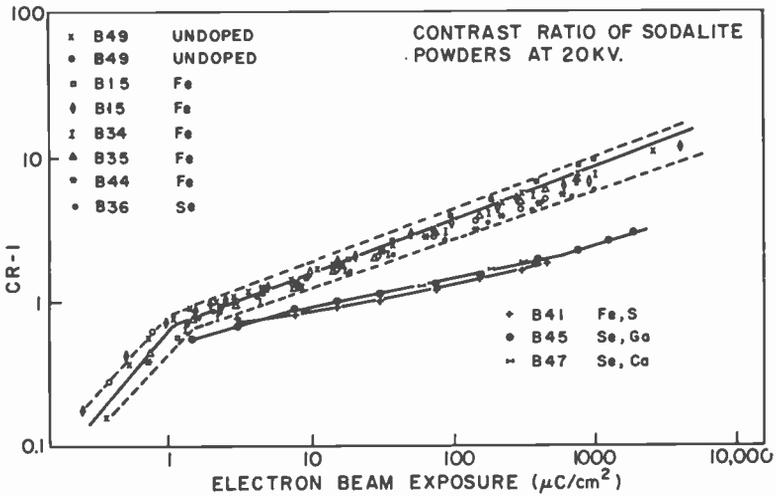


Fig. 4—CR minus one versus ϕ for loose sodalite powders ($V = 20$ KV).

increased penetration depth. This means that the energy density/cm³ for a given flux is actually less at higher electron beam energies and so the color center production rate per electron is higher. The increased rate occurs because of the nonlinear or saturating nature of the coloration curve. Actually, the increase in CR in going from 10 to 20 kV is less than might be expected, probably because in going

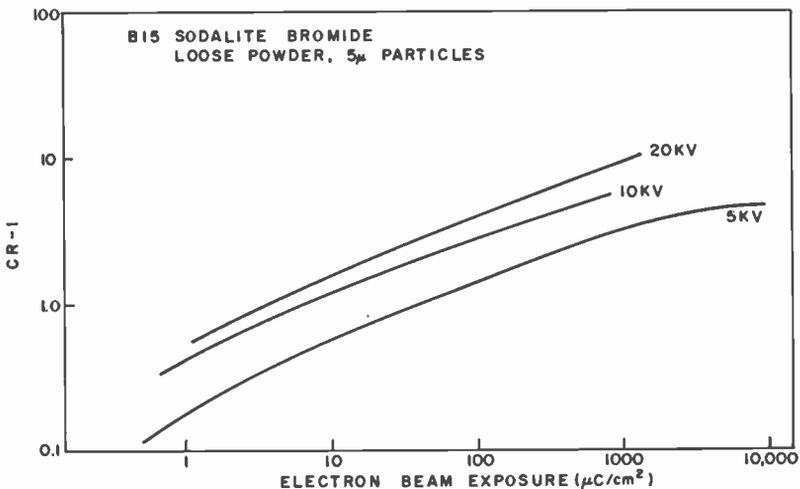


Fig. 5—CR minus one versus ϕ for loose sodalite powders showing effect of varying electron-beam voltage.

from 10 to 20 kV the penetration depth for electrons goes from less than the particle diameter to just over that. Thus, the colored region is not quite as effective as it would be for larger particles, since some light gets scattered out of the material after passing through only one layer. It is interesting that only at 5 kV and $\phi > 3000 \mu\text{C}/\text{cm}^2$ is it possible to approach saturation of the coloration. The effect of particle size is shown in Fig. 6 where contrast ratio for a given flux

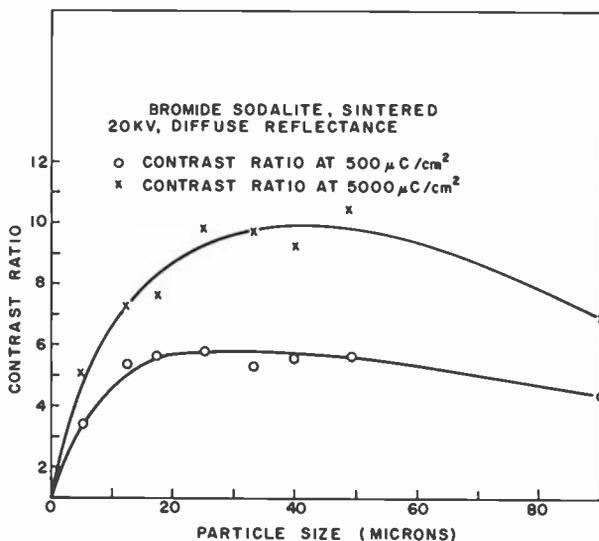


Fig. 6—Contrast ratio versus particle size at fixed electron-beam flux for a sintered sodalite powder.

is plotted versus particle size of a loose powder. These measurements were actually made on a sintered sodalite material, but that does not affect the nature of the results. We see that CR increases with particle size up to about $15 \mu\text{m}$ and then remains fairly flat out to $90 \mu\text{m}$. For particle diameters less than $5 \mu\text{m}$, the reduction in CR is very great in the reflection mode.

Method III

Now consider the transmission slide contrast ratio measurements. This method yields the highest values of contrast ratio. Since the preliminary survey work had already been done, only the most promising materials were studied by this method. CR-versus- ϕ curves for one such powder are shown in Fig. 7 for a series of electron voltages. They

show considerably better sensitivity than any of the preceding figures. A contrast ratio of 10:1 is reached after approximately $40 \mu\text{C}/\text{cm}^2$ of flux at 25 kV. All the data on transmission slides were taken using a 100-Å interference filter, centered at the bromide absorption peak. This improved the sensitivity by approximately a factor of two.

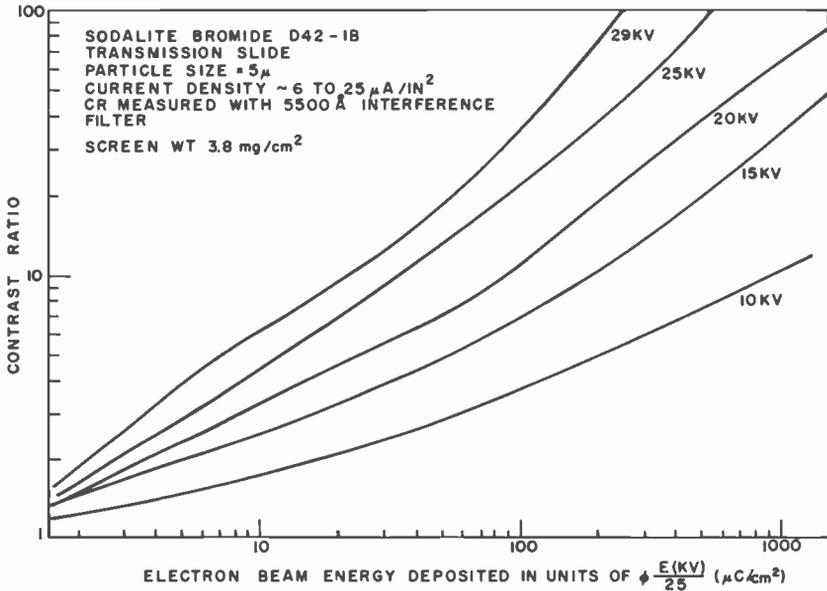


Fig. 7—CR (transmission) versus ϕ of bromide sodalite for different electron beam voltages.

The sodalite screen density varied between $4\text{--}6 \text{ mg}/\text{cm}^2$, the higher density corresponding to larger powder particle sizes. Since sensitized sodalite has an approximate density of $2.36 \text{ gm}/\text{cm}^3$,³ several layers of material were deposited on the slides.

The contrast ratio is much less sensitive to particle size in the transmission mode. When a particle size is specified, say between $20\text{--}30 \mu\text{m}$ for example, what is meant is that the sodalite from which the slide is made passes through a 30μ sieve, but not through a 20μ sieve. Scanning-electron-microscope pictures reveal that such a powder is actually made up of aggregates of smaller sized particles. This suggests that the intrinsic material properties of the powder are independent of aggregate size, but the light scattering properties may not be. A series of 40 runs was performed at 25 kV using 8 different slides made up from three different but similarly grown materials and with

particle sizes $< 10\mu\text{m}$, $10\text{-}20\mu\text{m}$, and $20\text{-}30\mu\text{m}$. All CR-versus- ϕ curves came out close to the 25 kV curve of Fig. 7 (within experimental error). This insensitivity to particle size and screen weight makes this technique ideally suited for studying the effects of more subtle variations in growth conditions.

5. Intrinsic Absorption and Effective Gain of a Sodalite Screen

The study of the optical properties of cathodochromic sodalite and the understanding of their device performance is hampered because one must work with powders. In practice this usually limits measurements to contrast ratio versus flux. But contrast ratio depends on the method of measurement, powder size, reflection versus transmission, and other factors. It would be desirable to have a direct measurement of the intrinsic absorption coefficient $k(\text{cm}^{-1})$.

k versus ϕ , the so-called growth curve, is an intrinsic property of the material alone. From it, one can determine such important quantities as initial efficiencies for the optical and thermal erase modes, and their respective saturation behavior. In addition to being important for predicting device performance, these experimental quantities are the ones needed for comparison with any theoretical model. From Smakula's equation, the density of color centers can be determined.⁵ Finally, if both k versus ϕ for a cathodochromic material and CR versus ϕ for a cathode-ray-tube screen constructed from the same material are known, the effective gain g of the screen, defined by the equation $\text{CR} = e^{gkD}$ (D , as before, is the electron penetration depth), can be computed. This concept is very useful since it relates the contrast ratio, which is the quantity important for device applications, to the product of three independent quantities, g , k , and D . The effective gain of the screen g is a function of method of observation (transmission versus reflection), particle size, packing density and thickness. The absorption coefficient k is a property of the material and the electron-beam energy flux and, of course, differs for optical and thermal modes in sodalite. Finally, the penetration depth for electrons D is a function of electron energy and material density. Actually D corresponds to the usually defined range for electrons for the thermal mode coloration, but may be longer for the optical erase mode (this point will be discussed later). The gain g also has a simple physical meaning. For a specular screen, e.g., single crystal or film with no particle scattering, $\text{CR} = e^{kD}$ in the transmission mode and $\text{CR} = e^{2gkD}$ in reflection when the viewing light passes through the colored material twice (reflective backing). Therefore, for a specular screen, $g_t = 1$,

$g_r = 2$. The magnitude of g found for a diffuse (particle) screen can be compared with these values. The higher g found, results from the extra path length the light rays travel, on the average, through the scattering medium. Total internal reflection inside a high index particle will be especially effective in increasing the effective gain.

Intrinsic Absorption

A variety of experimental techniques must be employed to obtain a complete k -versus- ϕ curve over the wide range of k (30 to 15,000 cm^{-1}) found in cathodochromic sodalite. These include (1) use of high energy electrons (300 Kev Van de Graaff) to irradiate a large volume for the low- F -center density region, (2) irradiation at 20 KV using the demountable system described previously to cover the high-absorption region. The absorption coefficient was determined in two ways—first, by calibrated EPR measurements of the number of F centers created and, second, by direct optical transmission measurements on a Cary 14 Spectrophotometer, using an index-matching technique to remove particle scattering. A detailed discussion of this experiment is given elsewhere.¹⁵ From these measurements, the efficiencies for creating F centers can be obtained. They are $\eta_{op} = 80 \text{ eV}/F\text{-center}$ and $\eta_{th} = 500 \text{ eV}/F\text{-center}$ for the optical erase and thermal erase modes, respectively. The latter number is typical for the most sensitive materials. Much higher values for η_{th} will, of course, be found for those materials that show little thermal mode coloration. Values of $\eta_{th} > 5 \times 10^4 \text{ eV}/F\text{ center}$ are possible. An example of a direct optical transmission measurement at 20 KV is shown in Fig. 8. This curve is measured point by point with the same electron-beam apparatus that was used to obtain the contrast ratio measurements. That is, a sodalite slide is irradiated to a certain flux ϕ_1 , removed from the demountable, and index matched, and the optical density through the colored raster is measured. The slide is returned to the demountable system and erased, and a second flux ϕ_2 is obtained. This cycle is repeated until a complete curve is traced out. The initial shape of the curve depends on whether the optical erase component is included, which in turn depends on the temperature of the slide. The optical component can be removed by heating the substrate or by employing a large current density.

Calculation of the Effective Gain "g"

The effective gain is easily obtained from the experimental procedure previously outlined for measuring k versus ϕ . For each flux point

$\phi_1, \phi_2 \dots$ the corresponding contrast ratio is measured before the slide is removed from the electron beam apparatus. Then in the equation $CR = e^{gkD}$, CR, k and D are all known quantities. One source of error is that CR decays somewhat with time after the beam is removed. To allow for this variation, the CR decay is observed for a few minutes before removing the slide. Any further decay in the 10-20 minutes it takes to complete the transmission measurement is negligible.

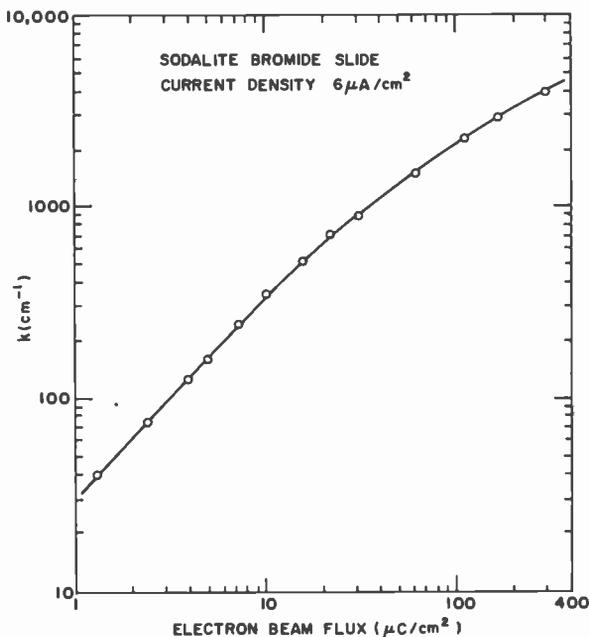


Fig. 8—Intrinsic absorption coefficient k versus ϕ for bromide sodalite: $V = 25$ KV; thermal erase mode only.

