RCA Review

December 1974 Volume 35 No. 4

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

RCA Review

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

Contents

December 1974 Volume 35 Number 4

System for Visualizing and Measuring Ultrasonic Wavefronts

R. S. Mezrich, K. F. Etzold, and D. H. R. Vilkomerson

RCA Laboratories, Princeton, N.J. 08540

Abstract—A system for the measurement and visualization of ultrasonic waves has been developed that features high sensitivity and acoustic-wavelength-limited resolution over apertures as large as 15 cm at frequencies up to 10 MHz. Acoustic-wave displacement amplitudes as small as 0.005 Å are measured by interferometrically detecting the motion of a thin, acoustically transparent, metallized pellicle as the ultrasonic wave passes through it. The basic arrangement is that of the Michelson interferometer with the addition of an open-loop method to stabilize the response and a deflection system in one leg of the interferometer to scan the pellicle. Presently, the system is used in the study of the interactions of ultrasonic waves with biologic tissue, the observation and measurement of radiation patterns of acoustic transducers, and as an aid in the design of basic acoustic elements for use in other ultrasonic devices.

1. Introduction

The interest in ultrasonic visualization for the examination of biological tissue arises from the possibility of observing tissue structures not observable by optical or radiological techniques. Recent investigations[1,2] on the potential of ultrasonic methods to distinguish between different pathological states, as between benign and malignant, have quickened this interest.

Before the full potential of ultrasonic diagnosis can be realized, there must be detailed knowledge of the characteristics of transducers as well as the radiation patterns they produce; there must also be

careful studies of acoustic wave interaction with biologic tissues, which must be accomplished by detailed measurements of these waves as they are transmitted through and reflected from tissues. Part of the value of ultrasonic observation, both for clinical use and laboratory investigations, depends on the ability to quantitatively characterize the interaction between ultrasound and biological tissue.

We have recently developed a system[3] which we refer to here as ultrasonovision, for the measurement and visualization of ultrasonic waves over a large aperture to quantitatively measure the response of biologic tissues to ultrasound and for the basic study of the generation and propagation of ultrasonic waves. Although primarily intended as a laboratory instrument for the general study of ultrasonic fields and their interactions and for use in the design of more sophisticated clinical instruments, it has sufficient sensitivity to be useful in some clinical applications.

The important characteristics of the system are:

Sensitivity—better than 5 nW/cm^2
Aperture—150 mm diameter
Dynamic Range—linear from 5 nW/cm^2 to 1 W/cm^2
Frequency Range—uniform (nonresonant) response to at least 10 MHz
Resolution—acoustic-wavelength limited
Angular response—flat to acoustic beam angles of incidence to $\pm 40°$

In addition, images can be obtained by using the acoustic wave either transmitted through or reflected from the insonified object in "near real time"—up to 2 frame/sec.

At the present time, this system is being used for the examination of tissue sections for pathology studies, for tissue visualization, and for basic measurements of acoustic wave propagation through, and reflection from, biological tissue. It is also being used for the analysis of transducer characteristics and the acoustic field patterns they produce.

The following discussion gives a complete analysis of the operation and characteristics of the system and several examples of its applicability. We start with a brief overview of the system and several results and then present the detailed analysis.

2. Basic Principle of Operation

The basic arrangement is that of the optical Michelson interferometer, shown schematically in Fig. 1. The key elements of the system are

a light source (laser), a beam-splitter, a reference mirror external to the sound field, and a thin flexible mirror through which the acoustic wave passes. In our experiments the flexible mirror is a thin (~6 μm) metallized plastic film (pellicle) located in the fluid through which the acoustic wave propagates.

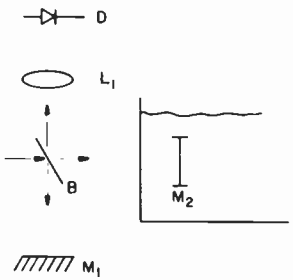


Fig. 1—Basic interferometric arrangement consisting of pellicle (M_2), reference mirror (M_1), beam splitter (B), light source, and detector.

The pellicle is so thin that it is essentially transparent to an ultrasonic wave for frequencies as high as 10 MHz and for angles of incidence from 0° to 40°. As used here, transparent means that the pellicle motion, or displacement, is almost exactly equal to the displacement amplitude of the acoustic wave.

The essential idea of the system is to accurately measure the displacement amplitude of the acoustic wave by interferometrically measuring the motion of the pellicle as the wave passes through it. A laser beam is scanned over the pellicle to measure the displacement amplitude at each point of the ultrasonic field; an image of the acoustic field is generated by processing the signal derived from the interferometer and using it to brightness-modulate a synchronously scanned cathode ray tube.

The displacement of the acoustic wave is related to the wave intensity by

$$I = \frac{1}{2} Z \omega^2 \Delta^2, \qquad [1]$$

where Δ is the displacement amplitude, $\omega = 2\pi f$ (with f the acoustic frequency), and Z is the acoustic impedance, e.g., 1.48×10^6 in metric units for water. Thus, the measurement of the pellicle motion is a precise measurement of the acoustic intensity. (A note on the sensitivity: at 1.5 MHz the displacement amplitude of an acoustic wave with an intensity of 1 μW/cm^2 is 0.1 Å.)

The image size of the system will be limited by the pellicle aperture (up to 150-mm diameter) and the ultimate lateral resolution by the size of the laser beam. (The practical limit below 5 MHz is the acoustic wavelength.) The acoustical numerical aperture, the sensitivity, and the dynamic range are determined by the pellicle response (discussed later) and the characteristics of the optical interferometer. Some results of imaging and measurement of acoustic wavefronts attained with the ultrasonovision system are shown in Figs. 2–7.

Fig. 2 is an image, at 4.5 MHz, of an infarcted myocardium with a fibrosed artery. The tissue section is approximately 1 cm thick. The attenuation through the clear area of the myocardium tissue was 7 dB

Fig. 2—Ultrasonic image at 4.5 MHz of infarcted myocardium with fibrosed artery.

while that of the artery (the hook-shaped structure) was 11 dB. The lateral size of the fibrosed artery is approximately 1 mm. Also seen on the figure are several screws used to hold the tissue in place during observation.

Fig. 3 shows an optical and acoustic image, to the same scale, of excised breast tissue with a benign tumor. The acoustic frequency was 1.5 MHz. The attenuation of the benign tumor was 6 dB/cm while that of the surrounding breast tissue was 1 dB/cm. The tumor can be clearly seen in both images and the shape and extent of the tumor under ultrasonic illumination is identical to that seen with optical illumination.

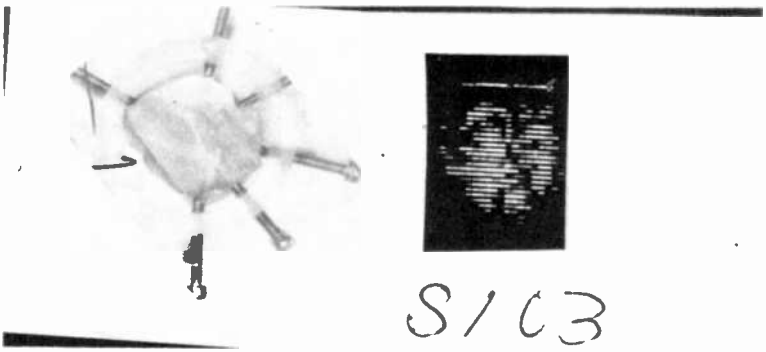
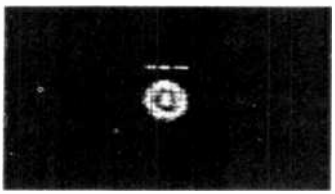
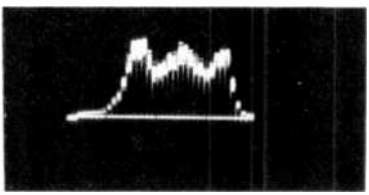


Fig. 3—Optical and acoustic image, both to same scale, of excised breast tissue with benign cystic tissue. Note identical structural appearance of cyst in both images.

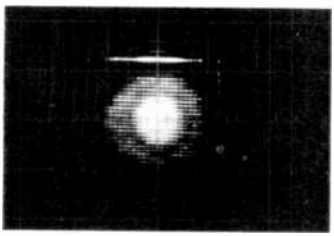
Fig. 4 shows images obtained with a commercial ultrasonic transducer operating in water at 2.25 MHz. Fig. 4a, which is an image of the face of the transducer, shows the mode pattern (which is expected from the theory of finite diameter acoustic transducers) and Fig. 4b



(a)



(b)



(c)

Fig. 4—Acoustic images of transducer surface and far-field radiation pattern. Mode pattern of transducer surface is shown. Peak surface displacement is 3 Å, frequency is 2.25 MHz.

shows the displacement amplitude along one scan across the face of the transducer. The method by which this scan is generated is described below. Clearly the transducer does not behave as a simple piston as some simple models assume. Fig. 4c shows the far-field acoustic pattern from the transducer. The images of Figs. 2, 3 and 4a were obtained with the aid of an acoustic polystyrene lens that imaged the object onto the pellicle. Fig. 4c was obtained without the use of the lens.

Fig. 5 shows the field pattern, with no acoustic lens used, of a "focused" acoustic transducer operating at 2.25 MHz. The nominal focal length of the transducer was 7.5 cm. The photographs show the change in field pattern as the transducer is moved to different distances from the pellicle, and it is apparent that the actual focus is at 5 cm, where the Bessel rings (or side lobes) are clearly seen.

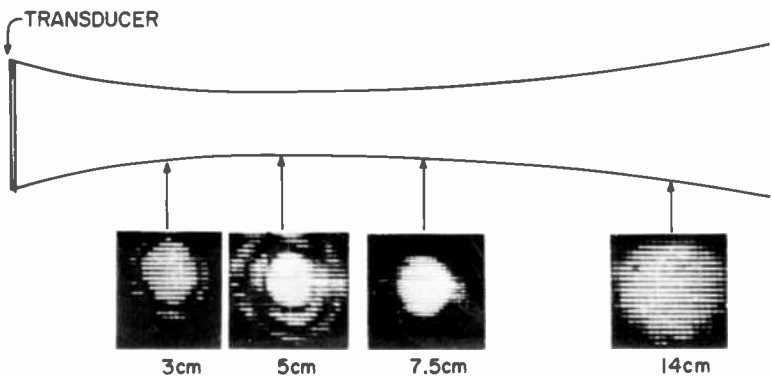


Fig. 5—Radiation pattern at different distances from transducer surface. Nominal transducer focal length is 7.5 cm, frequency is 2.25 MHz.

As an example of the utility of the ultrasonovision system in examining and checking transducers, Fig. 6 shows images from a clinical diagnostic transducer that was still in active use at the time of the experiments, although the user had noticed that "something was wrong." From the acoustic image of the transducer surface it can be seen that sound is only being emitted from some regions around the edges. The pattern in the far field is uniform (as could be predicted). The cause for the improper sound emission was that the protective coating over the transducer had delaminated, a fact that was not evident by looking at the surface.

Due to the rigors of constant clinical use, all transducers can be ex-

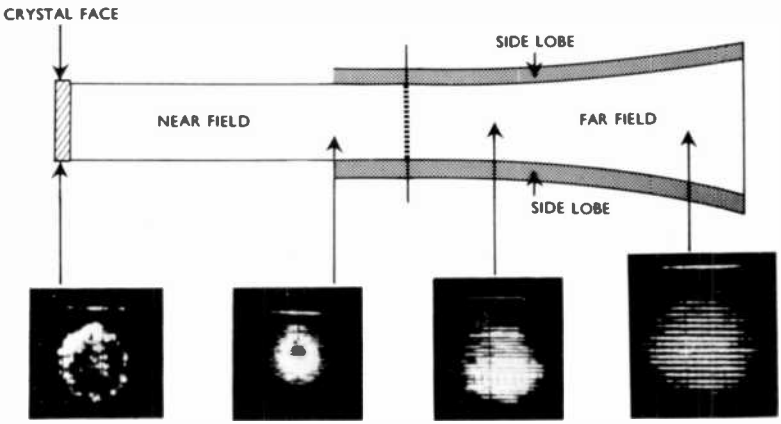


Fig. 6—Transducer surface and radiation pattern of damaged, but still used, diagnostic transducer. Bright spots on surface indicate only radiating regions of transducers.

pected to change their characteristics, sometimes drastically, without the user being aware of the changes; ultrasonovision would be useful in performing rapid and simple periodic checks to maintain the performance of diagnostic instruments.

Fig. 7 is a further example of acoustic imaging of soft tissue, in this case an adult hand. The region imaged includes the thumb and part of the palm. This example indicates the ability to image living tissue in "real time."

The important components of the ultrasonovision system are the pellicle, the optical arrangement of the interferometer, and the electronics system (which include the means for detection and display of the acoustic image). Important considerations are the sensitivity (or signal to noise ratio), the need for (and the means used to obtain) stability in the interferometer, and the effects of component characteristics used in a practical system. In the following section we give a detailed analysis of these and other matters.

3. Analysis of the Ultrasonovision System

3.1 Pellicle

The basis for the operation of the system is that, at every point, the motion of the pellicle is, as mentioned previously, very nearly equal to the displacement amplitude of the acoustic wave passing through it. By measuring the motion of the pellicle the acoustic wave is mea-

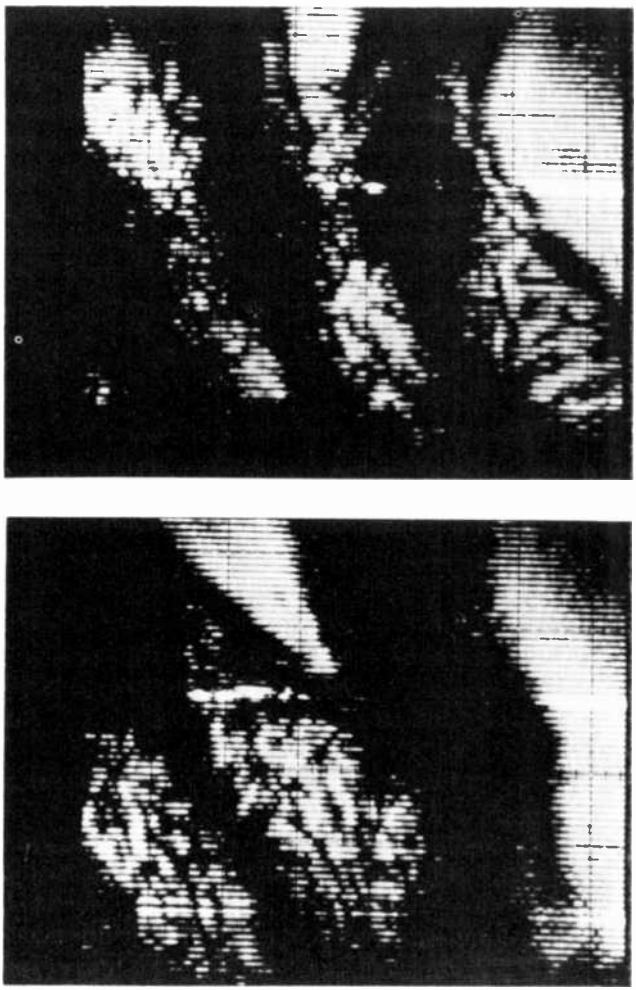


Fig. 7—Acoustic image of hand showing (top) region of palm near knuckles of first three fingers and (bottom) the palm and thumb.

sured. The pellicle is a thin (~6 μm) metallized plastic film that is suspended in the water through which the acoustic wave passes. Pellicles as large as 6 inches in diameter have been used; the measured optical flatness has been of the order of 1 wave/inch.

The motion of the pellicle in response to an acoustic wave has been calculated following an analysis of Brekhovskikh.[4] The result is that the angular response is flat over acoustic beam angles of ±40° and to acoustic frequencies at least as high as 10 MHz. The ratio of the nor-

mal component of the motion of the pellicle to the displacement amplitude of the acoustic wave (again in the direction normal to the pellicle) is 0.99.

These results were verified experimentally by measuring the waves transmitted through and reflected from the pellicle. For example, Fig. 8 shows a graph of the relative sensitivity of the pellicle (mea-

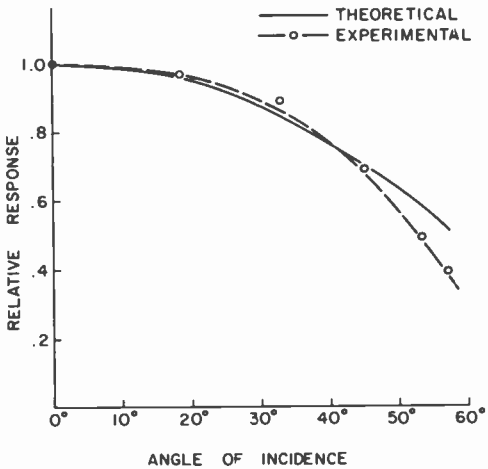


Fig. 8—Pellicle response, theoretical and experimental, as a function of angle of incidence of acoustic beam.

sured interferometrically) as a function of the angle of incidence θ of the acoustic beam. Also plotted is the theoretically expected response. Note that in both cases the response falls as $\cos\theta$, since the normal (perpendicular) component of the incident wave is proportional to $\cos\theta$. The agreement between the expected and experimental results is good over the angular range 0 to $\pm 40°$.

3.2 Basic Interferometer

The operation of the interferometer (Fig. 1) may be understood by the following analysis. Let an acoustic wave of displacement amplitude Δ pass through the pellicle. The deviation of the water, and hence the pellicle at that point is

$$\xi = \Delta \sin(\omega_s t). \qquad [2]$$

When a light beam is reflected by the pellicle its phase will be

changed in proportion to the pellicle motion, and the amplitude of the reflected optical wave becomes

$$A_p = A_{po} \exp\left\{j\left[\varphi_p + 2\Delta\frac{2\pi}{\lambda} \sin(\omega_s t)\right]\right\}. \quad [3]$$

when φ_p is an arbitrary constant phase and λ is the optical wavelength. The factor of 2 is due to the doubling of the relative phase shift on reflection.* The light amplitude reflected from the reference mirror is

$$A_r = A_{ro} \exp\{j\varphi_r\}, \quad [4]$$

where φ_r is again an arbitrary phase term. (Both φ_p and φ_r are actually measures of how far both mirrors are from the beam splitter.) The total amplitude of light incident on the photodetector is

$$A_T = A_r + A_p \quad [5]$$

and the total intensity is then

$$I = A_T \cdot A_T^* \quad [6]$$

$$= |A_{ro}|^2 + |A_{po}|^2 + 2A_{ro}A_{po}$$

$$\times \cos\left[\varphi_p - \varphi_r + 2\Delta\frac{2\pi}{\lambda} \sin(\omega_s t)\right] \quad [7]$$

$$= I_r + I_p + 2\sqrt{I_r I_p} \cos\left[\varphi_p - \varphi_r + \frac{4\pi\Delta}{\lambda} \sin(\omega_s t)\right] \quad [8]$$

We recognize that $|A_{ro}|^2 = I_r$ is the light intensity from the reference mirror alone and $|A_{po}|^2 = I_p$ is the light intensity from the pellicle.†

The photodiode generates a current i_s proportional to the incident light intensity

$$i_s = \eta I$$

where η, the quantum efficiency, is of the order of 0.3 A/W (for red

* There is an additional phase shift, due to the acousto-optic interaction of the light with the ultrasonic wave passing through the water after it passes the pellicle, which should be added to the phase term in Eq. [3]. To avoid complicating the analysis we do not include this effect here, but discuss it fully in Appendix 6.

† It is assumed that the two optical waves are coherent and overlap completely—conditions that can be readily met. If this condition is not met a factor $K(|K| \leq 1)$ multiplies the expression $\sqrt{I_r I_p} \cos[\varphi_p - \varphi_r + (4\pi\Delta/\lambda)\sin(\omega_s t)]$.

light incidence on a silicon photodiode). A high-pass filter is used to suppress the low-frequency terms. The signal after the filter is

$$i_s = 2\eta\sqrt{I_r I_p}\ \cos\left[\varphi_p - \varphi_r + \frac{4\pi\Delta}{\lambda}\ \sin(\omega_s t)\right] \qquad [9]$$

Assume that $\varphi_p - \varphi_r = 90°$; the general case where the relative phase shift can have arbitrary values will be considered below. Then

$$i_s = 2\eta\sqrt{I_r I_p}\ \cos\left[\frac{\pi}{2} + \frac{4\pi\Delta}{\lambda}\ \sin(\omega_s t)\right] \qquad [10]$$

$$= 2\eta\sqrt{I_r I_p}\ \sin\left[\frac{4\pi\Delta}{\lambda}\ \sin(\omega_s t)\right] \qquad [11]$$

If $\Delta/\lambda \ll 1$, which holds for acoustic intensities of less than 1 W/cm^2 in the frequency range 0.5–10 MHz, we may approximate

$$\sin\left[\frac{4\pi\Delta}{\lambda}\ \sin(\omega_s t)\right] \approx \frac{4\pi\Delta}{\lambda}\ \sin(\omega_s t),$$

and the measured electrical signal becomes

$$i_s = 8\pi\eta\frac{\Delta}{\lambda}\sqrt{I_r I_p}\ \sin(\omega_s t). \qquad [12]$$

This signal is directly proportional to the displacement amplitude of the acoustic wave.

Once the light intensities incident on the photodiode from the reference mirror and pellicle are measured, which can be done with great precision, the electrical signal is an accurate measure of the displacement amplitude and, hence, wave intensity at all frequencies and all intensities. The system can be accurately calibrated.

3.3 Sensitivity and Dynamic Range

The sensitivity and dynamic range are limited by noise. The major components of noise are thermal noise generated in the load resistor and shot noise. The shot noise is due to the steady component of the light incident on the photodiode ($I_r + I_p$ in Eq. [8]).

At low light levels thermal noise limits the sensitivity and from Eq.

[12] and the expression for thermal noise ($i_n^2 = 4kTf/R$), the signal-to-noise ratio is

$$\frac{i_s^2}{i_n^2} = \eta^2(8\pi)^2\frac{\Delta^2}{\lambda^2}\frac{I_r I_p}{4kTf}R.$$

Typical values are $\eta^2 = 10^{-1}$, $R = 10^3$ ohms, $f = 10^6$ Hz, $\lambda = 6.3 \times 10^3$ Å, $I_r = I_p$, and $4kT = 16 \times 10^{-21}$ joule. Then

$$\frac{i_s^2}{i_n^2} = 1 \times 10^{11}\Delta^2 I_p^2. \quad [13]$$

If we assume $I_p \approx 1 \times 10^{-4}$ watt (i.e., a 1 mW laser and a 20% system efficiency), the value for Δ with a signal-to-noise of ratio of unity is

$$\Delta \simeq .03 \text{ Å} \quad [14]$$

We note from Eq. [13] that the signal-to-noise ratio of the system, for low light intensities, is proportional to the square of the light intensity (the acoustic intensity is proportional to Δ^2). As the power of the laser, or the optical efficiency of the system, is increased the minimum detectable acoustic intensity increases as the square of the light intensity until, at higher light levels, shot noise becomes dominant. The shot noise current is

$$i_{ns}^2 = 2e\eta If \quad [15]$$

where $e = 1.6 \times 10^{-19}$ C is the electronic charge, η is the photodiode quantum efficiency, I is the incident light ($\approx I_p + I_r$), and f the system bandwidth. The value of light power at which shot noise becomes dominant may be found by equating the shot and thermal noise,

$$2e\eta If = \frac{4kTf}{R} \quad [16]$$

or

$$I = \frac{4kT}{2e\eta R} \quad [17]$$

$$= \frac{5 \times 10^{-2}}{\eta R} \quad [18]$$

again letting $\eta \approx 3 \times 10^{-1}$ and $R = 10^3$ ohms

$$I \simeq (I_p + I_r) \simeq 0.16 \text{ mW}, \quad [19]$$

which is essentially the same value as was used for the previous thermal noise calculation.

Experimentally, with approximately 2.0 mW incident on the photodiode and the other parameters as given in the examples, the minimum detected displacement was 0.005 Å. This corresponds to an acoustic intensity of 5 nW/cm^2 at 1.5 MHz.

When shot noise is dominant, and values of light intensity are greater than given by Eq. [19], the expression for the signal to noise ratio becomes

$$\frac{i_s^2}{i_n^2} = \frac{\eta(8\pi)^2}{4ef} \frac{\Delta^2}{\lambda^2} \frac{I_r I_p}{I_r + I_p}. \qquad [20]$$

Since the acoustic intensity is proportional to Δ^2, the minimum detectable acoustic intensity increases linearly with the available light.

The ultimate limit on the sensivity appears to be given by the maximum dissipation allowed in the photodiode, which for commercially available silicon photodiodes is about one watt. By Eq. [20] we can estimate the ultimate sensitivity of the ultrasonovision system to be of the order of 10^{-10} W/cm^2 at 1.5 MHz. While this is not as sensitive as can be achieved with piezoelectric detectors, the sensitivity is adequate for *in vitro* studies of biologic tissue and for the *in vivo* examination of the appendages and external organs such as arms, legs, and breasts.[5]

The maximum level of acoustic intensity that can be measured is determined by Eq. [11]. At high values of intensity the condition $\Delta/\lambda \ll 1$ no longer obtains and nonlinear effects, notably harmonics, appear. Eq. [11] may be expanded in a Fourier–Bessel series expansion as[6]

$$i_s = 2\eta\sqrt{I_r I_p}\, \sin\left[\frac{4\pi\Delta}{\lambda}\, \sin(\omega_s t)\right] \qquad [21]$$

$$= 2\eta\sqrt{I_r I_p}\left\{2\sum J_{2n+1}\left(\frac{4\pi\Delta}{\lambda}\right)\, \sin[(2n+1)\omega_s t]\right\}. \qquad [22]$$

For small values of the argument, $J_1(x) \approx x/2$, which gives the expression of Eq. [10]. As the argument increases the approximation becomes less valid. We may arbitrarily set the upper limit of validity at $x = 1/4$, where the expansion of Eq. [12] is valid to better than 1%. Then

$$\Delta \leq \frac{\lambda}{16\pi} \simeq 125\ \text{Å} \qquad [23]$$

At a frequency of 1.5×10^6 Hz, this corresponds to an acoustic inten-

sity of about one W/cm^2. Above this limit, the ultrasonovision system does not fail to operate, but simply becomes less linear in its response.

The range of accurate measurement, then, extends from 10^{-10} W/cm^2 to 1 W/cm^2.

3.4 Scanning

In the previous discussion we described the measurement of the acoustic intensity at a single point of the pellicle illuminated by the light beam. The displacement amplitude and, hence, acoustic intensity at every point of the pellicle may be measured by scanning the light beam over the pellicle. A method that maintains the Michelson arrangement and permits scanning the light beam over a large aperture with a minimum of large elements is shown schematically in Fig. 9. The deflection system is put in one arm of the interferometer and is designed so that the light beam, at every position, is normally inci-

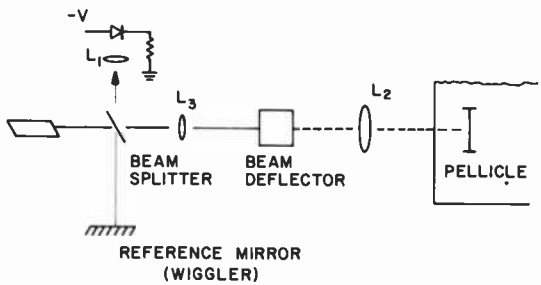


Fig. 9—Basic ultrasonovision system, with means for scanning included in one path of interferometer.

dent on the pellicle. The light beam is reflected from the pellicle and passes back through the deflection system to the beam splitter, where it recombines with the beam from the reference mirror. The advantage of this arrangement is that the beam splitter and reference mirror need be no larger than the diameter of the light beam incident on the interferometer—on the order of one millimeter—while the scanned aperture may be much larger.

A practical arrangement for the deflector, which uses mirror galvanometers for the deflection elements, is shown in Fig. 10. Lenses L_2 and L_3 form a telescope that images the first galvanometer (G_1), used for vertical deflection, into the second galvanometer (G_2) which deflects the beam horizontally. Since galvanometer G_2 is at the front

focal plane of lens L_4, the beams exiting from L_4 are parallel to the axis of L_4. As the galvanometers rotate, the exit beam is displaced proportionately (by the factor $d = f_4 \cdot \theta$, where $\theta/2$ is the mirror rotation) and at every position will be normally incident on the pellicle. Lenses L_1, L_2, L_3 together form an equivalent lens of focal length

$$f = \frac{f_1 f_3}{f_2} \qquad [24]$$

with back focal plane at galvanometer G_2. The size of the beam at the pellicle is proportional to the ratio f_4 to this effective focal length. With $f_1 = f_4$ and $f_2 = f_3$ the size of beam at the pellicle is approximately equal to the beam incident on L_1.

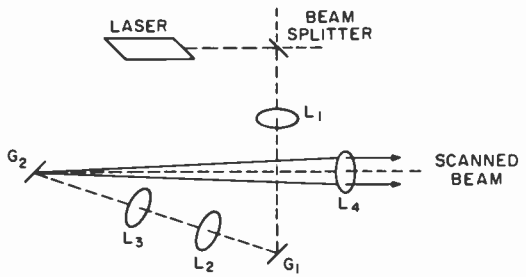


Fig. 10—Optical arrangement used to provide uniform laser beam scanning over the pellicle. It consists of vertical galvanometer (G_1), horizontal galvanometer (G_2), telescope lenses (L_2, L_3) and relay lenses (L_1, L_4).

With this arrangement, and with the pellicle perpendicular to the axis of lens L_4, the beam reflected back through the system will coincide exactly in position, size, and angle with the incident beam. This is true regardless of the position of the scanned beam on the pellicle.

The aperture of L_4 must be at least as large as the pellicle; the other lenses may be much smaller. The only requirement on the other lenses is that their f numbers (the ratio of focal length to aperture) must be no larger than the number of lens L_4. The f number of L_4 is determined by the angular range of the galvanometers and the high cost of very small f number lenses. In practice L_4 is a 6-inch aperture f:5 lens and the commercially available galvanometers are capable of deflecting the beam over at least 12°. With these, a pellicle size and, hence, a system aperture of 6 inches (150 mm) is possible.

For high-speed deflection, galvanometer G_2 could be replaced by an acousto-optic deflector, with minor changes in the optical arrange-

ment. Since, as we will see below, the frame rate is limited by other considerations to about four frames/sec, the speed of the galvanometers is adequate.

As the beam scans over the pellicle, the motion at every point is measured as described above. Thus the acoustic field over a large aperture is accurately mapped. The field may be qualitatively visualized by using the signal to brightness-modulate a cathode ray tube that is scanned in synchronism with the laser beam.

Several additions to the basic optical arrangement of Fig. 10 should be included to avoid spurious signals caused by light from the interferometer being reflected back to the laser and to minimize geometric distortions caused by nonlinearities in the galvanometers. These are described in Appendix 2.

3.5 Stability ("Wiggler")

In Eq. [10] it was assumed that the relative phase difference $\varphi_p - \varphi_r$ was 90°. The reason for this may be qualitatively seen in Fig. 11

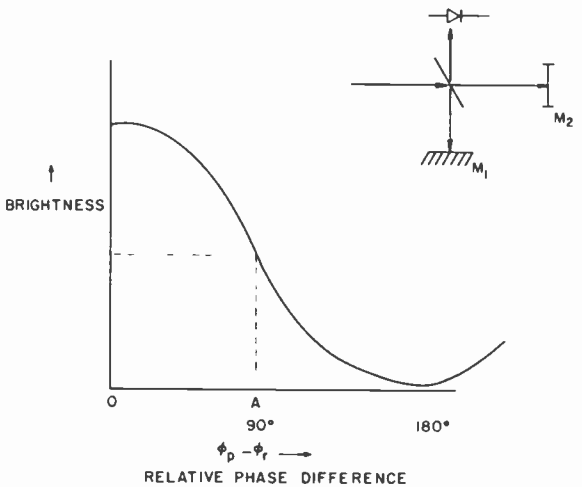


Fig. 11—System response as a function of relative phase shift between pellicle and reference beam mirror.

which is a plot of Eq. [10] (with $\Delta \approx 0$) as a function of the relative phase difference. (In the following analysis, for convenience, we define $\varphi_{pr} \triangleq \varphi_p - \varphi_r$.) At $\varphi_{pr} = 90°$ the slope of the curve—the ratio of

optical intensity change to phase difference change—is maximum. At this point the response of the system, which is proportional to the slope, is also maximum. At other values of the relative phase, the slope and, hence, the response is smaller; for example, at $\varphi_{pr} = 0°$ and $\varphi_{pr} = 180°$ it is zero.

Practical arrangements of interferometers are susceptible to mechanical and thermal drift, so that it is difficult to maintain the condition $\varphi_{pr} = 90°$. Further, in large-aperture systems such as that described here, it is difficult to achieve sufficient flatness, less than $\frac{1}{8}$ wave, over the entire aperture. This lack of flatness would cause such large variations in the output as to make a practical system useless. While feedback methods can be used to avoid these problems, we employ a simpler open-loop method.

The relative phase difference is purposely varied over at least 180° by vibrating, or "wiggling" the reference mirror. (Other methods of varying the phase are possible but wiggling the reference mirror is simplest.) The effect is that at least once per cycle of the wiggler (the vibrating mirror) the operating point of the system will pass through the point $\varphi_{pr} = 90°$, despite large mechanical or thermal variations and to a great extent, regardless of lack of exact flatness of the pellicle.

Since the system response is maximum at $\varphi_{pr} = 90°$, the peak value of the measured signal will give the same value as if the system were always adjusted to $\varphi_{pr} = 90°$.

As long as the net phase change in the system, including effects of external disturbances and the wiggler motion, is greater than $\frac{1}{4}$ wave over the period of the wiggler, the system will be stable. It shows none of the effects that plague ordinary interferometers, such as drift due to thermal or mechanical changes or air currents. It is stable even in the presence of violent mechanical shock due to shaking of the table or water waves set up when samples are placed in the water tank.

There are several constraints on the range of the allowable wiggler frequencies.

The lower limit of the wiggler frequency is determined by the scan rate of the deflector; the wiggler frequency must be sufficiently high to allow at least one cycle of the wiggler during each spot (or resolution element) of the scan. If the frame time of the system is T and the number of elements is N^2 then the wiggler frequency is constrained to

$$f_w > \frac{N^2}{T}. \qquad [25]$$

As an example, with $N = 100$, $T = 0.5$ sec (i.e., 2 frames/sec), the lower limit of the wiggler frequency is 20 kHz.