As shown in Fig. 9, the effective gain (transmission mode) is a smooth monotonically decreasing function of the contrast ratio. We have verified experimentally, at least in the thermal erase mode, that g does not depend on the electron-beam voltage, current density, or particle size (at least in the $5\text{-}40\mu\text{m}$ range). A detailed study of the gain in the reflection mode has not been carried out, but our earlier results show that it increases with particle size up to particle diameters of about $10\mu\text{m}$ and remains constant thereafter at a value that is always below the transmission gain.

From the known values of g and D and k versus ϕ , it is possible to

construct a family of curves of CR versus ϕ for different electron-beam voltages. Fig. 7 shows a family of such curves obtained experimentally. To obtain agreement with experiment in the low contrast ratio region, we find that it is necessary to assume that the optically reversible color centers are produced beyond the normal penetration depth for electrons. This could result, for example, if electron-hole pairs (exciton) could diffuse a few micron before recombining or producing F center.

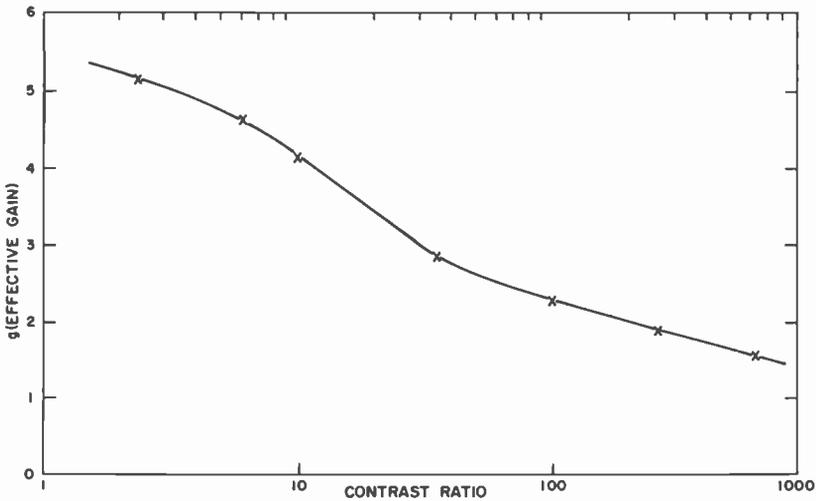


Fig. 9—Effective gain g of a settled sodalite slide versus contrast ratio (data taken at 25 KV).

6. Discussion

Another plot of interest for cathodochromic tube performance is CR versus ϕ for three different F -center-production efficiencies (see Fig. 10). The calculated curves assume no saturation in the F -center-production efficiency. Superimposed on the calculated curves are two different experimental curves taken on the same sodalite screen. The low-current-density curve includes the optically reversible component. The high-current curve consists only of the thermal erase component, because the screen heats up sufficiently to thermally bleach the optical part of the coloration (250 milliwatt are incident on a 0.25cm^2 area of an isolated mica slide $75\mu\text{m}$ thick).

The low current curve starts out close to the theoretical curve for $\eta = 100 \text{ eV}/F$ center. In fact the efficiency derived from this early part

of the curve is $\eta = 80$ eV/F center, in remarkable agreement with the 300 KV van de Graaff measurements.¹⁵ It rapidly falls below the theoretical $\eta = 100$ eV/F curve (saturation) and meets the high-current curve at CR = 13:1. The high-current curve (thermal erase part) closely follows the $\eta = 500$ -eV/F-center curve and doesn't show saturation until approximately CR = 10:1. Fortunately, most anticipated applications do not require contrast ratios in excess of this value, so no loss in sensitivity results from saturation in this case. A continuous series of curves intermediate between the two experimental ones shown can be obtained simply by varying the electron-beam current or, equivalently, by heating the mica substrate to reduce the contribution of the optical mode coloration.¹⁹

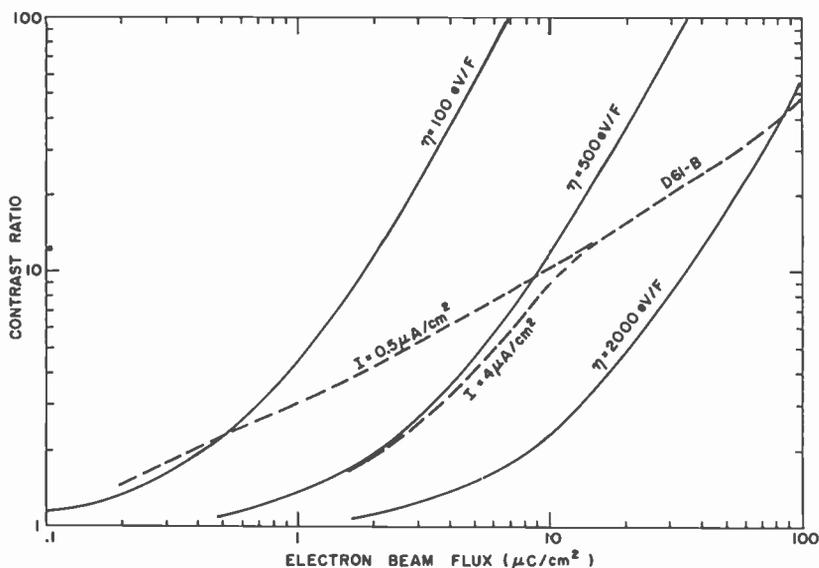


Fig. 10—Calculated CR versus ϕ curves for a sodalite slide for $\eta_{th} = 100$, 500 and 2000 eV/F center ($V = 25$ KV). Also included are experimental data for a bromide sodalite slide at low and high current density.

An example of the excellent γ characteristic that can be obtained by the proper combination of optical and thermal erase modes is shown in Fig. 11. Log CR is plotted versus log ϕ and results in a straight line with unity slope, i.e. $\gamma \approx 1.0$ up to contrast ratios of about 25:1. The addition of some optical erase mode coloration to the thermal mode compensates for the bend in the curve normally found at lower contrast ratios.

Finally, some room-temperature decay rates are shown in Fig.

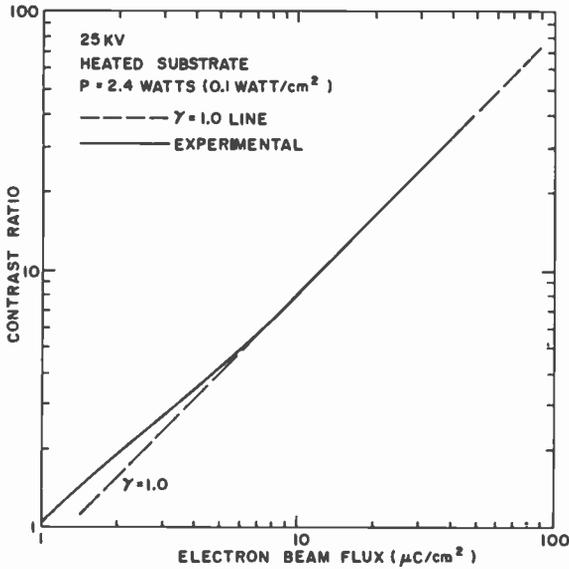


Fig. 11—CR versus ϕ for bromide sodalite slide showing gamma-characteristic for a heated substrate ($V = 25$ KV).

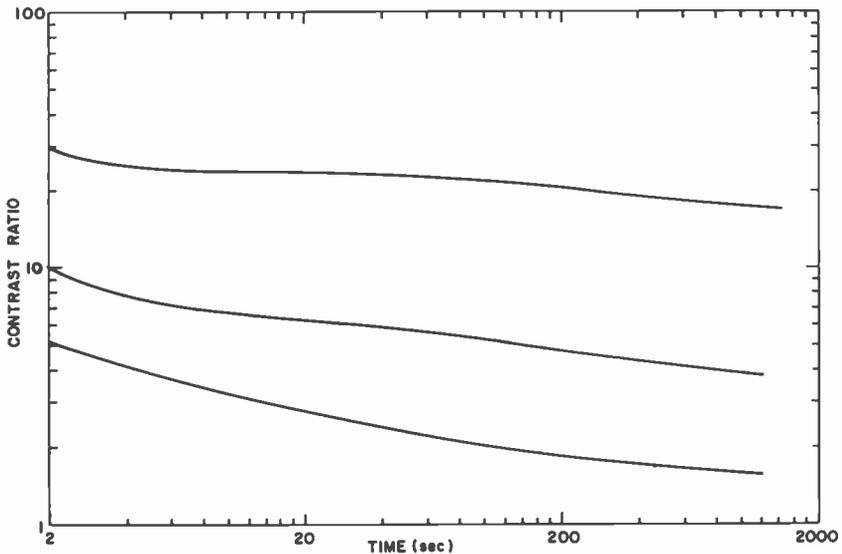


Fig. 12—Room-temperature decay of contrast ratio for a bromide sodalite slide ($V = 25$ KV).

12. They are most conveniently plotted on a log t scale. The theory of the optical decay will not be discussed here.

All the relevant properties of cathodochromic sodalite are summarized in Table 1. The pertinent data for KCl is included, since KCl is the only other material that has been suggested and used for dark-trace tubes. The entries for KCl are our own data which agree with published data on KCl single crystals, where appropriate. A similar table appears in Ref. [5]. However, Table 1 reflects the most recent results for sodalite, as well as the results of the intrinsic absorption measurements.

Some of the values in Table 1 are only approximate. In some cases accurate data for KCl is not available. Sometimes it is difficult to describe a property by one number. The following comments should help clarify the table (letters related to Table 1) :

- A. The values for sodalite are accurate measured values. While the efficiencies for KCl cathodochromic tubes have not been measured, a good estimate can be obtained from KCl single-crystal data, as well as from a comparison with sodalite tubes. F -center production efficiency for the early stage coloration in KCl is found to be approximately 100 eV/ F center,²⁰ which is close to that of sodalite. Values between 1500-6000 eV/ F center have been reported for the initial efficiency of late-stage coloration,^{21,22} and these values seem consistent with our KCl slide measurements.
- B.,C. The sodalite figures have been checked for a number of different materials and slides. It is of course possible to find less sensitive sodalite. The KCl data was measured on only one evaporated KCl slide* and may not represent the best possible results. However the numbers are consistent with the data in row A in Table 1 when the earlier saturation behavior of KCl (see row D) is taken into account.
- D. The maximum absorption coefficient for sodalite is several times that of KCl. In fact, in sodalite there are no appreciable saturation effects up to a contrast ratio of 10:1. This is definitely not true for KCl.
- E. It is difficult to give an exact figure for the lifetimes, but the values quoted are representative. The initial rapid drop when the electron beam is turned off is not counted in the estimate of lifetime in Table 1. It is perhaps true that it should also not be counted in the sensitivity data, since color centers which decay rapidly are not available for storage.

* We thank P. Heyman and D. Hoffman for the KCl slide.

Table 1—Cathodochromic Properties of Bromide Sodalite and KCl

Property	Comments*	Sodalite		KCl	
		Optical Erase Mode	Thermal Erase Mode	Optical Erase Mode	Thermal Erase Mode
Initial F-Center-Production Eff. η	A	100 eV/F	500 eV/F	same as sodalite	1500-6000 eV/F
Experimentally Measured Sensitivity at 25 KV	B	CR = 2:1 at $\phi = .35\mu\text{C}/\text{cm}^2$	CR = 2:1 at $2.0\mu\text{C}/\text{cm}^2$ 10:1 at $10\mu\text{C}/\text{cm}^2$ 100:1 at $250\mu\text{C}/\text{cm}^2$	CR = 2:1 at $2.0\mu\text{C}/\text{cm}^2$	CR = 4:1 at $\phi = 2.5\mu\text{C}/\text{cm}^2$ = 10:1 at $\phi = 1500\mu\text{C}/\text{cm}^2$
Maximum Contrast Ratio	C	2.3:1	CR = 10:1 no saturation CR > 100:1 in saturated region	2:1	10:1 but only at $\phi = 1500\mu\text{C}/\text{cm}^2$ $\eta \approx 75,000\text{ eV/F}$
Maximum Absorption Coefficient	D	$k = 230\text{ cm}^{-1}$	$k = 1000\text{ cm}^{-1}$ (linear region) $k > 10,000\text{ cm}^{-1}$ for $\phi 1500\mu\text{C}/\text{cm}^2$	$k \approx 200\text{ cm}^{-1}$	$k \approx 4000\text{ cm}^{-1}$
Lifetime (thermal decay) at 50%	E	3 min.	months	minutes	months
Effect of Visible Light	F	Bleaches	Will not bleach	Bleaches	Some Bleaching, complex centers formed

* See text for comments.

Sodalite shows many interesting photochromic and cathodochromic properties. The most promising material at this time for storage and display applications is bromide sodalite when used in the thermal erase mode, in cathodochromic cathode-ray-tubes. The advantages of this system, as well as the device configuration, have been previously described.⁵ They are high contrast ratio, rapid erase, long storage time unaffected by high light ambient, no fatigue, and mechanical simplicity using standard cathode-ray tube configurations.

The most serious drawback is the relatively low sensitivity to coloration. It is not possible at present to use television scan rates and applications are confined to areas that do not require large information flow rates. Particularly appropriate are applications where the cathodochromic tube is used for both storage and display, as exemplified by a remote graphics technique where information is transmitted over phone lines.⁶ If further material improvement proves possible, the applications will grow correspondingly.

Acknowledgments

We acknowledge the able technical assistance of R. Nielen, J. McGowan, and L. Blaich in preparing and measuring the sodalite materials. We thank Ross Shrader for the use of his demountable apparatus for the rapid evaluation of loose powders and Phil Heyman for the development of the In_2O_3 coated mica slides. Finally we thank the various groups at RCA Laboratories responsible for sputtering In_2O_3 slides, evaporation of gold contacts, and analytical measurements including x-rays, scanning electron microscopy, IR spectroscopy, and wet-chemical analysis.