There are three constraints on the upper limit of the wiggler frequency. These can be explained as follows. From Eq [9], the expression for the detected signal is

$$i_s = 2\eta\sqrt{I_r I_p}\cos\left[\varphi_{pr} + \frac{4\pi\Delta}{\lambda}\sin(\omega_s t)\right]$$

$$= 2\eta\sqrt{I_r I_p}\left\{\cos\varphi_{pr} - \frac{4\pi\Delta}{\lambda}\sin(\omega_s t)\sin\varphi_{pr}\right\}. \quad [26]$$

where the condition $\Delta/\lambda \ll 1$ has been invoked. Let

$$\varphi_{pr} = \varphi_0 + \frac{\pi}{2}\sin(\omega_w t), \quad [27]$$

where $\omega_w = 2\pi f_w$ and φ_0 is an arbitrary phase that in general is a slowly varying function of time. The detected signal becomes

$$i_s = 2\eta\sqrt{I_r I_p}\left\{\cos\left[\varphi_0 + \frac{\pi}{2}\sin(\omega_w t)\right]\right.$$

$$\left. - \frac{4\pi\Delta}{\lambda}\sin(\omega_s t)\sin\left[\varphi_0 + \frac{\pi}{2}\sin(\omega_w t)\right]\right\}. \quad [28]$$

The expansion of the first term of Eq. [28] gives

$$\cos\left[\varphi_0 + \frac{\pi}{2}\sin(\omega_w t)\right] = \left[J_0\left(\frac{\pi}{2}\right)\right.$$

$$\left. + 2\sum_{n=1}^{\infty} J_{2n}\left(\frac{\pi}{2}\right)\cos(2n\omega_w t)\right]\cos\varphi_0$$

$$- 2\left[\sum_{n=1}^{\infty} J_{2n+1}\left(\frac{\pi}{2}\right)\sin(2n+1)\omega_w t\right]\sin\varphi_0. \quad [29]$$

Higher harmonics of the wiggler frequency fall near the ultrasonic signal frequency range. The first constraint on the wiggler frequency is that the amplitude of this harmonic be less than the amplitude due to the ultrasonic signal. This may be written

$$J_m\left(\frac{\pi}{2}\right) < 4\pi\frac{\Delta}{\lambda}, \quad [30]$$

where $m \approx \omega_s/\omega_w$. From our previous discussions (Eq. [14] and the arguments following Eq. [20]) the smallest value of Δ is of the order

of 10^{-4} Å, so the condition is

$$J_m\left(\frac{\pi}{2}\right) < 2 \times 10^{-7}. \qquad [31]$$

Examination of Bessel functions[7] shows that

$$m \geq 10 \qquad [32]$$

will satisfy the condition.

If the lowest expected ultrasonic frequency is of the order of 1 MHz, the highest allowable wiggler frequency is 100 kHz. A high-pass filter, with a cutoff frequency below 1 MHz can be used to completely suppress the effect of the wiggler on the detected signal. (A narrow bandpass filter with center frequency at the ultrasonic frequency could also be used.) The signal after the filter is

$$i_s = 8\eta\pi\frac{\Delta}{\lambda}\sqrt{I_r I_p}\,\sin(\omega_s t)\,\sin\left[\varphi_0 + \frac{\pi}{2}\sin(\omega_w t)\right] \qquad [33]$$

The ideal response of the peak detector is $8\,\eta\pi(\Delta/\lambda)\sqrt{I_r I_p}$, but due to the randomness of φ_0 there is an unavoidable error in the actual response. This error may be minimized if the ratio of the acoustic frequency to the wiggler frequency is sufficiently high. (Intuitively, this increases the probability that the relative phase difference will be 90° at the same time that $\sin(\omega_s t) = 1$.) A detailed analysis is given in Appendix 4. Fig. 12 is a curve of the percent error of an ideally fast peak detector as a function of the frequency ratio (ω_w/ω_s). The error will be less than 1% if the frequency ratio is $\leq 1/15$.

A final constraint on the wiggler frequency is imposed by the use of pulsed acoustic signals. The period of the wiggler must be less than the duration of the acoustic pulse. In practice, wiggler frequencies between 25 and 80 kHz are used.

While there are a number of ways to achieve the necessary phase modulation, a particularly simple method is to mount the reference mirror on a piezoelectric disc that is chosen to be resonant at the desired wiggler frequency. The peak displacement of the disc must be $\lambda/4$, or approximately 1500 Å. If the thickness of the disc is greater than 0.015 cm, this motion is well within the elastic limits of piezoelectric materials.[8]

3.6 Optical Flatness

The criterion for maximum response is that the relative phase differ-

ence between the reference and pellicle beams be 90°. It has already been demonstrated that the use of a wiggler allows this criterion to be satisfied even in the presence of thermal and mechanical disturbances. In this section the effect on the wiggler of the requirement of optical flatness of the pellicle is discussed. The remarks also pertain to the flatness of the other components of the interferometer.

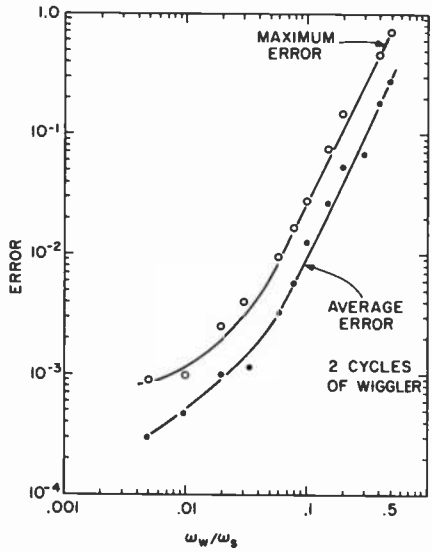


Fig. 12—The maximum and average error as a function of ω_w/ω_s ratio. The number of wiggler cycles in a burst is two.

If the laser beam is infinitesimally small, there is no stringent requirement on the pellicle flatness; because of the action of the wiggler, the relative phase difference will be 90° at some instant, regardless of surface deviations on the pellicle.

With finite sized laser beams, the effect of surface flatness becomes manifest. To find the response in this case the variation in phase due to deviations in the pellicle surface must be included by summing the total phase shift over the diameter of the beam. This is done in Appendix 1. Fig. 13 shows the response as a function of the ratio of the beam diameter to the period of the surface ripple of the pellicle. (As is usual in optics, the deviation from flatness of the pellicle is assumed to be sinusoidal.) The response is within 90% of the ideal, even with pellicles having 3 waves/inch. This is well within the experimen-

tally measured values of 1 to 2 waves per inch with commercially available pellicles.

It is clear that the use of the wiggler not only removes the problems of mechanical and thermal disturbances normally found in interferometers, but also considerably relaxes the usual stringent criteria on optical flatness of the elements used.

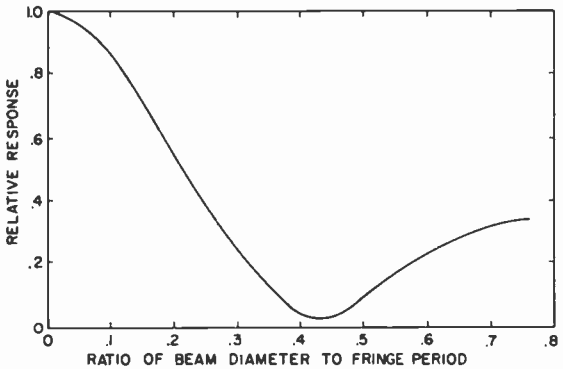


Fig. 13—Effect of wiggler on interferometric response as a function of the ratio of laser beam diameter to period of surface ripple of pellicle.

3.7 Complete System

The complete system, incorporating the deflection optics, wiggler, signal display, and visual display is shown in Fig. 14. The pellicle is mounted in a large water tank, the front wall of which is made of reasonably good optical quality glass. Insonified objects may be acoustically imaged onto the pellicle, as shown on the figure, by acoustic lenses made of Styrene.[9] For the determination of radiation patterns, for example from transducers, the lenses are removed.

The optical signal is detected by a silicon photodiode, converted to a proportional current which is then filtered, amplified, and displayed on an oscilloscope (the signal oscilloscope). Once the system has been calibrated, which means that the light from the reference mirror and pellicle is measured and the amplifier gain known, the voltage on the oscilloscope is an accurate representation of the acoustic displacement amplitude at every point on the pellicle. The galvanometers may be stopped to accurately measure the intensity at a particular point or may be scanned over any desired portion of the pellicle.

The acoustic field is visualized by using the signal to brightness modulate a second CRT. This CRT is scanned synchronously with the galvanometers so that the brightness of every point of the display is proportional to the displacement amplitude of the acoustic wave at the conjugate point on the pellicle. Since the brightness of available CRT's is not a linear function of the voltage applied to the cathode, means for correcting the gamma of the system are employed; this is briefly described in Appendix 5.

Thus with the ultrasonovision system, the acoustic wave can be measured quantitatively (on the signal scope) and observed qualitatively (on the display scope).

The limit on the frame time is set by geometrical considerations of the distance between the pellicle and the front wall of the tank. As seen in Fig. 14, a wave passing through the pellicle will travel to the

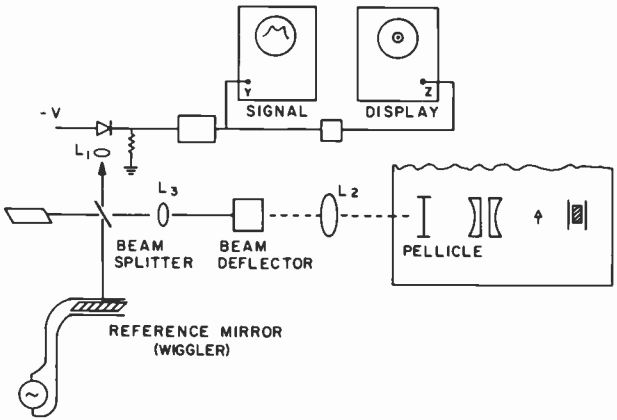


Fig. 14—Schematic of complete system, including oscilloscope and wiggler. Also indicated is acoustic lens arrangement to image insonified object onto pellicle.

front wall, be reflected, and travel back to the pellicle. The returning wave will interfere with the incident wave. To avoid this, pulsed acoustic waves are used and the duty cycle adjusted so the reflected wave and the following acoustic wave do not pass through the pellicle at the same time. This duty cycle is the main consideration in determining the frame rate of the ultrasonovision system.

The electronics for the system are schematically shown in Fig. 15. The detection electronics has a receive gate synchronized to the

pulsed acoustic wave so the signal detected and processed is due to the transmitted wave and not the reflected wave, or other reverberating waves in the tank. The repetition rate of the pulses depends on the spacing between the pellicle and the tank wall, while the duration of each pulse is determined by the scan rate. Typical values used are: a scan rate of 50 μsec/mm, a pellicle-to-wall spacing of 60 mm, and a duty cycle of $1/7$, with a period between pulses of 150 μsec.

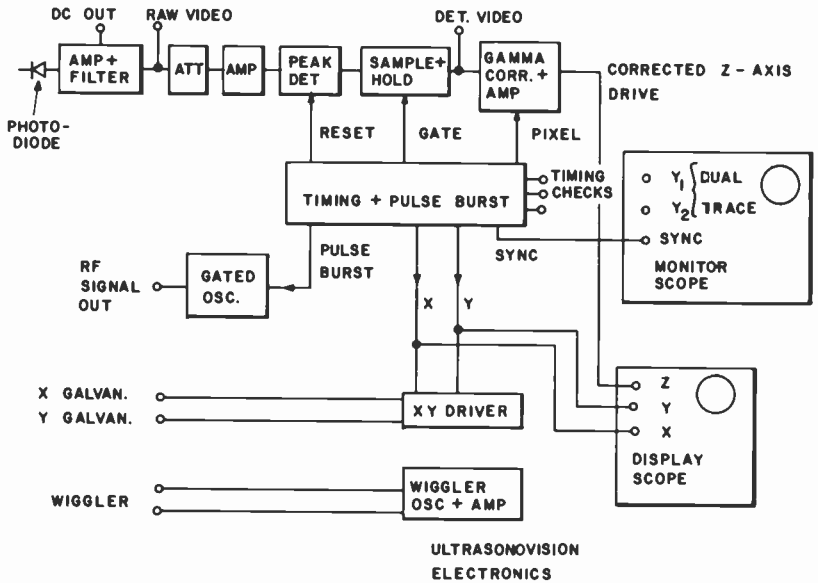


Fig. 15—Block diagram of electronic system.

The gate width, or sample time, is made approximately equal to the acoustic pulse width (~20 μsec in the example above). Since this is less than the time that any spot is illuminated by the scanner, the detected signal is "held" electrically for the balance of the period.

By synchronizing the initiation and rate of the laser scanner with the acoustic pulses, we can take advantage of the finite propagation time (1.5 mm/μsec in water) with the result that the acoustic signal can be range gated. By proper synchronization of the scanner and adjustment of the detector gate, small volumes of the insonified object can be examined while signals from other parts of the object are rejected.

The primary attribute of the system is that it allows the quantitative and qualitative evaluation of ultrasonic images with high resolution. Several examples of imaging have been given above. As a further

example, illustrating quantitative analysis, Fig. 16 shows an image of a test object, which is a piece of aluminum caning containing large (5 mm) and small (2.5 mm) holes. The object was insonified at 3.0 MHz and imaged onto the pellicle with an f:2 acoustic lens. The image is well resolved.

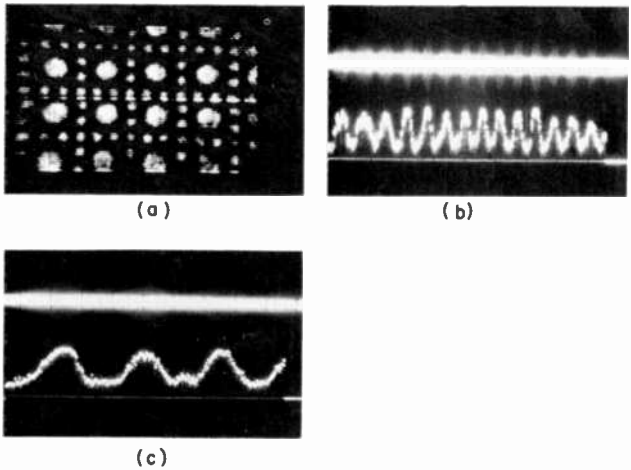


Fig. 16—Ultrasonic image of test object, aluminum sheet with small (2.5 mm) and large (5 mm) holes: (b) shows signal from one horizontal scan through line of small holes and (c) is expanded view of (b), demonstrating the system resolution. The frequency is 3 MHz and an f:2 lens was used.

The quantitative aspects of the system are shown in the lower part of Fig. 16, which is the electrical signal from one scan line through the image of the small holes of the pattern. The measurement of the electrical signal gives a rapid and accurate measure of the acoustic intensity at every and any point of the image.

Another example is shown in Fig. 17, which shows the image of an excised breast tissue with a malignant tumor. Again the ultrasonic image is seen in the top part of the image and the relative acoustic transmission (again measured across one line of the image) is shown in the lower part of the figure. It has been found by such measurements that some types of malignant tissues have a relatively low transmission as compared to normal tissue.

Summary

The ultrasonovision system is an accurate, high-resolution system for the measurement and visualization of ultrasonic images. It is easily

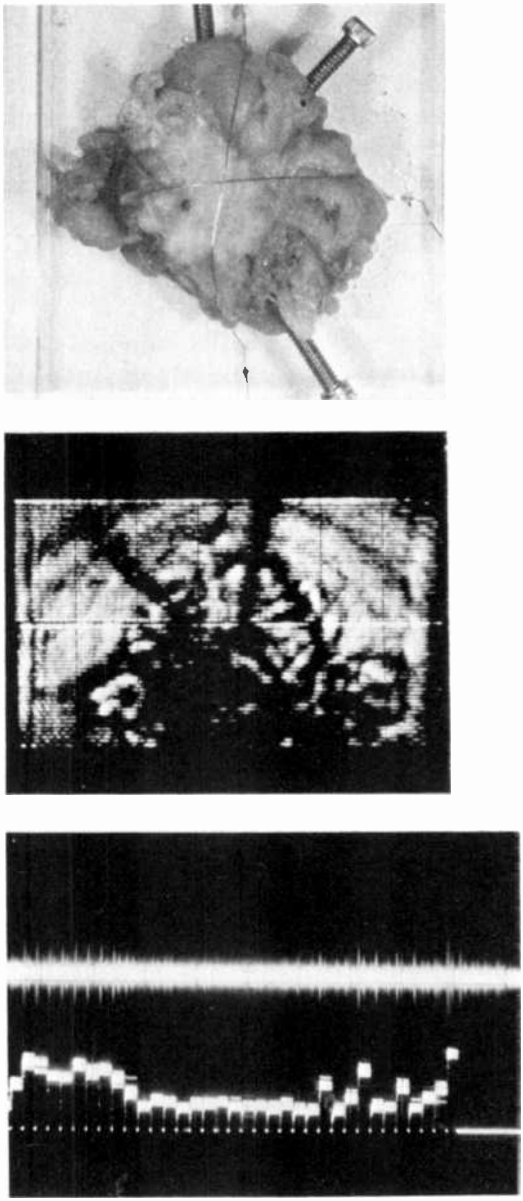


Fig. 17—Excised breast tissue with malignancy: top is optical picture; middle shows ultrasonic image (dark area in center with screws protruding is the malignant region); bottom is electronic signal from one horizontal scan through acoustic image.

calibrated and accurate enough to serve as a primary standard in the measurement of ultrasonic intensity. It has a uniform frequency response, extending to at least 10 MHz. It has an angular response that is flat to ±40° and wavelength limited in resolution. The sensitivity is linearly dependent on the intensity of the laser used (5 nW/cm^2 with a 15 mW laser) and its response is linear over 8 orders of magnitude. Although primarily intended as a laboratory instrument for the study of ultrasonic interactions and propagation characteristics, with the use of large lasers its sensitivity is adequate for some clinical applications (e.g., the imaging of breasts and appendages).

These attributes, together with its relative simplicity of construction and use, should ease and encourage the serious study of ultrasonic fields and increase the eventual utility of ultrasonic methods.

Acknowledgments:

We gratefully acknowledge the advice and generous assistance of several members of the staff of Hahnemann Medical College and Hospital, in particular Dr. M. Haskin, Chairman of Department of Diagnostic Radiology, Dr. C. Calderone, Ass't. Chief of Pathology, and Dr. B. Kingsley, head of ultrasonic methods in the cardiology department. We are indebted also to M. Haran of the Bureau of Radiological Health, FDA, who first suggested consideration of the acousto-optic interaction. Last, we want to thank Dr. Jan Rajchman of the RCA Laboratories for his constant support and encouragement.

Appendix 1—Effects of Departure from Flatness in the Pellicle

The expression for the signal from the interferometer of ultrasonovision is

$$i_s(t) = 2\eta\sqrt{I_r I_p}\ \cos\left(\varphi_p - \varphi_r + \frac{4\pi}{\lambda}\Delta\ \sin\omega_s t\right), \qquad [34]$$

in which φ_p and φ_r are phase terms that are measures of how far the pellicle and reference mirror is from the beam splitter. In the case of ideally flat components, $\varphi_p - \varphi_r$ is not a function of position on the pellicle. In the practical case, due to departures from ideal surface flatness, φ_p is a function of position. In the case of an infinitesimally narrow laser beam this is of no consequence, since the action of the wiggler will make $\varphi_p(x) - \varphi_r = 90°$ at some instant during the period of the wiggler. In the case of a finite-diameter laser beam at the pellicle, the lack of flatness is important and the actual signal must be

found by summing the response over the diameter of the laser beam. The output in this case is

$$i_s(t) = 2\eta\sqrt{I_r}\int\sqrt{I_p}\cos\left[\varphi_p(x) - \varphi_r + \frac{4\pi\Delta}{\lambda}\sin(\omega_s t)\right]dx. \quad [35]$$

To calculate this, assume the surface of the pellicle may be described by

$$h = h_o + \delta\sin\frac{2\pi x}{\Lambda}, \quad [36]$$

where δ is the deviation from flatness and Λ is the spatial period of the deviation. Then with

$$\varphi_p(x) = \frac{4\pi h_o}{\lambda} + \frac{4\pi\delta}{\lambda}\sin\frac{2\pi x}{\Lambda} \quad [37]$$

$$= \varphi_{po} + \frac{4\pi\delta}{\lambda}\sin\frac{2\pi x}{\Lambda}. \quad [38]$$

Eq. [35] becomes

$$i_s(t) = 2\eta\sqrt{I_r}\int\sqrt{I_p}\cos\left[\varphi_{po} - \varphi_r + \frac{4\pi\delta}{\lambda}\sin\left(\frac{2\pi x}{\Lambda}\right) + \frac{4\pi\Delta}{\lambda}\sin(\omega_s t)\right]dx. \quad [39]$$

Defining $\varphi_{po} - \varphi_r = \varphi_{pr}$ and using trigonometric identities,

$$i_s(t) = \left[2\eta\sqrt{I_r}\cos\left(\frac{4\pi\Delta}{\lambda}\sin(\omega_s t)\right) \times \int\sqrt{I_p}\cos\left(\varphi_{pr} + \frac{4\pi\delta}{\lambda}\sin\left(\frac{2\pi x}{\Lambda}\right)\right)dx\right] + 2\eta\sqrt{I_r}\left[\sin\left(\frac{4\pi\Delta}{\lambda}\right) \times \sin(\omega_s t)\right)\int\sqrt{I_p}\sin\left(\varphi_{pr} + \frac{4\pi\delta}{\lambda}\sin\left(\frac{2\pi x}{\Lambda}\right)\right)dx\right]. \quad [40]$$

With the approximation $4\pi\Delta/\lambda \ll 1$, and remembering that we are only interested in components of the signal at the acoustic frequency (ω_s), the first term of Eq. [40] may be neglected.

Assuming that

$$I_p = \sqrt{I_{po}}\cdot\frac{1}{a}\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{2x^2}{a^2}\right\}, \quad [41]$$

that is, a Gaussian beam (which is the case in practice), the expres-

sion for the signal is

$$i_s(t) = 2\eta\sqrt{I_r I_{po}}\ \sin\left(\frac{4\pi\Delta}{\lambda}\ \sin(\omega_s t)\right)\frac{1}{a}\sqrt{\frac{2}{\pi}}$$
$$\int \exp\left\{-\frac{x^2}{a^2}\right\}\ \sin\left(\varphi_{pr} + \frac{4\pi\delta}{\lambda}\ \sin\left(\frac{2\pi x}{\Lambda}\right)\right)dx. \qquad [42]$$

Recognizing that the expression

$$2\eta\sqrt{I_r I_{po}}\ \sin\left(\frac{4\pi\Delta}{\lambda}\ \sin(\omega_s t)\right) \approx 8\eta\frac{\pi}{\lambda}\sqrt{I_r I_{po}}\Delta\ \sin(\omega_s t) \qquad [43]$$

is the signal in the ideal case of a vanishingly narrow light beam, the term

$$\frac{1}{a}\sqrt{\frac{2}{\pi}}\int \exp\left\{\frac{-x^2}{a^2}\right\}\ \sin\left(\varphi_{pr} + \frac{4\pi\delta}{\lambda}\ \sin\left(\frac{2\pi x}{\Lambda}\right)\right)dx \qquad [44]$$

is the degradation, or departure from ideal, obtained with a pellicle that is not perfectly flat and with a Gaussian light beam.

Eq. [44] has been numerically integrated with the aid of a digital computer; the results are plotted in Fig. 13. With the ratio of the period of the surface ripple of the pellicle (Λ) to the beam diameter greater than 10:1, the departure of the response from the ideal is less than 10%. Thus for a beam diameter of 1 mm, a pellicle flatness of 2–3 waves/inch will cause only a 10% departure from the ideal response. Commercially available pellicles have been experimentally shown to have flatness on the order of one wave/inch even when used under water; the error will then be less than 3%.

Clearly the action of the wiggler is to greatly reduce the usually stringent requirement for flatness of elements used in interferometers. It must be noted that the results derived here hold equally well for the effects of deviations from flatness of the other elements used in the optical arrangement.

Appendix 2—Modifications to the Optical Arrangement

The arrangement of Fig. 10 has two fixed points through which the pellicle beam must pass. The first is the beam splitter and the second is the galvanometer mirror, which is in the back focal plane of lens L_4. Because of this, half of the light reflected from the pellicle (that portion split by the beam splitter that does not go to the photodiode) must return to the laser. Further, since the reference beam is arranged to be collinear with the pellicle beam at the photodiode, half

of the reference beam must also return to the laser. Thus in the properly arranged system, approximately half the light from the laser is reflected back to the laser. This must be avoided for several reasons. When a powerful laser is used the returned beam can damage the laser. If the returned light re-enters the laser cavity, fluctuation in the light output will occur. Finally, some of the returned light will be re-reflected by the laser optics and eventually lead to spurious detected signals due to interference with the primary beam from the laser.

There are several ways to avoid this. The simplest is with the use of an isolator, which consists of a linear polarizer oriented to pass the light from the (polarized) laser followed by a quarter wave plate. The light entering the interferometer will be circularly polarized and the reflected beam, still circularly polarized will emerge from the quarter wave plate linearly polarized at right angles to the preferred direction of the linear polarizer. The light will not re-enter the laser.

In a second arrangement the linear polarizer may be replaced by a polarizing beam splitter. This is a device that totally transmits light of one polarization and completely reflects (in a direction at right angles to the transmitted beam) light of the orthogonal polarization. In this case the reflected beam again does not reach the laser, but is reflected (at 90° to the laser beam direction) away from the laser. This light, containing both light from the pellicle and light from the reference mirror, also shows the interference between the pellicle and reference beams and, with the aid of a photodiode, can be used to detect the pellicle motion due to the ultrasonic wave. Further, it can be shown that this signal is 180° out of phase with the signal from the primary interference (from the original beam splitter). When a differential amplifier is used to combine the signals from the primary interference and this second interference, the detected signal will be doubled and common-mode noise will be suppressed.

There is a third possible arrangement that combines the features of the first two—namely, isolation of the laser from the interferometer and differential detection with common-mode suppression—and that further allows "beam balancing" between the pellicle and reference beams, which is necessary to achieve maximum light efficiency. This arrangement is shown schematically in Fig. 18.

A polarizing beam splitter is used as the primary beam splitter. Light from the laser is incident on the beam splitter after first passing through the polarization rotator (which can consist of two quarter-wave plates). By this means the ratio of light intensities in the two legs of the interferometer may be varied depending on the angle of the plane of polarization of the light incident on the beam splitter.

Quarter-wave plates are used in both legs of the interferometer and are arranged so the light transmitted through them will emerge circularly polarized. The reflected beams from the reference mirror and pellicle will, after again passing through the quarter wave plates, once more be linearly polarized, but with directions of polarizations at right angles to the incident planes of polarizations. At the beam splitter, all of the returning light will be directed towards the photo-

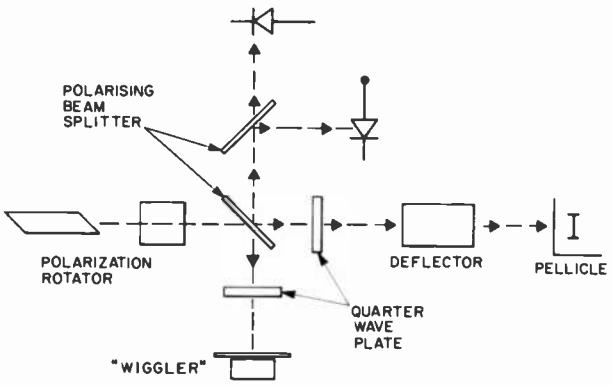


Fig. 18—Arrangement to isolate laser from interferometer, which allows maximum light utilization and suppression of common-mode noise.

diodes, with none of the light returning to the laser. (The light initially reflected by the beam splitter towards the reference mirror will, due to the change in polarization, be completely transmitted towards the photodiodes on its return. Similar results obtain for the light in the pellicle direction.)

The light beams from the two legs of the interferometer are polarized at right angles to each other and will not interfere. If an analyzer with its axis at 45° to the angles of polarization is used, half the light (with components of polarization along the axis of the analyzer) will interfere. By using a polarizing beam splitter as the analyzer, the other half of the light will also interfere (and be reflected at 90° to the original beam) and can be detected by a second photodiode. Again there will be a 180° phase difference between the two components, which allows detection with suppression of common-mode noise.

Because of the ability to vary the ratio of the intensities of light in

the two legs of the interferometer and thus, to an extent, compensate for the unavoidable losses due to the many elements in the pellicle leg of the interferometer, this last arrangement allows better light utilization, or higher system efficiency, than the other arrangements.

Appendix 3—Correction of Galvanometer Nonlinearities.

During experiments on the ultrasonovision system, it was found that the galvanometer response (angle vs. applied voltage) was not linear, mainly due to inertial effects. A simple means was found to compensate for these nonlinearities and thereby minimize geometric distortion in the displayed image. The arrangement used is schematically illustrated in Fig. 19.

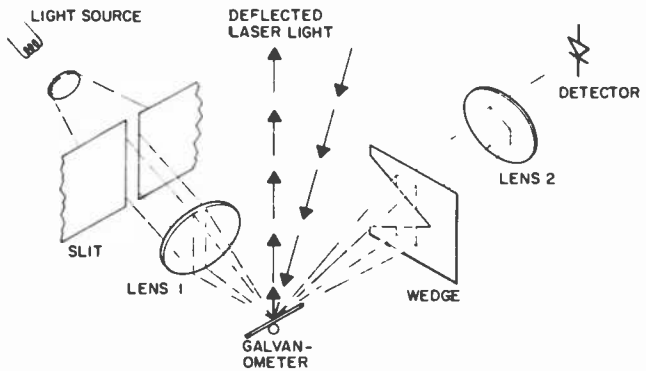


Fig. 19—Arrangement to measure and minimize effect of galvanometer nonlinearity on geometric fidelity of displayed image.

A light source, powered by a regulated dc supply, illuminates a slit cut into an opaque shield. A lens (lens 1) is placed so that it projects a demagnified image of the light source on the galvanometer and at the same time (since the light is reflected by the galvanometer) projects a 1:1 image of the slit onto a wedge. The wedge has been cut into a second opaque shield. As the galvanometer rotates, more or less of the light in the image of the slit will pass through the wedge. This light is collected by a second lens (lens 2) and focused onto the photodiode detector. In this way the light on the photodiode and, therefore, its photocurrent, will be linearly proportional to the angle of rotation of the galvanometer. The photocurrent produced may be used to generate the error signal in a feedback system to correct the nonlinearities

of the galvanometer or, more simply, it may be used to drive the deflection plates of the display oscilloscope. The latter course has been found to be adequate in practice.

This arrangement is quite simple and cheap, requiring only simple condensing lenses and a slit and wedge cut into pieces of cardboard for its implementation. The elements are laid out so as not to block any of the light of the interferometer.

Although the correction may be used for both galvanometers of the deflector, it has been seen in practice that correcting the nonlinearities of only the horizontal galvanometer (fast axis) is sufficient.

Appendix 4—Error Due to Wiggler

The measured amplitude of the signal is determined by the peak-detector as the maximum value of the product of the sensitivity curve and the signal. The sensitivity curve varies sinusoidally between plus and minus one due to the wiggler. The phase of the sensitivity curve varies as air currents, thermal drifts, etc., move the bias point of the system. The signal on which the peak-detector works is

$$S(t) = \{\sin[\varphi + (\omega_w t)]\} A \sin(\omega_s t), \qquad [45]$$

where the term in curly brackets is the sensitivity change at the wiggler frequency ω_w and the second term is the ultrasonic signal at frequency ω_s.

The error in determining the amplitude A of the signal is defined as A minus the maximum positive value of $S(t)$ during the signal burst. Examination of Eq. [45] shows that the error is a function of three variables: the ratio of wiggler to signal frequency, ω_w/ω_s; the phase, φ; and the number of cycles of the wiggler during the signal burst over which the peak detector can work.

One would expect that the lower ω_w/ω_s, the less would be the error, because the sensitivity would be changing slowly compared to the ultrasonic signal and, therefore, the peak values of both would more nearly coincide; the chances of the peaks coinciding also increases as the number of cycles of the wiggler per signal burst increases.

As the phase is a random function, the error dependent upon it is random as well. We will describe the inaccuracy in terms of the average error,

$$\text{Av. Error} = \sum_{i=1}^{N} \text{error}(\phi_i)/N$$

where N is the number of random phases examined. Of even greater interest is the maximum error that can occur for any phase.

We simulated the action of an ideal peak detector on the signal of Eq. [45]. By taking small increments in the phase φ we could determine the error over the range of this random variable $0-2\pi$. We varied the ω_w/ω_s ratio and the number of wiggler cycles in the detected burst. The average and maximum error were computed.

The relation between ω_w/ω_s ratio and the error for 2 cycles of the wiggler is shown in Fig. 12. The error increases with increasing ω_w/ω_s, as expected.

Fig. 20 is a plot of the error as a function of the number of cycles. (An irrational ω_w/ω_s ratio was used to ensure no harmonic relations

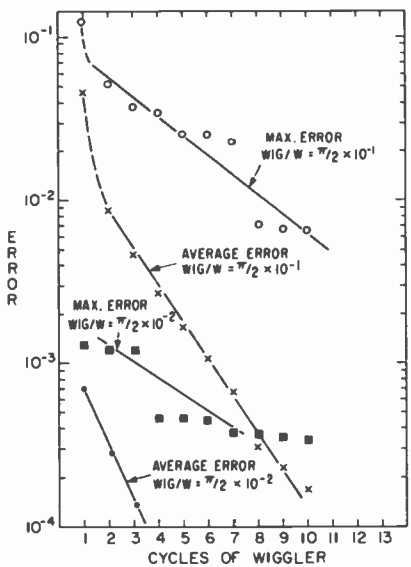


Fig. 20—Error due to wiggler as a function of number of cycles.

would skew the data). The error goes down with increasing number of cycles, again as expected. However, the maximum error falls more slowly with increasing number of cycles than with decreasing ω_w/ω_s ratio. This means that, in general, for a given signal burst length, the lower the wiggler frequency the lower will be the error. (For *slow* peak-detectors this may not hold true; the peak-detector might need several tries at catching the peak.)

We also see that for a ω_w/ω_s ratio $<1/15$, the maximum error is less than 1% in amplitude or 0.09 dB. As the requirement for keeping the higher harmonics of the Bessel functions (introduced by the phase modulating action of the wiggler) out of the signal band also requires a similar ratio (as described in Eq. [32]), the requirement for ratios of $\omega_w/\omega_s < 1/15$ is doubly strong. If the requirement is met, the error will be less than 0.1 dB.

Appendix 5—Gamma Correction

The input to the display device must be predistorted to obtain a display that has a maximum of visual information. If a CRT is used, for example in a TV monitor or an oscilloscope, the display can have approximately 10–12 different gray levels. To match the CRT input voltage for each gray level to a desired signal level, a gray-scale or gamma correction circuit is used. The eye perceives brightness changes logarithmically. The display, therefore, is adjusted to have a brightness step of 2 dB for a linear increment of $1/12$ of full scale (there are 12 allowed levels in the device).

Thus the signal processing first requires level-slicing and then assigning each of the 12 levels a specific, logarithmically scaled brightness level. The first operation is accomplished with an 8-bit A/D converter, and the second by assigning to 12 levels of the 16 (4-bit) levels output voltages that produce the required brightness. The voltage for each level can be adjusted and thus a logarithmic transfer function can be obtained. If desired, another transfer function can be chosen, such as that required for recording the image on film or Polaroid pictures. Using an A/D converter has the additional advantage that the picture elements become available in digital form for further processing, such as averaging, calculation of dB values, etc. Finally, the data can also be used in TV rate playback (scan conversion) if it is stored appropriately.

Appendix 6—Effect of Acousto-Optic Interaction

In the ultrasonovision system the motion of the pellicle gives rise to a phase shift on the light beam reflected from the pellicle. Due to the arrangement used, with the pellicle suspended in the fluid through which the acoustic wave passes, the light beam passes through the ultrasonic wave itself as the ultrasonic wave "emerges" from the pellicle. The acousto-optic interaction, in which the pressure of the acoustic wave causes a change in the index of refraction of the fluid, leads

to an additional phase shift on the light beam. In this appendix we consider the magnitude of this phase shift compared to that due to the motion of the pellicle alone.

The change in index of refraction Δn due to a change in pressure in a fluid may be derived, to a good approximation, from the Lorentz–Lorenz condition[10,11] with the result that

$$\Delta n = \frac{(n^2 - 1)(n^2 + 2)}{6n\rho_o c^2} p, \qquad [46]$$

where n is the static index of refraction, ρ_0 is the static density, c is the acoustic velocity, and p is the "excess pressure" due to the acoustic wave.

The total phase shift due to the index change is found by summing over the interaction length (i.e., the path the light beam takes through the ultrasonic wave), or

$$\varphi_n = \frac{4\pi}{\lambda} \int_o^L \Delta n dx, \qquad [47]$$

where λ is the optical wavelength.

The expression for an acoustic traveling wave is

$$p = \sqrt{2\rho_o c I} \cos\left\{\frac{2\pi}{\Lambda}(x - ct)\right\}, \qquad [48]$$

where Λ is the acoustic wavelength, c again is the acoustic velocity, and I is the acoustic intensity. Combining Eqs. [46], [47], and [48], the expression for the phase shift becomes

$$\varphi_n = \frac{(n^2 - 1)(n^2 + 2)}{6n\rho_o c^2} \frac{4\pi}{\lambda} \sqrt{2\rho_o c} \sqrt{I} \int \cos\left\{\frac{2\pi}{\Lambda}(x - ct)\right\} dx. \qquad [49]$$

If $t = 0$ is the time when the acoustic wave is just incident on the pellicle, the interaction length L is $L = ct$. The phase shift becomes

$$\varphi_n = \frac{(n^2 - 1)(n^2 + 2)}{6n\rho_o c^2} \frac{4\pi}{\lambda} \sqrt{2\rho_o c} \sqrt{I} \frac{\Lambda}{2\pi} \sin(\omega t). \qquad [50]$$

If the acoustic wave is a gated sine wave, as is used in this system, the duration of the signal due to the acousto-optic interaction is as long as the duration of the acoustic signal. Once the wave has com-

pletely passed the pellicle there is no further change in the phase of the light beam, since the interaction length remains constant.

The phase shift due to the acousto-optic interaction may now be compared to that due to the pellicle motion. The pressure of the traveling acoustic wave is, again, as given in Eq. [48]. The displacement amplitude is

$$\Delta = \frac{\sqrt{2\rho_o c I}}{\rho_o c^2} \frac{\Lambda}{2\pi} \sin\left\{\frac{2\pi}{\Lambda}(x - ct)\right\}. \quad [51]$$

or, if the pellicle is at $x = 0$,

$$\Delta = \frac{\sqrt{2}}{c\sqrt{\rho_o c}} \frac{\Lambda}{2\pi}\sqrt{I} \sin(\omega t). \quad [52]$$

The phase shift due to the pellicle motion is

$$\varphi_p = -\frac{4\pi}{\lambda}\Delta. \quad [53]$$

The ratio of the magnitude of the phase shift due to the pellicle motion and acousto-optic interaction is, from Eqs. [50], [52], and [53],

$$\frac{|\varphi_p|}{|\varphi_n|} = \frac{\frac{4\pi}{\lambda c}\frac{\sqrt{2}}{\sqrt{\rho_o c}}\sqrt{I}\frac{\Lambda}{2\pi}}{\frac{4\pi}{\lambda}\sqrt{2\rho_o c}\sqrt{I}\frac{\Lambda}{2\pi}M} \quad [54]$$

$$= \frac{1}{\rho_o c^2 M}. \quad [55]$$

where

$$M = \frac{(n^2 - 1)(n^2 + 2)}{6n\rho_o c^2}.$$

Taking the values for water of $n = 1.33$, $\rho_0 = 1$, and $c = 1.5 \times 10^5$ cm/sec, we obtain

$$M = 1.64 \times 10^{-11}$$

Substituting in Eq. [55], we obtain

$$\frac{|\varphi_p|}{|\varphi_n|} = 2.7.$$

Thus the phase shift due to the pellicle motion is 2.7 times as large as that due to the phase shift from the acousto-optic interaction.

The sign of φ_p is opposite that of φ_n, which means the phase shifts subtract, and so the pellicle motion as calculated by Eq. [12] should be increased (by about 30% in the case where water is the fluid).

References:

[1] K. R. Erikson, F. J. Fry, and J. P. Jones, "Ultrasound in Medicine—A Review," *IEEE Trans. Sonics and Ultrasound*, **SU-21**, p. 144 (1974). This paper is a general review of present ultrasound methods, with an excellent bibliography that includes studies of acoustic properties of normal and pathological tissue.

[2] T. Kobayashi, P. Takatani, N. Hattori, and K. Kimora, "Differential Diagnosis of Breast Tumors," *Cancer*, **33**, p. 940 (1974).

[3] R. Mezrich, K. F. Etzold, and D. H. R. Vilkomerson, "Ultra-sonovision: Interferometric Device for Ultrasonic Visualization and Measurement," 87th Meeting of ASA (1974).

[4] L. Brekhovskikh, *Waves in Layered Media*, p. 61, Academic Press, New York (1960).

[5] D. H. R. Vilkomerson, *Acoustical Holography*, Vol. 4, p. 401, Plenum Press, New York (1972).

[6] A. Gray, G. Matthews, and T. Macrobert, *Bessel Functions*, p. 32, Macmillan Co., London (1952).

[7] E. Jahnke and F. Emde, *Tables of Functions*, Dover Press, New York (1945).

[8] D. Berlincourt, D. Curran, and H. Jaffe, *Physical Acoustics*, Vol. 1A, p. 169, Academic Press, New York (1964).

[9] D. Folds, "Focusing Properties of Solid Ultrasonic Cylindrical Lenses," *J. Acoust. Soc. Amer.*, **53**, p. 826 (1973).

[10] M. Born and E. Wolf, *Principles of Optics*, p. 593, Pergamon Press, London (1965).

[11] T. Heuter and R. Bolt, *Sonics*, p. 352, John Wiley and Sons, New York (1955).

Electronic Processes in Oxide Cathodes

T. N. Chin

RCA Electronic Components, Princeton, N.J. 08540

R. W. Cohen and M. D. Coutts

RCA Laboratories, Princeton, N.J. 08540

Abstract—The rising interest in the use of high-current-density emitters in electron-beam devices prompts further consideration of the electronic processes in an oxide cathode. This paper begins with a brief review of the relevant characteristics of the alkaline-earth oxides used for thermionic emission. The proposed, idealized model in which ion adsorption occurs on the crystalline surfaces accounts for the emission characteristics and the conductivity of the oxide coating. It is possible to correlate the measured values from oxide crystals with those of oxide coatings by using a simplified analysis. Also, some of the underlying limitations in the cathode operations are revealed and discussed in connection with this model.

1. Introduction

There has been extensive work on oxide cathodes concerning various aspects of the operating properties. Most of the early efforts were directed toward those cathodes coated with the crystalline oxide powders. Later investigations on single crystals of barium oxide provided much of the understanding of the semiconducting properties of this compound. A summary of the experimental progress and the theories advanced to explain the operating mechanisms may be obtained from Refs. (1)–(8).

Most of the recent studies pursued the subject from the point of

view of defect chemistry.[9] While experiments are performed to examine the chemical properties of these oxides, the correlation with the operation of a real oxide cathode is yet to be completely determined. The lack of a widely accepted physical model indicates the complexities involved in the cathode operation.

For many years oxide-coated cathodes have been almost universally used in the cathode-ray[10] and other electron-beam[11] vacuum tubes. At present, the oxide cathode is still of considerable practical importance as an electron emitter. In this paper, a simple model is proposed to explain the electronic processes in an oxide cathode. Through a simplified analysis, we estimate and correlate the observed characteristics with the experimentally obtained parameters.

2. Physical Background

2.1 Oxide Coating

Although it is not necessary here to describe the detailed preparation of an oxide-coated cathode, some important characteristics of the oxides are outlined in order to deduce an idealized model. At present the oxide coatings are commonly prepared from the triple carbonates $(BaSrCa)CO_3$. Generally, they consist of about equal amounts of the barium and the strontium carbonates with a small quantity of calcium carbonate. Two types of carbonate particles precipitated from two different solutions are shown in Figs. 1 and 2. The coating density on the cathode surface ranges from 4 to 8 mg/cm^2. For a coating of 2–3 mils thick, there may be 15 to 30 layers of carbonate particles.

The porosity of such an oxide coating usually varies between 20 and 50%. The texture of a typical carbonate coating is given in Fig. 3. Since the carbonates are highly insulating, some charging problems are noticeable in the scanning electron micrograph (SEM). By using the conventional procedure, the carbonate coating is then converted and activated. At the end of activation, the electron current density reaches about 0.5 ampere/cm^2 at 1100°K. When the cathode is at high temperatures, the contrast of SEM pictures is inadequate to show the oxide crystallites. The photograph in Fig. 4 was taken a few seconds after the cathode heating was turned off in order to reveal the activated oxide crystallites.

2.2 Conductivity Data

In addition to the electron emission measurement, the conductivity of the oxide coating has been repeatedly measured by different ap-

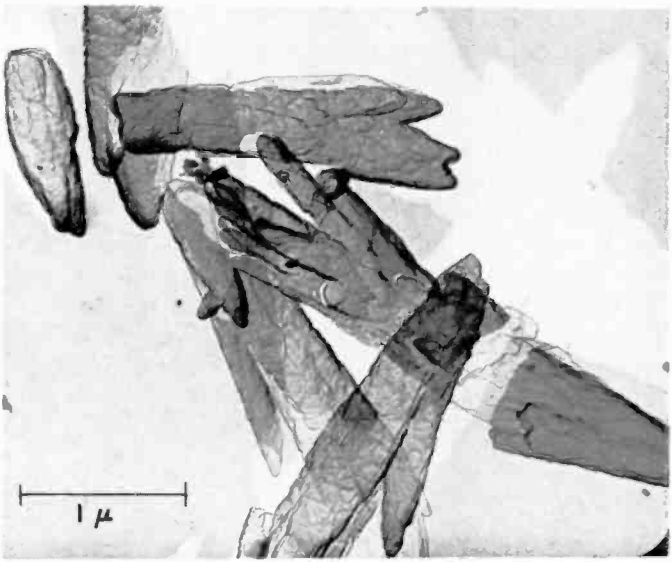


Fig. 1—Sodium-precipitated triple-carbonate crystallites photographed using the carbon-replica technique.

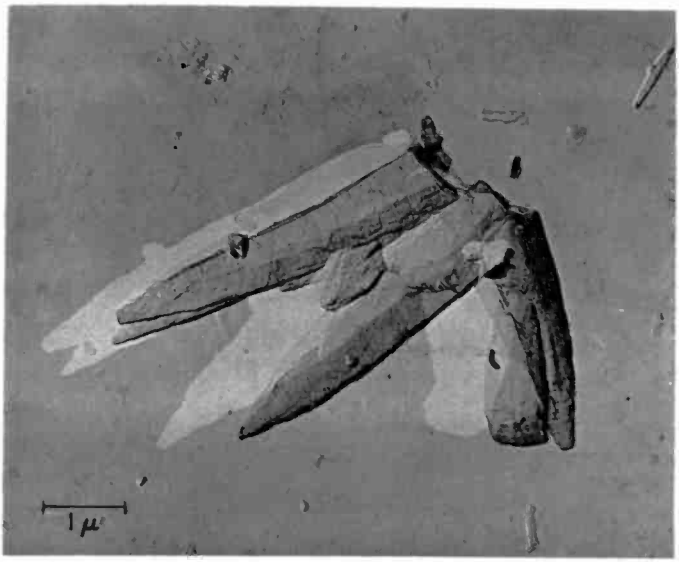


Fig. 2—Ammonium-precipitated triple-carbonate crystallites photographed using the carbon-replica technique.

Fig. 3—SEM photograph of a triple-carbonate coating on a planar cathode before activation; original magnification ×10000 (photo reduced by approximately 15% in printing).

proaches as reported in the literature. Selected values from these measurements appear in Table 1, which gives a list of mean values for the conductivity of single crystals,[12–15] BaO crystallites,[16,17] and (Ba,Sr)O coatings.[18] As with many other polycrystalline materials, the measurements to determine the conductivity are difficult to per-

Table 1—Values of Electrical Conductivity at 1100°K of Alkaline-Earth Crystals and Coatings

	Conductivity $(ohm\text{-}cm)^{-1}$
BaO Single Crystal[12,13]	$1\text{–}2 \times 10^{-4}$
SrO Single Crystal[14]	10^{-7}
CaO Single Crystal[15]	10^{-8}
Unactivated BaO Coating[16,17]	$10^{-5}\text{–}10^{-4}$
Activated BaO Coating[5,16]	$1\text{–}5 \times 10^{-2}$
Unactivated (Ba,Sr)O Coating[18]	$3 \times 10^{-6}\text{–}10^{-4}$
Activated (Ba,Sr)O Coating[18]	$2 \times 10^{-3}\text{–}10^{-2}$

Fig. 4—SEM photographs of triple-oxide crystallites on a planar cathode after activation; original magnifications: (a) ×5000, (b) ×20000 (photo reduced by approximately 15% in printing).

form, and some deviations may occur because of variations in the coating, such as the chemical inhomogeneity and the porosity or contaminations. However, we note for future reference that the conductivity of BaO crystals is about two orders of magnitude lower than that of activated BaO layers.

2.3 Emission vs. Conductivity

In the case of oxide coatings, several investigators[18] have concluded that the electron emission and the electrical conductivity of an oxide-coated cathode are linearly related at each state of activation. For this discussion, it is not necessary to include the pore conduction process, which becomes significant at very high temperatures. Considerable work on photoemission has been reported, but far fewer investigations have been carried out on the thermionic emission properties of BaO crystals. A similar relationship between the thermionic electron emission and the electrical conductivity was sought in the experimental study of BaO crystals. However, little correlation[19] could be established between the electron emission and the electrical conductivity for BaO crystals. This may be additional experimental evidence that the electrical conductivity measured from the oxide coating cannot be interpreted as a bulk property of these oxides.

2.4 Ion Adsorption Effects

In the early studies on the emission characteristics from monatomic layers, such as Cs, Th, and Ba on tungsten emitters, it was deduced[20] that any damage to the surface film, either by evaporation or by oxidation, destroys the emission. Because of the complexity involved in an actual oxide cathode, it is only recently[8] that the adsorption of barium has been identified as an essential factor in thermionic cathodes. Since the evaporation of BaO takes place at a much faster rate than that of SrO and CaO, the adsorption of barium ions on the strontium–calcium oxide is believed to be the activation process in a triple-oxide cathode.

From the texture and activation of the triple oxides, it is evident that barium adsorption can also take place on the interfaces between oxide crystallites. The porosity in the oxide coating allows the formation of such adsorbed interfaces in the same manner as that of the emitting surface. Consequently, electron accumulation will occur at these ion-adsorbed interfaces as well as in the emitting area of the oxide crystallites. As a result the barium adsorption serves not only to increase the electron emission at the cathode surface but also to

enhance the electron conductivity between the oxide crystallites. With the above assumption it is conceivable that the rate of barium-ion adsorption at the interfaces is equal to that at the emitting surfaces. This would explain the linear relation of electron emission and conduction found in these experiments.[18] Since the electron conductivity of BaO crystals cannot be increased through this adsorption process, the relationship of the thermionic emission with the electron conductivity is not expected to be linear. Hence, it is not surprising that the conductivity of the activated-oxide layer is two orders of magnitude higher than that of BaO crystals at high temperatures.

3. Simplified Analysis

3.1 An Idealized Model

With the adsorbed ions throughout the interfaces, the conducting paths along the crystallites are illustrated schematically in Fig. 5. The

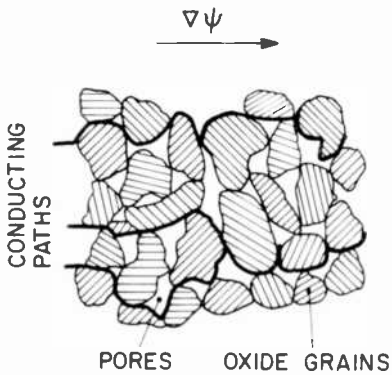


Fig. 5—The structure of oxide grains in a real cathode.

observed conductivity is the sum of these paths along the interfaces of the oxide crystallites. In order to facilitate the estimation of the conductivity of the coating, the oxide crystallites are depicted in Fig. 6 as simple cubes with an edge width D. For simplicity, these identical cubes are stacked in a regular array with some separation between them in each layer. The separations, as indicated in this model, simulate the porosity of a real oxide coating on the cathode. Also, separations are shown only between grains in each layer and not between layers. This is really a simplification rather than an indication of any anisotropy in the formation of the coating.

From this idealized model, it is obvious that the lengths of the conducting paths are not greatly increased from the apparent thickness of the oxide coating. The conductance G of the coating is a sum of the contributions of all the paths i,

$$G = \sum_i \frac{\sigma a_i}{l_i} \cong \frac{\sigma}{l} \sum_i a_i,$$

where the conductivity σ along the path is considered the same for all

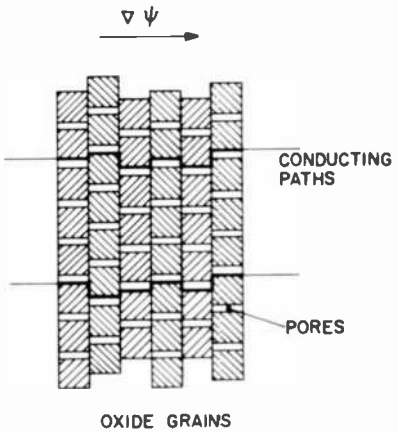


Fig. 6—A simplified model for the oxide grains in the cathode coating.

paths, l is the thickness of the coating, and a_i is the cross-sectional area of the conducting part of the grains. Although this estimate is valid only for cubes, the same reasoning can be employed to approximate the conductivity for grains of other shapes.

In accordance with the simplified model, the path of conductance G is parallel to the electric field, as shown in Fig. 6. A schematic diagram, Fig. 7, of the cross section taken through the oxide grains in a layer illustrates the conductive surface layer on the cubic grains. If the surface layer[21] of cubes is the major conductive part, the cross-sectional area of the conducting path is actually $4L(D-L)$ instead of D^2, the geometric cross section of the cube. Here, L is the approximate depth of the surface layer where the conductivity is greatly enhanced due to the ion adsorption. It will be shown later that for an unactivated grain the effective length L may reach $D/2$, so that electron conduction can be nearly uniform throughout the grain. The

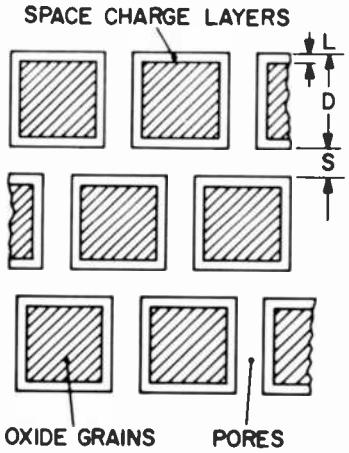


Fig. 7—A cross-sectional view of the oxide grains in a well-activated cathode.

thickness L of the surface layer will decrease as the barium adsorption progresses. In a well-activated coating, the value of L is expected to be only a small fraction of the grain size D.

3.2 Calculation of Space-Charge Accumulation

When the cathode is well activated, the free carriers accumulated at the surface will reach a quasi-equilibrium[22] condition. In Fig. 8, the

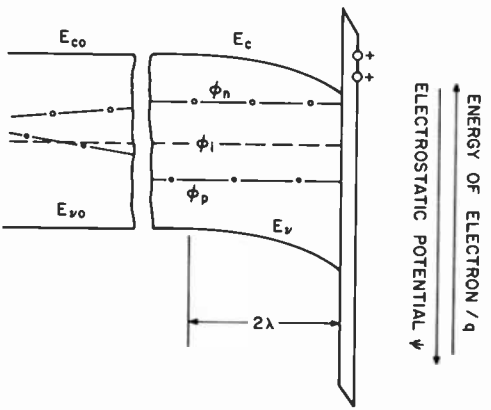


Fig. 8—The energy-level diagram at the surface of a well-activated oxide grain.

energy band edges E_c and E_v near the surface illustrate the effect of space charge accumulation, i.e.,

$$E_c(x) = E_{co} - q\psi(x).$$

and

$$E_v(x) = E_{vo} - q\psi(x).$$

where E_{co} and E_{vo} are the conduction and valence band edges in the bulk, q is the electronic charge, and ψ is the electrostatic potential. In this one-dimensional case, it is convenient to have the coordinate x directed toward the outside surface and to arbitrarily choose

$$\psi = 0.$$

and

$$\frac{d\psi}{dx} = 0$$

at the left boundary, not shown in Fig. 8, where $x = 0$.

In the nondegenerate case, the electron density at the surface can be described by

$$\begin{aligned} n &= n_i \exp\left(\frac{\phi_n - \phi_i}{kT}\right)\exp\left(\frac{q\psi}{kT}\right) \\ &= n_i \exp\left(\frac{q\psi}{kT}\right). \end{aligned} \qquad [1]$$

where n_i is the effective electron density in the bulk after the activation. As shown in Fig. 8, ϕ_i is the intrinsic Fermi level, and ϕ_n and ϕ_p are the quasi-Fermi levels for electrons and holes and in the well-activated condition. For a semiconductor with a dielectric constant κ, Poisson's equation is

$$\frac{d^2\psi}{dx^2} = \frac{4\pi q}{\kappa} n_i \exp\left(\frac{q\psi}{kT}\right). \qquad [2]$$

After substituting the dimensionless quantity

$$V = \frac{q\psi}{kT},$$

Eq. [2] can be simplified in the form

$$\frac{d^2V}{dx^2} = \lambda^{-2}\exp(V), \qquad [3]$$

where

$$\lambda = \sqrt{\frac{\kappa kT}{4\pi q^2 n_o}}. \qquad [4]$$

Here the characteristic quantity λ is the usual Debye length. After the first integration, Eq. [3] may be reduced to

$$\left(\frac{dV}{dx}\right)^2 = 2\lambda^{-2}\exp(V) + C_1. \qquad [5]$$

Evaluating C_1 with the boundary condition,

$$\left(\frac{dV}{dx}\right) = 0 \quad \text{at } V = 0,$$

one obtains

$$\frac{dV}{dx} = \frac{\sqrt{2}}{\lambda}(e^V - 1)^{1/2}. \qquad [6]$$

After the second integration, Eq. [6] becomes

$$2\sin^{-1}(e^{-V/2}) = \sqrt{2}\frac{x}{\lambda} + C_2. \qquad [7]$$

After using the boundary condition $V = 0$ at $x = 0$ to determine C_2, Eq. [7] can be written as

$$V = -2\ln\left[\cos\left(\frac{x}{\sqrt{2}\lambda}\right)\right] \qquad [8]$$

for

$$0 < x < \frac{\pi\lambda}{\sqrt{2}}$$

With this potential distribution at the surface, the potential gradient in the space-charge region is given by

$$\frac{dV}{dx} = \frac{\sqrt{2}}{\lambda}\tan\left(\frac{x}{\sqrt{2}\lambda}\right). \qquad [9]$$

and the space-charge density ρ in coulombs per cm^3 may be obtained from

$$\rho = \frac{qn_v}{\cos^2\left(\frac{x}{\sqrt{2}\lambda}\right)}. \qquad [10]$$

The general characteristics of the functions in Eqs. [8], [9], and [10] are shown graphically in Fig. 9. After integrating the charge density ρ

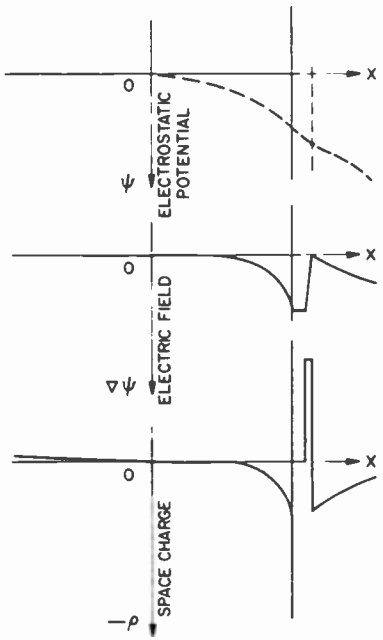


Fig. 9—The electrostatic potential, the electric field, and the space-charge distributions at the oxide surface.

up to the surface with the coordinate X, the total space charge Q in coulombs/cm^2 can be expressed as

$$Q = \sqrt{2}\lambda q n_v \tan\left(\frac{X}{\sqrt{2}\lambda}\right). \qquad [11]$$

Since the free charge in the grain before the activation is small, Eq. [11] is an approximate measure of the total adsorbed ions at the sur-

face. The maximum concentration of adsorbed barium ions would depend upon the chemical conditions of the oxide surface. Physically, it is the adsorbed positive charges that create the charge accumulation and the band bending in the surface layer.

3.3 Numerical Evaluations

Based upon the foregoing treatment, the characteristics of these oxides can be examined by using the experimentally obtained values from the literature.[23,24] From the measurements on BaO single crystals, the electron mobility μ in these oxides will be assumed[12] to be 5 cm^2/volt-sec.

(a) Unactivated Oxides

At the operating temperature of 1100°K, the electron density evaluated from the measured conductivity of unactivated oxides is not greatly increased from the intrinsic value of the material at that temperature as indicated in Table 2. Since there is no charge accumula-

Table 2—Physical Constants and Parameters for Alkaline-Earth Oxides

	BaO	SrO	CaO
Bandgap E_g (eV)	$4.4-0.001\,T$	$5.9-0.001\,T$	$7.6-0.001\,T$
Dielectric Constant (κ)	34	13.1	11.1
n_i at 1100°K (cm^{-3})	4.7×10^{12}	1.69×10^{9}	2.1×10^{5}
Unactivated σ_o $(ohm\text{-}cm)^{-1}$	10^{-5}	3×10^{-6}	
n_o (cm^{-3})	1.25×10^{13}	3.75×10^{12}	
$E_c - \phi_o$ (eV)	1.56	1.65	
$\phi_o - \phi_i$ (eV)	0.09	0.75	
λ (μm)	3.78	4.28	

tion at the crystallite surface, the energy-band diagram is that shown in Fig. 10. If the bandgap energies[8,25,26] given in Table 2 are used in the calculation, the quasi-Fermi level will be more than 1.5 eV below the conduction band. This deduction agrees with the views commonly expressed that these oxides are slightly n-type[13] with only small concentration of ionized donors at the operating temperature.

Before the oxides are activated, the electron conduction is assumed to be nearly uniform throughout the grain. The apparent electron density n_o will be equal to the effective density n_e in the grain. The

calculated densities at 1100°K are listed in Table 2. The case of (BaSr)O crystallites may be considered as an intermediate case between that of BaO and SrO as shown here. If the dielectric constants given in Table 2 are used, the corresponding Debye lengths λ can also be computed and are listed for the cases in consideration.

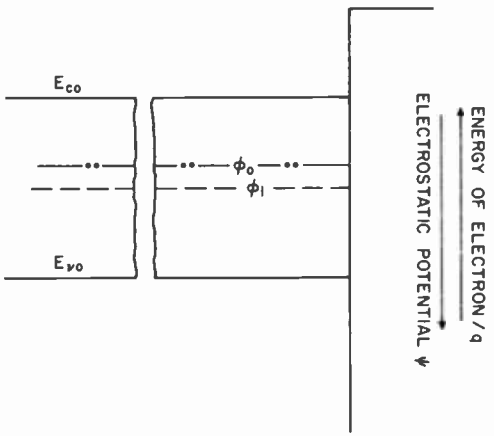


Fig. 10—The energy-band diagram at the surface of an unactivated oxide grain.

(b) Activated Oxides

For well-activated crystallites, the conductivity (Table 1) is about 2×10^{-3} (ohm-cm)$^{-1}$ at a temperature of 1100°K. Consequently, the apparent electron density n_a in the well-activated coating becomes

$$n_a = \frac{\sigma}{\mu q} = 2.5(10)^{15}\text{cm}^{-3}.$$

As a result of the space-charge accumulation on the oxide crystallites, the electron density n_b at the beginning of the surface layer where $x = 0$ may be estimated from

$$n_b = \frac{n_a D^2}{4L(D - L)} \qquad [12]$$

according to this idealized model.

For an average size $D = 5\ \mu$m and $n_a = 2.5 \times 10^{15}$ cm^{-3}, the values of n_b for five L's are listed in Table 3. For each of these values of n_b,

Table 3—Estimated Layer Thickness and Debye Length

L (μm)	n_b (cm^{-3})	λ (μm)
1	3.9×10^{15}	0.13
0.4	8.5×10^{15}	0.09
0.2	1.6×10^{16}	0.066
0.1	3.2×10^{16}	0.046
0.05	6.3×10^{16}	0.033

the corresponding Debye length λ can be calculated from Eq. [4] and is given in the last column. In view of the potential distribution, Eq. [8], it is well to approximate L to be twice the Debye length. For this activation condition, the correct thickness L is then between 1000 and 500 Å.

To examine the space-charge accumulation layer, Eqs. [8]–[11] are now evaluated in terms of the dimensionless quantity (X/λ). Values for four different surface layers are listed in Table 4 to indicate the increasing extent of the ion adsorption effects. If the adsorbed ions are singly charged, the last entry in the table represents the complete coverage by a monolayer of ions at the surface of the grain. According to the electrostatic fields, the voltage breakdown at the surface would occur before reaching the condition of the last entry. This may have been one of the sources that caused the voltage breakdowns[27] in the operation of oxide-coated cathodes.

While the electron density n_b is relatively moderate, Table 4 indicates that very high electric fields at the surface can occur as a result of positive ion adsorption. In taking λ = 500 Å, the major portion of

Table 4—Electronic Parameters for Space-Charge Surface Layers

$\frac{X}{\lambda}$	V	ψ (volts)	$\left(\frac{d\psi}{dx}\right)\left(\frac{\text{volt}}{\text{cm}}\right)$	$\frac{\rho(X)}{qn_e}$	$\frac{Q}{q}$ cm^{-2}
2.00	3.71	0.35	1.6×10^5	41	3×10^{12}
2.20	8.32	0.79	1.7×10^6	4.3×10^3	3.1×10^{13}
2.21	9.62	0.91	3.1×10^6	1.5×10^4	5.8×10^{13}
2.22	13.80	1.31	2.5×10^7	10^6	4.7×10^{14}

the sharp bending of energy band takes place only in the last 100 Å. If the value $n_b = 10^{16}$ cm^{-3} is taken at the left boundary of the accumulation layer, the possible electron density at the surface may reach 10^{20} cm^{-3}. Although impurities may contribute to the conductivity of unactivated crystallites, it is unlikely that they play a significant role if the electron density at the surface is so high.

4. Conclusion

Since Ba is the least electronegative (0.9) element of the alkaline-earth metals, the exposed Ba on the oxide crystallites will probably become a positive ion on the surface. In n-type semiconducting material, the adsorption of the positive ion will cause the energy levels to bend downward (Fig. 8) toward the surface. A similar condition was pointed out that the adsorption of O_2 in the Cu_2O lattice[28] attracts mobile holes toward the surface and thus enhances[29] the conductivity. It is also likely that Zn in hot-pressed ZnSe[30] diodes may play the same role in facilitating electronic injection and in assisting electrical conduction through the samples.

In the simple energy diagram, Fig. 8, the energy levels of the adsorbed ion are schematically indicated to be well above ϕ_n, the quasi-Fermi level. This is intended to show that the electron transfer between the space-charge layer and the adsorbed ions at the surface reaches detailed balance in the steady-state condition. As a consequence, a constant number of adsorbed ions remain at the surface, thus maintaining the space-charge accumulation. It may be expected that the ion adsorption effects should be similar in the cases of Ba on SrO, Ba on CaO,[31] and Ba on (BaSrCa)O. The work function of the oxides increases in the order from BaO, SrO, to CaO. Based upon the experimental findings, SrO with Ba ion adsorption seems to produce the highest emission density from the well-activated crystallites.

In the photoemission experiment,[32] the threshold energy of BaO was determined to be approximately 5.1 eV. For BaO, the bandgap energy at 300°K is 4.1 eV and therefore the electron affinity is about 1 eV. From this simplified analysis, this quasi-Fermi level is about 0.9 eV below the conduction band, and the band bending due to positive ion adsorption may be as high as 0.9 eV. If the experimental value[33] of the thermionic work function of this surface is 2.2 eV, then the effective electron affinity is about 1.3 eV, a reasonable value for SrO.

In the case of metallic emitters where the free electron density is of the order of 5×10^{22} cm^{-3}, the electron emission is most sensitive to the effective work function at the operating temperature. The adsorption phenomena[34] on metal (e.g., Cs on W, Ba on W) can be ex-

perimentally demonstrated with a true maximum emission region (S-shaped curve) as the operating temperature is varied. For Ba adsorption on BaO and SrO, only an abrupt change of the emission characteristics was obtained in the experiment as shown in Fig. 3 of Ref. [33]. This may be explained by the interdependence of the space-charge density and the adsorbed ions at the surface. As the temperature is raised to the extent that some desorption of Ba takes place, there is still a continuous increase in the free-electron density so that no maximum emission appears in the measured characteristics.

Since mobile carriers are attracted toward the surface of the oxide crystallite, the electrical conductivity is principally determined by that of the space-charge layer. By using this adsorption hypothesis, shallow donors are no longer required to explain the electrical conductivity of the oxide coating. As illustrated in the numerical example, the electronic properties of the activated oxide layer—one or two Debye lengths[35] at the surface—become most important in the operation. In future developments on oxide emitters, this area deserves more effort in the pursuit of improvements.

If oxygen vacancies exist at the surface, they would behave as donors and become the source of electrons for the space-charge layer. But if barium vacancies occur at the surface, they would act as acceptors and hence produce a compensation[36] effect. In essence, the acceptors would reduce the density of the free electrons at the surface space-charge layer. Since these barium vacancies will drift towards the surface due to the electrostatic field, this type of lattice defect might be an important deterioration factor to the operation of an oxide-coated cathode.

Since the electron density at the surface can be very high, a small density of surface states does not alter the qualitative description presented here. At present, the surface properties of these alkaline-earth crystals are not clear; it is difficult to explore the conditions of possible stable sites for the adsorbed ions. Further knowledge of the surface[37] layers will be valuable in clarifying the details of the operating mechanisms and in deducing the atomic structure—SrO, O, and Ba—at the surface of oxide crystallites.

Acknowledgment

The authors are indebted to E. P. Bertin for the use of unpublished pictures in Figs. 1 and 2, and to R. Williams for a critical reading and comments concerning the manuscript.