References:

- ¹ W. Phillips and Z. Kiss, "Photo Erasable Dark Trace Cathode-Ray Storage Tube," *Proc. IEEE*, Vol. 56, p. 2072 (1968).
- ² W. Phillips, "Properties of Cathodochromic Sodalite," *J. Electrochem. Soc.*, Vol. 117, p. 1557 (1970).
- ³ M. J. Taylor, D. J. Marshall, P. A. Forrester, and S. D. MacLaughlan, *Radio Electron. Eng.*, Vol. 40, p. 17 (1970).
- ⁴ I. Gorog, "Cathodochromic Image Display Applications," *Appl. Optics*, Vol. 9, p. 2243 (1970).
- ⁵ P. M. Heyman, I. Gorog, and B. Faughnan, "High Contrast Thermal-Erase Cathodochromic Sodalite Storage-Display Tubes," *IEEE Trans. Elect. Dev.*, Vol. ED-18, p. 685 (1971).
- ⁶ I. Gorog, P. M. Heyman, H. K. Kim and M. D. Lippman, "An Experimental Low Cost Graphic Information Distribution System Terminal," 1971 S.I.D. Intl. Symposium, May 4-6, 1971, Phila, Pa.
- ⁷ I. Shidlovsky, B. W. Faughnan, P. M. Heyman and I. Gorog, "Recent Developments in Cathodochromic Materials and Devices," 140th National Meeting, The Electrochemical Soc., Cleveland, Ohio, Oct. 3-7, 1971.
- ⁸ P. T. Bolwijn, D. J. Schipper and C. Z. van Doorn, "Cathodochromic Properties of Sodalite," *J. Appl. Phys.*, Vol. 43, p. 132-137 (1972).

- ⁹ H. W. Leverenz, "Luminescence and Tenebrescence As Applied in Radar," *RCA Review*, June 1946, Vol. 7, No. 2.
- ¹⁰ D. B. Medved, "Hackmanite and its Tenebrescent Properties," *Am. Mineralogist*, Vol. 39, p. 615 (1954).
- ¹¹ E. F. Williams, W. G. Hodgson, and J. S. Brinnen, "Synthetic Photochromic Sodalite," *J. Am. Ceramic Soc.*, Vol. 52, p. 139 (1969).
- ¹² W. G. Hodgson, J. S. Brinnen, and E. F. Williams, "Electron Spin Resonance Investigation of Photochromic Sodalites," *J. Chem. Phys.*, Vol. 47, p. 3719 (1967).
- ¹³ M. J. Taylor, D. J. Marshall and H. Evans, "Infra-Red Spectra of Photochromic Sodalite," *J. Phys. Chem. of Solids*, Vol. 1, p. 1-6 (1971).
- ¹⁴ S. D. McLaughlan and D. J. Marshall, "Paramagnetic Resonance of Sulfur Radicals in Synthetic Sodalites," *J. Phys. Chem.*, Vol. 74, p. 1359 (1970).
- ¹⁵ B. W. Faughnan, I. Shidlovsky and A. Rothwarf, "An Experimental and Theoretical Study of Electron Beam Coloration in Sodalite," to be published.
- ¹⁶ James H. Schulman and W. Dale Compton, *Color Centers in Solids*, Pergamon Press, 1963, Chap. VII, pp. 209-238.
- ¹⁷ D. W. G. Ballentyne and K. L. Bye, "The Nature of Photochromism in Chlorosodalites from Optical Data," *J. Phys.*, Vol. D3, p. 1438 (1970).
- ¹⁸ K. L. Bye and E. A. D. White, "Hydrothermal Growth of Sodalite Single Crystals," *J. Cryst. Growth*, Vol. 6, p. 355 (1970).
- ¹⁹ P. Heyman and I. Gorog, Private Communication.
- ²⁰ H. Rabin and C. C. Klick, *Phys. Rev.* 117, p. 1005 (1960).
- ²¹ D. Pooley, "The Saturation of F-Center Production in Alkali Halides under Proton Irradiation," *Brit. J. Appl. Physics*, Vol. 17, p. 855 (1966).
- ²² F. Herrmann and P. Pinard, "The Penetration of Electrons in Alkali Halide Single Crystals and the Effect of Nonuniform Irradiation on Color Center Growth Curves," *J. Phys.*, Vol. C3, p. 1037 (1970).

Liquid Crystals for Electro-Optical Application*

Joseph A. Castellano

RCA Laboratories, Princeton, N.J.

Abstract—A review of recent studies in the synthesis and physical chemistry of mesomorphic (liquid crystal) materials is presented. The materials parameters that affect both the mesomorphic and electro-optical behavior of the compounds are discussed. Compounds that possess a resultant dipole moment oblique to the long axis of the molecule exhibit flow alignment or dynamic scattering, while those compounds that have a dipole moment along the molecular axis can be used to align various guest molecules.

1. Introduction

With the object of developing new techniques for light modulation, a systematic investigation was begun of the relationship between molecular structure and the electro-optic properties of liquid crystals in early 1965.[†] Up to that time, it had been known¹ that nematic liquid crystals could be oriented by externally applied electric and magnetic fields and that this orientation produced a change in the optical properties of the material. These experiments, which were largely carried out in the early 1930's by L. S. Ornstein and W. Kast, were performed with *p*-azoxyanisole at elevated temperatures (115-134°C).

[†] The group engaged in this investigation included G. H. Heilmeyer, J. E. Goldmacher, A. Sussman, W. Helfrich, L. A. Zanoni, E. F. Pasierb, M. T. McCaffrey, C. S. Oh, L. A. Berton, and R. N. Friel, in addition to myself.

* Some of the material presented here has been published in *Ferro-electrics*, 3(1) (1971).

Our research led to the discovery of several new electro-optic effects in nematic liquid crystals and mixtures of nematic and cholesteric liquid crystals. In addition, we developed a variety of new liquid crystalline materials that enabled these effects to be applied to practical display devices.

2. Liquid Crystallinity or Mesomorphism

Mesomorphism has been described² as a new state of matter intermediate between a crystalline solid and a normal isotropic liquid. The phenomenon is generally exhibited by long, rod-shaped organic molecules that contain dipolar and polarizable groups. The mesophase exists over a very specific temperature range. Below this range, the material is a solid and at higher temperatures it becomes an isotropic liquid. Both of these transitions are sharp and reproducible.

Friedel³ carried out extensive optical studies on a number of mesomorphic materials, discovering three main types of mesomorphic states which he designated as the smectic, nematic, and cholesteric mesophase.

The smectic mesophase is a turbid, highly viscous state with certain properties similar to those found for soaps. The term smectic is, in fact, derived from the Greek word, *smectos* or soaplike. The smectic structure is stratified with the molecules arranged in layers; their long axes are parallel to each other in the layers and approximately normal to the plane of the layers. The molecules can move in two directions in the plane and they can rotate about one axis. Within the layers, which are approximately 20 Å thick, the molecules can either be arranged in neat rows or randomly distributed. In addition, the planes can slide without hindrance over similar neighboring layers.

The term nematic was derived from the Greek word meaning thread, because it describes the thread-like nature of the material as seen under the microscope. This liquid is characterized by a turbid, mobile appearance. The long axes of the molecules in this structure maintain a parallel or nearly parallel arrangement to each other. They are mobile in three directions and can rotate about one axis. This can be compared to a long cylinder of round pencils; the pencils can roll and slide back and forth but remain parallel to one another in the direction of their long axes.

The cholesteric mesophase is found primarily in derivatives of cholesterol, especially the esters. The structure consists of parallel, monomolecular layers in which the direction of the long axes of molecules in a chosen layer is slightly displaced from the direction of the axes of molecules in an adjacent layer. This displacement of direction

continues from one layer to another resulting in a helical structure. This helical molecular ordering gives rise to a circular dichroism for light propagating parallel to the helical axis.³ Thus, one particular wavelength of light of one sense of circular polarization is transmitted without attenuation, while light of the opposite sense is totally reflected. Changes in the color of light reflected from the material as a function of temperature are attributed to changes in the pitch of the helix. This property makes these materials useful as temperature indicators in many unique applications.⁴

This review will deal primarily with the nematic and cholesteric mesophase, since several interesting electro-optic effects have been discovered with these materials. These include: (1) dynamic scattering, (2) electronic color switching, (3) reflective optical storage, and (4) field-induced phase changes.

3. Nematic Materials

A detailed discussion of the various textures that nematic materials adopt in thin layers is presented elsewhere¹ and, therefore, an extensive treatment of this subject will not be considered here. For the purposes of this discussion, however, it is appropriate to briefly describe the so-called homeotropic and homogenous textures that are observed in thin layers between glass surfaces. The homeotropic texture is optically extinct between cross polarizers, while the homogenous texture is not. Both of these textures possess the optical properties of a positive uniaxial crystal. The texture observed depends upon the nature of the compound, the conditions used to obtain the mesophase, and the nature of the supporting surface.

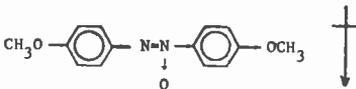
In general, however, molecules that contain groups that permit the long axis to be strongly attracted to the glass surfaces will exhibit the birefringent, homogenous texture, while those molecules that are weakly attracted to the surface are most likely to form the homeotropic texture. Molecules that possess the former characteristic generally have a strong dipole moment along the long axis. In this case, the dielectric constant parallel to the molecular axis is greater than the dielectric constant in the perpendicular direction, and the molecule is said to possess *positive dielectric anisotropy* (PDA). Conversely, molecules that have their dipole moment operating across the molecular axis generally exhibit the homeotropic texture and are said to possess *negative dielectric anisotropy* (NDA).⁵⁻⁷ In this paper, an attempt will be made to explore, albeit qualitatively, the interrelationships among dielectric anisotropy, molecular structure, and electro-optic behavior.

3.1 Dynamic Scattering

In any discussion of electro-optic effects in liquid crystals, one must include a description of the basic fabrication techniques used in the construction of display devices. The basic cell consists of a parallel plate capacitor with liquid crystalline material acting as the dielectric. The plates are merely two pieces of glass, each having a thin conductive coating, such as tin or indium oxide. In fabrication, a drop of the liquid crystalline compound is placed on one of the plates and a sandwich is formed by placing another plate over it. The thickness of active area, which is generally in the range 6-25 microns, may be controlled by the use of appropriate polymeric spacers.

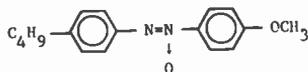
Application of a dc or low-frequency ac signal across the plates changes the material into a milky white liquid. This appearance is due to the creation of scattering centers within the liquid and not to a chemical reaction. When the voltage is removed, the panel returns to its transparent state. The mechanism of this process, although not completely understood, is presently believed to involve the disruption of an orientation pattern by ions in transit. Refractive-index gradients resulting from this disruption would give rise to 1-5 micron-sized scattering centers.

The initial studies on the electrical behavior of nematic compounds were conducted with *p*-azoxyanisole (I) and other azoxy compounds at elevated temperatures.^{1,9-11} These compounds have strong negative



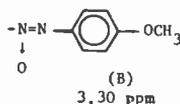
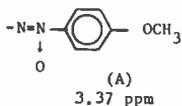
I

dielectric anisotropy due to the dipolar nitrogen-oxygen group in the center of the molecule. Williams⁹ discovered that thin layers of these materials form regular patterns or domains at low voltages. At higher voltages, these stationary patterns gave way to a turbulent condition. Similar observations have subsequently been reported in other papers.¹⁰⁻¹² Most of the definitive synthetic work on this class of materials was performed some years ago.¹³ Attempts to prepare azoxy compounds that exhibit nematic properties at ambient temperatures have been reported only very recently. Steinsträsser et al¹⁴ reported the preparation of a series of *p*-alkyl-*p'*-alkoxy- and *p*-alkyl-*p'*-

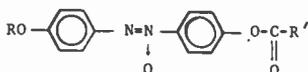


II

acyloxyazoxybenzenes which had melting points below 40°. One member of this series, II, exhibited nematic behavior in the range 16-76°. Nuclear magnetic resonance spectroscopy revealed that these materials were not single compounds but instead consisted of mixtures of two azoxy isomers. The protons of the methoxy group in each isomer showed two different signals.



The preparation of a series of *p*-alkoxy-*p'*-acyloxyazoxybenzenes (III) was also reported recently.¹⁵ The azoxy compounds of this series were generally lower melting than the corresponding azo compounds from which they were derived due to the fact that the former are mixtures of isomers.

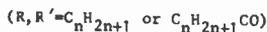
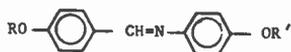


III

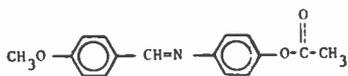
The higher nematic thermal stability of the azoxy materials compared to the azo compounds and to analogous Schiff base compounds (*vide infra*) was attributed to broadening of the molecule as a result of the presence of a lateral oxygen atom.

Although nematic mesomorphism has been observed in a variety of molecular structures, the large majority of compounds that exhibit the phase are aromatic Schiff base derivatives. These compounds are generally prepared by condensation of ring-substituted aldehydes with similarly substituted amines. The discovery that certain of these compounds exhibit the electro-optic effect known as dynamic scattering¹⁶ prompted studies directed toward the synthesis of materials with low melting points. Initial studies¹⁷ involved the preparation of ben-

zylideneaniline substituted with a variety of groups in the para positions of both rings. However, only nine of the 21 new compounds exhibited nematic behavior. On the basis of a comparison with three-ring anils, it was concluded that a critical balance of lateral and terminal intermolecular attractive forces must be attained in order for mesomorphism to occur in these two-ring anils. A favorable balance is created by the presence of alkyl, alkoxy, and acyloxy groups in the para positions of the rings, and a number of compounds containing these groups were prepared. The first Schiff base materials to exhibit nematic behavior at ambient temperatures¹⁸ were mixtures of compounds of this type (IV). For example, an equimolar mixture of compounds V, VI, and VII was found to have a nematic range of 22-105°C and to exhibit very efficient dynamic scattering. Additional low-melting binary and ternary mixtures from this series were subsequently reported.¹⁹

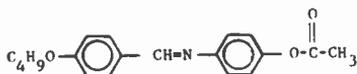


IV



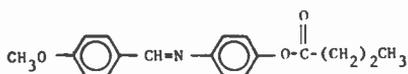
(81-110°C)

V



(82-113°C)

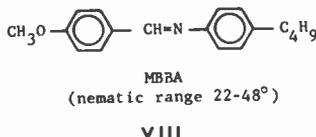
VI



(50-113°C)

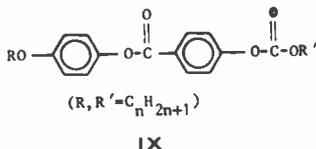
VII

The first example of a single Schiff base compound to exhibit nematic behavior at ambient temperatures was VIII, prepared by Kelker and Scheurle.²⁰ The electro-optic properties of the new compound, MBBA (*p*-methoxybenzylidene-*p*'-butylaniline), were studied shortly thereafter and it was found²¹ that the material exhibited dynamic scattering. This compound has since become a model in a number of theoretical studies of the dynamic scattering effect.²²



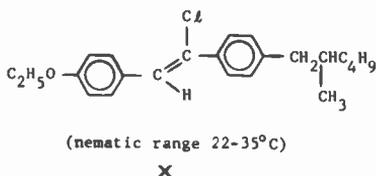
A number of other Schiff base derivatives with NDA have been reported in recent years including the *p*-alkoxybenzylidene-*p*'-aminocinnamates,²³ heterocyclic benzylidene-4-amino-4'-methoxybiphenyls,²⁴ and 4-alkoxybenzylidene-2'-alkoxy-5'-aminopyridines.²⁵ The electro-optic properties of these materials, however, have not been reported.

In addition to azoxy and Schiff base compounds, a third group of materials, namely the *p*-alkylcarbonato-*p*'-alkoxyphenyl benzoates (IX), have been prepared and their electro-optic properties measured. Of



the 48 compounds that were synthesized in this series, 41 exhibited nematic mesomorphism. The compound that had $\text{R} = \text{C}_7\text{H}_{15}$ and $\text{R}' = \text{C}_6\text{H}_{13}$ had a nematic range of $36-54^\circ$. In addition, nine mixtures of compounds from the series had crystal \rightarrow nematic transition temperatures below 30°C . These carbonato esters exhibited spontaneous homeotropic alignment between transparent electrodes. This is presumably due to a strong attraction between the terminal carbonyl group ($-\text{C}=\text{O}$) and the metal oxide surface. The fact that rather high dc fields (150 kV/cm) are required to produce light scattering effects appears to confirm this.

A fourth series of nematic compounds which possess NDA are the nonplanar, *trans*-stilbenes such as X. These low melting materials, which were prepared by Young et al,²⁶ appear to have higher chemical stability than the Schiff base materials. They are also reported to exhibit the dynamic scattering effect.



Dynamic scattering has also been found to occur in mixtures of two different structural types of nematic materials with NDA. For example, Toriyama, et al⁶ have found that a binary mixture of butyl-*p*-(*p*-ethoxyphenoxy-carbonyl)phenyl carbonate and anisylidene-*p*-aminophenyl acetate, which is nematic between 40 and 84°C, shows dynamic scattering at moderate field strengths (1.3-2.6 kV/cm). The scattering characteristics of the mixture appear to be different from those of each component.

It is clear from these results that nematic materials with a variety of structures may exhibit the dynamic scattering effect. A quantitative relationship between molecular structure and the electro-optical properties (i.e., scattering efficiency, threshold, speed, etc.) of these materials remains to be developed.

3.2 Electronic Color Switching

It was found^{27,28} that cooperative alignment of certain nematic compounds with positive dielectric anisotropy could be used to orient pleochroic dye molecules. The optical absorption spectrum of a pleochroic dye molecule is a function of its molecular orientation with respect to the polarization of the incident light. Materials that exhibit pleochroism are usually long, cylindrically shaped molecules containing chromophoric groups which form part of an extended aromatic system. Thus, if the pleochroic molecule is oriented with its long axis parallel to the electric vector of the incident polarized light, absorption of light by the molecule (low energy transition) occurs and the characteristic color of the dye is observed.

Conversely, orientation of the molecule with its long axis perpendicular to the electric vector results in little or no absorption by the visible transition, and the incident light is transmitted unchanged.

The effect of electric fields on mixtures of these dyes with nematic hosts is illustrated in Fig. 1. The cell is constructed in the form of a parallel plate capacitor with transparent electrodes. The nematic material containing a pleochroic dye serves as the dielectric. Alignment of the molecules with their long axes parallel to the electric vector of the polarized light occurs by stroking the Nesa-coated surface prior to cell fabrication. This alignment procedure therefore results in a cell that has a color characteristic of the dissolved dye (Fig. 1a).

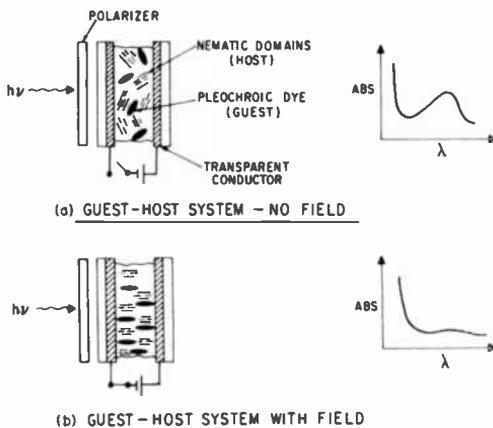
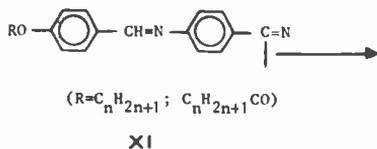


Fig. 1—Schematic representation of electro-optic cells exhibiting electronic color switching.

The very strong permanent dipole moment operating along the long molecular axis enables the molecules to align in the direction of an applied electric field and in turn to orient the dissolved dye molecules with their long axes perpendicular to the electric vector of the incident polarized light. This produces a large decrease in the optical density and hence in the disappearance of color (Fig. 1b).

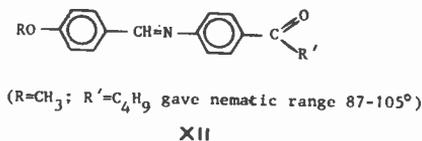
Initial studies of this phenomenon were conducted with *p*-butoxybenzoic acid and *p*-methoxycinnamic acid, but the high operating temperatures required by these nematic compounds prompted a search for other materials with lower melting points. The class of Schiff base compounds of structure XI was found quite suitable and a number of derivatives and mixtures were prepared.²⁹ The lowest melting compound ($R = C_7H_{15}CO$) in this series had a nematic range of 52-94°C, while one ternary mixture was nematic between 26 and 88°C. The



cyano group (C ≡ N) provides the molecule with a strong dipole moment along the molecular axis.

Several classes of organic dyes have been found to exhibit pleochroism including methyl red, indophenol blue, and isolar green M, among others. In principle, any color can be obtained by suitable choice of dye. All of the dyes are long, rod-like molecule that possess highly polarizable terminal groups and aromatic rings. These features resemble those of mesomorphic compounds and thus make the molecules compatible with nematic hosts.

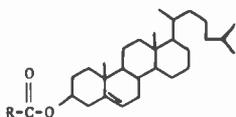
Another series of compounds that have been used to orient pleochroic dye molecules are the *p*-alkoxybenzylidene-*p'*-aminoalkylphenones (XII).³⁰



Between Nesa-coated slides, these molecules are oriented with their long axes perpendicular to the electrode surfaces. This orientation presumably occurs as a result of the attraction between the carbonyl group and the metal oxide coating. Application of electric fields (0.1-0.5 kV/cm) to mixtures of pleochroic dyes and XII produced a reorientation of the molecules as they realigned with their dipole moments in the direction of the applied field. Thus, it was possible to produce an electronic color switching effect that was the opposite of that described above.

4. Cholesteric Materials

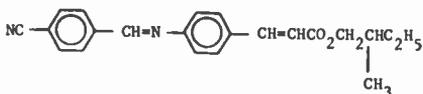
Compounds that exhibit the cholesteric mesophase are of two types. The first and most common type is the cholesteryl ester with the steroid structure XIII. The R group in this structure is generally a long



XIII

alkyl chain (C_nH_{2n+1}) containing from 1 to as many as 18 carbon atoms. These molecules are all optically active (*vide infra*) since they are derived from the natural product, cholesterol.

The second class of compounds that exhibit the "cholesteric" mesophase are not derivatives of cholesterol. These nonsteroidal cholesteric compounds have structures that are nearly identical to those of nematic compounds, but with one important exception, that is they possess an asymmetrically substituted carbon atom, which is a carbon atom bonded to four different atoms or groups. If a molecule has such an asymmetric center, it will be optically active, since the molecule will be nonidentical with its mirror image. Molecules that have two mirror image forms are said to possess chirality or handedness, because they rotate the plane of polarized light in a left or right direction. However, the chemical and physical properties of each form are identical. A nonsteroidal cholesteric compound that has been known for some time² is represented by structure XIV.



XIV

* asymmetric carbon atom

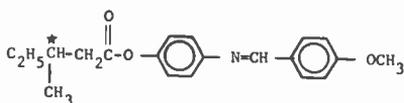
There are two types of electro-optic effects that occur with cholesteric materials. The materials that possess negative dielectric anisotropy exhibit the optical storage mode, while materials with positive dielectric anisotropy undergo field-induced cholesteric-nematic phase changes.

4.1 Optical Storage Mode

The addition of optically active molecules to nematic compounds produces the cholesteric mesophase, because the asymmetric center of the chiral molecule forces the molecules to adopt the helical conformation characteristic of the cholesteric mesophase. Mixtures of nematic

materials with cholesteryl esters that also possess negative dielectric anisotropy have been found³¹ to exhibit an optical storage effect when they are used as the dielectric in the cell configuration described above. The initially transparent panel becomes milky white, similar in appearance to the dynamic scattering mode, when subjected to a dc or low-frequency ac signal (generally 25-30 volts for a 6 micron thick active layer). In contrast to dynamic scattering, however, this highly scattering appearance remains after the electrical excitation is removed. The transparent state can be abruptly (milliseconds) restored by application of an ac signal in the kilohertz range. The mechanism of this process appears to involve the formation of scattering centers in which the helical axes of the cholesteric molecules have a distribution of angles with respect to the electrode surfaces. The high-frequency erasure field produces molecular alignment of the molecules so that the transparent, planar state is restored.

Recent work in this area³² has led to the development of materials that exhibit this effect at ambient temperatures. These materials consist of mixtures of Schiff bases taken from the series represented by IV and long chain cholesteryl esters such as cholesteryl oleate. In addition, non-steroidal cholesteric compounds such as XV were used in conjunction with the nematic Schiff bases to produce these materials.



(cholesteric range 41-78°C)

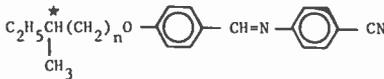
XV

* asymmetric carbon atom

4.2 Field-Induced Phase Changes

The last group of materials that will be considered here are mixtures of nematic and nonsteroidal cholesteric compounds that possess positive dielectric anisotropy. When subjected to electric fields of 200 kV/cm, mixtures of XIV and compounds from among series XI were found³³ to undergo a transition from the cholesteric to the nematic mesophase. This transition was observed as a change from the bright-yellow iridescence characteristic of the focal-conic texture¹ of the cholesteric mesophase to the transparent, homeotropic texture of the

nematic mesophase. The mechanism of this effect appears to involve field-induced unwinding of the helical structure³⁴⁻³⁸ followed by orientation of the resulting nematic molecules with their long axes in the direction of the applied field. Recent research on materials that exhibit this phenomenon has been directed toward new nonsteroidal Schiff base compounds such as series XVI. These compounds exhibit field-induced phase changes at lower voltages and at lower temperatures than those previously reported.



XVI

* asymmetric carbon atom

5. Applications

Although the application of the various electro-optic effects to practical display devices is still some years away, several experimental devices that demonstrate the versatility of the new concepts have been constructed. These include electronically controlled windows and optical filters, simple numeric indicators, an all-electronic clock, and a portable, desk-top calculating machine.

The electronically controlled window is the simplest device and can be used in the dynamic scattering mode, the color-switching mode, or with the field-induced texture change. A "stained-glass" type window could be constructed by using a mosaic pattern of cells that exhibit the color-switching mode. Application of an electric field to each segment would then produce an unusual color display.

Frosted glass door panels that could be "opened" at the flick of a switch are possible with field-induced texture changes. The possibility also exists for the use of those effects in conjunction with a photoconductor or other photosensor to provide electronic curtains that would control the amount of light admitted to a room.

Electronically tuned optical filters have been fabricated using the color-switching mode of operation. The filters could be used in conjunction with other filters to provide appropriate color changes and are expected to be useful in portable photographic equipment.

Numeric indicators have been fabricated with materials that exhibit the dynamic-scattering mode or the color-switching effect. The com-

plete range of numbers (0 through 9) is possible by photoetching the usual seven-segment pattern onto one of the tin oxide coated surfaces. An all-electronic clock that has no moving mechanical parts was constructed using these seven-segment numeric indicators. Four of these cells arranged in succession are used to present the time in hours and minutes. The time reference is the 60-Hz line frequency which is divided by a 60 to 1 counter to produce pulses at 1 Hz. Since the liquid-crystal cell has low power and flat construction characteristics, it may be possible to extend this concept to timing devices with novel configurations.

With eight of the seven-segment numeric indicators arranged in succession, it was also possible to build a solid-state, portable, desk-top calculating machine using the logic and push-button systems that are commercially available. The low power characteristics of liquid crystal displays promise to yield the first practical hand-held portable calculator.