References:

[1] S. Dushman, "Thermionic Emission," *Rev. Mod. Phys.*, **2**, p. 381, (1930).

[2] A. L. Reimann, *Thermionic Emission*, p. 188, Chapman and Hall, Ltd., London, (1934).

[3] J. P. Blewett, "The Properties of Oxide-Coated Cathodes," *J. Appl. Phys.*, **10**, p. 668 and p. 831, (1939).

[4] A. S. Eisenstein, "Oxide Coated Cathodes," in *Advances in Electronics*, Academic Press, Inc., New York, (1948).

[5] G. Hermann and S. Wagener, *The Oxide Cathode*, Vols. I. and II, Chapman and Hall, Ltd., London, (1951).

[6] W. B. Nottingham, "Thermionic Emission," p. 97 in *Encyclopedia of Physics*, Springer-Verlag, New York, (1956).

[7] L. S. Nergaard, "Electron and Ion Motion in Oxide Cathodes," p. 154 in *Halbleiterprobleme*, Vieweg, Braunschweig (1956).

[8] P. Zalm, "Thermionic Cathodes," p. 211 in *Advances in Electronics and Electron Physics*, Academic Press, Inc., New York, (1968).

[9] F. Kroger, *Chemistry of Imperfect Crystals*, p. 564, John Wiley (1964).

[10] I. G. Maloff and D. W. Epstein, *Electron Optics in Television*, McGraw-Hill Book Co., New York, (1938); V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope*, John Wiley and Sons, New York, (1954).

[11] J. R. Pierce, *Theory and Design of Electron Beams*, p. 145, Van Nostrand (1949); G. R. Brewer, "High-Intensity Electron Guns," p. 23 in *Focusing of Charged Particles*, Academic Press, Inc., New York, (1967).

[12] E. M. Pell, "The Hall Effect in Single Crystals of Barium Oxide," *Phys. Rev.*, **87**, p. 457, (1952).

[13] R. T. Dolloff, "Electrical Conductivity of Barium Oxide Single Crystals as a Function of Temperature and Excess Barium Density," *J. Appl. Phys.*, **27**, p. 1418, (1956).

[14] W. D. Copeland and R. A. Swalin, "Studies on the Defect Structure of Strontium Oxide," *J. Phys. Chem. Solids*, **29**, p. 313, (1968).

[15] N. A. Surplice, "The Electrical Conductivity of Calcium and Strontium Oxides," *Brit. J. Appl. Phys.*, **17**, p. 175, (1966).

[16] W. Meyer and A. Schmidt, "Über die Elektrizitatsleitung von Bariumoxyd in Zusammenhang mit der Elektronenmission," *Z. Tech. Phys.*, **13**, p. 137, (1932).

[17] H. J. Spanner, "Über die Thermische Emission Elektrisch Geladener Teilchen," *Ann. Phys.*, **75**, p. 609, (1924).

[18] N. B. Hannay, D. Macnair, and A. H. White, "Semiconducting Properties in Oxide Cathodes," *J. Appl. Phys.*, **20**, p. 669, (1949); R. Loosjes and H. J. Vink, "The Conduction Mechanism in Oxide-Coated Cathodes," Phillips Res. Rep. 4, p. 449, (1949).

[19] E. O. Kane, Thesis, Cornell Univ., Ithaca, N.Y. (1953).

[20] L. R. Koller, "Electron Emission from Oxide Coated Filaments," *Phys. Rev.*, **25**, p. 671, (1925).

[21] A. R. Hutson, "Semiconducting Properties of Some Oxides and Sulfides," p. 543 in *Semiconductors*, Reinhold Pub. Corp. (1959).

[22] W. Shockley, "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," *Bell Systems Tech. J.*, **28**, p. 435, (1949).

[23] R. S. Bever and R. L. Sproull, "The Dielectric Constant of Barium Oxide," *Phys. Rev*, **83**, p. 801, (1951).

[24] J. L. Jacobson and E. R. Nixon, "Infrared Dielectric Response and Lattice Vibrations of Calcium and Strontium Oxides," *J. Phys. Chem. Solids*, **29**, p. 967, (1968).

[25] G. A. Saum and E. B. Hensley, "Fundamental Optical Absorption in the IIA-VIB Compounds," *Phys. Rev.*, **113**, p. 1019, (1959).

[26] R. C. Whited and W. C. Walker, "Exciton and Interband Spectra of Crystalline CaO," *Phys. Rev.*, **188**, p. 1380, (1969).

[27] E. A. Coomes, "The Pulsed Properties of Oxide Cathodes," *J. Appl. Phys.*, **17**, p. 647, (1946); B. Ya. Moizhes and S. Sh. Rutshtein, "Breakdown (Sparking) of Oxide Cathodes," *Sov. Phys.-Tech. Phys.*, **16**, p. 1748, (1972).

[28] S. R. Morrison, "Surface Barrier Effects in Adsorption Illustrated by Zinc Oxide," p. 259 in *Advances in Catalysis*, Vol. 7, Academic Press, Inc., New York, (1955).

[29] W. G. Garner, F. S. Stone, and P. F. Tiley, "The Reaction between Carbon Monoxide and Oxygen on Cuprous Oxide at Room Temperature," *Proc. Roy. Soc.*, **A211**, p. 472, (1952); M. Zouaghi, B. Prevot, C. Carabatos and M. Sieskind, "Near Infrared Optical and Photoelectric Properties of Cu_2O," *Phys. Stat. Sol.*, **All**, p. 449, (1972).

[30] T. N. Chin and L. A. Boyer, "D.C. Electroluminescence in Hot-Pressed ZnSe Diodes," *Solid-State Electron*, **16**, p. 143, (1973).

[31] H. H. Glascock, Jr., "Thermionic Emission from CaO in Ba Vapor," *Surface Sci.*, **29**, p. 291, (1972).

[32] L. Apker, E. Taft, and J. Dickey, "On the Photoelectric Emission and Energy Structure of BaO," *Phys. Rev.*, **84**, p. 508, (1951).

[33] K. S. Beynar and B. P. Nikonov, "The Emission and Adsorption Properties of the BaO-Ba, SrO-Ba and CaO-Ba Systems," *Radio Eng. Electron, (USSR)*, **10**, p. 408, (1965), (english translation).

[34] J. B. Taylor and I. Langmuir, "The Evaporation of Atoms, Ions and Electrons from Caesium Films on Tungsten," *Phys. Rev.*, **44**, p. 438, (1933); T. Smith, "Cesium Adsorption on W:Ellipsometry, Auger Spectroscopy, and Surface Potential Difference Studies," *J. Appl. Phys.*, **43**, p. 2964, (1972).

[35] A. Rose, *Concepts in Photoconductivity and Allied Problems*, p. 134, John Wiley and Sons, New York, (1963).

[36] J. S. Blakemore, *Semiconductor Statistics*, p. 175, Pergamon Press, Inc., Elmsford, New York, (1962).

[37] J. W. Gadzuk and E. W. Plummer, "Field Emission Energy Distribution (FEED)," *Rev. Mod. Phys.*, **45**, p. 527, (1973).

A Membrane Page Composer—Further Developments*

L. S. Cosentino and W. C. Stewart

RCA Laboratories, Princeton, New Jersey 08540

Abstract—Membrane page composers were made and were evaluated in a simulated holographic optical memory system. Calculated and experimentally determined electromechanical and optical characteristics of the circular membrane light valves used on the arrays are shown to be in close agreement. Several operating prototypes of 8 × 8 and 16 × 16 elements were produced. Measurements were made of switching time, optical contrast, and dynamic storage time of many cells on the devices. Digital patterns were stored in the arrays. The performance required of the page composer as a component of an optical memory system is considered. The fabrication techniques used can be easily extended to larger arrays.

1. Introduction

An earlier paper,[1] in which the membrane page composer concept was introduced, forms the background for the work reported here. The previous investigation demonstrated that membrane elements have useful light-valve properties. Arrays of 64 elements were made on substrates with feedthroughs allowing access to individual elements from the back side of the substrate. Single light valves on such arrays were operated with transistors designed and produced for selection and storage at each bit location. This simulated the operation of a prototype page composer with semiconductor chips beam-lead bonded to the back of the substrate on which there would also be ap-

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

propriate conductor patterns for addressing the array. We have extended that effort by making and evaluating 8 × 8 and 16 × 16 element membrane page composers.

The first portion of this paper describes experiments and computer analyses that were performed to determine the electromechanical and optical responses of circular membrane elements, which are utilized in all of the arrays. System requirements are then considered. Next, the fabrication and evaluation of page composer devices are detailed, followed by conclusions in the final section of the paper.

2. Circular Membrane Elements

2.1. Electromechanical Response

In Ref. (1), some results for the deflection and the dynamic response of circular membrane elements were arrived at and are repeated here for convenience. The deflection equations are

$$z = \frac{\epsilon V^2}{8T_0 d^2}\left(r^2 - \frac{D^2}{4}\right) \qquad [1]$$

and

$$z_0 = -\frac{\epsilon V^2 D^2}{32 T_0 d^2}. \qquad [2]$$

where z is the deflection at a point on the membrane, z_0 is the deflection of the center point of the element, ϵ is the permittivity of the medium, D is the diameter of the element, d is the membrane-to-electrode separation, V is the applied voltage, and T_0 is the tension in the membrane. For the case in which the electrode diameter S is smaller than D, the appropriate equations are*

$$z = \frac{\epsilon V^2}{8T_0 d^2}\left(r^2 + \frac{S^2}{4} - \frac{SD}{2}\right). \qquad \frac{S}{2} \geq r \geq 0 \qquad [3]$$

and

$$z = \frac{\epsilon V^2 S}{8T_0 d^2}\left(r - \frac{D}{2}\right). \qquad \frac{D}{2} \geq r \geq \frac{S}{2} \qquad [4]$$

With an electrode one-half the width of an element, $S = D/2$, z_0

* In Ref. (1) Eq. [7] for a line element should read

$$z = \frac{\epsilon V^2 S}{4T_0 d^2}\left(x - \frac{W}{2}\right), \text{with } \frac{W}{2} \geq x \geq \frac{S}{2}.$$

would be three-quarters of the central deformation with a full-width electrode, $S = D$; while for the case of $S = 2D/3$, z_0 would be about nine-tenths of the value with $S = D$.

The dynamic response of a circular membrane element is characterized by a resonant frequency in vacuum given by

$$f = \frac{u}{\pi D}\sqrt{\frac{T_0}{\delta t}}. \qquad [5]$$

where δ is the density of the membrane material, t is its thickness, and u equals the values at which the zero-order Bessel function of the first kind $J_0(u)$ equals zero ($u_1 = 2.40$, $u_2 = 5.53$, $u_3 = 8.65$, etc.).

The fundamental mode is therefore given by

$$f_1 = \frac{0.764}{D}\sqrt{\frac{T_0}{\delta t}}. \qquad [6]$$

2.2. Optical Response

Fig. 1 is a diagram of the optical system described previously,[1] which simulates the imaging characteristics of a holographic memory and is used in experimental tests of membrane arrays. The cross section of an isolated circular membrane element, whose central deformation measured with respect to its supported perimeter is z_0, is shown to exaggerated scale at the far right of the figure. The flat portion of the profile represents the undeformed part of the membrane over the rigid support structure. The lens projects an image of the membrane with a magnification $M = -(x' + f)/(x + f)$ onto the detector aperture whenever the imaging condition $xx' = f^2$ is fulfilled. With collimated laser light (not shown) incident on the membrane, and the membrane undeformed, all reflected rays from the element are focused by the lens in the small circular opening of the filter aperture. Parabolic deformation of the membrane, which is assumed throughout this analysis, changes the collimation of the reflected light so that the focal point occurs in a different plane, and very little light is transmitted by the filter aperture. The filter aperture, which also determines the resolution of the imaging system, simulates the storage hologram in a memory system.

a. *Ray-Tracing Analysis*—In the geometric optics approximation, deformation of the membrane by an amount z_0 at the center gives a parabolic mirror of focal length approximated by

$$f_m = \frac{D^2}{16 z_0} \qquad [7]$$

when $z_0 \ll D$, where D is the diameter of the element. It is easily shown that the diameter d of the region of uniform illumination in the plane of the filter aperture is independent of the object spacing x in Fig. 1, and is given by

$$d = fD/f_m, \qquad [8]$$

where f is the focal length of the projection lens. Let the reflected light flux from the membrane be I_0, and assume that the lens is lossless. For d less than the diameter d_a of the filter aperture, the light flux I transmitted by the aperture is

$$I = I_0, \quad d \leq d_a. \qquad [9a]$$

Otherwise, the fraction of the incident light flux that is transmitted is given by the ratio of the filter aperture area to the area of uniform illumination incident on the plane of the aperture;

$$I = I_0(d_a/d)^2, \quad d > d_a. \qquad [9b]$$

We assume that the detector aperture is at least as large as the projected image of the element, so that the detector collects all the light flux transmitted by the filter aperture and gives a signal proportional to I. It is convenient to express d_a in terms of a proportionality constant α, defined by the relation

$$d_a = 2.44\alpha\lambda f/D. \qquad [10]$$

where λ is the wavelength of the illumination. Combining Eqs. [7] through [10] gives

$$I/I_0 = 1, \quad z_0 \leq 2.44\alpha\lambda/16 \qquad [11a]$$

and

$$I/I_0 = (2.44\alpha\lambda/16z_0)^2, \quad z_0 > 2.44\alpha\lambda/16. \qquad [11b]$$

These simple considerations predict that the image contrast between deformed and undeformed states increases with the square of the central displacement, and that the contrast is decreased by increasing the diameter of the filter aperture.

b. *The Physical Optics Approach*—We next turn to a more accurate analysis, which includes the effects of diffraction, and which is the basis of our subsequent numerical calculations. The constant α in Eq. [10] is recognized as the ratio of the filter aperture diameter to the diameter of the central Airy disc in the diffraction pattern that occurs for an isolated undeformed membrane element of diameter D. The value of α also provides a rough measure of the importance of diffraction effects in the optical system; the geometric optics approximation requires $\alpha \gg 1$ in order to be strictly valid.

The coherent image amplitude at the detector can be expressed in general[2] as the two-dimensional spatial convolution of the reflected light amplitude from the membrane with the amplitude point spread function produced by the filter aperture. In the present case, in which the object and the point spread function have circular symmetry, the image also has circular symmetry. This permits reducing the calculation of the image to a one-dimensional superposition integral. A detailed derivation follows.

We assume unity magnification in order to simplify the notation, since diffraction effects relative to the image are independent of the magnification. Let r_1 be the radial polar coordinate in the plane of the membrane, and the function $g(r_1)$ represent the complex amplitude of the light from the membrane. A Fourier transform relationship exists between $g(r_1)$ and $G(p)$, the light amplitude incident on the filter aperture. With circular symmetry,

$$G(p) = 2\pi \int_0^{\infty} r_1 g(r_1) J_0(2\pi r_1 p) dr_1, \qquad [12]$$

where p is proportional to the radial coordinate in the filter aperture plane. The filter aperture truncates $G(p)$ at the value

$$p_a = 3.832\alpha / \pi D \qquad [13]$$

in accordance with Eq. [10]. There is a further Fourier transform relationship between the transmitted wave amplitude at the filter aperture and $g'(r_2)$, the image amplitude at the detector plane.

$$g'(r_2) = 2\pi \int_0^{p_a} p G(p) J_0(2\pi p r_2) dp, \qquad [14]$$

where r_2 is the radial coordinate in the detector plane. Combining Eqs. [12] and [14] and exchanging the order of the definite integrals gives

$$g'(r_2) = 4\pi^2 \int_0^{\infty} dr_1 r_1 g(r_1) \int_0^{p_a} dp\, p J_0(2\pi r_2 p) J_0(2\pi r_1 p). \qquad [15]$$

Denoting the innermost integral in Eq. [15] by $K(r_1, r_2)$, evaluating the integral, and combining with Eq. [13] gives

$$K(r_1, r_2) = [1.916\alpha/\pi^2 D(r_1^2 - r_2^2)][r_1 J_1(qr_1) J_0(qr_2) - r_2 J_1(qr_2) J_0(qr_1)], \qquad [16a]$$

for $r_1 \neq r_2$ and

$$K(r_1, r_2) = (3.832\alpha/\sqrt{2}\pi D)^2[J_0^2(qr_1) + J_1^2(qr_1)], \qquad [16b]$$

for $r_1 = r_2$, where $q = 7.664\alpha/D$. The image amplitude is therefore

$$g'(r_2) = 4\pi^2 \int_0^{\infty} r_1 g(r_1) K(r_1, r_2) dr_1. \qquad [17]$$

The detector signal is proportional to the intensity of the image integrated over the radius r_d of the detector aperture;

$$I = 2\pi \int_0^{r_d} r_2 |g'(r_2)|^2 dr_2. \qquad [18]$$

The fact that the input plane to the imaging system is not in the front focal plane of the lens in Fig. 1 produces spherical phase factors,

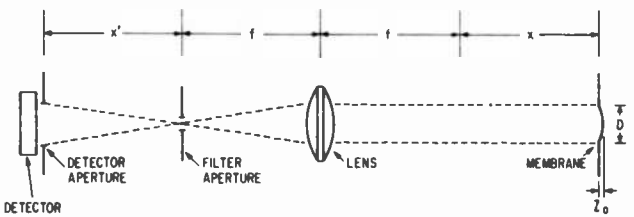


Fig. 1—Diagram of the coherent optical system.

equivalent to wavefront curvatures, in the wave amplitudes at the filter plane and image plane. These factors, combined with the imaging condition given earlier, establish the Fourier transform relationship between the wave amplitudes in the filter and image planes. Omission of these factors in Eqs. [14] and [16] has no effect on the image

intensity $|g'(r_2)|^2$ and simplifies the computations to be made. Finally, the reflection of a collimated wave from a paraboloid of diameter D and central deformation z_0 is expressed by the complex amplitude

$$g(r_1) = \exp\left\{\frac{j4\pi z_0}{\lambda}\left[1 - \left(\frac{2r_1}{D}\right)^2\right]\right\}, r_1 \leq \frac{D}{2},$$
$$g(r_1) = 1 \text{(uniform illumination)}, r_1 > \frac{D}{2}, \qquad [19]$$
$$g(r_1) = 0 \text{(local illumination)}, r_1 > \frac{D}{2}.$$

where $j = \sqrt{-1}$. Numerical results have been obtained by combining Eqs. [17], [18], and [19], and carrying out the integrations on a digital computer.

c. *Numerical Results*—Figs. 2 through 5 show the detector signal as a function of the membrane displacement up to 1.5 λ for a variety of

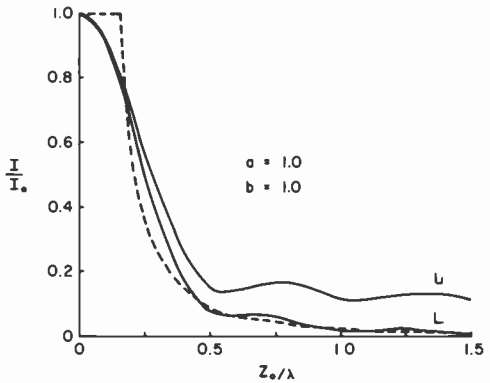


Fig. 2—Calculated detector signal versus membrane displacement (inverse contrast curve) for filter and detector apertures of conventional size. Curve U represents uniform illumination, L represents local illumination, and dashed curve is the geometric optics prediction.

conditions. The parameter b is the ratio of the radius of the detector aperture to the radius of the geometric image of the membrane element,

$$b = 2r_d/D; \qquad [20]$$

the parameter α is defined in Eq. [10] and has been discussed pre-

viously. The curves labeled U are for the case of uniform illumination and reflectance of the membrane element and its surrounding supported region. Localized illumination only of the deformable membrane region is designated by the label L. In the design of holographic memory systems, α and b are conventionally set equal to unity.[3] This case is shown in Fig. 2. The result of the geometric optics approximation, Eq. [11], is depicted by the dashed curve, whose general trend is similar to that obtained for local illumination by exact calculation. The contrast is 50:1 at $z_0 = \lambda$. For uniform illumination, the contrast is considerably poorer because light from the rigidly supported periphery spreads by diffraction into the detector aperture.

When the filter aperture is enlarged to coincide with the second minimum in the Airy diffraction pattern, the results are as shown in Fig. 3. For local illumination, the contrast is not as high as before since more of the light from the deformed element passes the filter

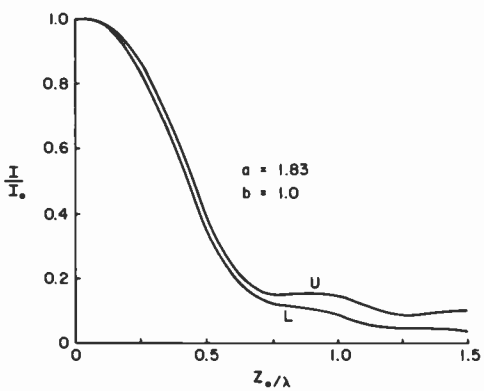


Fig. 3—Inverse contrast computed for a larger filter aperture.

aperture. For uniform illumination, however, there is a slight improvement in contrast for large deformations because of the reduced diffraction spreading from the illuminated area outside the rim. The case of a small filter aperture is shown in Fig. 4. Diffraction spreading of the image severely degrades contrast with uniform illumination. With local illumination, the minima of the detector signal are much lower than for $\alpha = 1$; the computed contrast at $z_0 = \lambda$ approaches 500:1. The subsidiary maxima around $3\lambda/4$ and $5\lambda/4$ are comparable in value to those for $\alpha = 1$. A reduction in detector aperture area by a factor of approximately two results in curves as shown in Fig. 5. The

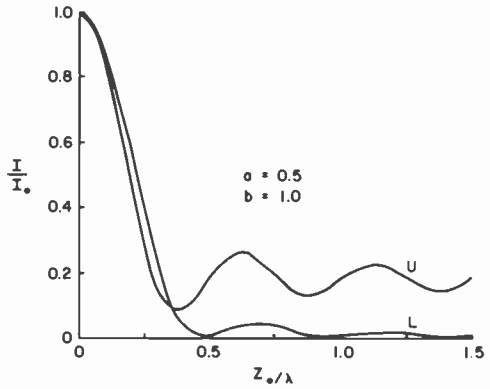


Fig. 4—Inverse contrast computed for a smaller filter aperture.

curve for uniform illumination continues to show low contrast, and there is essentially no change for local illumination.

From these and other results of the numerical calculations, several trends can be identified for membrane displacements limited to 1.5λ:

(1) the use of uniform illumination limits the available contrast to approximately 10.
(2) With local illumination and apertures of conventional size, contrast values of 25 to 50 are typical and 100 is possible. The geometric optics approximation is surprisingly good for this case.
(3) Contrast values over 100 require a reduction in filter aperture diameter and control of the membrane displacement to better than 10%.

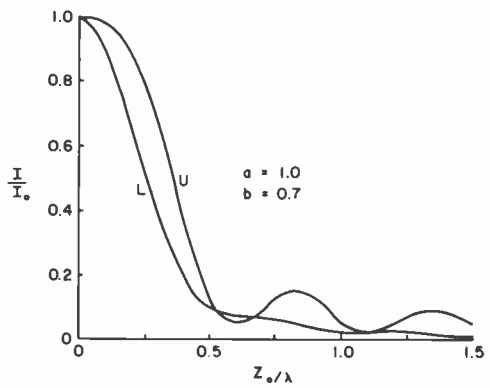


Fig. 5—Inverse contrast computed for a smaller detector aperture.

2.3. Experimental Results

Experiments were performed to characterize the operation of circular membrane elements. An 8 × 8 array of elements 0.75 mm in diameter with individual access to each element was used for the tests. The membrane on the sample is electroplated nickel 0.6 μm thick, while the membrane-to-electrode spacing is about 3 μm.

With the sample in the simulated optical memory system, detector light was measured as a function of applied voltage using various filter and detector apertures and illumination from a helium–neon laser (λ = 0.6328 μm). Some typical results are shown in Fig. 6.

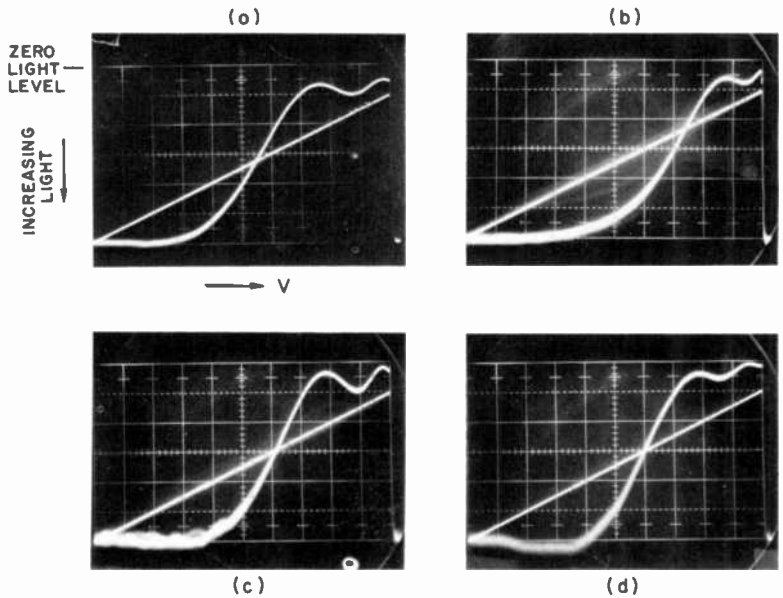


Fig. 6—Detector output vs. applied voltage. The straight line shows the sawtooth voltage while the curve shows the light variations. The voltage scale is 10 V/div; the marks on the traces are for measurement convenience. (a) Uniform illumination (α = 1, b = 1); (b) uniform illumination (α = 1.9, b = 1); (c) uniform illumination (α = 1, b = 0.707); and (d) local illumination (α = 1, b = 1).

Measurements of the central deflection, z_0, of circular membrane elements were made interferometrically. A Michelson–Morley interferometer was set up on an optical bench using the 0.6328-μm illumination of a helium–neon laser. Displacements smaller than 10^{-10}

meter can be detected with systems of this type. For our measurements the sensitivity was adjusted to resolve 8×10^{-9} meter. The data for one cell are plotted in Fig. 7 on log-log scales together with a line calculated for a cell from Eq. [2]. The value of T_0 was obtained by using Eq. [6] with $\delta = 7.6$ gm/cm^3 for electroplated nickel, $t = 0.6$ μm, $D = 0.75$ mm, and an observed resonant frequency, $f_1 = 125$ kHz.

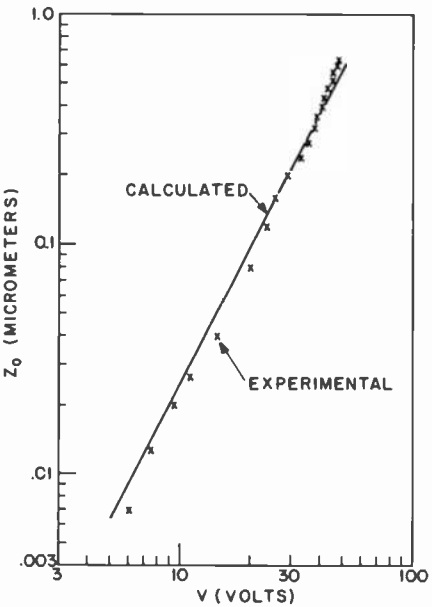


Fig. 7—Central deflection vs. voltage.

The data from the interferometric tests were correlated with the data from the tests measuring detector light vs. applied voltage to obtain light at the detector as a function of central deflection, z_0. The resultant curves are shown in Figs. 8, 9, and 10.

2.4. Discussion of Experimental Results

The curves of detector light vs. applied voltage with various filter and detector apertures are in general agreement with the results of the optical response analyses. Fig. 6(a) shows a maximum contrast of about 12 with $\alpha = 1$, $b = 1$, with about 48 V applied. With a larger filter aperture, $\alpha = 1.9$, $b = 1$, Fig. 6(b) shows that the contrast is ap-

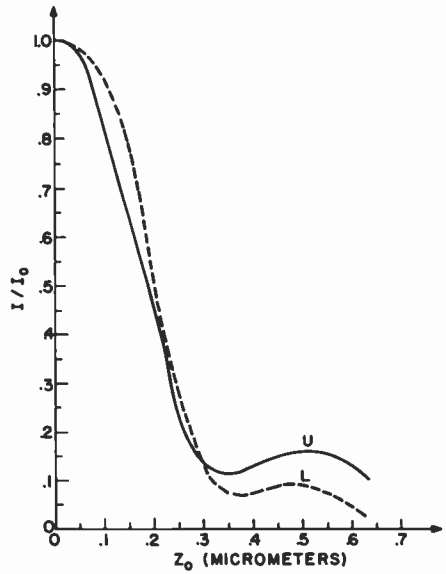


Fig. 8—Inverse contrast with $\alpha = 1$, $b = 1$.

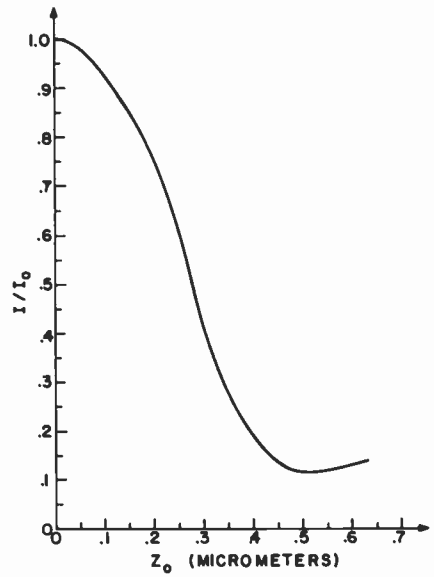


Fig. 9—Inverse contrast with $\alpha = 1.9$, $b = 1$.

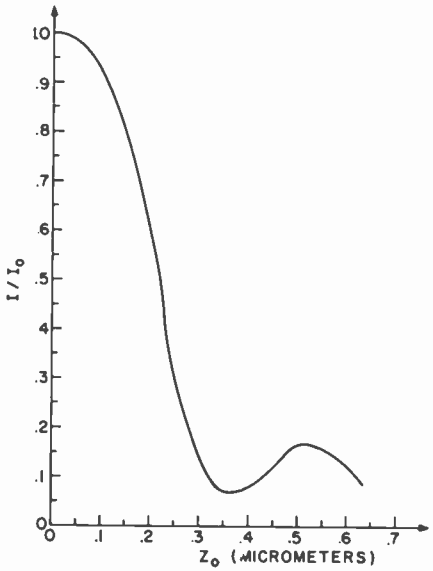


Fig. 10—Inverse contrast with $\alpha = 1$, $b = 0.7$.

preciably lower than with $\alpha = 1$ until voltages become quite large, at which point contrast peaks are similar. With $\alpha = 1$, $b = 0.7$, the contrast is lower than with $b = 1$ until higher voltages are reached when contrast peaks are considerably higher with the smaller detector, showing a value of about 30 at 48 V in Fig. 6(c). With $\alpha = 1$, $b = 1$, and local illumination, the maximum contrast at 48 V increases to about 60 as seen in Fig. 6(d). Direct comparisons of the I/I_0 vs. z_0 curves generated from the computer numerical results with those of Figs. 8, 9, and 10 show similar qualitative agreement. Discrepancies in the location of the peaks and valleys of the oscillatory portion of the curves may be caused by nonparabolic deformation of the real element, differences in the reflectivity of the membrane and support structure, and the actual placement of components such as the filter aperture in the experiment.

The deflection vs. voltage data of Fig. 7 are seen to be in very good agreement with the static deflection analysis corresponding to the case of constant built-in tension, T_0, essentially unaffected by deformation. For small values of deflection z_0 is approximately proportional to V^2, while larger values of z_0 have stronger exponential dependence on the applied voltage because of the reduced spacing between membrane and electrode. It is interesting that T_0, found from Eq. [6] and the experimentally observed value of f_1, when substituted

in Eq. [2] with other known values, produces calculated values of deflection that are nearly equal to those found by direct interferometric measurement for small values of z_0 when d in Eq. [2] is close to the nominal value. The exercise serves to show that the model of membrane element operation conforms well to the actual case. In addition, the plots of Fig. 7 indicate the degree to which actual operation deviates from that of an element strictly conforming to Eq. [2].

3. System Considerations

The size of a page-composer element and the spacing between elements are dictated by the system requirements, and therefore cannot be arbitrarily chosen.[3] These dimensions will in turn be involved in determining other system values. Since there were no rigid system specifications for the membrane page composer, dimensions of a reasonable and practical size were used as judged from experience with optical memory systems such as the one described in Ref. (3). The general properties of the device can then be determined, but the detailed performance will still depend on the system in which it is used.

As previously stated, the value of α is usually taken to be 1. However, if the page composer is not flat enough, the light from all of the unenergized elements will not get through the aperture and a larger one will be necessary. This acts to reduce optical contrast, as shown by the analysis, and also acts to reduce the capacity of the optical memory, since the aperture simulates a hologram in which a page of information is stored. It is therefore important to make the membrane surface as flat as possible over the entire active region.

As explained in Ref. (1) a single MOSFET switch controls each membrane element in the array. Four transistors are included on each semiconductor chip. The page composer is activated word at a time. A word line is connected to one row of transistor gates, while a digit line is connected to one column of drains. The source of each MOSFET is connected to its respective feedthrough and electrode. In operation, a word line is selected causing the MOSFET channels in that row to be conductive. The desired voltage pulse ("1" or "0") on each digit line then appears at the corresponding electrode and acts on the grounded membrane of that element accordingly, to leave it flat or to deform it. With the gate disabled before the data pulse ends, the charge on the capacitance element formed by the electrode–membrane combination will leak off slowly. If the cycle time of the page composer is short compared to the leakage time, effective memory is obtained.

The time required to load information into the page composer is determined by the time it takes a transistor to energize an element on one line to the drive voltage, the number of lines in the array, and the response time of the elements. After the last line is addressed and sufficient time is allowed for the slowest element to reach its proper state, the page is ready to be optically written into the memory. All elements must retain their given state until this is done. Consider an array with 50 word lines. If 1 μs is required to fully charge an element, then 50 μs will be needed to address the array. If a membrane element switches states in 50 μs, then that must be added to the addressing time. Thus, 100 μs is needed before all the information is in the page composer. This is also the minimum cycle time for the device. Elements in the first line reach their new states in about 50 μs and must then hold them for another 50 μs. The time required to write the page into the memory must be added to the storage time of the elements and will also determine the actual cycle time of the page composer. A page composer cycle time of 1 ms or less is presently envisioned, so that total voltage decay times of about 100 ms should ensure that little change in the stored state occurs. To obtain long storage times for slower memories, either refresh techniques with the present drivers or static storage using flip-flops would be needed with membrane elements.

The response time of a membrane element in vacuum is determined by its dimensions, its mass, and the tension in the metal film. In air, as explained in Ref. (1), an important determinant is the number of microscopic holes in the membrane that allow for the easy passage of air into and out of the membrane-to-electrode space. Rise times as low as 10 μs have been observed with circular membrane elements although 50 μs to several hundred microseconds are more typical values for recent samples. If the time for a membrane element to change states is critical, it can be tailored by controlling the cited factors, although this was not assigned high priority in our work.