References:

- ¹ G. W. Gray, "Molecular Structure and the Properties of Liquid Crystals," **Academic Press**, 1962, and references therein.
- ² G. Friedel, **Ann. Physique**, Vol. 18, p. 273 (1922).
- ³ F. J. Kahn, "Cholesteric Liquid Crystals for Optical Applications," **Appl. Phys. Letters**, Vol. 18, p. 231 (1971) and references therein.
- ⁴ J. L. Ferguson, "Liquid Crystals," **Sci. Amer.**, Vol. 211, No. 2, p. 76 (1964).
- ⁵ W. Helfrich, "Conduction-Induced Alignment of Nematic Liquid Crystals: Basic Model and Stability Considerations," **J. Chem. Phys.**, Vol. 51, p. 4092 (1969).
- ⁶ K. Toriyama, T. Aoyagi, and S. Nomura, "Mixed Liquid Crystals with New Electro-Optic Effect," **Japan J. Appl. Phys.**, Vol. 9, p. 584 (1970).
- ⁷ G. H. Heilmeyer and W. Helfrich, "Orientational Oscillations in Nematic Liquid Crystals," **Appl. Phys. Letters**, Vol. 16, p. 155 (1970).
- ⁸ G. H. Heilmeyer, L. A. Barton, and L. A. Zanoni, "Dynamic Scattering in Nematic Liquid Crystals," **Appl. Phys. Letters**, Vol. 13, p. 46, No. 1 (1968).
- ⁹ R. Williams, "Domains in Liquid Crystals," **J. Chem. Phys.**, Vol. 39, p. 384 (1963).
- ¹⁰ A. P. Kapustin and L. S. Larinova, "Behavior of Anisotropic Liquids in an Electric Field," **Soviet Phys. Cryst.**, No. 9, p. 235 (1964).
- ¹¹ G. Elliott and J. G. Gibson, "Domain Structures in Liquid Crystals Induced by Electric Fields," **Nature**, Vol. 205, p. 995 (1965).
- ¹² R. Williams and G. H. Heilmeyer, "Possible Ferroelectric Effects in Liquid Crystals and Related Liquids," **J. Chem. Phys.**, Vol. 44, p. 638 (1966).
- ¹³ C. Weygand and R. Gabler, **Ber.**, Vol. 71B, p. 2399 (1938); **J. Prakt. Chem.**, Vol. 155, p. 332 (1940); **Z. Physik. Chem.**, Vol. B46, p. 270 (1940).
- ¹⁴ R. Steinsträsser and L. Pohl, **Tetrahedron Letters**, Vol. 1971, p. 1921. H. Kelker, B. Scheurle, R. Hatz, and W. Bartsch, **Angew. Chem.**, Vol. 82, p. 987 (1970).
- ¹⁵ M. T. McCaffrey and J. A. Castellano, Abstracts of Papers presented at 162nd Nat. ACS Meeting, Washington, D.C., 1971.
- ¹⁶ G. H. Heilmeyer, L. A. Barton, and L. A. Zanoni, "Dynamic Scattering: A New Electro-Optic Effect in Certain Classes of Nematic Liquid Crystals," **Proc. IEEE**, Vol. 56, p. 1162 (1968).
- ¹⁷ J. A. Castellano et al., "Liquid Crystals. II—Effects of Terminal Group Substitution on the Mesomorphic Behavior of Some Benzylideneanilines," **J. Org. Chem.**, Vol. 33, p. 3501 (1968).

- ¹⁸ J. A. Castellano and J. E. Goldmacher, U. S. Patent 3,540,796 (1970). Applied for June 9, 1966.
- ¹⁹ L. Pohl and R. Steinsträsser, III Int. Liq. Cryst. Conf., Berlin, 1970, p. 70.
- ²⁰ H. Kelker and B. Scheurle, *Angew. Chem.*, Vol. 81, p. 903 (1969).
- ²¹ D. Jones, L. Creagh, and S. Lu, "Dynamic Scattering in Room-Temperature Nematic Liquid Crystals," *Appl. Phys. Letters*, Vol. 16, p. 61 (1970).
- ²² Abstracts of Papers presented at III Liq. Cryst. Conf., Berlin, 1970. I. Haller and J. D. Litster, p. 26; Orsay Liquid Crystal Group, p. 98; W. Haas, J. Adams, and J. B. Flannery, p. 109; H. Koelmans and A. M. van Bortel, p. 113; R. Borel and F. Tobazeon, p. 117.
- ²³ G. W. Gray and K. J. Harrison, *Mol. Cryst. and Liq. Cryst.*, Vol. 13, p. 37 (1971).
- ²⁴ W. R. Young, I. Haller, and L. Williams, *Liquid Crystals and Ordered Fluids*, Plenum Press (1970), p. 383.
- ²⁵ C. S. Oh, Abstracts of 162nd Nat. ACS Meeting, Washington, D.C., 1971.
- ²⁶ W. R. Young, A. Aviram, and R. J. Cox, Abstracts of Papers presented at 162nd Nat. ACS Meeting, Washington, D.C., 1971.
- ²⁷ G. H. Heilmeyer and L. A. Zanoni, "Guest-Host Interactions in Nematic Liquid Crystals, A New Electro-Optic Effect," *Appl. Phys. Letters*, Vol. 13, p. 91 (1968).
- ²⁸ G. H. Heilmeyer, J. A. Castellano, and L. A. Zanoni, *Mol. Cryst. and Liq. Cryst.*, Vol. 8, p. 293 (1969).
- ²⁹ J. A. Castellano, U. S. Patent 3,597,044 (1971); Applied for Sept. 1, 1968.
- ³⁰ J. A. Castellano and M. T. McCaffrey, *Liquid Crystals and Ordered Fluids*, Plenum Press (1970), p. 293.
- ³¹ G. H. Heilmeyer and J. E. Goldmacher, "A New Electric Field Controlled Reflective Optical Storage Effect in Mixed Liquid Crystal Systems," *Proc. IEEE*, Vol. 57, p. 34 (1969).
- ³² J. A. Castellano et al., Reports to Air Force Materials Laboratory (MAYH), Wright-Patterson AFB, Dayton, Ohio, Contract F33615-70-C-1590.
- ³³ G. H. Heilmeyer, L. A. Zanoni, and J. E. Goldmacher, *Liquid Crystals and Ordered Fluids*, Plenum Press, 1970, p. 215.
- ³⁴ E. Sackmann, S. Meiboom, and L. Snyder, "On the Relation of Nematic to Cholesteric Mesophases," *J. Am. Chem. Soc.*, Vol. 89, p. 5981 (1967).
- ³⁵ F. J. Kahn, "Electric-Field-Induced Color Changes and Pitch Dilation in Cholesteric Liquid Crystals," *Phys. Rev. Letters*, Vol. 24, p. 209 (1970).
- ³⁶ P. G. DeGennes, *Mol. Cryst. and Liq. Cryst.*, Vol. 7, p. 325 (1969).
- ³⁷ R. Meyer, "Effects of Electric and Magnetic Fields on the Structure of Cholesteric Liquid Crystals," *Appl. Phys. Letters*, Vol. 12, p. 281 (1968).
- ³⁸ W. Helfrich, "Deformation of Cholesteric Liquid Crystals with Low Threshold Voltage," *Appl. Phys. Letters*, Vol. 17, p. 531 (1970).
- ³⁹ J. A. Castellano, E. F. Pasierb, C. S. Oh, and M. T. McCaffrey, Reports to NASA, Langley Research Center, Hampton, Va., Contract NAS 1-10490.

RCA Technical Papers Fourth Quarter, 1971

October

- "Authors' Reply to Comments on Electronic Image Storage Utilizing a Silicon Dioxide Target," R. S. Silver and E. Luedicke, **IEEE Trans. GED**
- "Cathodoluminescence Efficiency of Tm^{+3} in Zinc Sulfide," R. E. Shrader, S. Larach, and P. N. Yocom, **Jour. Appl. Phys. (Communications)**
- "Charge-Coupled Digital Circuits," W. F. Kosonocky and J. E. Carnes, **IEEE Jour. Solid-State Circuits**
- "Current Status of Negative Electron Affinity Devices," B. F. Williams and J. J. Tietjen, **Proc. IEEE**
- "Decoder for Delay-Modulation Coded Data," J. Lewin, **IEEE Trans. GCOM (Concise Papers) (Part I)**
- "Drift-Aiding Fringing Fields in Charge-Coupled Devices," J. E. Carnes, W. F. Kosonocky, and E. G. Ramberg, **IEEE Jour. Solid-State Circuits**
- "With Feedback, Isolation Amp Gives Better-Than-Unity Gain," R. J. Turner, **Electronics (Designer's Casebook) (October 11)**
- "Film Technology in Microwave Integrated Circuits," M. Caulton, **Proc. IEEE**
- "A Motion-Picture Looping System Employing Computer Techniques," D. R. Brewer, **Jour. SMPTE**
- "Nonlinear Optical Susceptibility of a Crystalline Coumarin Dye," P. D. Southgate and D. S. Hall, **Jour. Appl. Phys.**
- "Recent Advances in Thin-Film Silicon Devices on Sapphire Substrates," R. S. Ronen and P. H. Robinson, **Proc. IEEE**
- "Self-Healing Breakdown Measurements of Pyrolytic Aluminum Oxide Films on Silicon," J. E. Carnes and M. T. Duffy, **Jour. Appl. Phys.**
- "Switchable Acoustic Surface Wave Sequence Generator," G. D. O'Clock, Jr., C. L. Grasse, and D. A. Gandolfo, **Proc. IEEE (Letters)**
- "Theory of Transverse Cavity Mode Selection in Homojunction and Heterojunction Semiconductor Diode Lasers," J. K. Butler, **Jour. Appl. Phys.**
- "Thermal Quenching of Tb^{+3} , Tm^{+3} , Pr^{+3} , and Dy^{+3} $4f^7$ Emitting States in La_2O_2S ," C. W. Struck and W. H. Fonger, **Jour. Appl. Phys.**

November

- Comments on "Angular Accuracy of Amplitude Monopulse Off-Boresight Radar," S. M. Sherman, **Proc. IEEE (Letters)**
- "Automatic Test Systems Dedicated or Integrated," A. M. Greenspan, **IEEE Trans. GIM**
- "Direct Evidence for Generation of Defect Centers During Forward-Bias Degradation of $GaAs_{1-x}P_x$ Electroluminescent Diodes," H. Schade, C. J. Nuese, and J. J. Gannon, **Jour. Appl. Phys.**
- "Liquid Crystal Matrix Displays," B. J. Lechner, F. J. Marlowe, E. O. Nester and J. Tufts, **Proc. IEEE**
- "Radiation Conductance of Open-Circuit Microstrip," H. Sobol, **IEEE Trans. GMTT (Correspondence)**
- "RUD: A Computer-Controlled Test-Data-Acquisition and Processing System," B. Mangolds, **IEEE Trans. GIM**
- "Self-Scanned Image Sensors Based on Charge Transfer by the Bucket-Brigade Method," P. K. Weimer, M. G. Kovac, F. V. Shallcross, and W. S. Pike, **IEEE Trans. GED**
- "Surface-State Lifetimes and Current Flow in Liquid Helium," R. Williams and R. S. Crandall, **Phys. Rev. A**
- "Technical Journals Can Be Produced Faster and For Less: A Case History," D. B. Dobson, **IEEE Trans. GAES**
- "Transcendent Silicon Power Rectifier," S. W. Kessler, Jr. and Coauthor, **IEEE Trans. GAES**
- "Exact Solution of the Korteweg-de Vries Equation for Multiple Collisions of Solitons," R. Hirota, **Phys. Rev. Letters (November 1)**
- "Cathodoluminescence of Magnetite," I. Balberg and J. I. Pankove, **Phys. Rev. Letters (November 15)**

December

- "Electrocoloration in Sr, TiO₃: Vacancy Drift and Oxidation-Reduction of Transition Metals," J. Blanc and D. L. Staebler, **Phys. Rev. B** (November 15)
- "Photochromism in Transition-Metal-Doped SrTiO₃," B. W. Faughnan, **Phys. Rev. B** (November 15)
- "Amplification—Modern Trends, Techniques, and Problems II," L. S. Nergaard, **RCA Review**
- "Codes for Error Correction in High-Speed Memory Systems Part II: Correction of Temporary and Catastrophic Errors," C. V. Srinivasan, **IEEE Trans. GC**
- "Damage and Recovery Characteristics of Lithium-Containing Solar Cells," T. J. Faith, **IEEE Trans. GNS**
- "Discussion of Various Views on Popcorn Noise," R. Oren, **IEEE Trans. GED** (Correspondence)
- "Electron Transmission Through a Metal Target from Vacuum to Air," R. C. Blosser, **IEEE Trans. GED**
- "High-Power Microwave Amplifier Using an Antiparallel Avalanche-Diode Pair," H. Kawamoto, **IEEE Trans. GMIT**
- "Low-Noise Punch-Through P-N- μ -P, P-N-P, and P-N-Metal Microwave Diodes," S. G. Liu and J. J. Risko, **RCA Review**
- "PCM Plus C/MOS Spells Reliable, Low-Dissipation Data Acquisition," A. Young, **Electronics** (December 20)
- "Pyrohydrolytic Al₂O₃ for MOS Applications," F. B. Micheletti, P. E. Norris, K. M. Schlesier, and J. M. Shaw, **Solid State Tech.**
- "Relaxation Phenomena Associated with Radiation-Induced Trapped Charge in Al₂O₃ MOS Devices," F. B. Micheletti and F. Kolondra, **IEEE Trans. GNS**
- "Resolving Power Functions and Integrals of High-Definition Television and Photographic Cameras—A New Concept in Image Evaluation," O. H. Schade, Sr., **RCA Review**
- "Switching Times of a Moderate-Power GaAs Field-Effect Transistor," L. S. Napoli, W. F. Reicher, R. E. DeBrecht, and A. B. Dreeben, **RCA Review**
- "Techniques for High-Data-Rate Two-Dimensional Optical Pattern Recognition," R. Croce and G. Burton, **RCA Review**
- "On the Transient Scattering of Light by Pulsed Liquid Crystal Cells," L. S. Cosentino, **IEEE Trans. GED** (Correspondence)
- "Transit-Time-Spread-Limited Time Resolution of Image Tubes in Streak Operation," I. P. Csorba, **RCA Review**
- "Vapor Deposited Tungsten for Silicon Devices," J. M. Shaw and J. A. Amick, **Solid State Tech.**

Patents Issued to RCA Inventors Fourth Quarter, 1971

October

- R. D. Altmanhofer** Color Temperature Control Circuits (3,612,754)
G. E. Anderson Color Television Kinescope Setup Apparatus (3,612,757)
R. E. Asker Tuning Arrangement for a Strip Transmission Line in a Hermetically Sealed Package (3,613,035)
J. Avins Shielded Semiconductor Device (3,614,546)
M. A. Blumenfeld and K. J. Sonneborn Method of Making a Phosphorus Glass Passivated Transistor (3,615,942)
T. A. Bridgewater Television Signal Seeking System with Horizontal Synchronizing Pulse Detector Circuit for Controlling the Signal Seeking (3,610,817)
S. Caplan Liquid Crystal Day/Night Mirror (3,614,210)
D. R. Carley Transistor Assembly (3,611,059)
R. M. Carrell Electronic Photocomposing System that Forms Characters of Different Point Sizes (3,614,767)
W. H. Cherry Homopolar Apparatus which Requires No Moving Parts for Producing Direct Current (3,611,113)
T. J. Christopher Frequency Controlled Oscillator (3,611,176)
M. S. Crouthamel Thermoelement Array Connecting Apparatus (3,615,870)
E. B. Davidson Photoresist Composition and Method of Forming and Using the Same (3,615,952)
F. R. Di Meo and N. W. Burwell Antenna (D 222,358)
E. C. Farnett and L. O. Upton, Jr. Maximum Length Pulse Sequence Generators (3,614,400)
J. R. Fendley, Jr. Double-Ended Ion Laser Tube (3,611,183)
D. L. Franklin and B. J. Fehder Power Transistors having Controlled Emitter Impurity Concentrations (3,614,553)
L. H. Fulton Squeeze Film Bearing Servosystem (3,614,579)
R. L. Giordano and D. J. Poitras Drift-Compensated Average Value Crossover Detector (3,610,956)
D. L. Greenaway Method of Generating High Area-Density Periodic Arrays by Diffraction Imaging (3,615,449)
W. J. Greig Process for Isolating Semiconductor Elements (3,616,348)
M. C. Guerrero, R. E. Justice, and R. M. Rudy Apparatus for Handling Endless Tape (3,613,976)
W. B. Hall and F. G. Block Constant Temperature Output Heat Pipe (3,613,773)
S. E. Harrison and J. E. Goldmacher Cyanine Dye Photographic Film (3,615,562)
S. R. Hofstein Scan Generator Circuit (3,610,960)
R. I. Klein and S. Caplan Liquid Crystal Display Device (3,612,654)
D. L. Kratzer Doppler Correlation Radar System (3,614,785)
R. C. Lemmon Power Supply (3,611,012)
M. E. Miller Stereophonic Phonograph Pickup with Single Pad for Piezoelectric Element Coupling, Support, and Damping (3,610,840)
D. W. Peterson Antenna for a Radio or Television Set (D 222,359)
J. A. Rajchman Digital Light Deflector Having Liquid and Vapor States (3,612,653)
K. Sadashige and M. Horii Servo System for Recorder-Reproducer Apparatus Utilizing Frequency and Phase Synchronizing (3,611,096)
R. S. Silver and J. Jaklik, Jr. Semiconductor Diode Array Vidicon Target Having Selectively Insulated Defective Diodes (3,612,954)
B. Silverstein Edge Connector with Polarizing Member (3,614,714)
L. E. Smith and A. Mayor Bilateral Pincushion Correction Circuit (3,611,004)
W. S. Sommers, Jr. Germanium-Tin Alloy Infrared Detector (3,615,856)
H. C. Stewart and L. S. Cosentino Holographic Memory with Illumination Hologram Providing Reference and Object Beams (3,614,189)

- G. A. Swartz** Bulk Semiconductor Negative Resistance Loaded Slow-Wave Device Amplifiers and Oscillators (3,611,192)
- G. W. Taylor** Light Valve Matrix (3,614,200)
- R. D. Thompson** Video Recording with Alternate Period Inversion and Low-Frequency Premphasis (3,610,819)
- J. E. Volkmann** Stereophonic Sound Enhancement System with Reverberation Chamber (3,614,202)
- C. P. Wen** Balanced Mixer Utilizing Strip Transmission Line Hybrid (3,611,153)
- C. F. Wheatley, Jr.** Differential Amplifier (3,614,645)
- C. F. Wheatley, Jr.** Bias Networks for Class B Operation of an Amplifier (3,611,170)

November

- A. A. Ahmed and M. B. Knight** Trigger Circuits Having Uniform Triggering Voltages (3,619,666)
- J. J. Amodè** Electro-Optic Parity Checker (3,619,031)
- J. J. Amodè and R. S. Mezrich** Hologram Memory (3,618,049)
- G. E. Anderson** Television Amplifier Circuits (3,624,280)
- L. K. Baker and L. R. Shardlow** Multiple Pellet Semiconductor Device (3,619,731)
- J. R. Barger and P. B. Scott** Broad Slope Determining Network (3,619,509)
- W. H. Barkow** Staircase Voltage Generators (3,619,647)
- J. C. Bleazey** Labyrinth for Unidirectional Microphone (3,619,517)
- W. C. Blumenstein** Capacitance Multiplication Network (3,624,425)
- F. Borgini** Clocked Set-Reset Flip-Flop (3,624,423)
- D. R. Bosomworth and Z. J. Kiss** Noise Cancellation in Video Signal Generating Systems (3,624,286)
- L. A. Boyer** Visual Displays Utilizing Liquid Crystals (3,623,392)
- F. W. Brill** Photographic Method for Producing a Metallic Pattern with a Metal Resinate (3,622,322)
- R. D. Browning** Endless Loop Tape Cartridge (3,618,841)
- N. W. Burwell and F. R. Di Meo** Collapsible Structure to Support Antenna Elements (3,623,117)
- D. E. Christensen** Sequencer (3,621,487)
- A. A. Clark** Canonical Orthogonal Filter (3,617,919)
- R. J. Clark** Color Display (3,624,634)
- L. A. Cochran** Color Compensating Circuits (3,617,621)
- A. L. Coen** Oval Loudspeaker Basket and Method of Manufacturing Same (3,618,185)
- E. W. Curtis** Color Television Receiver Hue Control (3,624,279)
- J. A. Dean** Monostable Multivibrator (3,621,297)
- R. H. Edmondson** Error Controlled Automatic Reinterrogation of Memory (3,619,585)
- W. G. Einthoven** High-Voltage Transistor Structure Having Uniform Thermal Characteristics (3,617,821)
- W. W. Evans** Automatic Frequency Control System (3,624,512)
- W. W. Evans** Automatic Fine Tuning Circuitry (3,619,492)
- N. Feldstein** Method of Making a Patterned Metal Film Article (3,619,285)
- M. S. Fischer** Two Terminal Current Regulator (3,624,490)
- D. L. Franklin and L. M. Balents** Mounting Structure for High-Power Semiconductor Devices (3,620,692)
- J. R. Frattarola** Packing of Holotape (3,619,024)
- H. F. Frohbach, A. Macovski, and P. J. Rice** Shadowing System for Color Encoding Camera (3,619,489)
- W. J. Grieg** Semiconductor Device with Multilevel Metalization and Method of Making the Same (3,619,733)
- J. M. Hammer and B. Vural** Amplifiers and Oscillators Comprised of Bulk Semiconductor Negative Resistance Loaded Slow-Wave Structure (3,621,462)
- J. J. Hanak** Method of Epitaxially Growing Single Crystal Films of Metal Oxides (3,617,381)
- W. J. Hannan** Elimination of the Effects of Scratches and Twisting in a Holographic Motion Picture Tape (3,619,025)

L. A. Harwood Oscillator Circuits for Providing a Variable Amplitude Output Signal under Control of an Injected Input Signal (3,617,622)

J. J. Hawley and F. H. Taylor Frequency Modulation System for Spreading Radiated Power (3,617,892)

S. E. Hilliker and J. B. George Ultra High Frequency Oscillator Utilizing Transmission Line Tunable Resonant Circuits (3,624,554)

W. F. Hingsdon Frequency Shift Oscillator which Avoids the Generation of Transients (3,618,132)

M. W. Hoelscher and P. D. Strubhar Method for Metallizing a Ceramic Body (3,620,799)

J. G. Hoehn and R. A. Mancini Optical Card Reading Apparatus (3,619,569)

S. R. Horstein Self-Scanned Phototransistor Array Employing Common Substrate (3,617,823)

W. R. Isom Endless Loop Tape Cartridge (3,620,467)

L. B. Juroff Threshold Digital Switch Circuit for Remote Control System (3,624,510)

D. B. Kaiser and J. A. Powell Gas Laser Tube Mount (3,619,811)

W. F. Kosonocky Balanced Optically-Settable Memory Cell (3,624,419)

W. F. Kosonocky Optically Settable Flip-Flop (3,619,665)

W. M. Lee Zinc Oxide and Titanium Oxide Sensitized by AZO Dyes (3,622,341)

A. Lichowsky Magnetic Drum System (3,623,120)

R. A. Mao First-In-First-Out Buffer Register (3,623,020)

T. R. Martin Cathode-Ray Tube with Electrode Supported by Straplike Springs (3,619,689)

D. L. Mathias Liquid Crystal Cells in a Linear Array (3,622,222)

R. M. Moore Heterojunction Semiconductor Transducer having a Region which is Piezoelectric (3,624,465)

R. M. Moore and C. J. Busanovich Device Employing Selenium-Semiconductor Heterojunction (3,622,712)

R. R. Norley and R. D. Altmanholzer Selectively Operated Tint Correction Circuit (3,619,487)

P. Nyul Light-Emitting Diode Array (3,622,906)

J. Rabinowitz Time-Frequency-Phase In-Band Coded Communications System (3,617,889)

A. Rosen and E. Mykityyn Frequency Multiplier Employing Input and Output Strip Transmission Lines without Spatially Coupling Therebetween (3,621,387)

M. R. Royce and J. S. Martin, Jr. Very Short Luminescent Decay-Time Phosphor (3,623,994)

M. R. Royce Phosphor Screen Comprising Two Kinds of Particles, Each Having Phosphor Core and Phosphor Coating (3,622,826)

T. A. Sautler Photographic Method for Producing a Cathode Ray Tube Screen Structure (3,623,867)

L. C. Scheeparkoetter, Jr. Phase or Frequency Modulator Using Pin Diodes (3,624,559)

J. F. Schanne Article Labeling and Identification System (3,622,758)

D. G. Shipley Frequency Modulated Phase-Locked Oscillator Having Low- and High-Frequency Response (3,622,913)

E. D. Sirmshauer Astigmatism Correction for Electron Beam Devices (3,619,707)

R. R. Soden and W. J. Greig Electrical Fuse Link (3,619,725)

S. A. Stecker High-Gain Differential Amplifier (3,622,903)

A. W. Stephens and J. J. Walsh Electronic Scanner Utilizing a Laser for the Simultaneous Scanning and Reproducing of Images (3,622,690)

A. I. Stotler and W. H. Schillp Method for Isolating Semiconductor Devices from a Water or Semiconducting Material (3,623,219)

G. W. Taylor and A. Miller Electro-Optical System (3,623,795)

P. K. Welmer and F. V. Shallos Electric Signal Processing Circuit Employing Capacitively Scanned Phototransistor Array (3,624,428)

D. H. Willis Video Amplifiers (3,619,488)

P. L. Wiseman Blue Lateral Magnet Structure (3,624,573)

H. A. Wittlinger Analog Multiplier in which One Input Signal Adjusts the Transconductance of a Differential Amplifier (3,621,226)

J. R. Zuber and H. Foxman Method of Branding Silicon Plastic Surface (3,617,326)

- D. I. Boswick Hologram Memory (3,628,847)
 D. J. Carlson Television Tuning Circuit Utilizing Voltage Variable Capacitance (3,628,152)
 B. R. Clay and W. A. Strickland Autocollimator including a Retroreflector Element (3,628,869)
 W. F. Dietz Linearity Correction Circuit Utilizing a Saturable Reactor (3,628,082)
 P. J. Donald Electrophotographic Recording Elements with Half-Tone Screen Coating Thereon (3,627,526)
 G. F. Fordyce and E. J. Mitchell Orthogonal Filters (3,629,721)
 G. Forster Thyristor Controlled Power Supply Circuits and Deflection Circuitry Associated with a Kinescope (3,626,238)
 E. C. Glatmo, Jr. and S. Larach Electrophotographic Recording Element Having Photoconductor with Quenched Luminescence During Charging and Method of Making the Photoconductor (3,627,528)
 W. E. Goldner and J. Pellegrino Cylindrical Magnetic Memory Construction (3,631,417)
 J. E. Goldmacher and O. E. Dow Method of Applying an N,N'-Diallylmethylamine Resist to a Surface (3,627,599)
 I. Gorog Holographic Scan Converter (3,630,594)
 R. B. Goyer Current Source with Positive Feedback Current Repeater (3,629,692)
 J. M. Hammer Optical Modulation System (3,628,177)
 J. M. Hammer Digital Light Deflector Using Electro-Optic Grating (3,626,511)
 J. J. Hanak Magnetic Heads with Poles Joined by Molecular Transport Bonding (3,629,519)
 D. D. Harbert Operation of Field-Effect Transistor Circuit Having Substantial Distributed Capacitance (3,629,612)
 J. R. Harford Wide-Band Amplifier (3,628,166)
 R. C. Heuner and S. J. Niemiec Voltage Reference and Voltage Level Sensing Circuit (3,628,070)
 R. N. Hurst Signal Processor for Dropout Correction Before Demodulation (3,629,494)
 P. Kokko Method of Reconstituting Unfiled, Case, Alumina Scrap (3,631,131)
 W. F. Kosonocky Electrically and Optically Accessible Memory (3,631,411)
 W. L. Lehmann Signal Distribution System (3,631,348)
 A. L. Limberg Electronic Processing Apparatus (3,629,611)
 S. Reich Power Supply Circuits (3,626,242)
 L. A. Rempert Digital Synchronization System (3,629,503)
 F. Ruppensburg Stereophonic Sound Reproducing Apparatus (3,627,392)
 W. W. Siekawicz and R. W. Paglione Broadband Double Ridge Waveguide Magic Tee (3,629,734)
 M. Slovinsky Magnetic Recording Medium with Lubricant (3,625,760)
 M. C. Steele LSA or Hybrid Mode Oscillator Started by Series Connected Gunn or Quenched Mode Oscillator (3,628,170)
 G. W. Taylor Electrical Control of Light Polarization Utilizing the Optical Property of Fluids (3,625,593)
 C. P. Wen Transmission Line Formed by a Dielectric Body Having a Metalized Non-planar Surface (3,629,737)
 C. F. Wheatley, Jr. Current Source (3,629,691)

Juan J. Amodel received his B.S. degree in Electrical Engineering from Case Institute of Technology in 1956 and his M.S.E.C. from the University of Pennsylvania in 1961. In 1965, he returned to the University of Pennsylvania for further study and received his Ph.D. degree from that institution in 1968. Dr. Amodel began his career in 1956 with the Philco Corporation, where he worked on transistor radio developments. In 1957, he joined the Advanced Development Department of RCA's Industrial Electronic Products Division in Camden, N.J. Here he had responsibility for the development of a high-accuracy analog-to-digital converter, the development of a novel pressure telemetering system for time that later became a product. He transferred to RCA Laboratories, Princeton, N.J., in 1959, working on the development of high-speed computer circuits and subsystems. The results of this work included the realization of shift registers capable of 100 and 200 megabits/second information transfer rates and binary full adder systems with fraction-of-a-nanosecond state delays. He later became engaged in research and development of integrated circuits. Since 1966, he has been engaged in research and development of materials and devices for optical information storage. The work included research on the optical storage properties of doped and pure SrTiO_3 crystals, leading to quantitative models that explained the thermal and optical coloring behavior of these materials. This work also led to the discovery and understanding of two-photon induced coloration effects. Later he was actively engaged in research and development in the field of optical recording in magneto-optic and metallic thin films as well as other approaches to optical information storage. In 1967, he studied and developed a model for electron migration effects in photochromic materials that led to the proposal of an electro-optic holographic storage device. He is presently the project scientist in a program aimed at developing better materials and techniques for high sensitivity, erasure resistant electro-optic storage systems.