As discussed above, a minimum storage time must be exhibited by every page-composer cell. If an element does not hold its state for at least this time, the result is a loss of contrast and possible loss of information. Any storage of longer duration is acceptable because the element is automatically reset to the new conditions upon rewriting of the page. The voltage decay time of an isolated cell with capacitance C and effective leakage resistance R has an associated time constant, RC. When the capacitive cell is combined with a transistor and selection lines, the decay time is reduced because of the leakage current of the transistor and the parasitic leakages through the selection circuitry. The effective decay time of this combination must then

be greater than the minimum storage time for the system to work properly.

One concern that arises in arrays of any kind is uniformity. An operating voltage range must be chosen to activate all cells. System requirements will specify the minimum contrast which an element must exhibit to be detectable. If any cells require more or less voltage than the selected value for optimum contrast, then contrast for those cells will be degraded accordingly. In addition, surface conditions at an element affect the light reflected from it which will also influence contrast. If the combined effects reduce the contrast at any cell below the minimum value, then the array cannot perform its function. Factors affecting the drive voltage required for a given deflection or contrast can be seen from Eq. [2] to include the tension in the membrane and the membrane-to-electrode spacing at the cell.

4. 8 × 8 Element Page Composers

4.1. Fabrication

The fabrication cycle involves the substrate with its associated feedthroughs, the transistor array on the back of the substrate, and the deformable membrane with its associated structure on the front of the substrate. Any processing of one part must be compatible with that done on other parts. The metallization of the pattern for the selection circuitry and the beam-lead-bonding technique require temperatures of 200°C or higher. Most epoxy materials, which are conveniently used to form feedthroughs, cannot withstand such high temperatures and remain conductive. In addition, the epoxies tend to be grainy and soft relative to the substrate so that when polished, they are typically undercut and not smooth and flat. Consequently, the epoxy feedthrough is usually offset from the electrode area so that it is not part of the optical element. An ideal feedthrough would not be temperature limited, would polish well, and could be located directly in the electrode area. Although such a feedthrough does not exist at present, one candidate is a metal (such as nickel) that can be plated through the holes while filling them in. For our purposes, a workable solution to the high-temperature requirement was found in the use of a conductive gold epoxy, which survived overnight bakes at 250°C and 2 hours, or longer in some cases, at 300°C. When substrates filled with the gold epoxy were baked out and checked, the feedthroughs remained intact with little or no change in resistance.

A fabrication sequence that has been found to give good results is as follows. First, a substrate is filled with gold epoxy, prebaked, and

checked for continuity of the feedthroughs. Next, the substrate is polished flat. The metallization pattern is then formed on the back of the substrate. Electrodes and support structure are formed on the front of the substrate. Then the transistor array is bonded to the substrate and checked out for continuity, minimum breakdown voltages, and maximum leakage currents. Protective resist is spread over the transistor array and remains there during the forming of the metal membrane on the front of the substrate, after which it is removed. The main features of producing the metal membrane are detailed in Ref. (1). After the semiconductor chips are on the sample, additional processing must be clean and kept to a minimum to avoid contamination leading to parasitic leakages in the selection circuitry. Precautions in handling and in wiring to a device are needed to avoid gate breakdowns caused by electrostatic charge build-up.

Each 8 × 8 page composer has membrane elements of 0.75 mm diameter spaced on 1.50-mm centers on a 5 cm square substrate. Feedthroughs are 0.25 mm in diameter and are offset from the optical element area. Electrodes extend beneath the entire membrane element and are covered with a blanket of silicon monoxide. The support structure is also made of SiO and is nominally 2 to 3 μm in height. The membrane is electroplated nickel about 8,000 Å thick.

4.2. Test Results and Discussion

Early 8 × 8 samples served to establish feasibility of the page composer concept and to indicate what improvements would be needed. Coincident selection of bits, dynamic storage of patterns, and 1-μs switching of the drive transistors to full voltage were demonstrated with these samples (see Fig. 11). Selection and storage at a cell with higher contrast and much longer memory are shown in Fig. 12

The problems encountered in making the first devices indicated where effort would be needed to improve the arrays. Some membrane elements were collapsed or torn. Other cells were inaccessible because of open word or digit lines. There were also defective transistors. Many feedthroughs were found to be discontinuous because the silver epoxy formerly used could not withstand high temperatures. All of these problem areas were attacked and considerable improvement was achieved when the proper fabrication sequence and techniques were worked out.

Three 8 × 8 devices with gold epoxy feedthroughs were processed with much better results. All membrane elements on the three samples were flat and intact although surface defects at some cells reduced contrast. All feedthroughs were continuous. All cells on each

array operated when accessed in parallel with dc through the diodes formed by the chip substrates (n) and the source regions (p) connected to the feedthroughs. Tests were made on each device in the simulated optical memory system to determine the rise time, contrast, and storage time of cells. Typical measurements were a rise time of a few hundred microseconds, a storage time of a few hundred

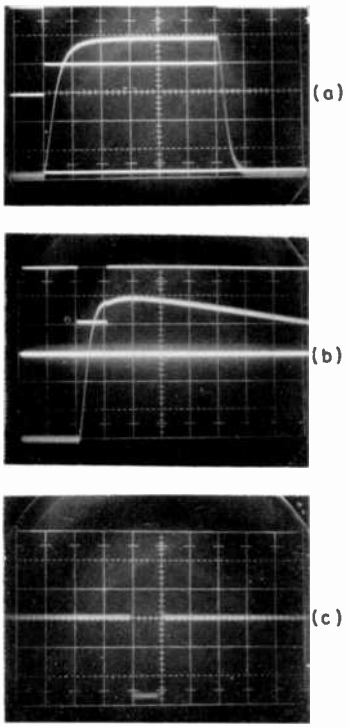


Fig. 11—Pulse activation of an 8 × 8 page composer element. The voltage scale is 20 V/div (α = 1.9, b = 1). (a) Coincident pulse operation (time scale is 0.5 ms/div); the top trace is the digit pulse, the middle trace is the word pulse, and the bottom trace is the detector output. (b) Memory action when the gate pulse (which is barely discernible) is made narrower than the digit pulse. (c) The gate pulse used in (b) on a time scale of 1 μs/div.

milliseconds, and a contrast of about 10. In some cases, when more than 50 V was required for cell operation, a positive dc bias voltage, which just barely affected contrast, was applied to the membrane so that the transistors never had to switch pulses more negative than 50 V to drive a cell. An electronics test panel with an internal 8 × 8

memory was designed and built to drive the page composer array. Photos of the images of some digital patterns stored in the page composers are shown in Fig. 13. A copper mask with an 8 × 8 array of holes was used to define circular beams of light from the laser output, which then illuminated only the active elements of the membrane. The photos were taken with 45 V applied to the array, with α = 1.9

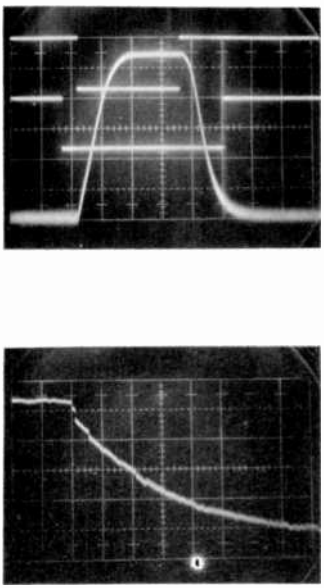


Fig. 12—Pulse activation of another 8 × 8 page composer element (α = 1.5, b = 1): (a) coincident pulse operation (the voltage scale is 20 V/div and the time scale is 0.2 ms/div) and (b) the light decay from the dark state to the light state after a single-shot store operation at the beginning of the trace (the time scale is 1 s/div).

and with an exposure time of 1/100 second. Note that cell (W8,D6) does not store properly in Fig. 13(b) because a leaky gate causes cell voltage to decay rapidly before the pattern can be refreshed, and the element is unenergized most of the time. Also note that because the drive voltage chosen for good response over most of the plane is not the optimum value for some cells near the (W8,D8) corner, these cells exhibit low optical contrast in Fig. 13(c).

Leaky MOSFET gates and short storage times were seen at several cells on the page composers. Since all gates were previously checked and were found to be working initially, it was decided that the faulty

ones were caused by breakdown from electrostatic charge build-up. No new cases of leaky gates were seen after appropriate precautions were instituted in handling and wiring to a sample. The short storage times resulted mainly from transistors with large leakage currents, or from contamination introduced during processing of the sample to form a membrane, or from a combination of the two. The solution is

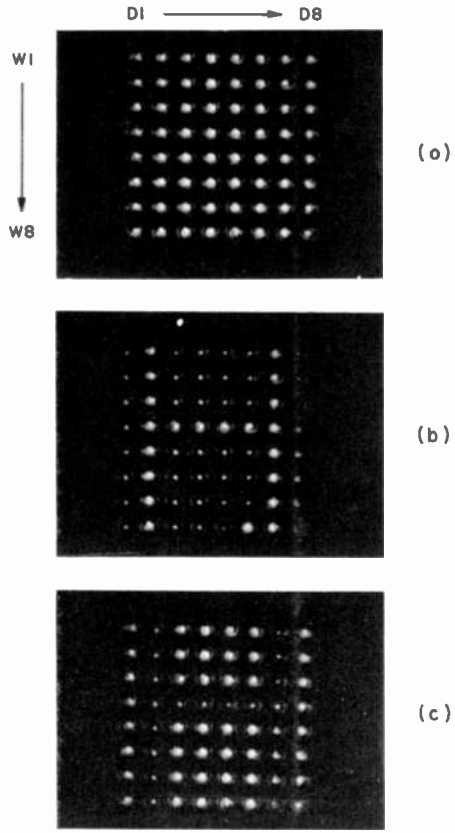


Fig. 13—Images of stored digital patterns (α = 1.9): (a) all cells bright, (b) H pattern bright, and (c) H pattern dark.

to select transistors with the lowest leakage currents and to keep substrate processing clean and to a minimum after the semiconductor chips are on the sample.

A study was also made of the factors affecting the storage time of a cell. The capacitance C of a typical cell is 2 pF. The leakage resis-

tance R is strongly dependent on conditions at the cell, including surface cleanliness and the possible presence of filaments that tend to bridge the gap between the electrode and membrane to cause short-circuits or lower resistance R. As the voltage at a cell decays, the detected light from that element changes from the dark state to the bright state value in accord with a curve such as those shown in Fig. 6. Decay times of photodetector light intensity from a few hundred milliseconds to a few seconds have been measured for clean isolated cells, with most of the cells having values at the lower end of the range. The apparent decay time of light intensity increases with drive voltage for two reasons. One is that as the membrane is deformed, C increases while R is not substantially changed, and the electrical time constant increases. The second reason is that a higher contrast and a flatter portion of the I vs. V curve is reached. Thus, as V decreases from a large value, there is little absolute change in the intensity of detected light until the shoulder of the curve is reached after which the intensity decreases rapidly to its value for the undeformed element [see Fig. 12(b)]. Because of this nonlinear curve, the light decay time would apparently increase with drive voltage even if the electrical time constant remained the same. Note too that the shape of the curve ensures that the light decay time is always less than the voltage decay time since the light intensity apparently returns to its original value at a point where V is still greater than zero.

The leakage current, I_l, of a transistor must be low enough to not reduce appreciably the decay time of an isolated cell. If we assume a constant value of I_l and a 2-pF element that is charged to 50 V, the decay time caused by I_l alone would be $t = CV/I_l = 10^{-10}/I_l$. Thus, for $t = 1$ second, I_l must be about 10^{-10} A. We have made complete arrays with $I_l \leq 0.5 \times 10^{-10}$ A, proving that the transistors need not degrade the storage time of a cell.

After the initial evaluation of the 8 × 8 page composers, further improvements were achieved on one sample by removing chips with defective transistors and epoxying good chips in their place. In this way a device was obtained with all cells independently accessible and operable. The only remaining operational flaw on this array was one cell with a short storage time.

5. 16 × 16 Element Page Composers

The design of the 16 × 16 element page composer is an extension of the 8 × 8 array design with no change in the element size or spacing or in the semiconductor chip design. The square substrate was enlarged from 5 cm to 6.35 cm on a side while the number of elements

was quadrupled (see Fig. 14). The techniques developed in making small arrays were carried over to the fabrication of large devices. Completed samples were evaluated in the simulated optical memory system. Wiring and switchboxes were arranged to select and drive each quadrant of a 16 × 16 as if it were an 8 × 8 element page composer. This method of testing allowed us to use the existing electronics test panel without modifications. Diodes were included on the

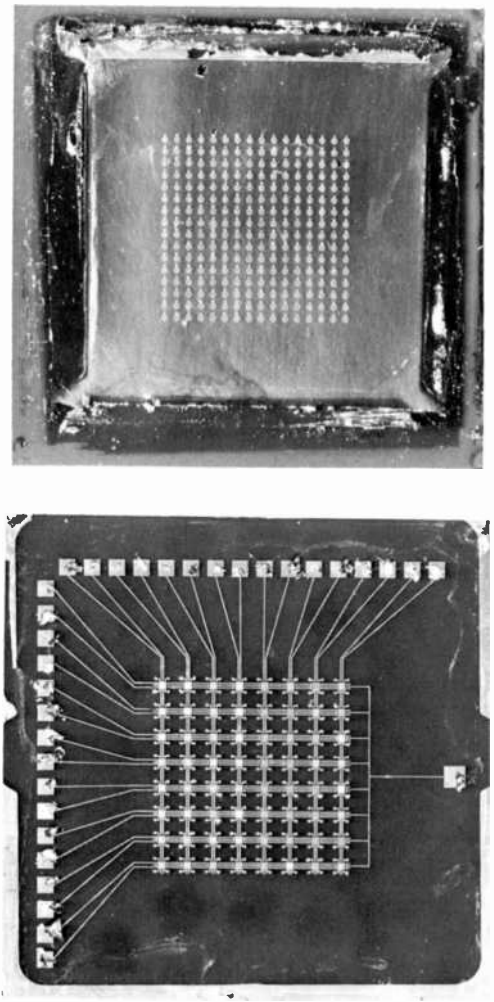


Fig. 14—A 16 × 16 element page composer. Top shows front side of membrane and bottom shows transistor array on back side.

sample holder for protection against electrical breakdown of the MOSFET gate insulation.

The 16 × 16 page composers were evaluated in the same manner as the smaller devices. Detailed measurements of rise time, optical contrast, and storage time were made on many cells. Photos of images of stored patterns were taken to illustrate the quality of the membrane surface and to demonstrate overall device operation. In order to illuminate only the active elements of the membrane, a hologram was generated using a copper mask with a 16 × 16 array of holes. When

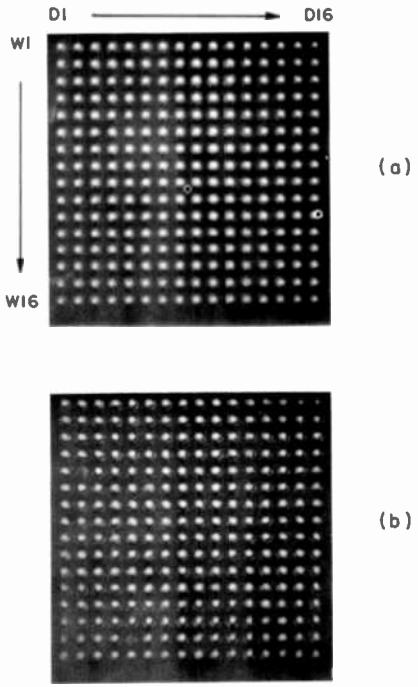


Fig. 15—Images of stored digital patterns: (a) all cells bright (no aperture) and (b) all cells bright (α = 1).

the laser output hits this hologram, an array of light beams plays out on a lens that collimates the beams and causes them to register with the page-composer elements. The hologram was made on dichromated gelatin using an argon laser and the appropriate dimensional compensation so that the pattern plays out properly with light from a helium–neon laser.

Four 16 × 16 element page composers were fabricated and tested. Three of the 4 samples had no collapsed elements. The two best samples had only a few defects and are described in more detail. One sample had a transistor array that was fully working with a maximum leakage current of 4.4×10^{-11} A after bonding. The membrane was quite flat, with no collapsed elements; two cells had surface point defects that reduced contrast substantially at those elements. In addition, five other cells were defective. Four showed a lack of electrical continuity between a drive transistor and an activating electrode and one had a short storage time. Thus, 249 of 256 cells were fully operable on this sample. Some representative photos are seen in Fig. 15. All such photos were made with an exposure time of $\frac{1}{50}$ second. An

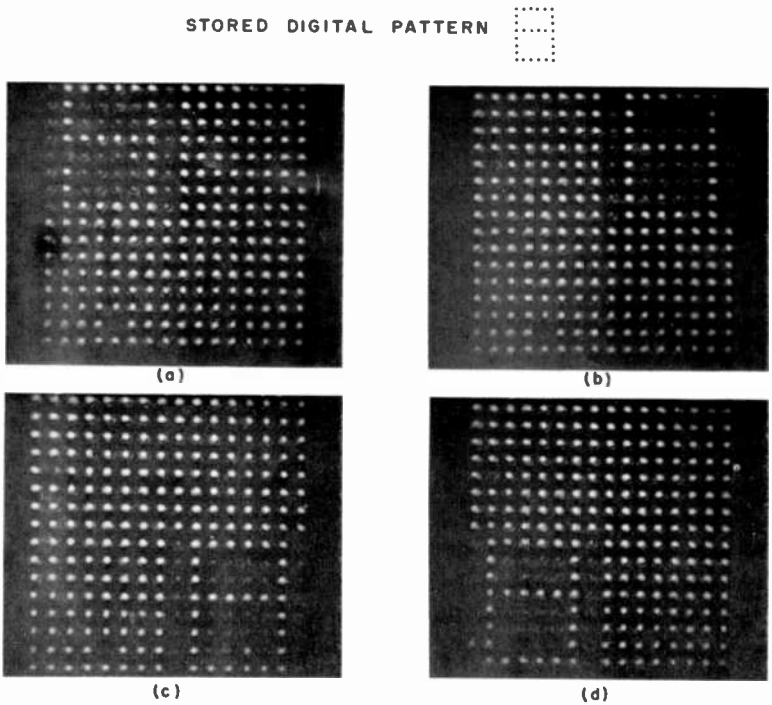


Fig. 16—Images of a stored digital pattern (α = 1): (a) pattern in quadrant 1, (b) in quadrant 2, (c) in quadrant 3, and (d) in quadrant 4.

image of the illumination dots on the circular elements of the page composer with no aperture in the optical system is shown in Fig. 15(a). Note the point defects in cells (W5,D2) and (W15,D5). Fig.

15(b) shows the image with an aperture in the system such that $\alpha = 1$. The light from the two cited cells is almost totally excluded from the aperture. A larger aperture would let more light from the two cells reach the image but contrast for all other cells would correspondingly be degraded. Some images of patterns stored in the page composer are illustrated in Fig. 16 with a few defective bits evident. A membrane bias of +20 V and a drive pulse of −45 V were applied. Typical cell measurements showed a rise time of a few hundred microseconds, a contrast of 12, and a storage time of 200 ms on this sample.

The other 16 × 16 page composer also had a completely working transistor array with a maximum leakage current of 5.5×10^{-11} A after bonding. There were no collapsed elements but there were some

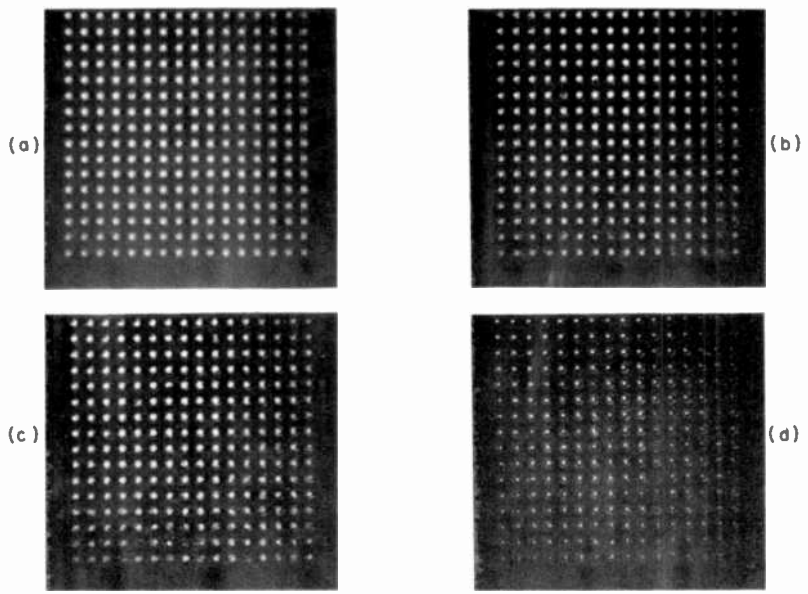


Fig. 17—Images of stored digital patterns: (a) all cells bright (no aperture), (b) all cells bright (α = 1.8), (c) all cells bright (α = 1.4), and (d) all cells dark (α = 1.8).

surface problems such as small tears or dimples, and the flatness of the membrane was not as good as that of the previous sample. All 256 elements operated with coincident voltage selection. Five cells had short storage times.

Thus, 251 of 256 cells were fully operable on this device. Some representative photos are shown in Fig. 17. The image with no aperture

in the system is shown in Fig. 17(a). The images with an aperture such that $\alpha = 1.8$ and $\alpha = 1.4$ are shown in Fig. 17(b) and 17(c), respectively. Even with $\alpha = 1.4$, many elements are dim in the image, so that a larger aperture is needed. However, contrast is thereby reduced. Fig. 17(d) shows the image with $\alpha = 1.8$ and 50 V dc applied to all elements in parallel. The drive voltage is less than that needed for

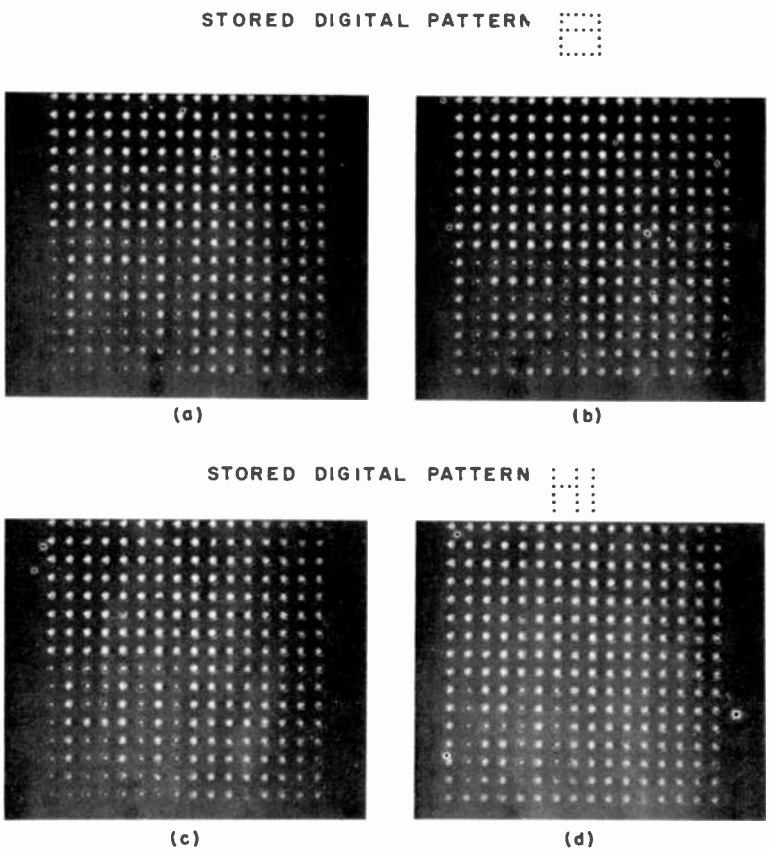


Fig. 18—Images of two different stored digital patterns in quadrant 4 ($\alpha = 1.8$): (a) and (c) bright; (b) and (d) dark.

maximum contrast at most cells, but (W6,D5), which has a surface defect, goes darkest at a lower voltage than that applied, which causes it to flatten somewhat with a reduction in contrast. Note that all elements change states when activated in this mode. Fig. 18 shows

images of patterns stored in one fully working quadrant of the sample. A membrane bias of +15 V and a drive pulse of −50 V were applied. Typical cell values measured were a rise time of 200 μs, a contrast of 6, and a storage time of 100 ms.

6. Conclusions

Operating 8 × 8 and 16 × 16 element membrane page composer prototypes were made and evaluated under realistic conditions. The electromechanical and optical operation of membrane light valves is well understood, as evidenced by the close agreement between analytical and experimental results. Fabrication technology has been developed to the point where complete devices can be made with few defects. Substrates with conductive epoxy feedthroughs are now processed routinely. Fully working transistor arrays with low leakage current for selection and storage of bits were successfully bonded on several samples. Membranes with no collapsed elements have been produced on both small and large devices. All of the techniques can be extended to arrays of larger sizes.

While the feasibility of producing a membrane page composer has been shown, no perfect device has been fabricated. Considering the complexity and technological expertise involved, this is not surprising, especially in view of the small number of devices actually made. Yield cannot be discussed realistically under such circumstances. The nature of the few problems encountered is not fundamental, and solutions exist to avoid defects or to repair those that occur.

This investigation has demonstrated that a membrane light valve page composer can serve effectively as a component of a holographic optical memory system.

Acknowledgments

It is a pleasure to acknowledge contributions to this project. J. N. Hewitt was instrumental in the fabrication of the page composers. Thanks are due many members of the RCA Solid State Technology Center at Somerville for the semiconductor work. E. M. Nagle was an able consultant on the test electronics equipment. Discussions with V. Christiano were helpful and informative.

References:

[1] L. S. Cosentino and W. C. Stewart, "A Membrane Page Composer," *RCA Rev.* 34, p. 45, March 1973.

[2] See, for example, J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill Book Co., New York, (1968).

[3] W. C. Stewart et al., "An Experimental Read-Write Holographic Memory," *RCA Rev*, **34**, p. 3, March 1973.

High-Efficiency GaAs Impatt Structures

L. C. Upadhyayula, S. T. Jolly and H. C. Huang

RCA Laboratories, Princeton, N. J.

B. J. Levin

Advanced Technology Laboratories, Camden, N. J.

Abstract—High-efficiency GaAs Impatt structures (high-low, low-high-low) were grown using the vapor hydride technique. Platinum Schottky-barrier, plated-heat-sink devices have been fabricated and evaluated. A 26.9% efficiency with 1.74 W output power and 25% efficiency with 2.1 W output power at 6.1 GHz were obtained from a low-high-low Impatt oscillator device.

Introduction

GaAs Impatt devices promise high efficiency, high power, and low noise. Because of this, considerable theoretical and experimental work has been carried out in the past few years, particularly on improving the efficiency. Read[1] has pointed out that dc-to-rf conversion takes place in the drift region, and that for maximum efficiency the voltage drop in the drift region V_D should be maximized and the voltage in the avalanche space V_A should be minimized. Assuming a voltage modulation of 50%, Read predicted a maximum efficiency of 32%. Scharfetter and Grummel,[2] using large-signal calculations and the same voltage modulation, showed that maximum efficiency for GaAs Impatts is 23%. Huang[3] and Su and Sze[4] proposed a modified Read (high-low) structure to eliminate the high field present at the interface in the Read structure and yet confine the avalanche zone to

a narrow region, thus retaining the Read efficiency for the Impatt. Salmer et al[5] analyzed the high-low and low-high-low (clump) structures and predicted a maximum efficiency of 29% for clump and 27–30% for high-low structures at 12.0 GHz. Efficiencies as high as 35–35.5% with output powers of 8.0 W at X- through Ku-band frequencies have already been reported using such high-efficiency structures.[6–8]

In order to realize these high-efficiency structures, efforts have to be redirected to the development of GaAs epitaxial growth systems. We have successfully grown high-efficiency Impatt structures by the vapor-hydride method. Integral heat sink devices were fabricated to minimize the thermal resistance and increase the device output power and efficiency. Premature burn-out in GaAs Impatts due to the low-frequency oscillation associated with large low-frequency negative resistance was discussed by Brackett.[9] Stabilizing circuits were used to minimize this problem. Small-signal device impedance measurements were made on active devices. Coaxial or radial waveguide cavities were used to provide the impedance match, and oscillator performance with each was studied. Efficiencies as high as 26.9% with 1.74 W output power and 25% with 2.1 W output power were realized at C-band.

Material Growth

Fig. 1 shows a typical doping profile required for a high-efficiency, low-high-low Impatt structure. Rapid changes of impurity concentration of three orders of magnitude within 0.1 to 0.5 μm thickness are required. With a typical growth rate of 25 μm/hr, the changes in doping concentration in the reactive gas stream occur within 15 to 20 sec. Also, the memory effects due to the adsorption of dopant gases by parts of the growth apparatus have to be avoided. The vapor hydride technique described by Tietjen et al[10] was used, with some major modifications to the growth system. Emphasis was placed on developing a capability for growing epitaxial layers having the doping profiles shown in Fig. 1. The proprietary technique developed allows us to obtain rapid changes in impurity concentration levels in the active region of the wafer. However, the transition from high to low carrier concentration at the buffer layer–active layer interface occurs in about 2 μm. By interrupting the growth process between the buffer layer and active layer, a sharper transition was obtained, as shown in Fig. 2. Several low-high-low and high-low structures were grown. 10-mil (250 μm) diameter aluminum Schottky barriers were evaporated and doping profiles were measured using analog techniques.[11] Fig. 3 shows the doping profile of a typical low-high-low epitaxial structure

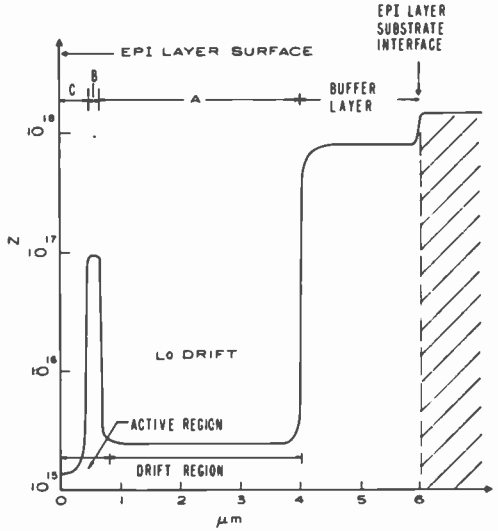


Fig. 1—Typical carrier concentration as a function of distance from the surface of the grown layer for low-high-low structures for an X-band device.

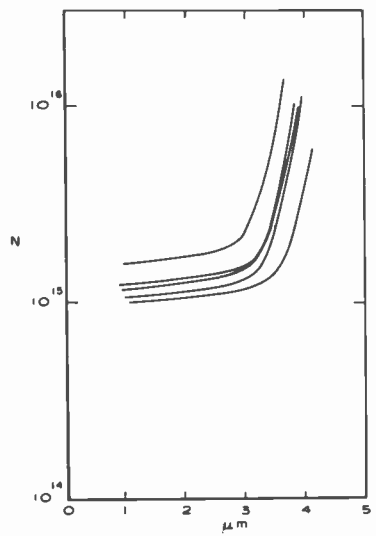


Fig. 2—Buffer/device layer interface profile with sharper transition caused by variation of growth.

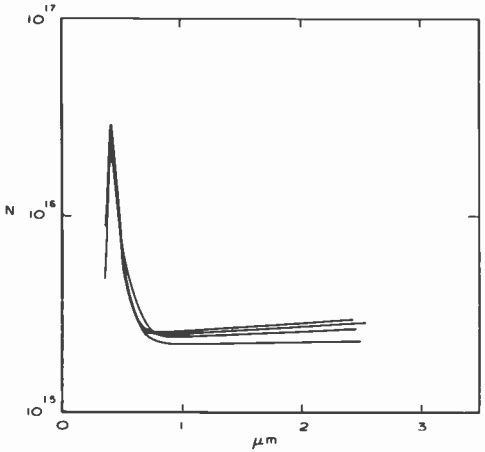


Fig. 3—Device layer profile.

that was grown. The initial low region is not seen here, as it is within the zero-bias depletion region. However, in a wafer with a wider low region (as shown in Fig. 4), the initial low regime can be clearly seen. The width of the high region at 10% of maximum carrier concentration is less than 0.25 μm.

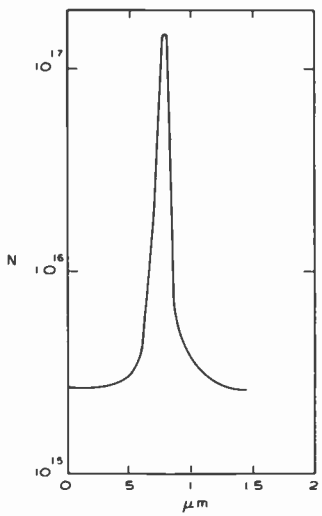


Fig. 4—Composite profile obtained by combining one profile having wider initial low region and the actual device profile.

Device Fabrication and Evaluation

Even though any metal can give a reasonably good Schottky barrier on n-GaAs, platinum Schottky barriers have been reported to be more stable and withstand higher operating junction temperatures. We prepared our Schottky barriers by sputtering platinum. Fabrication of plated heat sink devices has been reported by several laboratories. The fabrication steps followed by us (see Fig. 5) are summarized below.

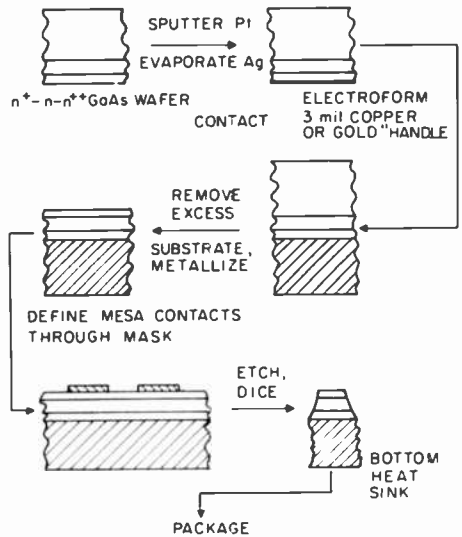


Fig. 5—Steps in fabrication process.

(1) The n^+-n-n^{++} wafer is chemically cleaned and up to 2000 Å of platinum is sputtered onto the epitaxial side.
(2) A few hundred angstroms of silver are vacuum-deposited on the sputtered platinum.
(3) The wafer is then sintered at 450°C for 5 to 10 minutes in a hydrogen ambient.
(4) 3 to 5 mils (75 to 125 μm) of gold are electroplated onto the Ag/Pt metallization.
(5) The substrate (n^{++}) is then thinned to less than 1 mil (25 μm) by chemical etching.
(6) 5- to 6-mil (125 to 150 μm) diameter Au:Ge/Au dots are vacuum-deposited on the substrate side.

(7) The gold dots are protected by photoresist techniques and mesas etched.
(8) The mesas are separated and bonded into varactor packages.

I-V characteristics and doping profile measurements were made on packaged devices to verify the material parameters. The *I-V* characteristics are somewhat softer for high-efficiency devices as compared to the conventional (uniformly doped active layer) devices.[7] Doping profiles on fabricated diodes are similar to those measured on device wafers. The doping density ratio for the high-to-low regions is about 30–40, which is desired for high-efficiency Impatt operation.

A knowledge of the active Impatt device impedance as a function of frequency is important for the design of oscillator and amplifier circuits. The rf impedance of several Impatt devices was measured on a semi-automatic network analyzer system in a manner described by Upadhyayula and Perlman.[12] Figure 6 shows the small-signal imped-

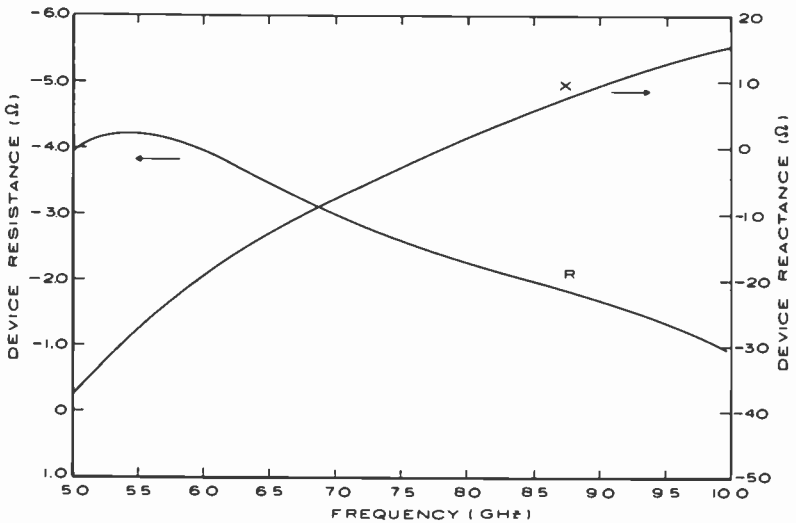


Fig. 6—Small-signal impedance data of an Impatt diode.

ance plot of a packaged GaAs Impatt diode. The device exhibited negative resistance over more than an octave band. The device Q is considerably larger at both low and high ends of the negative resistance bandwidth than in the center.

Circuit Development

Obtaining the best possible performance from Impatt oscillators requires that the microwave cavity be optimized. Impatt diodes, particularly GaAs devices, exhibit a large low-frequency (<20 MHz) negative resistance in addition to the rf negative resistance. In the past, the existence of low-frequency oscillations in Impatt diodes was found to be responsible for premature diode burn-outs. (Thermal- and tuning-induced burn-outs were also reported.) Hence, both bias and rf circuits must be designed for proper Impatt oscillator performance.

Bias Circuit Design

Brackett[9] has discussed in detail low-frequency negative resistance in GaAs Impatts and a means of eliminating the burn-out problem due to the low-frequency bias-circuit oscillations. We adopted, essentially, the bias circuit design considered by Brackett, as illustrated in Fig. 7. A 470-ohm noninductive resistor in parallel with a 50-μH in-

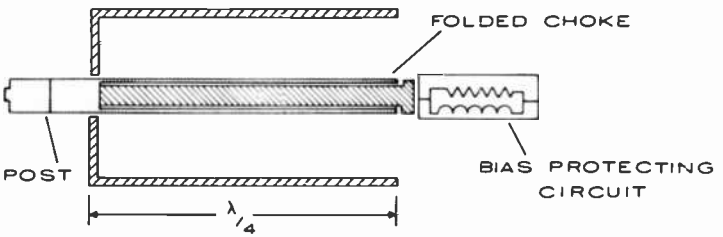


Fig. 7—Bias circuit arrangement.

ductance was used in series with the diode in the biasing circuit. To minimize the shunt capacitance seen by the diode, the bias circuit arrangement was included in the walls of the waveguide or coaxial cavity itself. A folded choke arrangement was used to establish an rf short circuit at the plane of the waveguide wall. This bias circuit design worked very well in C through X bands, and practically eliminated device burn-out due to low-frequency bias-circuit oscillation. This allowed us to operate the Impatt devices at high current densities (the low-frequency negative resistance increases with an increase in current density).

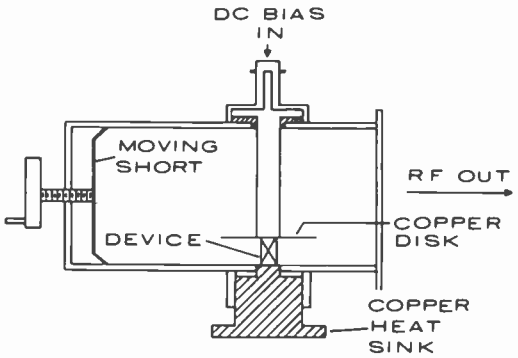


Fig. 8—Radial waveguide cavity.

RF Circuit Design

To extract the maximum available microwave power and efficiency from an Impatt diode, an optimum oscillator circuit must be provided to match both the real and imaginary parts of the diode impedance. The microwave impedance of the oscillator test circuit is designed to be adjustable over a wide range to account for the differences in the devices. Both a radial waveguide cavity and a coaxial cavity have been used successfully for device evaluation. The radial cavity has higher circuit Q and is less critical in tuning, but the tunable impedance range is smaller than that of the triple-slug coaxial cavity. Fig. 8 shows a radial waveguide cavity. The frequency of oscillation is determined by the disk diameter, the spacing between the disk and the ground plane, and the diode-junction capacitance. The

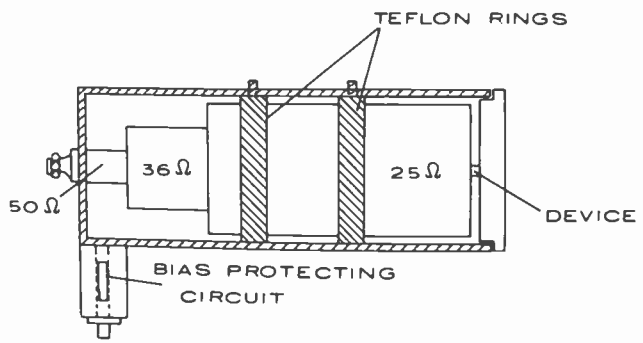


Fig. 9—Coaxial cavity.

adjustable short circuit helps match the radial-cavity impedance to the high waveguide impedance. Movement of the adjustable short circuit varies the oscillator frequency by about 2%, while changing the disk diameter and height can tune the frequency over as much as an octave.

Figure 9 shows a coaxial cavity. Step transformers and movable slugs were employed to obtain the impedance transformation re-

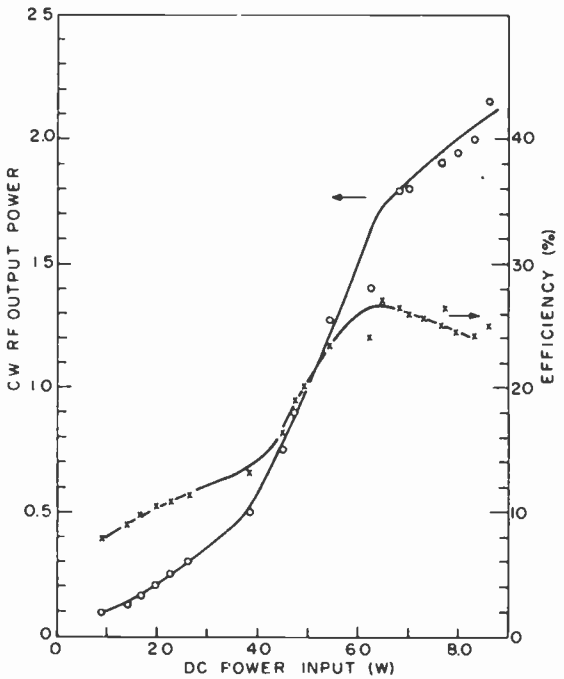


Fig. 10—RF output power and dc-to-rf conversion efficiency as a function of dc input power for a low-high-low Impatt diode.

quired to match the device, while the range of diode impedances measured under large-signal conditions served as a guideline in the transformer design. A computer program[13] was used in circuit optimization.

Oscillator Performance

Schottky-barrier Impatt devices fabricated from many wafers were rf-tested. Both waveguide and coaxial test circuits were used. The os-

cillator frequency was measured on a spectrum analyzer, and the power output was measured on a power meter. The losses in the test setup were calibrated and accounted for. The best performance was obtained on the platinum Schottky-barrier low-high-low junction Impatt.

Fig. 10 shows the rf output power and efficiency as a function of dc input power for a low-high-low Impatt device. A sudden increase in output power and efficiency occurs at 4–5 W input power levels, corresponding to a current density of about 230 A/cm^2. A 26.9% efficiency at 1.74 W output power and 25% efficiency at 2.1 W output power were obtained at 6.1 GHz. The device current and bias voltage of this Impatt device are shown in Fig. 11. Output power and efficiency are

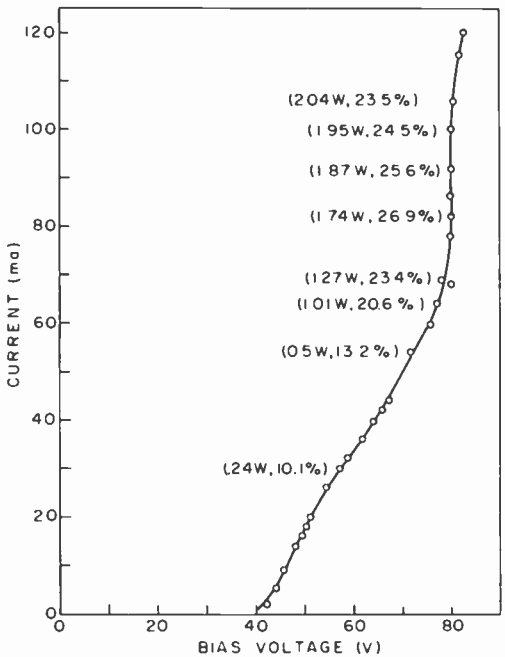


Fig. 11—Current-voltage characteristic of a low-high-low Impatt diode when operating in the high-efficiency mode. The output power and efficiency are indicated along the curve for 6.1 GHz operation.

indicated along the curve for 6.1 GHz operation. The soft I-V curve is a characteristic of high-low and low-high-low Impatt devices. At about 230 A/cm^2 current density, the high efficiency mode of operation occurs. A 20% efficiency with 1.0 W output power was measured

at this point. For current densities higher than this, the bias voltage remained practically constant. Maximum efficiency of 26.9% was measured at 305 A/cm^2 current density and maximum output power of 2.1 W was measured at 395 A/cm^2 current density.

Summary

Epitaxial GaAs structures suitable for high efficiency Impatt devices have been grown using the vapor hydride technique. Platinum Schottky-barrier devices with integral heat sink have been fabricated and evaluated.

A low high-low GaAs Impatt device with a high-to-low doping ratio of about 30–40 has yielded a dc-to-rf conversion efficiency of 26.9% with 1.74 W output power at C band. The theoretical efficiency for a high-low Impatt with such a doping profile is about 30%. The experimental result is therefore in good agreement with theory. Other researchers[6,7] working with similar doping profiles have shown that the high efficiency operation of Impatts occurs at current densities between 1000 and 1600 A/cm^2. In our case, we observed that high efficiency operation can be achieved at current densities as low as 300–400 A/cm^2.

Acknowledgments

The authors gratefully acknowledge the helpful suggestions and encouragement received from S. Y. Narayan. The authors also wish to thank K. Pinkerton for the fabrication of the devices and James F. Wilhelm for the fabrication of circuits. This work was supported by Electronics Command, Fort Monmouth, N.J., under Contract No. DAABO7-72-C-0308.

References:

[1] W. T. Read, "A Proposed High-Frequency Negative Resistance Diode," *Bell Syst. Tech. J.*, **37**, p. 401, March 1958.

[2] D. L. Scharfetter and K. H. Gummel, "Large Signal Analysis of a Silicon Read Diode Oscillator," *IEEE Trans. Elec. Devices*, **ED-16**, p 64, Jan. 1969.

[3] H. C. Huang, "A Modified GaAs Impatt Structure for High Efficiency Operation," *IEEE Trans. Elec. Devices*, **ED-20**, p. 482, May 1973.

[4] S. Su and S. M. Sze, "Design Considerations of High-Efficiency GaAs Impatt Diodes," *IEEE Trans. Elec. Devices*, **ED-20**, p. 541, June 1973.

[5] G. Salmer et al, "Theoretical and Experimental Study of GaAs Impatt Oscillator Efficiency," *J. Appl. Phys.*, **44**, p. 314, Jan. 1973.

[6] K. Kim and W. G. Matthei, "GaAs Read Impatt Diode Oscillators," *Proc. Fourth Biennial Cornall Electrical Engineering Conf.*, p. 299 (1973).

[7] H. W. Tserng, W. R. Wisseman and T. E. Hasty, "Characterization of GaAs Schottky-Read Impatt Diodes." Technical Digest of Intl. Elec. Devices Meeting, Wash. D.C., p. 483 (1973).

[8] R. E. Goldwasser and P. E. Rosztoczy, "High Efficiency Low-High-Low Impatt Devices by Liquid Phase Epitaxy for X-band," *Applied Phys. Lett.*, **25**, No. 1, p. 92, 1 July 1974.

[9] C. A. Brackett, "The Elimination of Tuning Induced Burnout and Bias Circuit Oscillations in Impatt," *Bell Syst. Tech. J.*, **52**, No. 3, p. 271, March 1973.

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

[11] Instruction Manual for Impurity Profile Plotter, Type 336, JAC Electronics, Ltd., Surrey, England.

[12] Chainulu L. Upadhyayula and B. S. Perlman, "Design and Performance of Transferred Electron Amplifiers Using Distributed Equalizer Networks," *IEEE J. Solid-State Circuits*, **SC-8**, No. 1, Feb. 1973.

[13] B. S. Perlman, private communication.

Empirical Relationships Between Thermal Conductivity and Temperature for Silicon and Germanium

Achilles G. Kokkas

RCA Laboratories, Princeton, N. J. 08540

Abstract—The temperature dependence of the thermal conductivity of Si and Ge in the range most important tc semiconductor devices can be accurately expressed in the form $A(T - B)^{-1}$ or aT^{-b}. Values are suggested for A, B, a, and b.

The ever increasing demand for more power from discrete and integrated devices and for higher packing densities in IC's has made thermal management the concern not only of heat specialists but of electronic engineers as well. Device engineers are expected to optimize the thermal capabilities of their components, while circuit engineers are expected to select the components most appropriate for their design and provide adequate means for the removal of the generated heat.

In the course of developing techniques for the thermal analysis of semiconductor structures, it became necessary to express in closed form the dependence on temperature of the thermal conductivity of silicon and germanium. An empirical relationship which was found to represent accurately the deviation of high-temperature thermal conductivity from the theoretical $1/T$ dependence is

$$k(T) = \frac{A}{T - B}, \qquad [1]$$

where k is the thermal conductivity in W/cm-°K, T is the tempera-

ture in degrees Kelvin, and A and B are constants. By choosing A and B respectively, to be 320 and 80 for pure silicon and 152 and 46.5 for pure germanium, excellent agreement is obtained between Eq. [1] and the data of Reference [1] (see Fig. 1) in the range 220 °K < T < 600 °K.

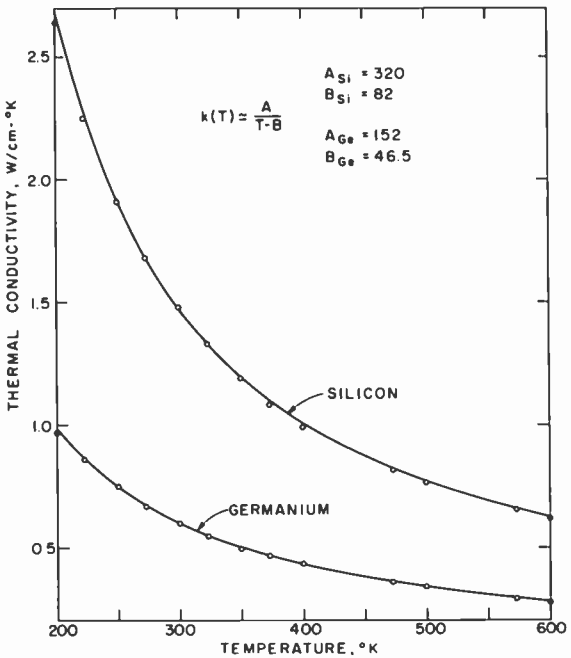


Fig. 1—Thermal conductivity of Si and Ge as a function of temperature. The curves are plots of Eq. [1] and the circles are data points from Reference [1].

Another empirical expression, particularly useful for temperatures less than about 400 °K, is

$$k(T) = aT^{-b}, \qquad [2]$$

where a and b are, respectively, 4,350 and 1.4 for silicon and 533 and 1.19 for germanium.

A linear relationship between $k(T)$ and T, of the type assumed for GaAs in References [2] and [3], yields poor results for Si and Ge. In contrast, although a "best fit" to experimental data has not been at-

tempted, it appears that the use of Eqs. [1] and [2] can be extended to the case of GaAs.

References

[1] C. Y. Ho, R. W. Powell, and P. E. Liley, "Thermal Conductivity of the Elements," *J. Phys. Chem. Ref. Data*, **1**, No. 2, p. 279 (1972). (For Si see p. 394 and for Ge p. 339)

[2] W. Fallmann, H. L. Hartnagel, and P. C. Mathur, "Experiments on Heat Sinking of Semiconductor Devices," *Electron. Letters*, **7**, No. 18, p. 512, 9 Sept. 1971.

[3] H. Hartnagel and V. C. Hutson, "Thermal Resistance of Planar Semiconductor Structures," *Proc. IEE*, (G.B.), **119**, No. 6, p 655, June 1972.

An Introduction to the Science and Technology of Liquid Crystals—III

The next five papers in this issue comprise the third installment in a series of lectures on liquid crystals being published in RCA Review. This series grew out of a weekly study seminar conducted by members of the liquid-crystal activity at RCA Laboratories to stimulate increased interaction among individuals from different disciplines who share a common interest in liquid crystals. Thus, these papers are intended as a primer in the science and technology of liquid crystals for those using or contemplating the use of these devices.

The two earlier installments of papers in this series appeared in the March and September 1974 issues of RCA Review:

March

Liquid Crystal Mesophases
E. B. Priestley

Structure-Property Relationships in Thermotropic Organic Liquid Crystals
Aaron W. Levine

Introduction to the Molecular Theory of Nematic Liquid Crystals
P. J. Wojtowicz

Generalized Mean Field Theory of Nematic Liquid Crystals
P. J. Wojtowicz

Hard Rod Model of the Nematic-Isotropic Phase Transition
Ping Sheng

Nematic Order: The Long Range Orientational Distribution Function
E. B. Priestley

September

Introduction to the Molecular Theory of Smectic-A Liquid Crystals
Peter J. Wojtowicz

Introduction to the Elastic Continuum Theory of Liquid Crystals
Ping Sheng

Electrohydrodynamic Instabilities in Nematic Liquid Crystals
Dietrich Meyerhofer

Liquid-Crystal Displays—Packaging and Surface Treatments
L. A. Goodman

Pressure Effects in Sealed Liquid-Crystal Cells
Richard Williams

Introduction to the Optical Properties of Cholesteric and Chiral Nematic Liquid Crystals

E. B. Priestley

RCA Laboratories, Princeton, N. J. 08540

Abstract—The wave equation is solved for the case of plane waves propagating parallel to the axis z of an ideal helix, which, for any plane z = constant, is characterized by two indices of refraction $n_x(n_\parallel)$ and $n_y(n_\perp)$. Four solutions are obtained, two corresponding to waves traveling in the positive z-direction and two to waves traveling in the negative z-direction. One of the two waves propagating in each direction is found to be strongly reflected when its wavelength becomes comparable to the pitch of the helix. The polarization vectors of the eigenmodes are determined, and it is shown that the strongly reflected wave is the one whose instantaneous spatial electric field pattern is a helix that is superposable on the cholesteric or chiral nematic helix. The other wave is found to propagate without significant reflection loss, independent of its wavelength. Finally, the optical rotatory power of the helical structure is discussed.

Introduction

The results derived here, and the related discussion, apply both to cholesteric and to chiral nematic liquid crystals; however, in the interest of brevity, we refer specifically only to cholesteric liquid crystals.

The helical arrangement of the molecules in a cholesteric phase has been described in an earlier paper.[1] On a sufficiently microscopic scale one cannot distinguish between cholesteric and nematic ordering. However, as we consider larger and larger volumes of the two types of material, a difference in the molecular ordering begins to be-

come apparent; we observe that the cholesteric director $\hat{n}$ follows a helix

$$\left.\begin{aligned}(\hat{n})_x &= \cos(q_0 z + \varphi)\\ (\hat{n})_y &= \sin(q_0 z + \varphi)\\ (\hat{n})_z &= 0\end{aligned}\right\}, \qquad [1]$$

as shown in Fig. 1, whereas this secondary, helical structure is absent in the nematic phase. In general, both the direction of the helix axis z in space and the magnitude of the constant φ are arbitrary. It is evident from Fig. 1 that the structure of a cholesteric liquid crystal is periodic with a spatial period

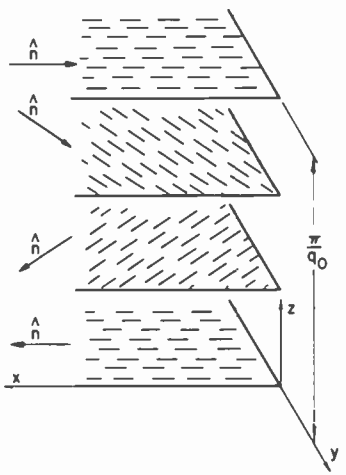


Fig. 1—A schematic representation of the helical arrangement of the constituent molecules in a cholesteric liquid crystal.

$$L = \pi/|q_0|. \qquad [2]$$

In a right-handed coordinate system, Eq. [1] describes a right-handed helix for positive q_0 and a left-handed helix for negative q_0. Thus the sign of q_0 determines the sense of the helix and its magnitude determines the spatial periodicity.

Several unique optical properties arise from this spatially periodic, helical structure of cholesteric liquid crystals.

(1) Bragg reflection of light beams is observed. For light incident parallel to the helix axis z, only the lowest-order reflection is al-

lowed; for oblique incidence the higher-order reflections also become allowed.

(2) The Bragg reflected light is circularly polarized if the incident wave propagates parallel to the z-axis and elliptically polarized for oblique incidence.

(3) Only the component of optical polarization for which the instantaneous spatial electric field pattern matches the spiraling cholesteric director is strongly reflected. The other component is transmitted with no significant reflection loss.

(4) Very strong rotatory power is observed. Rotations of tens of revolutions per millimeter are typical, compared to the fraction of a revolution per millimeter characteristic of isotropic, optically active liquids.

A detailed treatment of the optical properties of cholesteric liquid crystals for obliquely incident light beams involves extensive numerical calculations[2] and is outside the scope of this paper. However, much of the physics underlying the observed optical behavior of these spiral structures can be understood by considering the more restricted case in which the wave vector of the incident light is everywhere normal to the local director, i.e., $\mathbf{k} \| z$. The development presented below parallels closely that of de Vries;[3] however, the approach is somewhat different.[4]

Maxwell's Equations

As noted above we treat only waves propagating along the helix axis z; $\mathbf{D}$ and $\mathbf{E}$ are therefore confined to the xy-plane and are related by a two-dimensional, second-rank dielectric tensor $\bar{\epsilon}$. In addition to this restriction, we neglect (1) the weak intrinsic optical activity of the constituent molecules which persists even in the isotropic phase, (2) energy dissipation by absorption, and (3) magnetic permeability ($\mu = 1$). Finally, we assume all waves to be of the form

$$\mathrm{Re}\{f(z) \exp(-i\omega t)\} \qquad [3]$$

(Re $\equiv$ real part) so that $\nabla^2 \rightarrow \partial^2/\partial z^2$ and $\partial^2/\partial t^2 \rightarrow -\omega^2$. With these assumptions, Maxwell's equations reduce to

$$\frac{\partial^2 \mathbf{E}}{\partial z^2} + \left(\frac{\omega}{c}\right)^2 \bar{\epsilon}\,\mathbf{E} = 0. \qquad [4]$$

One could at this point proceed to solve Eq. [4] with $\bar{\epsilon}$ suitably expressed as a periodic function of z in the fixed laboratory frame of

reference. Then, by Floquet's theorem,[6] we know that there exist solutions to Eq. [4] such that

$$\mathbf{E}(z + L) = K\mathbf{E}(z) \qquad [5]$$

where K is a constant that may be complex. However, it is somewhat simpler to solve Eq. [4] if we first transform it to a coordinate system that rotates with the cholesteric helix. In the rotating frame $\vec{\epsilon}$ has a simple diagonal form for all values of z. In making this transformation we will utilize the Pauli matrices and we therefore digress briefly to consider the properties of these matrices.

Pauli Matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are known as the Pauli matrices. Their principal properties, which can be deduced from their explicit form, are summarized by

$$\sigma_j\sigma_k = i\epsilon_{jkl}\sigma_l + \delta_{jk}\sigma_0 \qquad [6]$$

where j, k and l can independently take on the values 1, 2 or 3. ϵ_{jkl} in Eq. [6] is the Levi-Cevita antisymmetric symbol,[6] which behaves as follows:

$$\epsilon_{jkl} = \begin{cases} 1 \text{ if } jkl = 123, 231, 312 \\ -1 \text{ if } jkl = 213, 321, 132 \\ 0 \text{ otherwise.} \end{cases}$$

σ_0 is the 2×2 unit matrix. Inspection of Eq. [6] reveals that the Pauli matrices anticommute, i.e., $\sigma_j\sigma_k + \sigma_k\sigma_j = 0, j \neq k$.

In transforming Eq. [4] to the rotating frame of reference we make use of the exponential operator, $\exp(-i\sigma_j\theta)$. Its properties can best be seen from a series expansion of the exponential, viz,

$$\exp(-i\sigma_j\theta) = \sigma_0 - i\sigma_j\theta + \frac{1}{2!}(i\sigma_j\theta)^2 - \frac{1}{3!}(i\sigma_j\theta)^3 + \frac{1}{4!}(i\sigma_j\theta)^4 + \ldots. \qquad [7]$$

Regrouping the terms in this expansion, and bearing in mind the properties of the Pauli matrices summarized in Eq. [6], we obtain

$$\exp(-i\sigma_j\theta) = \sigma_0\left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - + \ldots\right)$$
$$- i\sigma_j\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - + \ldots\right), \qquad [8]$$

where the two power series in parentheses will be recognized as expansions of $\cos\theta$ and $\sin\theta$. Thus

$$\exp(-i\sigma_j\theta) = \sigma_0\cos\theta - i\sigma_j\sin\theta. \qquad [9]$$

The reader can check that for $j = 2$, Eq. [9] becomes

$$\exp(-i\sigma_2\theta) = \begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}, \qquad [10]$$

which is the rotation matrix for a vector in a plane. The utility of Eq. [10] in our present problem results from the ease and compactness with which the coordinate transformation can be made using the exponential operator notation. Notice that the z-dependence is contained in the relationship

$$\theta = 2\pi z / P. \qquad [11]$$

where $P = 2L$ is the pitch of the cholesteric structure.

Maxwell's Equations (Cont'd)

As mentioned previously, $\overleftrightarrow{\epsilon}$ can be written in diagonal form in the rotating coordinate system. Letting $\epsilon_\parallel$ and $\epsilon_\perp$ represent the dielectric constant parallel and perpendicular to the local director $\hat{n}$, respectively, it follows that

$$\overleftrightarrow{\epsilon} = \begin{pmatrix}\epsilon_0 + \epsilon_1 & 0 \\ 0 & \epsilon_0 - \epsilon_1\end{pmatrix},$$

which, in terms of σ_0 and σ_3, is simply

$$\overleftrightarrow{\epsilon} = \epsilon_0\sigma_0 + \epsilon_1\sigma_3. \qquad [12]$$

$\epsilon_0 = (\epsilon_\parallel + \epsilon_\perp)/2$ is the mean dielectric constant, and $\epsilon_1 = (\epsilon_\parallel - \epsilon_\perp)/2$ is a measure of the dielectric anisotropy, in the xy plane.

The wave equation in the rotating frame of reference is (see Appendix 1)

$$\left(\frac{\partial^2 \mathbf{E}'}{\partial z^2}\right) - \left(\frac{4\pi i \sigma_2}{P}\right)\left(\frac{\partial \mathbf{E}'}{\partial z}\right) + \left[\left(\frac{\omega}{c}\right)^2(\epsilon_0\sigma_0 + \epsilon_1\sigma_3) - \left(\frac{4\pi^2\sigma_0}{P^2}\right)\right]\mathbf{E}' = 0 \quad [13]$$

where the x and y axes of the rotating coordinate system have been fixed parallel and perpendicular to $\hat{n}$, respectively. We try as a solution to Eq. [13]

$$\mathbf{E}' = \mathrm{Re}\left\{\hat{u}E_0 \exp\left[-i\left(\omega t - \frac{2\pi m \epsilon_0^{1/2} z}{\lambda}\right)\right]\right\}. \quad [14]$$

where

$$\hat{u} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad [15]$$

is a (complex) two-component vector that describes the state of polarization of the wave in the rotating frame, E_0 is a real constant that gives the amplitude of the wave, and λ is the wavelength in *vacuo* of the wave. The product $m\,\epsilon_0^{1/2}$ plays the role of the refractive index of the cholesteric material; however it is not strictly correct to think of it as such. As we shall see later, the m values are complicated functions of the pitch and dielectric anisotropy of the medium and of the wavelength of the electromagnetic wave. ϵ_0 in Eq. [12] is an average (optical frequency) dielectric constant defined above. Eq. [13] then reduces to

$$\left[-\frac{4\pi^2 c^2}{\omega^2 \epsilon_0}\left(\frac{\sigma_0}{P^2} - \frac{2\sigma_2 m \epsilon_0^{1/2}}{P\lambda} + \frac{m^2 \epsilon_0 \sigma_0}{\lambda^2}\right) + \sigma_0 + \left(\frac{\epsilon_1}{\epsilon_0}\right)\sigma_3\right]\hat{\mu} = 0. \quad [16]$$

Notice that $4\pi^2 c^2/\omega^2$ is just the square of the vacuum wavelength λ. Defining $\alpha = \epsilon_1/\epsilon_0$ and $\lambda' = \lambda/\epsilon_0^{1/2}P$ simplifies Eq. [16] to

$$[(1 - (\lambda')^2 - m^2)\sigma_0 + \alpha\sigma_3 + 2m\lambda'\sigma_2]\hat{\mu} = 0. \quad [17]$$

which can be written explicitly as

$$\begin{pmatrix} 1 - (\lambda')^2 - m^2 + \alpha & -2im\lambda' \\ 2im\lambda' & 1 - (\lambda')^2 - m^2 - \alpha \end{pmatrix}\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = 0. \quad [18]$$

The two simultaneous equations represented by Eq. [18] have a nontrivial solution only if

$$\begin{vmatrix} 1 - (\lambda')^2 - m^2 + \alpha & -2im\lambda' \\ 2im\lambda' & 1 - (\lambda')^2 - m^2 - \alpha \end{vmatrix} = 0. \qquad [19]$$

The resulting fourth-order equation in m

$$m^4 - 2[1 + (\lambda')^2]m^2 + [1 - (\lambda')^2]^2 - \alpha^2 = 0 \qquad [20]$$

has solutions

$$m^2 = 1 + (\lambda')^2 \pm [4(\lambda')^2 + \alpha^2]^{1/2}. \qquad [21]$$

It will be useful to have more explicit expressions for the roots given by Eq. [21] in two limiting cases. In both limits we will expand Eq. [21] and keep only the lowest-order terms.

(a) The $4(\lambda')^2/\alpha^2 \ll 1$ *Limit*

In this limit the pitch is large compared to the wavelength of the light. The square root in Eq. [21] can be expanded to give

$$m^2 = 1 + (\lambda')^2 \pm \left(\alpha + \frac{2(\lambda')^2}{\alpha} + \ldots\right) \qquad [22]$$

whence

$$m_1 = -m_3 = (1 - \alpha)^{1/2} + \frac{(\lambda')^2(\alpha - 2)}{2\alpha(1 - \alpha)^{1/2}} + \ldots \qquad [23]$$

and

$$m_2 = -m_4 = (1 + \alpha)^{1/2} + \frac{(\lambda')^2(\alpha + 2)}{2\alpha(1 + \alpha)^{1/2}} + \ldots \qquad [24]$$

The positive and negative roots are associated with waves traveling through the cholesteric medium in the positive and negative z directions, respectively.

(b) The $4(\lambda')^2/\alpha^2 \gg 1$ *Limit*

In this case the pitch is small compared to the wavelength of the light and the appropriate expansion of Eq. [21] is

$$m^2 = 1 + (\lambda')^2 \pm \left(2\lambda' + \frac{\alpha^2}{4\lambda'} + \ldots\right), \qquad [25]$$

from which it follows that

$$m_1 = -m_3 = 1 - \lambda' - \frac{\alpha^2}{8\lambda'(1 - \lambda')} + \ldots \qquad [26]$$

and

$$m_2 = -m_4 = 1 + \lambda' + \frac{\alpha^2}{8\lambda'(1 + \lambda')} + \cdots. \qquad [27]$$

Again the plus and minus signs correspond to waves traveling in opposite directions along the helix axis.

The results we have obtained above are identical to those given by de Vries.[3]

Now that we have explicit expressions for the roots m, it is simple to determine the polarization vectors $\hat{\mu}$ of the corresponding modes. We begin by rewriting Eq. [18] schematically as

$$\begin{pmatrix} a + \alpha & -ib \\ ib & a - \alpha \end{pmatrix}\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = 0, \qquad [28]$$

where $a = 1-(\lambda')^2-m^2$ and $b = 2m\lambda'$. Expanding Eq. [28], we have

$$(a + \alpha)\mu_1 - ib\mu_2 = 0, \qquad [29]$$

and

$$ib\mu_1 + (a - \alpha)\mu_2 = 0. \qquad [30]$$

Eqs. [29] and [30] require

$$\mu_2\mu_2^* = \left(\frac{a + \alpha}{a - \alpha}\right)\mu_1\mu_1^*, \qquad [31]$$

which, when combined with the normalization condition

$$\mu_1\mu_1^* + \mu_2\mu_2^* = 1,$$

leads to

$$\mu_1\mu_1^* = \left(\frac{a - \alpha}{2a}\right), \qquad [32]$$

and hence

$$\mu_2\mu_2^* = \left(\frac{a + \alpha}{2a}\right). \qquad [33]$$

The relative phase of μ_1 and μ_2 is fixed by Eq. [29], viz

$$\frac{\mu_2}{\mu_1} = -i\left(\frac{a + \alpha}{b}\right). \qquad [34]$$

Substituting for a and b in Eqs. [32] through [34] yields the following

expressions which, together with the appropriate m values, determine the polarization vectors $\hat{\mu}$:

$$\mu_1\mu_1^* = \frac{1 - (\lambda')^2 - m^2 - \alpha}{2[1 - (\lambda')^2 - m^2]}, \quad [35]$$

$$\mu_2\mu_2^* = \frac{1 - (\lambda')^2 - m^2 + \alpha}{2[1 - (\lambda')^2 - m^2]}, \quad [36]$$

$$\frac{\mu_2}{\mu_1} = -i\left(\frac{1 - (\lambda')^2 - m^2 + \alpha}{2m\lambda'}\right). \quad [37]$$

For the case $4(\lambda')^2/\alpha^2 \ll 1$, we have

$$\hat{\mu}(m_1) = \hat{\mu}(m_3) \approx \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad [38]$$

$$\hat{\mu}(m_2) = \hat{\mu}(m_4) \approx \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad [39]$$

These vectors describe linearly polarized light. Thus, the modes are essentially linearly polarized near the origin of the λ' axis in Fig. 2. In the other extreme, $4(\lambda')^2/\alpha^2 \gg 1$, the eigenvectors are

$$\hat{\mu}(m_1) = \hat{\mu}(m_4) \approx \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix} \quad [40]$$

$$\hat{\mu}(m_2) = \hat{\mu}(m_3) \approx \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, \quad [41]$$

which describe circularly polarized light. Thus the modes are essentially circularly polarized for large values of λ'.