Since 1962, Dr. Amodel has been a lecturer at LaSalle College in Philadelphia, where he has been Chairman of the Electronic Physics Department of the Evening Division since 1967. Dr. Amodel is a member of the American Physical Society, the IEEE, and Eta Kappa Nu.

Robert A. Bartolini received a B.S. in Electrical Engineering

from Villanova University in 1964 and an M.S. in Electrical Engineering at the University of Pennsylvania. Prior to joining the staff of RCA Laboratories in 1966, he held the position of Teaching Assistant on the faculty at Case Institute of Technology. His work at RCA includes experimental studies of interaction between drifting carriers in semiconductor with slow EM waves carried by metal structures. Currently he is engaged in holographic recording techniques for data storage and retrieval systems. Mr. Bartolini is a member of the IEEE, Eta Kappa Nu, Tau Beta Pi, Sigma Xi and the Optical Society of America.



Joseph Bordogna received the BSEE and Ph.D. degrees from the University of Pennsylvania in 1955 and 1964, respectively, and the SM degree from the Massachusetts Institute of Technology in 1960. His background includes service as a line officer in the U.S. Navy and professional association with RCA Corporation and the University of Pennsylvania. He has worked on laser communication and radar systems, electro-optic modulators, holographic millimeter wave imaging, holographic pre-recorded video techniques, and computer-communications based education systems. Presently, he is Associate Professor at the Moore School of Electrical Engineering, University of Pennsylvania and a Consultant to RCA Laboratories, Princeton. He holds membership in the IEEE, Optical Society of America, American Society for Engineering Education, Tau Beta Pi, Eta Kappa Nu, Sigma Tau, and Sigma Xi.



Joseph A. Castellano received a B.S. degree from the City College of New York in 1959, with a major in chemistry and a minor in physics and mathematics. In June of 1959 Dr. Castellano was employed by Witco Chemical Co., Paterson, New Jersey, where he devised new analytical techniques for detergent compositions and related chemicals employing infrared and ultraviolet spectrophotometry as well as gas chromatography. Dr. Castellano left Witco in April 1962 to join the Thokol Chemical Corp., Reaction Motors Division, Denville, N.J., where he rose to senior chemist. His studies were primarily in synthetic and mechanistic organic chemistry under a contract to the Advanced Research Projects Agency.

During the time that Dr. Castellano was employed he attended evening classes at the Polytechnic Institute of Brooklyn and received the M.S. degree in June 1964. In 1966 he received an RCA Doctoral Study Award which permitted him to obtain the Ph.D. degree in June 1969.

In 1965, he joined RCA Laboratories as a Member of the Technical Staff and has worked in the fields of organic photochemistry, organic semiconductors (and insulators) and liquid crystals. He is presently Project Leader—Liquid Crystal Research. His work in liquid crystals led to the development of the first useful room-temperature nematic material. As a result of this work, Dr. Castellano was co-recipient of an RCA Laboratories Outstanding Achievement Award in 1967 and the David Sarnoff Outstanding Team Award in Science in 1969. This top 100 new products of 1968.

Dr. Castellano is a member of the American Chemical Society, the American Association for the Advancement of Science, and the Society of the Sigma Xi, as well as a Fellow of the Chemical Society of London and the American Institute of Chemists.



Mr. Duncan received an RCA Laboratories Achievement Award in 1963. He is a member of the American Physical Society and Sigma XI.

Since 1962, Mr. Duncan has carried out both fundamental and applied research on the optical properties of insulating crystals, particularly rare-earth-doped calcium fluoride. This work has included development of several new optically-pumped laser materials, demonstration of the first-band-pumped and laser-pumped infrared quantum counters, and, most recently, contributions to the understanding, development, and evaluation of inorganic photochromic crystals and powders.



Robert Duncan holds B.S. (1951) and M.S. (1958) degrees in physics from Union College and Cornell University, respectively. For the M.S. degree he carried out thesis research using the total reflection of x-rays to study the structure of evaporated metallic films. Mr. Duncan joined the research staff of the RCA Laboratories in 1957, and from 1957 to 1962 did both experimental and theoretical work on the electronic and microwave properties of semiconductors. This work included studies of space-charge regions in semiconductor surfaces and cyclotron resonance of negative-effective-mass charge carriers in germanium.

Mr. Duncan has carried out thesis research in semiconductor surfaces and cyclotron resonance of negative-effective-mass charge carriers in germanium.

Thomas L. Credelle received the BSEE degree from Drexel Institute of Technology in 1969 and the MSEE degree from Massachusetts Institute of Technology in 1970. He joined the technical staff of RCA Laboratories in September, 1970. As a member of the Research Training Program, he worked on thin film heterojunction diodes and thermoplastic holography. At present he is with the Electro-Optic Systems Research Group engaged in research on thermoplastic media for holographic recording.



Mr. Credelle is a member of IEEE, Eta Kappa Nu, Tau Beta Pi, and Phi Kappa Phi.

Dr. Cohen is a member of the American Physical Society and Sigma XI. recently investigated the magnetic properties of small metallic particles. In the area of solid-state physics, his research activities have included experimental and theoretical work on solar energy conversion, excess current in semiconductor tunnel diodes, electron-hole plasmas in solids, thermoelectric devices, thermal conductivity of intermetallic compounds, transport and optical effects in granular metals, and possible superfluidity in liquid He³. He has been very active in the field of superconductivity, having studied the effect of microwave phonons on superconductive tunneling, the enhancement of the transition temperature in superconductors, fluctuation phenomena in granular superconductors, the normal state properties of beta-tungsten compounds, and the relationship of lattice instabilities to high-temperature superconductivity. He has been a member of the technical staff at RCA Laboratories since that time, where he is presently head of the Magnetism and Superconductivity Research Group in the Solid State Research Laboratory.



Roger W. Cohen joined the staff of RCA Laboratories in 1960 after receiving the B.S. degree in physics from M.I.T. He participated in the Graduate Study Program that provided for study leading to a Masters Degree while maintaining research activities at the Laboratories. This program was completed in 1962 with the award of the M.S. degree in physics from Rutgers University. In 1966 he received the Ph.D. degree in physics from Rutgers. Since that time, he has been a member of the technical staff at RCA Laboratories where he is presently head of the Magnetism and Superconductivity Research Group in the Solid State Research



E. C. Fox received a bachelor and masters degree in Electrical engineering from MIT in 1957. In 1957 he joined RCA Laboratories where his work has included high-frequency magnetic measurements, research on a magnetic scanning head for video tape recording, disc-recording research (including the invention and development of tracing distortion cancelling for stereo records), work on television color encoding systems, and, most recently, redundancy devices for holography. Mr. Fox is a member of IEEE and the Audio Engineering Society.



Arthur H. Firester received a B.A. degree, cum laude, with Honors in Physics from Brandeis University in 1962, and an M.A. and Ph.D. degrees from Princeton University in 1964 and 1967, respectively. His doctoral dissertation was based upon research performed on the modulation of light by optically pumped alkalali metal vapors. Before joining the staff of RCA Laboratories, he held the position of instructor on the faculty of Princeton University for the academic year 1966-67. His recent research has been in the area of nonlinear optical phenomena and their possible application to image processing, and coherent light optical problems and Dr. Firester is a member of The Society of Sigma Xi and the American Physical Society, the Optical Society of America, and the IEEE.



Brian W. Faughnan graduated from McGill University in 1955 with a B. Eng. in Engineering Physics. He attended the Massachusetts Institute of Technology from 1955 to 1959; he was awarded an M.S. degree in Electrical Engineering in February 1957 and a Ph.D. in Physics in June 1959. His thesis work was on spins and phonons in paramagnetic crystals. In September 1959 he joined RCA Laboratories in Princeton, New Jersey. Since that time he has carried out research in a variety of topics in solid-state physics. These include the study of negative mass cyclotron resonance effects in germanium, radiation damage to solar cells, and the study of radiation damage centers in silicon by electron spin resonance techniques. From 1963 to 1964, he spent a year at the Tokyo Laboratories of RCA studying the propagation of Alfvén waves in the semimetal bismuth. Upon returning to Princeton, he continued work on semimetals including device applications. From 1966 to 1969 he engaged in fundamental studies of various photochromic single crystals, especially SrTiO₃, using combined electron-spin resonance and optical techniques. For the past three years he has been studying the electron-beam coloration of materials (cathodochromism), especially sodalite, with the goal of understanding the physical mechanism and optimizing the materials for device applications. Dr. Faughnan is a member of Sigma Xi and the American Physical Society.

Thomas E. J. Gayeski received a B.S. degree in electrical engineering and an A.B. degree in physics from the University of Pennsylvania in 1969 and a M.S. in electrical engineering from Princeton University in 1971. He joined the staff of RCA Laboratories in June, 1969. His first assignment involved diffraction printing and its application to the silicon storage tube. Since October 1969, he has been working in the area of holographic information storage as it applies to the Holotape system.

Mr. Gayeski is a member of Tau Beta Pi, Eta Kappa Nu, the IEEE, and the Optical Society of America.



William J. Hannan received the ESEE from Drexel Institute in 1954 and the MSEE from the Polytechnic Institute of Brooklyn in 1955. From 1951 to 1956 he worked in RCA's Industrial Products Division where he contributed to the design of RCA's industrial television systems and the "Walkie Lookie", the first man-portable television station. From 1956 to 1966, he worked in the Applied Research Section of RCA's Defense Electronic Products Division. In 1953 he was promoted to Group Leader and spent the following seven years working mainly on laser systems. His group was responsible for developing the first room-temperature injection-laser communication system, NASA's first sun-pumped laser, one of the early electro-optic reading machines, and the first laser transmitter sent into space. He developed the first laser cane for the blind and received an IR 100 Award for his effort in 1967. Since 1966 he has been a Group Head at RCA Laboratories, working in the field of holography. In 1970 he received an RCA Achievement Award for his holographic research and in 1972 his research group received the David Sarnoff Outstanding Team Award in Science for the development of a holographic precordded video system.

He is a senior member of the IEEE and a member of the OSA.



Dainis Karlsons received his BSEE in 1958 from Drexel Institute of Technology, and his MSEE in 1960 from the University of Pennsylvania. Mr. Karlsons worked on early solid-state microwave cavity masers and participated in the design and evaluation of comb and meander-line traveling-wave microwave masers using ruby and rutile as the active materials. In addition, he developed a 5.8 Gc. comb traveling-wave maser to be used in conjunction with a low-noise satellite tracking system. Mr. Karlsons was instrumental in the development of an early working model of a ruby laser. Shortly afterward he designed and operated a ruby laser that utilized an elliptical reflector to increase pumping efficiency. He participated in a study evaluating the possibility of the laser's application to ranging and communications and in the development of a Q-controlled laser operating at liquid nitrogen temperature that should be capable of producing nanosecond pulses in the infrared. Mr. Karlsons, who was a member of Tau Beta Pi, and Phi Kappa Phi, died on October 4, 1971, after a long illness.



Dietrich Meyerhofer received his Bachelor of Engineering Physics from Cornell University in 1954 and his Ph.D. degree in physics from MIT in 1958. Since 1958 he has been a Member of the Technical Staff at RCA Laboratories. His research activities have included the study of electrical properties of semiconductors and insulators, light emission from CO₂ laser. From 1966 to 1968 he was associated with a laboratory doing applied research in support of the RCA Graphic Systems Division. He investigated the application of electronics and laser techniques to the printing industry. Since then he has been with the Electronic Printing Group and has studied holographic recording in dichromated gelatin and the physical properties of liquid crystals. Dr. Meyerhofer is a member of the American Physical Society, the Optical Society of America, and the IEEE.



M. Lurie received a B.S. in physics from Brooklyn College in 1959, an M.S. in physics from Rutgers University in 1961, and a Dr. Eng. Sci. in E.E. from Newark College of Engineering in 1967, he worked on ceramic materials and components, aerospace power systems and associated test equipment, and television circuits for TOS. He was an instructor in E.E. at NCE and did research in physiological optics in addition to holography. After receiving his doctorate, he spent a year and a half as a Senior Research Fellow at the National Physical Laboratory in England, working on spatial filtering, holography, and interferometry. In 1968 he joined RCA Laboratories, where he has been developing a holographic video storage system.