Discussion

Eq. [21] has been evaluated numerically with α arbitrarily set equal to 0.1. The results for m_1 and m_2 are plotted in Fig. 2. The qualitative features of Fig. 2 are apparent from the approximate expressions for the roots, Eqs. [23], [24], [26], and [27]. For example, in the limit of very small λ', Eqs. [23] and [24] reduce to

$$m_1 = (1 - \alpha)^{1/2} \text{ and } m_2 = (1 + \alpha)^{1/2}.$$

These equations also show that the initial dependence on λ' is qua-

dratic. For large values of λ', on the other hand, Eqs. [26] and [27] become linear in λ', as observed in Fig. 2.

$$m_1 = 1 - \lambda' \text{ and } m_2 = 1 + \lambda'.$$

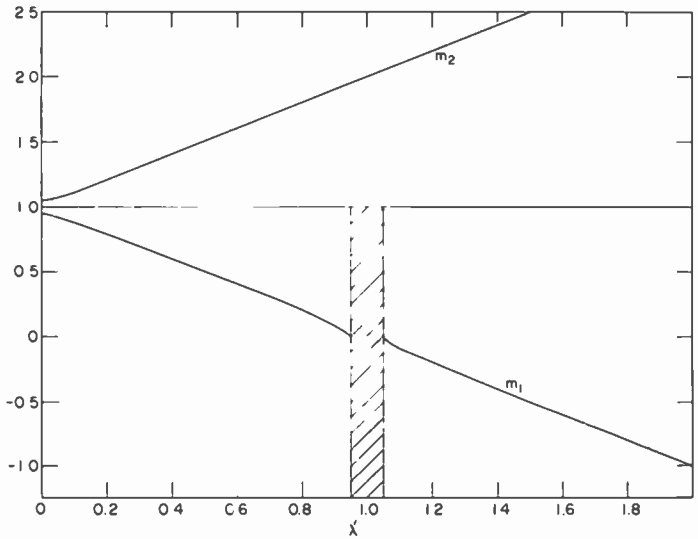


Fig. 2—Solutions given by Eq. [21], with $\alpha = 0.1$ as a function of reduced wavelength λ' for the two waves propagating in the positive z direction. In the shaded region the m_1 wave is strongly reflected.

We have seen by Eqs. [38] and [39] that the normal waves are linearly polarized in the local frame of reference for $\lambda' \to 0$. As the local frame rotates ($\theta = 2\pi z/P$), so do the polarization vectors of the normal waves. This is the "waveguide" regime discussed by de Gennes[7] and is also the regime in which twisted nematic field-effect devices[8] operate (values typical of such devices are $P \sim 50\ \mu$m, $\epsilon_0^{1/2} \sim 1.65$ and $\lambda \sim 0.5\ \mu$m, leading to a value of $\lambda' \sim 0.006$).

For values of λ' near unity, i.e., within the shaded region of Fig. 2, m_1 and m_3 are imaginary and the corresponding waves are nonpropagating. It is apparent from Eq. [21] that this reflection band extends over the range of λ'

$$(1 - \alpha)^{1/2} < \lambda' < (1 + \alpha)^{1/2}.$$

The m_2 and m_4 waves are unaffected and are observed to propagate freely for all values of λ'. Considering now only the waves traveling in

the plus z-direction, we see from Eq. [41] that the m_2 wave is left circularly polarized and that its instantaneous electric field pattern is of opposite sense to the (right-handed) cholesteric helix (see Appendix 2). The strongly reflected m_1 wave, on the other hand, is right circularly polarized and has an instantaneous electric field pattern that is superposable on the cholesteric helix. Thus, for $\lambda' \sim 1$, a right-handed cholesteric liquid crystal reflects right circularly polarized light and transmits left circularly polarized light. The reverse is true for a left-handed cholesteric material.

For larger values of λ', the normal waves are nearly circularly polarized (see Fig. 3), have opposite signs of rotation, and propagate

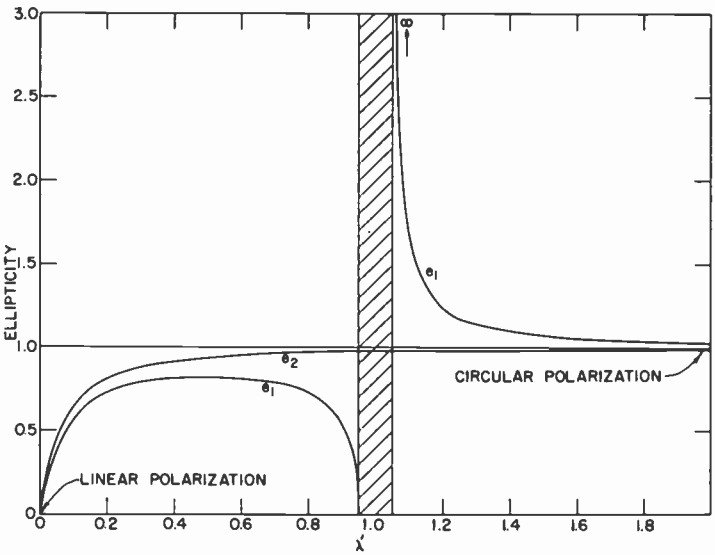


Fig. 3—The ellipticities $e_1 = |\mu_1(m_1)/\mu_2(m_1)|$ and $e_2 = -|\mu_2(m_2)/\mu_1(m_2)|$ as a function of reduced wavelength λ' for the two waves propagating in the positive z direction.

with different phase velocities. A superposition of two waves of opposite circular polarization can be thought of as a linearly polarized wave whose plane of polarization lies along the bisector of the instantaneous angle between the two rotating vectors. Since the two circularly polarized waves travel at different velocities in the cholesteric medium, there is a net rotation of the plane of polarization when a linearly polarized wave is passed through a slab of the cholesteric material. We can estimate the magnitude of this rotation in a straightforward manner using the results derived above.

Again we consider only the m_1 and m_2 waves and we restrict ourselves to the $4(\lambda')^2/\alpha^2 \gg 1$ limit. We have seen that in this limit the m_1 wave has the form

$$E_1' = \mathrm{Re}\left\{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}E_0 \exp\left[-i\left(\omega t - \frac{2\pi m_1 \epsilon_0^{1/2} z}{\lambda}\right)\right]\right\} \quad [42]$$

in the rotating coordinate system, with components

$$(E_1')_x = \frac{E_0}{\sqrt{2}} \cos\left(\omega t - \frac{2\pi m_1 \epsilon_0^{1/2} z}{\lambda}\right) \quad [43]$$

and

$$(E_1')_y = -\frac{E_0}{\sqrt{2}} \sin\left(\omega t - \frac{2\pi m_1 \epsilon_0^{1/2} z}{\lambda}\right). \quad [44]$$

Thus, by Eq. [52], we can write $(E_1)_x$ in the laboratory frame of reference as

$$(E_1)_x = (E_1')_x \cos\left(\frac{2\pi z}{P}\right) - (E_1')_y \sin\left(\frac{2\pi z}{P}\right), \quad [45]$$

which, upon substitution of Eqs. [43] and [44], reduces immediately to

$$(E_1)_x = \frac{E_0}{\sqrt{2}} \cos\left[\frac{2\pi z}{P} - \left(\omega t - \frac{2\pi m_1 \epsilon_0^{1/2} z}{\lambda}\right)\right] \quad [46]$$

From Eq. [46] we see that the phase angle θ_1 of the m_1 wave is

$$\theta_1 = \frac{2\pi z}{P}\left(1 + \frac{m_1}{\lambda'}\right) - \omega t \quad [47]$$

in the laboratory frame of reference. Similarly we find for the m_2 wave

$$\theta_2 = \frac{2\pi z}{P}\left(1 - \frac{m_2}{\lambda'}\right) + \omega t. \quad [48]$$

The angular position of the resultant linear polarization vector is determined by

$$\psi = \frac{1}{2}(\theta_1 + \theta_2), \qquad [49]$$

which is the bisector of the instantaneous angle between the two rotating electric vectors. Combining Eqs. [47], [48] and [49], and substituting the results of Eqs. [26] and [27], we find

$$\frac{d\psi}{dz} = -\frac{2\pi}{P} \frac{\alpha^2}{8(\lambda')^2[1 - (\lambda')^2]} \qquad [50]$$

for the rotation per unit length. Taking $\alpha = 0.1$, $p = 0.5$ μm, $\epsilon_0^{1/2} = 1.65$ and $\lambda = 1$ μm, $d\psi/dz$ is calculated to be approximately 36 revolutions cm^{-1}.

Finally, we plot in Fig. 3 the ellipticities $e_1 = |\mu_1(m_1)/\mu_2(m_1)|$ and $e_2 = -|\mu_2(m_2)/\mu_1(m_2)|$ of the waves corresponding to m_1 and m_2 as a function of reduced wavelength λ'. These have been defined so that linear polarization is associated with a value of zero; one could equally well have chosen to plot the reciprocal relations in which case linear polarization would correspond to infinite ellipticity. With either definition, the ellipticity of circularly polarized waves is unity. It is apparent that the ellipticity of the m_1 wave is anomalous in the region of the reflection band whereas the m_2 wave is "well behaved" for all values of λ'.

Conclusion

The optical properties of cholesteric liquid crystals have been examined in the restricted case of waves propagating parallel to the helix axis. Solutions to the wave equation were obtained in order to determine the polarization states of the normal waves in the medium. It was found that the waves corresponding to two of the four solutions are strongly reflected when the wavelength becomes comparable to the helix pitch. Closer examination revealed that the instantaneous electric field pattern of these waves is superposable on the cholesteric helix. That is, the instantaneous electric field pattern of the reflected waves is a right-handed helix that, when the wavelength is correct, matches exactly the right-handed cholesteric helix. An expression was derived for the optical rotatory power (rotation per unit length), that correctly accounts for the sign and magnitude of observed optical rotations in cholesteric liquid crystals.

Acknowledgement

I am indebted to P. Sheng for helpful discussions.

Appendix 1

In this appendix we transform the wave equation to a coordinate system that rotates with the cholesteric helix described by Eq. [1]. Let q_0 be positive so the resulting helix is right handed. Then, an arbitrary electric field vector **E** with components E_x and E_y in the laboratory frame of reference has components E_x' and E_y'

$$\left.\begin{aligned} E_x' &= E_x\cos\theta + E_y\sin\theta \\ E_y' &= -E_x\sin\theta + E_y\cos\theta \end{aligned}\right\} \qquad [51]$$

in the $z = P\theta/2\pi$ plane of the rotating coordinate system. Thus, by Eq. [9], we can write

$$\left.\begin{aligned} \mathbf{E} &= \exp(-i\sigma_2\theta)\mathbf{E}' \\ \text{and} \quad \mathbf{D} &= \exp(-i\sigma_2\theta)\mathbf{D}' \end{aligned}\right\}. \qquad [52]$$

Also, in the rotating frame of reference, we know that

$$\mathbf{D}' = \overleftrightarrow{\epsilon}_{\text{local}}\mathbf{E}', \qquad [53]$$

where $\overleftrightarrow{\epsilon}_{\text{local}}$ has the simple diagonal form given by Eq. [12]. Hence the wave equation

$$\frac{\partial^2\mathbf{E}}{\partial z^2} + \left(\frac{\omega}{c}\right)^2\mathbf{D} = 0 \qquad [54]$$

in the laboratory frame of reference becomes

$$\frac{\partial^2\mathbf{E}'}{\partial z^2} - \left(\frac{4\pi i\sigma_2}{P}\right)\frac{\partial\mathbf{E}'}{\partial z} + \left[\left(\frac{\omega}{c}\right)^2(\epsilon_0\sigma_0 + \epsilon_1\sigma_3) - \frac{4\pi^2\sigma_0}{P^2}\right]\mathbf{E}' = 0 \qquad [55]$$

in the rotating frame of reference.

Appendix 2

Here we consider the spatial and temporal variation of the electric vector for a circularly polarized light wave. We begin with a plane wave propagating in the positive z direction

$$\mathbf{E}(t, z) = \text{Re}\left\{\hat{\mu}E_0\exp\left[-i\omega\left(t - \frac{z}{c}\right)\right]\right\} \qquad [56]$$

where Re ≡ real part, $\hat{\mu}$ is a vector describing the polarization state of the wave, and E_0 is the amplitude of the wave. Eq. [56] can be rewritten

$$\mathrm{E}(t,z) = \mathrm{Re}\left\{\hat{\mu}E_0\left[\cos\left[\omega\left(t - \frac{z}{c}\right)\right] - i\,\sin\left[\omega\left(t - \frac{z}{c}\right)\right]\right]\right\}. \qquad [57]$$

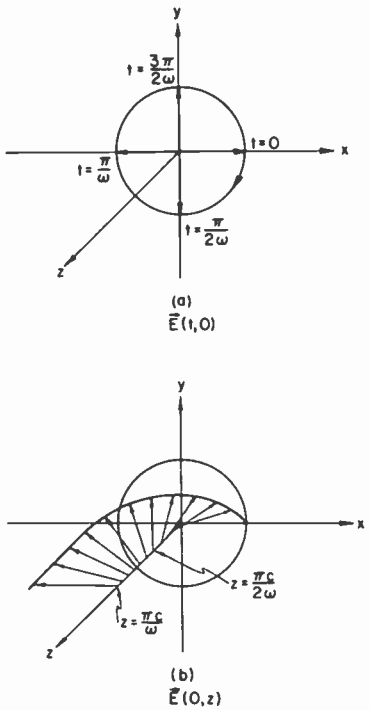


Fig. 4—Behavior of the electric vector for a right circularly polarized light wave: (a) rotation of the electric vector as a function of time *t* at a fixed position on the *z* axis; (b) position of the electric vector as a function of *z* at a fixed instant in time.

Choosing

$$\hat{\mu} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}, \qquad [58]$$

it is apparent that

$$E_x(t,z) = \frac{E_0}{\sqrt{2}}\cos\left[\omega\left(t - \frac{z}{c}\right)\right] \qquad [59]$$

and

$$E_y(t,z) = -\frac{E_0}{\sqrt{2}} \sin\left[\omega\left(t - \frac{z}{c}\right)\right]. \qquad [60]$$

Consider first the temporal development of the wave at some fixed value of z, say $z = 0$ for convenience. Examination of Eqs. [59] and [60] shows that at $t = 0$, the electric vector lies along the x axis while at a later time $t = \pi/2\omega$ it is along the negative y axis and at a still later time $t = \pi/\omega$ it is along the negative x axis, and so on. Thus, as viewed from a point on the positive z axis, looking back toward the origin, the electric vector in the $z = 0$ plane rotates in a clockwise sense from left to right, and by convention this is called a right circularly polarized wave (see Fig. 4a).

Next we consider the spatial distribution of the electric vector along the z axis at one instant of time, e.g., at $t = 0$. We see from Eqs. [59] and [60] that at $z = 0$ the electric vector points along the x axis, as before. As we move along the positive z axis we find the electric vector rotates first to the positive y direction at $z = \pi c/2\omega$ and then to the negative x direction at $z = \pi c/\omega$, etc. Thus the instantaneous electric field pattern traces out a right-handed spiral (see Fig. 4b). All directions are reversed for left circularly polarized light, which is described by the polarization vector

$$\hat{\mu} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}. \qquad [61]$$

These properties of circularly polarized light are important in understanding the optical properties of cholesteric liquid crystals.

References:

[1] E. B. Priestley, "Liquid Crystal Mesophases," *RCA Rev.*, 35, p. 81 (1974).

[2] See, for example, D. W. Berreman and T. J. Scheffer, "Reflection and Transmission by Single-Domain Cholesteric Liquid Crystal Films: Theory and Verification," *Mol. Cryst. and Liquid Cryst.*, 11, p. 395 (1970); R. Dreher and G. Meier, "Optical Properties of Cholesteric Liquid Crystals," *Phys. Rev.*, A8, p. 1616 (1973).

[3] This problem has been studied in detail by H. De Vries, "Rotatory Power and Other Optical Properties of Certain Liquid Crystals," *Acta Cryst.*, 4, p. 219 (1951).

[4] The approach presented in this lecture follows that contained in some unpublished lecture notes of P. S. Pershan on the optical properties of cholesteric liquid crystals.

[5] See, for example, A. Messiah, *Quantum Mechanics*, Vol. II, pp. 544–549, John Wiley & Sons, Inc. (1966).

[6] J. Mathews and R. L. Walker, *Mathematical Methods of Physics*, W. A. Benjamin Inc., N. Y. (1965).

[7] P. G. de Gennes, *The Physics of Liquid Crystals*, p. 228, Oxford University Press, Oxford (1974).

[8] L. A. Goodman, "Electro-optic Effects and Addressing Techniques," *RCA Rev.*, 35, p. 613 (1974).

Electrochemistry in Nematic Liquid-Crystal Solvents

Alan Sussman

RCA Solid State Division, Somerville, N. J. 08876

Abstract—Compounds capable of ionizing in nematic solvents produce electrolytic solutions with electrically normal properties once the solvent anisotropies are accounted for. The conductance in the bulk depends on the viscosities and dielectric constants of the solvent, orientation of the solvent, and on the concentration and molecular dimensions of the solute. Temperature and concentration variation of the conductance depend on ion-pair equilibrium behavior. Calculated double-layer dimensions are finite with respect to cell dimensions and lead to unusual transient effects. Some of the solved and unsolved problems of ion transport in thin layers are discussed, and the role of electrochemical bulk and electrode processes are considered with respect to dynamic-scattering operating life.

Introduction

This paper discusses the electrolytic-solution properties of low-dielectric-constant nematic solvents. Dissolved substances, if electrolytes, can contribute only a fraction of their ions to the conductance because of equilibrium between the free ions and ion pairs. If the solute forms ions through intermediate charge-transfer reactions, additional equilibria must be considered. For nematics, the solvent fluidity is anisotropic, and the conductance depends on the direction of current flow with respect to the orientation of the fluid. The variation of the conductance with temperature is directly related to the variation with temperature of both the ionic equilibrium and the fluidity.

A considerable background of both theoretical and experimental work is available.

The properties of the interface between conductors and ordinary electrolytic solutions are exceedingly complex; for low-dielectric-constant solvents, details of the double layer are lacking, but dimensions may be estimated from simple theory. In cells of the dimensions of the usual liquid-crystal devices, many of the properties of the interface can assume an increased significance as the usual dimensional differences between the bulk and the interface become less distinct.

Many problems of charge transport are incompletely solved, but through the use of carefully purified solvents, specially prepared electrodes, and well-defined experimental conditions, it is possible to separate the contributions of bulk processes, electrode processes, and diffusion. Some kinetic studies of transport phenomena, operating life, and a few electrochemical reactions are discussed.

The relationship between electrolytic and hydrodynamic solution properties is still under intensive study and is not treated in this paper. Many instances of specific electrolyte–low-dielectric solvent interaction need to be investigated fully. The equivalent problems in anisotropic solvents are not completely understood. This review is presented with that thought in mind.

Equilibrium Properties of Bulk Solutions

The equivalent conductance of a solution is a convenient chemical quantity. It is defined as the hypothetical conductance of one chemical equivalent of a dissolved substance; $\Lambda = \alpha N^{\circ} \mu e$, where α is the fraction of the dissolved substance (solute) in the ionic form, and N°, μ, and e are, respectively, Avogadro's number, the ionic mobility, and the elementary charge. The equivalent conductance is related to the conductance $\sigma = \mu e$ by the relation $\Lambda = 1000\ \sigma/c$, where c is the solute concentration in moles per liter. With solvents of dielectric constant greater than 30, solutions of simple electrolytes generally may be expectd to be fully ionized at all concentrations, i.e., $\alpha = 1$. Upon dilution of concentrated solutions in which the mobility of the ions is reduced by interionic forces, the variation of Λ with concentration follows the general limiting law:[1]

$$\Lambda = \alpha(\Lambda_0 - S\sqrt{\alpha c}) \qquad [1]$$

where S is the limiting slope and $\alpha \rightarrow 1$. Λ_0 is obtained by extrapolation of Eq. [1] to "infinite" dilution; Λ_0 is about 50% greater than Λ at 0.01 m/l in a typical simple electrolyte.

In solvents of low dielectric constant, on the other hand, the principal variation of Λ on dilution results from an increase in free ion concentration due to the dissociation of ion pairs:[2]

$$A^+B^- \rightleftharpoons A^+ + B^- \qquad [2]$$

where A^+, B^- and A^+B^- are the positive, negative, and paired ions, respectively. The pairs are the direct result of interionic attraction, since the electrostatic force between ions is shielded less in the low dielectric constant solvent than in the more usual solvents. There is no charge transfer in the formation of ion pairs. The value of α is less than unity, and the variation of Λ is almost entirely controlled by the dependence of the free-ion concentration on the solution concentration.

Writing the equilibrium constant for the reaction of Eq. [2], using brackets to denote concentrations,

$$K_1 = [A^+B^-]/[A^+][B^-] = \frac{c(1-\alpha)}{\alpha^2 c^2} \approx \frac{1}{\alpha^2 c}, \alpha \ll 1 \qquad [3]$$

$$\alpha = \sqrt{\frac{1}{K_1 c}}$$

we see that the fraction in ionic form depends inversely on the square root of the solute concentration. The simplifying assumption of $\alpha \ll 1$ breaks down at concentrations near K^{-1}; this usually occurs at dilutions outside experimental range.

A second set of reactions involving the ion pairs and free ions gives conducting triplets and still further clustering at higher concentrations, resulting in a nonlinear increase in the equivalent conductance; the presence of these reactions results in a minimum in the conductance curve (see Ref. [2], Chap. 8). Taking the formation of ion triplets into account, the equivalent conductance may be written:

$$\Lambda = \Lambda_0\sqrt{1/K_1 c} + \lambda_0\frac{\sqrt{c/K_1}}{K_2}, \qquad [4]$$

where λ_0 is the limiting equivalent conductance of the ion triplets, and K_2 is the equilibrium constant for the reactions $AB+A^+ = A_2B^+$ and $AB+B^- = AB_2^-$, assumed for simplicity to have identical equilibrium properties. At concentrations below the minimum, and for $c \geq 1/K_1$, the functional variation of Λ is

$$\Lambda = \Lambda_0 \sqrt{1/K_1 c}. \qquad [6]$$

Substituting Eq. [6] into the relation between σ and Λ, we find the current; it is proportional to the square root of the solute concentration,

$$J = \sqrt{\frac{c}{K}} \frac{\Lambda_0}{1000} \frac{VA}{d}, \qquad [7]$$

where V is the voltage, A the area, and d the cell thickness. Fig. 1 shows the equivalent conductance as a function of concentration for

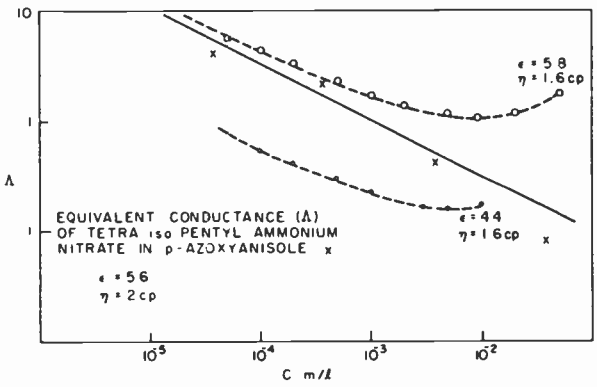


Fig. 1—Equivalent conductance of tetra-*iso*-pentyl ammonium nitrate in isotropic *p*-azoxyanisole at 152°C, (solid line). The equilibrium constant for ion-pair formation is 2×10^{-6} m/l. The data is bracketed between calculated values of Eq. [4] for two values of the dielectric constant. The variation of the equivalent conductance with dielectric constant is found in Eq. [8]. (Ref. [3])

the solute tetra-*iso*- pentyl ammonium nitrate in the nematic *p*-azoxyanisole.[6] The problem of obtaining values of Λ_0 in this case by extrapolation to infinite dilution are experimentally complicated because of the very low concentrations needed to insure that $\alpha = 1$. Use can therefore be made of the semi-empirical Walden's rule:

The product of the limiting equivalent conductance and the viscosity is a constant for each solute, almost independent of temperature.

It is particularly accurate for large ions, but must be corrected slightly for dielectric constant.[4] In an anisotropic solvent, the constant may be calculated by using the appropriate viscosity. Fig. 2 shows[5], for *p*-azoxyanisole, the reciprocal viscosity (fluidity) parallel to the orien-

tation, perpendicular to the orientation, and for a nonoriented sample plotted against the reciprocal temperature.

In the nematic range, the large variations in fluidity on the nematic side of the nematic–isotropic transition may be related to pretransitional phenomena; when the parallel orientation becomes more disordered, the flow becomes more difficult, while for the perpendicular

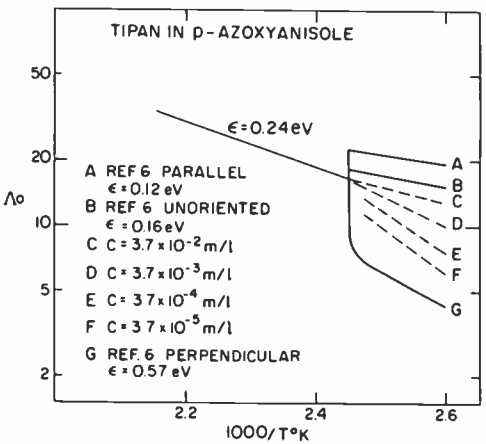


Fig. 2—Fluidity of *p*-azoxyanisole versus the reciprocal temperature for flow parallel to the orientation, perpendicular to the orientation, and for a nonoriented sample. The dashed line shows the current data of Fig. 3 normalized to the isotropic range. It is suggestive to consider that the variation of the activation energy of conduction with concentration depends on the solute influence on the orientation. (Ref. [5])

orientation, disordering makes the flow easier. This change in orientation can be noticed in the current–reciprocal-temperature graph of Fig. 3, for the tetra-*iso*-pentyl ammonium nitrate/*p*-azoxyanisole system. Some effects on the fluidity may be related to the concentration of solute. During the measurements the sample birefringence indicated that the solvent was ordered perpendicular to the electric field, requiring that the appropriate mobility be for flow perpendicular to the orientation. With increasing solute concentration, the fluidity properties begin to resemble these for flow parallel to the orientation. This may be either a direct interaction between the ions and the solvent or may be due to the influence of the ions on the way the surface affects the bulk orientation.

Ionic equilibrium was also observed in 1-cm-thick cells in the system tetra-*iso*-pentyl ammonium tetra-phenyl borate/methoxy-ben-

zylidene *p-n*-butylanaline (MBBA).[6] Walden's rule gave good agreement in the isotropic region, i.e., the product of the conductance and the viscosity was constant. The samples were in the nonaligned state. In the nematic range, the experimental values of the mobility near the transition temperature were five times too low, increasing to ten times too low at the lowest temperature, compared to the calculated values. Although no attempt to order the sample was made, ordering

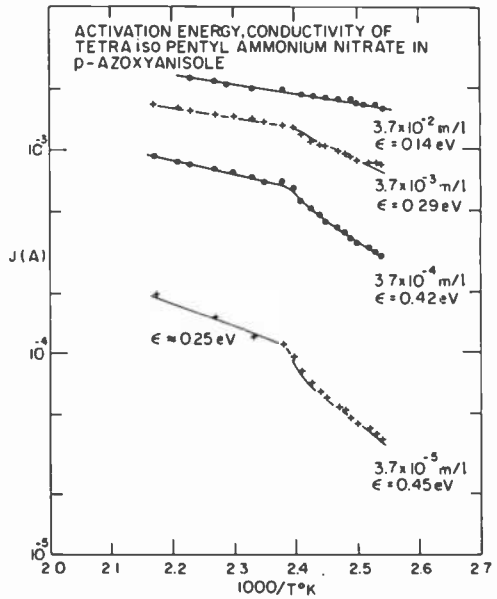


Fig. 3—Current as a function of the reciprocal temperature in the system tetra-*iso*-pentyl ammonium nitrate/*p*-azoxyanisole. Note how the activation energy in the nematic range decreases with increasing solute concentration. Compare this variation with the influence of the orientation (Fig. 2). (Ref. [3])

of such samples has been observed under similar experimental conditions.[7] Nonapplicability of Walden's rule has been noted under conditions of unipolar charge injection, where measured values were up to ten times lower than the calculated ones.[8]

The conductance of a solution and, therefore, the current will depend on the square root of the solute concentration over the range for which ion-pair equilibrium operates. In those devices for which a finite conductance is required, such as dynamic scattering and the storage-effect devices,[10] departures from Ohm's law may be observed because of the reorientation of the fluid by field and hydrodynamic

effects. If the voltage is slowly raised in a dynamic scattering cell that was originally in the perpendicular homeotropic orientation, the current will be controlled by the fluidity parallel to the field, $\mu_{\parallel}$, until the threshold voltage for the reorientation to a birefringent condition is exceeded. The current than becomes dependent on the value of the fluidity perpendicular to the field, $\mu_{\perp}$. Between the threshold for reorientation (a dielectric effect) and the onset of domain formation (a hydrodynamic effect) at another threshold voltage, the application of a voltage pulse will result in a current transient, as shown by the arrows in Fig. 4. When the voltage is raised still further, the domains

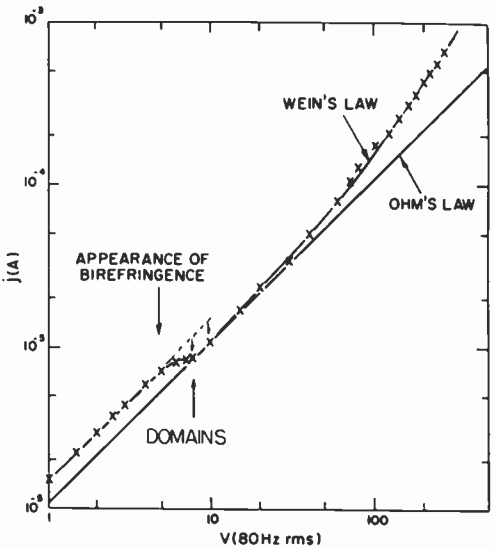


Fig. 4—Current–voltage characteristic for a dynamic scattering device, original orientation perpendicular homeotropic. The Ohm's law line is drawn for mobility based on fluidity perpendicular to the orientation, $\mu_{\perp}$.

become unstable and the turbulent dynamic scattering regime is entered. An average mobility,

$$\mu_{ave} = \frac{1}{3}(2\mu_{\perp} + \mu_{\parallel})$$

now controls the current. At sufficiently high fields, the ion-pair equilibrium is disturbed in favor of dissociation. This, the second Wein effect, is most easily observed in solvents of low dielectric constant (see Ref. [1], Chap. 4, Sec. 7).

Measurements of conductance are usually made using alternating voltage of a frequency that falls between the dielectric relaxation frequency (see Ref. 1, Chap. 4, Sec. 1), above which ions cannot contribute to the current, and the inverse transit time,[11] below which, as will be seen later, there are complications due to polarization of the electrodes.

Variation of the conductance with temperature, both in the tetra-*iso*-pentyl ammonium nitrate/*p*-azoxyanisole and tetra-*iso*-pentyl ammonium phenyl borate/MBBA systems, depends almost entirely on the fluidity,[3] since ionic equilibrium contributes only a small factor over the ranges of the studies. Because this may not always be the case, the variation of the equilibrium constant with temperature will be considered.

For the association–dissociation reaction Eq. [1], the equilibrium constant is given approximately as[4]

$$K_1^{-1} \approx \text{const.} \exp\left(\frac{e^2}{a\epsilon kT}\right) \qquad [8]$$

The exponential term is the ratio of the electrostatic to thermal energy, with a the distance at which the ions can be considered paired, k is Boltzmann's constant, and ϵ is the average dielectric constant of the solvent ϵ_{ave}. The ion size parameter a is a constant with temperature showing little variation with solvent. The variation of K_1^{-1} does depend on the variation of the dielectric constant(s) with temperature. Once the values for a, $\epsilon(T)$, $\mu(T)$, and μ (orientation) are known for a given solute/solvent system, the current as a function of temperature and concentration may be considered determined. Experimental and calculated results for the system tetra-*iso*-pentyl ammonium bromide in a mixed solvent containing compounds of alkoxy benzylidine-*p*-amine phenyl esters (APAPA family)[12] are shown in Fig. 5.

There is another equilibrium system that can lead to ionic conduction. When an electron-accepting compound is introduced into a solvent that is an electron donor, the reaction to form a donor–acceptor pair ensues: D + A $\rightleftharpoons$ (DA). This donor–acceptor pair may subsequently dissociate to give a pair of ions: (DA) $\rightleftharpoons$ $D^+ + A^-$. This reaction is distinct from ion-pair reactions, since charge transfer does occur formally, but the role of the low-dielectric solvent is similar in affecting the equilibrium. A typical electron acceptor, chloranil (tetrachloro 1,4 benzoquinone), in reacting with MBBA in the dual role of solvent and electron donor, was able to alter the conductance.[13] In another case, equal parts of the donor hydroquinone and

the acceptor *p*-benzoquinone were dissolved in MBBA;[14] this resulted in the formation of ionizable charge transfer complex: HQ + *p*-BQ $\rightleftharpoons$ HQ · *p*-BQ $\rightleftharpoons$ $HQ^+ + p\text{-}BQ^-$. The conductance of the solution became sufficient to produce dynamic scattering. In a 1:1 mixture of MBBA and *p*-methoxy *p*-*n*-hexyl aniline, carefully purified, the current remained constant at 0.15 $\mu A/cm^2$ for 4000 hours in hermetically sealed cells of thickness 25 μm, probably because the conducting species can undergo oxidation and reduction reactions by electron transfer alone at the appropriate electrodes, thereby revers-

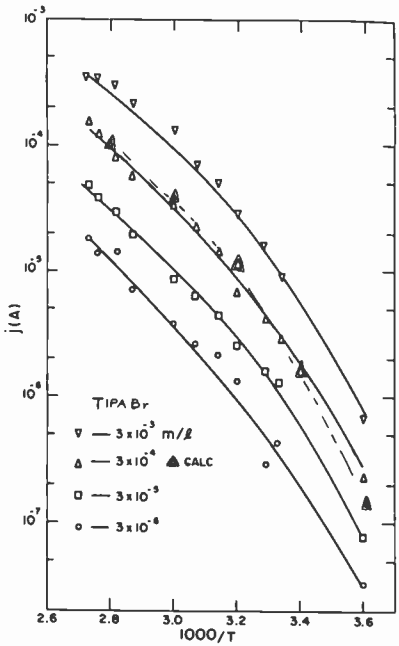


Fig. 5—Current as a function of the reciprocal temperature in the system tetra-*iso*-pentyl ammonium bromide/APAPA RT mixture, with concentration as a parameter. The calculated current shows good agreement with the experimental data.

ing the formation reactions. The accumulation of neutral products at the respective electrodes results in a concentration gradient, and eventually a steady state is set up, allowing reformation of the charge transfer complex in the bulk. No irreversible reactions can be expected except as side reactions. With unpurified MBBA, increased currents were obtained with continued operation.