Scott A. Keneman received his undergraduate education and early graduate training at the Massachusetts Institute of Technology. In June 1967, he received both SBEE and SIMEE degrees. Since 1967, Mr. Keneman has been a Member of the Technical Staff at RCA Laboratories. He has been engaged in research on computer memory systems. Initially this research centered on cryoelectric content-addressable memory devices. More recently, Mr. Keneman has been engaged in research on optical memory systems. He performed theoretical and experimental studies on light-valve and image-storage devices using ferroelectric bismuth titanate.



Mr. Keneman is a member of the IEEE, Eta Kappa Nu, and Tau Beta Pi, and an associate member of Sigma Xi.

Dr. Lurie is a member of OSA, IEEE, Eta Kappa Nu and Tau Beta Pi.

Reuben S. Mezrich received the B.S.E.E. degree in 1963 and the M.S.E.E. and Ph.D. degrees in 1967 and 1970 respectively, all from the Polytechnic Institute of Brooklyn. He joined the RCA Laboratories in 1963 and has worked in the fields of cryogenic electronics, medical electronics, the fabrication parameters of thin films, magnetic holography, and holographic storage systems. Dr. Mezrich has won two IR-100 Awards and a RCA Laboratories Achievement Award for his work in holographic memories and magnetic holography. He is a member of Eta Kappa Nu, Tau Beta Pi, the IEEE and the O.S.A.



William Phillips received an A.B. degree in physics from Columbia University in 1958. He did his graduate work at the Carnegie Institute of Technology, receiving an M.S. degree in physics in 1961 and a Ph.D. in electrical engineering in 1964. From 1958 to 1961, he was a graduate research assistant specializing in the properties of single-crystal alkali halides. During the summers of this period, he was employed as an engineer at the Westinghouse Bettis Atomic Power Division, working on problems in nuclear reactor design. From 1961 to 1964, Dr. Phillips conducted research on crystal growth and optical properties of potential laser materials. His thesis research was concerned with the optical properties of the tetraalent neptunium ion.



Dr. Phillips joined the technical staff of RCA Laboratories in 1964, and from 1964 to 1966 was engaged in research on the synthesis and properties of crystal-line materials for laser applications. In 1966 he became involved in work on the synthesis and evaluation of improved photochromic materials, particularly rare-earth-doped calcium fluoride, and sodalite. His work on sodalite led to the demonstration of its cathodochromic properties in 1967, and included the development of the bromide and iodide modifications of this material. He received RCA Laboratories Achievement Awards in 1968 and 1970 and shared in an IR-100 Award for development of the cathodochromic CRT in 1968. Dr. Phillips is currently engaged in research on improved materials for the storage of volume phase holograms.

Dr. Phillips is a member of the American Physical Society, the IEEE, the American Ceramic Society, and Sigma Xi.

Edward G. Ramberg received the A.B. degree from Cornell University in 1928 and the Ph.D. degree in theoretical physics from the University of Munich in 1932. After working on the theory of x-ray spectra as research assistant at Cornell, he joined the Electronic Research Laboratory of the RCA Manufacturing Company in Camden in 1935. He has been associated with the RCA Laboratories in Princeton since their establishment in 1942. He has worked primarily on electron optics as applied to electron microscopy and television, and various phases of physical electronics, thermoelectricity, and optics. In 1949 he was visiting professor in physics at the University of Munich and, in 1960 and 1961 Fulbright Lecturer at the Technische Hochschule, Darmstadt.



Dr. Ramberg is a Fellow of the Institute of Radio Engineers and the American Physical Society and a member of Sigma Xi and the Electron Microscope Society of America.



Igal Shidlovsky graduated from the Israel Institute of Technology in Haifa, Israel with a B.S. in chemistry in 1961. He obtained his M.S. in Inorganic Chemistry from Hebrew University in Jerusalem in 1963. His Ph.D. in Inorganic Chemistry was obtained from Hebrew University in 1968. His Ph.D. thesis was in the area of silicones and germanides of the earth metals. He concurrently served as an instructor in the chemistry department. During the academic year 1968-69, he was a Research Associate in the Physical Chemistry Department at the University of Pittsburgh working on magnetic properties of intermetallic compounds. In August 1969, he joined RCA Laboratories in the Luminescent and Electro-optics group of the Materials Research Laboratory. His primary interest has been in the area of inorganic photochromics and cathodochromic materials.

F. W. Spong received his B.S. degree in physics from the University of Utah in 1958 and the Ph.D. in physics from the University of California at Berkeley in 1964. His dissertation topic was an investigation of the Fermi surface of aluminum by cyclotron resonance. In 1964 Dr. Spong joined the RCA Laboratories, Princeton, N.J., as a member of the Systems Research Laboratory. He has worked on improving power, efficiency, and coherence of gas lasers in the Display Devices Group. Recently he has been concerned with recording color-television signals on film by optical, electronic, and holographic means.



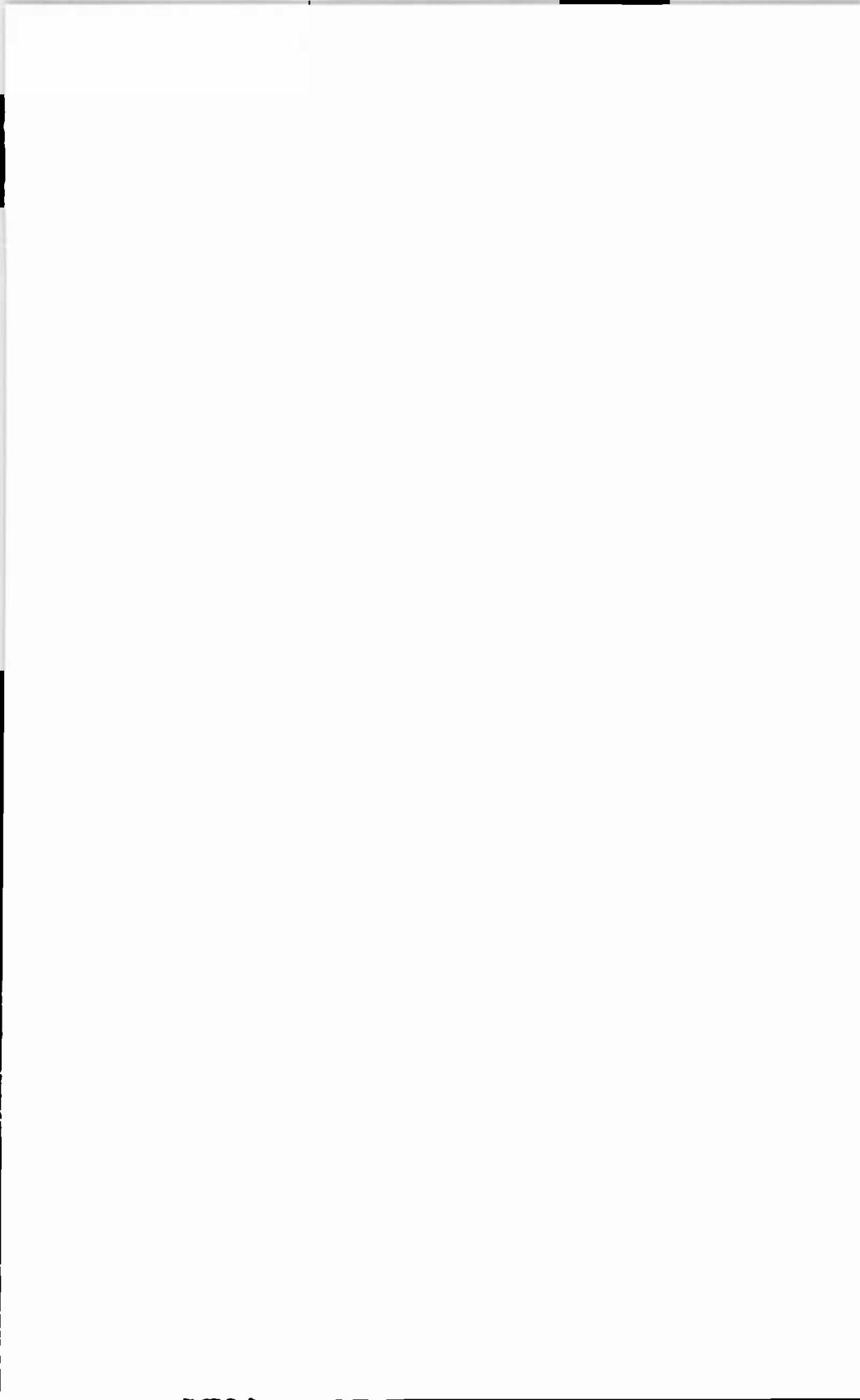
Dr. Spong is a member of the American Physical Society, Sigma Xi, and Phi Beta Kappa.



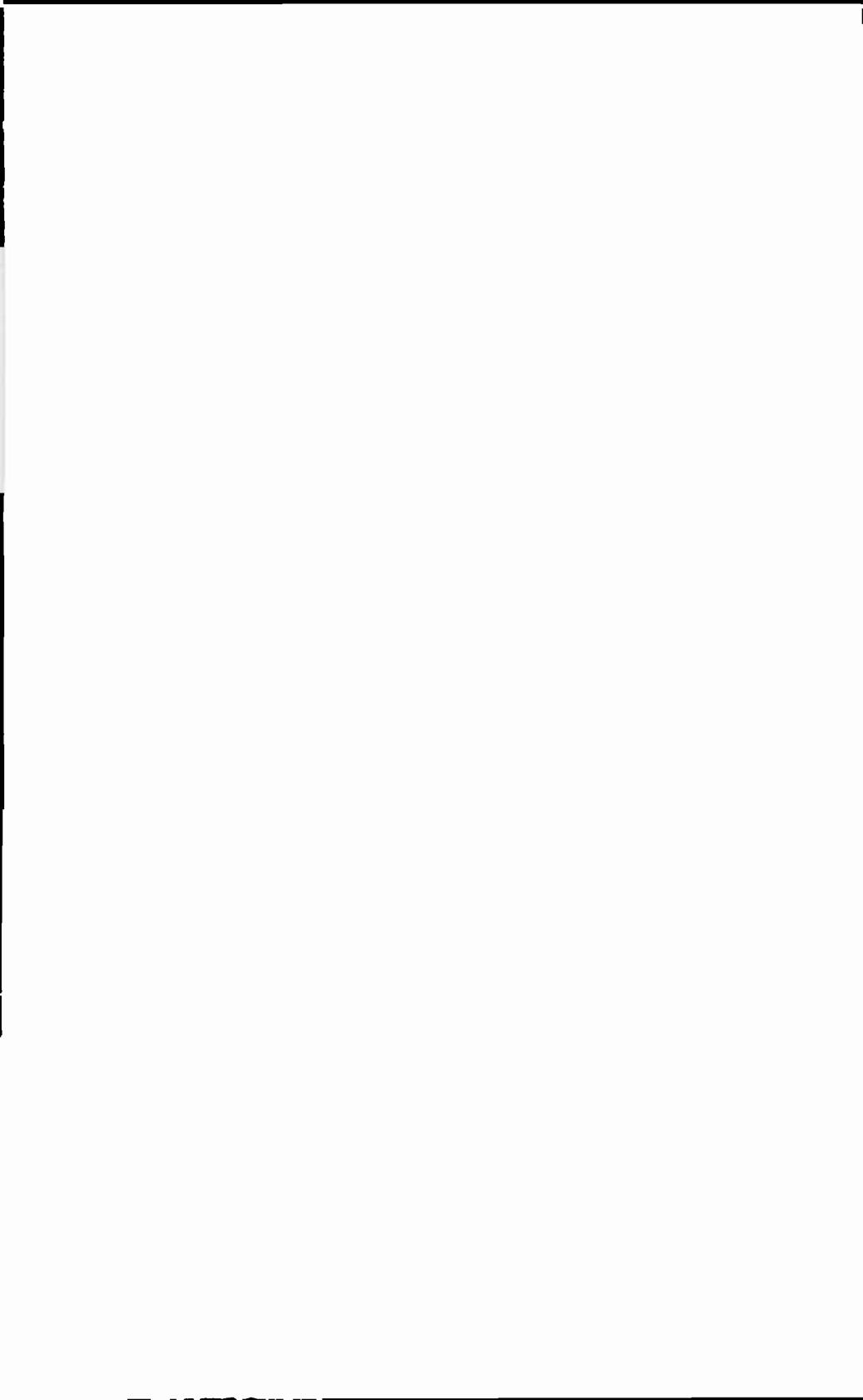
David L. Staebler received the B.S., E.E. with Distinction and the M.S., E.E. from the Pennsylvania State University in 1962 and 1963, respectively. He then joined the research training program at RCA Laboratories and worked with narrow band phosphors, laminated ferrite memories, and injection diode lasers. In 1964, upon completion of the program, he began studies on oxidation-reduction processes in gamma-irradiated rare earth doped fluorites. This work led to his subsequent investigation of photochromic behavior in CaF_2 . In 1966, he received an RCA Laboratories Graduate Study Award and entered Princeton University, graduating in 1970 with a Ph.D. in Electrical Engineering. His doctoral research involved optical studies of rare earth associated color centers in CaF_2 . Upon returning to RCA Laboratories, he has worked on electric field coloration of transition element doped titanites, and is now involved in the study of electro-optic materials for holographic storage.

Dr. Staebler has received an RCA Laboratories Achievement Award for his work on photochromic materials. He is a member of Eta Kappa Nu, Sigma Tau, Tau Beta Pi, and the IEEE.









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 15, 1971. 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 the headquarters or general business offices of the publisher and Editor: Publishers, RCA Laboratories, Princeton, New Jersey, 08540. 6. Names and addresses of the publishers: 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; Kane & Co., c/o Chase Manhattan Bank, 1 Chase Plaza, New York, New York; Cudd & Co., c/o Chase Manhattan Bank, 1 Chase Plaza, New York, New York; Continental Illinois National Bank & Trust Company of Chicago, Lock Box H, Chicago, Illinois 60690; Martin B. Serefsan c/o Coronet Industries, Inc., Box 1248, Dalton, Georgia 30720; Burl Jackson Bandy, c/o Coronet Industries, Inc., Box 1248, Dalton, Georgia 30720.

I certify that the statements made by me above are correct and complete.

Ralph F. Ciafone, Editor