Electrochemical Reactions

Kinetic studies of some electrochemical reactions in nematic solvents have been made, as well as studies of those solvents dissolved in more polar solvents. It was noticed early that under dc excitation of dynamic-scattering cells,[15] the currents remained relatively constant and the cells continued to operate for orders of magnitude longer than would be expected for a simple faradic process, i.e., where an electrochemical reaction was controlled by the net charge passed through the cell.

In *p*-methoxy benzylidene-*p*-amino phenyl acetate, the dynamic scattering life was limited by a faradic reaction of low efficiency which produced polymeric anode films, probably by a free radical mechanism.[16] The probable source of the ionic conduction was ionization of intrinsic impurities. Electron injection by a Schottky mechanism was proposed,[9] but could not be correlated with the activation energy of the conductance. A possible paradox exists, since in order to have a high enough field at the electrodes to cause charge injection, a concentrated electrolytic solution is required, and any injection would be masked by ionic currents. A dilute solution, with no overshadowing ionic currents, would naturally have a lower field across the double layer, so in order to achieve high enough fields for injection, large voltages must be applied to the entire cell.

The thickness[17] of the double layer, $K^{-1} = [4\pi\alpha c/(\epsilon kT)]^{-1/2}$, for a solvent of dielectric constant 5.5 appears in Fig. 6 as a function of concentration, assuming an equilibrium constant for ion-pair dissociation of 10^{-7}, which is appropriate for MBBA[6] or *p*-azoxyanisole.[3] Estimates of the double-layer field can be made assuming that the double-layer potential saturates at a few volts. Note that the double-layer thickness, which depends on $(\alpha c)^{-1/2} = C^{-1/4}$ is almost constant and that its thickness is not insignificant when compared to typical liquid-crystal-device cell thicknesses (10^{-3} cm).

An alternative theory to Schottky emission is the creation of negative and positive ions from the solvent by oxidation and reduction reactions at the appropriate electrodes, followed by recombination after diffusion into the bulk. This mechanism was suggested to explain the long life of dynamic scattering in compounds of the APAPA series.[18] The role of added ionic compounds was unstated.

In undoped purified samples of MBBA, low-field conductance was attributed to thermal dissociation of trace impurities,[19] but at fields greater than 1500 V/cm, electrode processes begin to interfere. Through the use of ion exchange membranes as an electrode coating, injection effects were supressed. Then one observes at low fields an

ohmic current due to the natural impurities, if the dissociation rate of the impurities is fast enough to overcome the rate at which the ions are deposited on the electrodes. Eventually, the current saturates as the deposition rate (related to the inverse transit time) exceeds the generation rate, and at the high fields, the current again begins to increase because of protons injected from the membrane, giving an imine which is believed to be indentical to that occuring in the first step of MBBA hydrolysis. That species may be one of the conducting impurities in undried MBBA

The effects of direct current on MBBA containing 300 parts per million of water and, therefore, traces of the hydrolysis products *p-n*- butyl analine and *p*- anisaldehyde were that, first, the *p-n* butyl aniline disappeared by anodic oxidation, producing blackening and

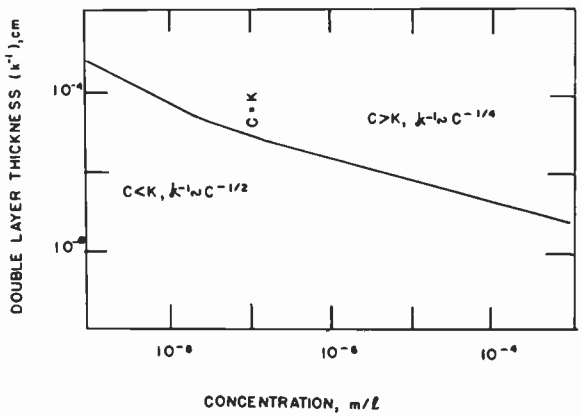


Fig. 6—Double layer thickness K^{-1} as a function of concentration for dielectric constant 5.5 at room temperature. The effect of ion pairing has been included. The thickness of the double layer over the range of interest varies only as the fourth root of the solute concentration.

then a rapid rise in current.[20] The reaction rate depended on the current density. Oxidation–reduction reactions of MBBA itself and of the hydrolysis products were studied in the solvent acetonitrile. It was found that the oxidation of MBBA and *p-n*- butyl aniline is irreversible, while the anisaldehyde is not oxidized; the reduction of MBBA and the aldehyde produce anion radicals by acceptance of single electrons in a reversible reaction, while the *p*- butyl aniline is not reduced. In the solvent dimethyl formamide,[24] reduction of the MBBA resulted in a radical ion with a measured half-life of 4 seconds.

Observations of phenomena related to space-charge accumulation at the electrodes during dc operation of dynamic scattering was reported and attributed to nonspecific electrode processes.[22] The electrode charging and discharging transients, i.e., the ionic charge accumulated in the double layers, was found to be approximately 10^{-5} coulomb per cm^2. The kinetics of the discharge were shown to depend on diffusion from the double layers into the bulk; the charge did not leave the cell via the electrodes. Fig. 7 shows the charging transients

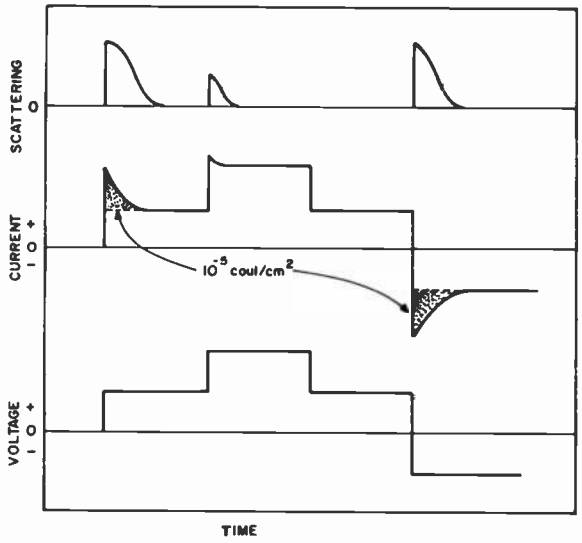


Fig. 7—Current and dynamic scattering transients upon application of voltage steps and voltage reversal (typical results for a low conductivity sample). The excess current above steady state corresponds to electrode charge of about 10^{-5} coulomb/cm^2.

most easily observed when the conductance of the fluid is less than 10^{-10} $(ohm\text{-}cm)^{-1}$. Then the dynamic scattering disappears, also as the charge accumulates in the double layers. If the voltage is increased or if the polarity is reversed, another transient is produced, the latter causing a reversal of the potential of the double layer and releasing the charge, which is transported to the other double layer. At high enough voltage, dynamic scattering resumes, either because of injection from the electrodes or field-assisted dissociation in the bulk.

References:

[1] H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions*, 3rd edition, Reinhold Publishing Co., New York (1958). (A general reference for bulk properties of electrolytic solutions.)

[2] R. M. Fuoss and F. Assascina, *Electrolytic Conductance*, Chap. XVI And XVII, Interscience Pub., Inc., New York (1959).

[3] A. Sussman, "Ionic Equilibrium and Ionic Conduction in the System Tetra-*iso*-pentyl Ammonium Nitrate *p*-Azoxyanisole," *Mol. Cryst. and Liq. Cryst.*, **14**, p. 182 (1971).

[4] J. Barthel, "Conductance of Electrolyte Solutions," *Agew. Chem. Internat. Edit.*, **7**, p. 260 (1968) (special reference to nonaqueous solvents).

[5] R. S. Porter and J. F. Johnson, "Orientation of Nematic Mesophases," *J. Phys. Chem.*, **66**, p. 1826 (1962).

[6] A. Denat, B. Gosse and J. P. Gosse, "Étude du Cristal Liquide *p*-Méthoxybenzilidène *p*-Butylaniline," *J. Chim. Phys;* **2**, p. 319 (1973).

[7] F. Gaspard, R. Herino and F. Mondon, "Electrohydrodynamic Instabilities in DC fields of a Nematic Liquid Crystal with Negative Dielectric Anisotropy," *Chem. Phys. Lett.*, **25**, p. 449 (1974).

[8] J. C. Lacroix and R. Tebazeon, "Sur la Mesure de Mobilitiés Ioniques dans un Cristal Liquide Nematique," *Comptes Rendus*, **278**, p. 623 (1974).

[9] G. H. Heilmeier, L. A. Zanoni, and L. A. Barton, "Dynamic Scattering—A New Electro-optic Effect in Certain Classes of Nematic Liquid Crystals," *Proc. IEEE*, **56**, p. 1162 (1968).

[10] G. H. Heilmeier and J. Goldmacher, "A New Electric Field Controlled Reflective Optical Storage Effect in Mixed Liquid Crystal Systems," *Proc. IEEE*, **57**, p. 34 (1969).

[11] A. Sussman, "The Electro-optic Transfer Function in Nematic Liquids," *RCA Review*, **35**, p. 176 (1974).

[12] A. Sussman, unpublished results.

[13] A. I. Baise, I. Teucher, and M. M. Labes, "Effect of Charge-Transfer Acceptors on Dynamic Scattering in a Nematic Liquid Crystal," *Appl. Phys. Lett.*, **21**, p. 142 (1972). See however: F. Gaspard and R. Herino, "Comments on 'Effect of Charge-transfer Acceptors on Dynamic Scattering in a Nematic Liquid Crystal'," *Appl. Phys. Lett.*, **24**, p. 252 (1974).

[14] Y. Ohnishi and M. Ozutsumi, "Properties of Nematic Liquid Crystals Doped with Hydroquinine and *p*-Benzoquinone: Long-term Dynamic Scattering Under DC Excitation," *Appl. Phys. Lett.*, **24**, p. 213 (1974).

[15] G. H. Heilmeier, L. A. Zanoni, and L. A. Barton, "Further Studies of the Dynamic Scattering Mode in Liquid Crystals and Related Topics," *IEEE Trans. Elec. Dev.* **ED 17**, p. 22 (1970).

[16] A. Sussman, "Dynamic Scattering Life in the Nematic Compound *p*-Methoxy-benzylidene-*p*-Amino Phenyl Acetate as Influenced by the Current Density," *Appl. Phys. Lett.*, **21**, p. 126 (1972).

[17] P. Delahay, *Double Layer and Electrode Kinetics, Chap. 3*, Interscience–John Wiley & Sons, Inc. New York (1965).

[18] M. Voinov and J. S. Dunnett, "Electrochemistry of Nematic Liquid Crystals," *J. Electrochem. Soc.*, **120**, p. 922 (1973).

[19] G. Brière, R. Herino, and F. Mondon, "Correlation Between Chemical and Electrochemical Reactivity in the Isotropic Phase of a Liquid Crystalline *p*-Methoxybenzilidene *p-n*-Butylaniline," *Mol. Cryst.*, **19**, p. 157 (1972).

[20] A. Denat, B. Gosse, and J. Gossé, "Chemical and Electrochemical Stability of *p*-Methoxybenzilidene-*p-n*-Butylaniline," *Chem. Phys. Lett.*, **18**, p. 235 (1973).

[21] A. Lomax, R. Hirasawa, and A. J. Bard, "The Electrochemistry of the Liquid Crystal N-(*p*-Methoxybenzilidene)-*p-n*-Butylanaline (MBBA)," *J. Electrochem. Soc.*, **119**, p. 1679 (1972).

[22] A. Derzhanski and A. G. Petrov, "Inverse Currents and Contact Behavior of Some Nematic Liquid Crystals," *Phys. Lett.*, **36A**, p. 307 (1971).

Liquid-Crystal Displays—Electro-Optic Effects and Addressing Techniques

L. A. Goodman

RCA Laboratories, Princeton, N. J. 08540

Abstract—The various electro-optic effects are described with particular emphasis on the twisted nematic field effect and dynamic scattering. Various approaches for matrix addressing are presented and their utility is discussed. Some methods for electron-beam and light-beam addressing of liquid-crystal displays are summarized.

1. Introduction

Many of the physical properties of mesomorphic materials, such as birefringence, optical activity, viscosity, and thermal conductivity are sensitive to relatively weak external stimuli. Electric fields, magnetic fields, heat energy, and acoustical energy can all be used to induce optical effects. At the present time, most of the display-related research is centered on the application of electro-optic effects because of the relative ease and efficiency of excitation with an applied voltage as compared with other means of stimulation. Liquid-crystal electro-optic effects are important because they do not require the emission of light; instead they modify the passage of light through the liquid crystal either by light scattering, modulation of optical density, or color changes. The salient properties are low-voltage operation, very low power dissipation, size and format flexibility, and washout immunity in high-brightness ambients.

This paper is divided into three major sections. The first describes

the various liquid-crystal electro-optic phenomena; the second discusses important display-related parameters; and the third describes the operation of liquid-crystal devices in matrix-addressed and beam-scanned modes of operation.

2. Electro-optic Phenomena

Liquid-crystal electro-optic phenomena can be divided into two categories—those caused only by dielectric forces and those induced by the combination of dielectric and conduction forces. The two conduction-induced phenomena discussed later are dynamic scattering and the storage effect. Four of the dielectric phenomena, or field effects as they are sometimes known, are discussed first: (1) induced birefringence, (2) twisted nematic effect, (3) guest–host interaction, and (4) cholesteric–nematic transition.

In all of the present theories about the excitation of nematic or cholesteric liquids by an electric field, the mesomorphic material is treated as a continuous elastic anisotropic medium. The Oseen[1]–Frank[2] elastic theory is used to describe the interaction between the applied field and the fluid. The application of an electric field causes the liquid crystal to deform. For a material with a positive dielectric anisotropy, $\Delta\epsilon = \epsilon_{\parallel} - \epsilon_{\perp} > 0$, the director aligns in the direction of the field; if the dielectric anisotropy is negative, the director tends to align perpendicular to the applied field. The elastic forces attempt to restore the field-driven fluid to the initial orientation, which is determined by the surface alignment. The interplay between dielectric and elastic torques leads to the occurrence of the threshold voltage or field. In another paper in this series,[3] the interaction between the liquid crystal and an applied field is discussed in detail. The interested reader can refer to that paper and to papers published elsewhere[4–6] for the calculations of the field–liquid-crystal interaction.

2.1 Field-Induced Birefringence

The first electro-optic effect is field-induced birefringence or deformation of aligned phases.[7–10] Schematic representations of the fluid with zero applied volts and for a voltage exceeding the threshold voltage are presented in Fig. 1. With no applied voltage, the nematic liquid is in the perpendicular state. In the discussion of induced birefringence and the other effects considered, the assumption is made that the surface orientation of the molecules remains constant even when the field is applied, while the voltage-induced deformation increases toward the center of the cell. When the applied voltage ex-

ceeds the threshold voltage, the liquid crystal distorts if it has negative dielectric anisotropy. A maximum director rotation of 90° is possible. The threshold voltage is given in m.k.s. units by

$$V_{TH} = \pi\sqrt{\frac{K_{33}}{\epsilon_0 \Delta\epsilon}}, \qquad [1]$$

where K_{33} is the bend elastic constant.[11,12]

The perpendicular texture is optically isotropic to light propagating perpendicular to the cell walls. Consequently, with crossed polarizer and analyzer, no light is transmitted through the analyzer. Dur-

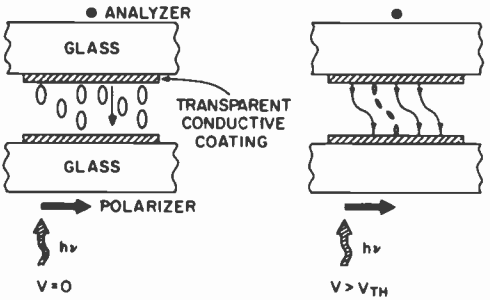


Fig. 1—Schematic illustration of the induced birefringence effect with and without an applied voltage. Thin arrows represent director orientation.

ing fluid deformation, the liquid crystal becomes birefringent to the transmitted light, and part of the light passes through the analyzer. The intensity of the emerging light is expressed by[10]

$$I = I_p \sin^2 2\phi \sin^2 \frac{\delta}{2} \qquad [2]$$

where $\delta = [2\pi d\,\Delta n(V)/\lambda]$, I_p is the light transmitted through two parallel polarizers, ϕ is the angle between the input-light optical vector and the projection of the director on the plane parallel to the cell walls, d is the cell thickness, $\Delta n(V)$ is the voltage-induced change in birefringence, and λ is the wavelength of the light.

The transmitted light intensity is maximum when $\phi = 45°$. Normally, the angle ϕ is not well-defined because of the cylindrical symmetry that results when a perpendicularly aligned fluid is deformed by the electric field (see Fig. 2b). However, as described in the surface

investigations of Creagh and Kmetz,[13] a preferential direction can be established in a plane parallel to the cell walls by grooving or rubbing the substrates. Samples prepared in this manner deform with a well-defined direction for the fluid director.[14] This preferred direction can be set at a 45° angle to the crossed polarizer and analyzer.

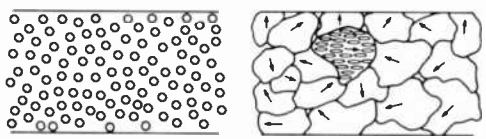


Fig. 2—Top views of perpendicularly oriented cells. (A) $V = 0$ and the long axes of the molecules are perpendicular to the electrodes. (B) The fluid is partially deformed with no prescribed direction.

The maximum value of $\Delta n(V)$ is the index of refraction anisotropy, $\Delta n = n_{\parallel} - n_{\perp}$. Typically, Δn is about 0.2 to 0.3 This anisotropy is so large that, for monochromatic radiation, the transmitted light intensity undergoes many maxima and minima as the voltage increases above threshold. With white light, variable colors can be observed as a function of voltage.

To first order, the frequency response of induced birefringence is constant in amplitude from low frequency to the molecular dispersion frequency in the dielectric constant where the dielectric anisotropy changes.[9] This property is typical of all the field effects.

For a nematic material with positive dielectric anisotropy, induced birefringence can also be observed. However, the liquid crystal must be in the uniform parallel orientation at zero volts.[15] Above the threshold voltage, the detector aligns itself parallel to the applied field. With crossed polarizer and analyzer, the voltage dependence of the light intensity is reversed from that described previously for a fluid of negative dielectric anisotropy.[5,11]

For materials with positive dielectric anisotropy, the threshold voltage can be as low as 1.0 volt, whereas devices using negative dielectric anisotropy fluids typically possess threshold voltages in the 4–6 volt range. Since the elastic constants are relatively independent of material, the difference in threshold voltages is ascribed to the much larger magnitude of anisotropy normally found in fluids with positive anisotropy as compared to that occurring in materials with negative anisotropy.

2.2 Twisted Nematic Effect

The twisted nematic field effect is probably the most important of the field effects because of its combined properties of very low voltage threshold, low resistive power dissipation, and relatively wide viewing angle in the reflective mode.

The typical cell structure used in the twisted nematic device is shown in Fig. 3.[16] The molecules in each surface layer of the liquid crystal are uniformly aligned in one direction, but with a twist angle of 90° between the preferred direction for the two surfaces. With no

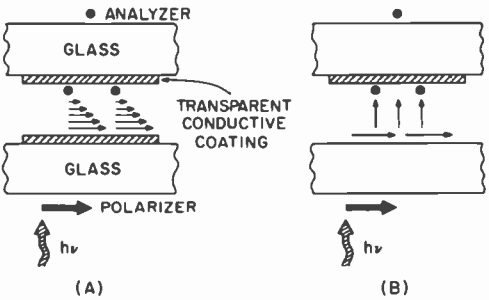


Fig. 3—Side view of twisted nematic effect for (A) $V = 0$ and (B) $V > V_{TH}$. The thin arrows represent the orientation of the nematic molecules.

applied voltage, the bulk fluid distorts so as to provide a gradual rotation of the molecular alignment from one cell wall to the other.[3] With a nematic fluid of positive dielectric anisotropy, voltages exceeding the threshold voltage cause the nematic director to become untwisted and to tend to align parallel to the applied field. The threshold voltage is

$$V_{TH}^2 = \frac{1}{\Delta\epsilon\epsilon_0}\left[K_{11}\pi^2 + (K_{33} - 2K_{22})\varphi_0^2\right], \qquad [3]$$

where K_{11} is the splay elastic constant, K_{22} is the twist elastic constant, and the twist angle φ_0 is equal to $\pi/2$. Eq. [3] is the corrected version of the expression derived by Schadt and Helfrich for the dielectric analog to the magnetic case originally solved by Leslie.[17] The theoretical dependence of the threshold voltage on the anisotropy has been verified.[18] Also, several mixtures with large dielectric anisotropy have been reported with threshold voltages less than 1.0 volt.[19–21]

The optical properties of the twisted nematic field effect are par-

ticularly interesting. Linearly polarized light propagating perpendicular to the cell is rotated by approximately 90° as it passes through the fluid when there is no applied voltage.[22,23] Maximum light transmission is obtained for the zero-field case by orienting the crossed polarizer and analyzer with the polarizer optic axis parallel to one of the preferred surface alignment directions in the cell. The transmitted light decreases when the applied voltage exceeds the threshold voltage, and the fluid starts to align in the perpendicular state. Fig. 4 presents the data obtained by Schadt and Helfrich[16] with parallel polarizer and analyzer, which can be used instead of crossed polarizer and analyzer. Extinction is obtained with no applied voltage, while light transmission occurs for voltages exceeding the threshold voltage. Hence, depending on the orientation of the polarizer and analyzer, either a black-on-white or white-on-black display can be obtained.

The formula presented in Eq. 3 indicates the threshold voltage at which the director starts to reorient. Gerritsma, DeJeu, and Van Zanten[24] have measured the magnetic threshold by both capacitive and optical techniques and found that the capacitive threshold is lower than the optical one. Van Doorn[25] has shown that this difference is to be expected, since the fluid starts to reorient by the tilting of the director toward the applied magnetic field before the twist has appreciably changed. Consequently the capacitive threshold, which occurs when the director starts to tilt toward the applied field, is lower than the optical threshold, which occurs when the twist becomes sufficiently nonuniform that the optical vector of the light does not "follow" the twist. A similar difference has been observed in twisted nematic devices excited with electric fields.[26,27] Berreman's[28] explanation of the static characteristics of electric-field-excited devices is similar to that of Van Doorn.[25]

The data presented in Fig. 4 and the optical results given by Gerritsma et al[24] were obtained with light normally incident upon the cell and coaxial with the detector. Recently, the angle and voltage dependence of the light transmission characteristics of twisted nematic devices prepared by the rubbing technique has been measured by several investigators. Kobayashi and Takeuchi[29] have obtained the data presented in Fig. 5. The transmitted light as a function of voltage is not symmetric with the viewing angle. The curves also demonstrate that the apparent optical threshold is a function of viewing angle. Both of these phenomena have been explained in terms of the relative ability of the incident light to "follow" the twist for different angles of incidence.[30] As a result, the true threshold is the capacitive threshold, and it is always lower than the angle-dependent optical threshold. Nonuniformities in cell appearance caused by reverse

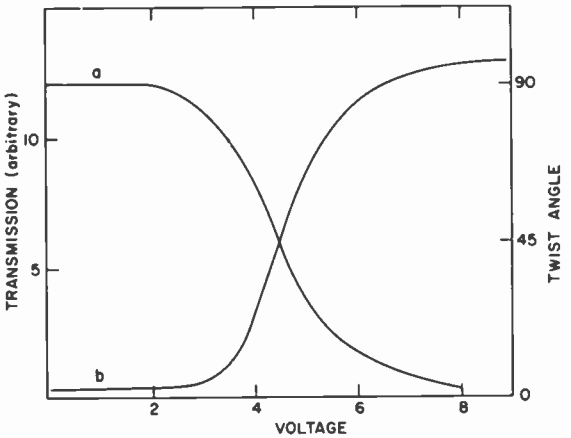


Fig. 4—Curve (a) shows rotation angle of linearly polarized light versus voltage for a nematic liquid crystal at room temperature and 1 kHz; curve (b) is transmission versus voltage with parallel polarizers (Ref. [16]).

tilt[3,31] and reverse twist[32,33] have been studied and can be eliminated by the addition of a cholesteric to the nematic and by proper surface preparation.[33,34]

With field-induced birefringence, variable colors are transmitted when the voltage exceeds the critical value. These angle-dependent color effects are unavoidable and constitute one of the main disadvantages of the induced birefringence effect. In the twisted nematic

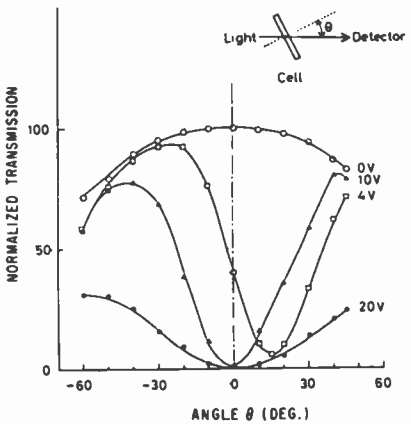


Fig. 5—Light transmission of a twisted nematic cell with crossed polarizers as a function of turning angle for various values of applied field at 5 kHz. Cell thickness is 30 μm (Ref. [29]).

phenomenon, relatively little birefringence occurs except at applied voltages just above the threshold voltage. For a twisted nematic device operating either with zero volts or with a voltage more than two or three times the threshold value, where most of the molecules are aligned perpendicular to the walls, the principal color phenomena are minor effects in the fieldless state associated with the inhomogenous thickness of the fluid.[35]

In reflective applications, the small angle visibility of twisted nematic devices about the normal to the cell can be much better than with scattering displays. For both scattering and nonscattering devices, a specularly reflecting mirror would normally be used behind the cell to obtain a bright, legible display. However, not only is the desired display information seen by the observer, but quite often unwanted specular glare can also be detected.

With nonscattering displays, such as the twisted nematic device, this glare can be prevented by the use of a diffuse reflector in place of the mirror reflector. However, there is a loss of brightness due to the depolarization of radiation reflected from the diffuse surface. The compromise is acceptable because the glare-free viewability is usually considered more important than high brightness.

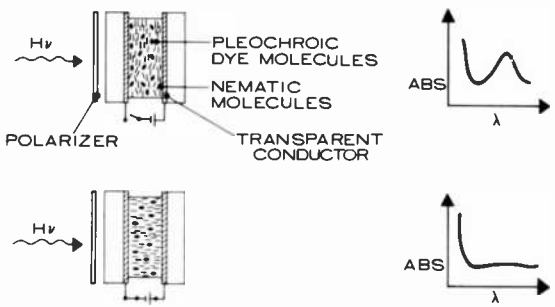


Fig. 6—Schematic diagrams of electronic color switching phenomenon.

2.3 Guest–Host Effect

A third phenomenon depending solely upon dielectric forces is the guest–host or electronic color-switching interaction[36,37] in which "guest" pleochroic dyes are incorporated within nematic "host" materials. The dyes have different absorption coefficients parallel and perpendicular to their optical axes. As illustrated in Fig. 6, the dye molecules can be oriented by the liquid crystal. With zero field, the liquid crystal is in the uniform parallel orientation and the dye mole-

cules are aligned with long axes parallel to the optical vector of the linearly polarized light. In this configuration, the dye molecules have absorption bands in the visible. Above the threshold voltage, the nematic fluid of positive dielectric anisotropy tends to align parallel to the field. This is the condition for low dye absorption. Consequently, a color variation can be observed between the two states. Only one polarizer need be used with this effect. With optimum dye concentrations, optical density changes as large as 1.5 have been measured and a threshold voltage of approximately 2.0 volts has been observed.[37]

2.4 Cholesteric-to-Nematic Transition

The electric-field-induced cholesteric-to-nematic phase transition was observed by Wysocki et al.[38] The magnetic analog had been previously measured by Sackmann et al,[39] and the theoretical magnetic and electric field dependents have been calculated by DeGennes[40] and Meyer.[41]

The phase transition, which is illustrated in Fig. 7, only occurs in an electric field with a cholesteric fluid of positive dielectric anisotro-

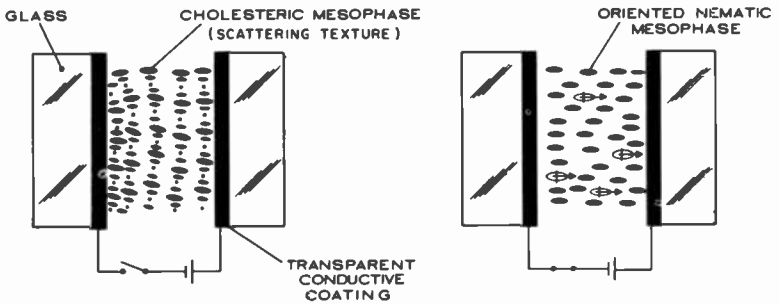


Fig. 7—Side view representation of the electric-field-induced cholesteric-to-nematic phase change.

py. The cholesteric planes are approximately perpendicular to the cell walls with zero applied field. The helical axes have random orientation and this state is strongly scattering in appearance.[42,43] As the electric field approaches the critical value, the helices begin to unwind and dilate. Above the threshold field, all the molecules, except for surface layers, are aligned parallel to the electric field. This latter condition is the perpendicular or homeotropic texture.[38] When the field is lowered below threshold, the scattering texture returns. The theoretical calculations performed by DeGennes[3,40] result in the fol-

lowing expression for the threshold field.

$$E_{TH} = \frac{\pi^2}{P_0}\left(\frac{K_{22}}{\Delta\epsilon\epsilon_0}\right)^{1/2}. \qquad [4]$$

where P_0 is the undeformed helix pitch.

Meyer[44] performed optical measurements of the magnetically induced pitch dilation. By mixing nematic and cholesteric materials, Durand et al[45] were able to vary the pitch and to verify that the threshold magnetic field was inversely proportional to the undeformed helix pitch.

In the original experiment of Wysocki et al, the threshold field was about 10^5 V/cm.[38] Heilmeier and Goldmacher[46] reduced the threshold field to 2×10^4 V/cm by using a mixture of cholesteric and high positive dielectric anisotropy nematic materials. More recently, the threshold field has been reduced to 5×10^3 V/cm by use of the positive biphenyl compounds.[20,47]

The model presented so far is accurate, but incomplete. In actuality, the scattering texture with the helical axes parallel to the cell walls is not a stable state without an applied field. Rather, it is a metastable state that has a lifetime of from minutes to months, depending upon the surface alignment, fluid thickness, and pitch.[48] The stable state is the planar or Grandjean texture. Kahn[43] made the initial measurements of the various steps in the texture changes. Additional experiments with both electric and magnetic fields have added more detail to the model.

With no applied electric field, the liquid crystal exists in the planar texture, in which it remains until the voltage exceeds the critical field, E_H, which is proportional to $1/\sqrt{P_0 L}$.[49–51] At E_H, a square grid pattern of periodic distortions of the helical axes occurs.[42,52–54] In these perturbations, the direction of the helix axis periodically varies between being perpendicular and somewhat nonperpendicular to the cell walls. As the field increases further, the distortions grow until the helix planes become perpendicular to the walls.[42,43] With further increase in voltage, the fluid becomes perpendicularly aligned as described previously. When the voltage is shut off, the fluid returns to the metastable scattering state. The transient decay from the perpendicular texture to the scattering state is retarded by the presence of a bias voltage that is below the threshold voltage for the field-induced transition.[55,56] For no bias level, the decay to the scattering texture proceeds rapidly with the natural relaxation time. As the bias level approaches the threshold voltage, the relaxation slows considerably.

Greubel[57] has recently analyzed the cholesteric-to-nematic transition for an applied voltage. With perpendicular alignment of the liquid crystal molecules at the cell surfaces, he found that the field transition for the cholesteric-to-nematic transition was higher than the opposite nematic-to-cholesteric transition. With one mixture, the ratio of field thresholds was approximately 2.5. This work points out the importance of the proper surface orientation and cell cleanliness in achieving the best bistability.

White and Taylor[58] have described a new device that combines both the guest–host effect and the cholesteric-to-nematic transition. The pleochroic dye is added to a cholesteric material and the cell transmission changes when the cholesteric undergoes the cholesteric-to-nematic transition. Because of the rotational symmetry of the long axes of the dye molecules about the cholesteric helical axes, contrast ratios of greater than 4 to 1 were obtained without the use of a polarizer.

2.5 Dynamic Scattering

In nematic materials with negative dielectric anisotropy and electrical resistivity less than $1–2 \times 10^{10}$ ohm-cm, conduction-induced fluid flow occurs during the application of an applied voltage. The wide-angle forward-scattering phenomenon known as dynamic scattering[59,60] is the most important manifestation of the turbulence that

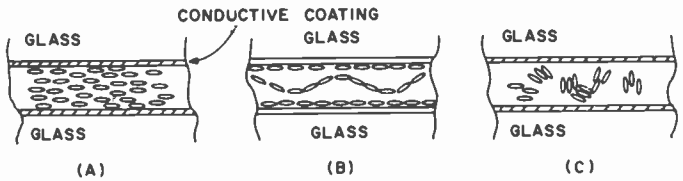


Fig. 8—Side view of the various steps in the formation of dynamic scattering for uniform parallel orientation. (A) $V = 0$, (B) $V = V_W$, and (C) $V > V_W$.

accompanies the electrohydrodynamic flow. The light scattering arises from micron-sized birefringent regions in the turbulent fluid.[61]

The various steps in the production of dynamic scattering are portrayed schematically in Fig. 8 for initial uniform parallel alignment. Below voltage V_W, no change in orientation occurs. At V_W, the fluid becomes unstable and it deforms into the periodic structure shown in Fig. 8.[62] Two similar mechanisms are responsible for the creation of the instability, one for ac voltages and a second for dc voltages. Both

interactions require the presence of space charge in the fluid, but differ in the means by which the space charge is generated.

In the ac case, space-charge separation perpendicular to the applied field is caused by the anisotropy in conductivity.[62,63] The applied field produces a force upon the liquid crystal because of the space charge. This stimulus drags the fluid toward the walls. The cells walls impose boundary conditions that necessitate vortical flow of the fluid. The fluid shear torque aligns the director in the direction of the fluid flow, while the dielectric and elastic forces oppose the fluid deformation. At the threshold voltage, the fluid becomes unstable and the periodic distortion takes place.

When the cell is observed through a microscope, the periodic deformation appears as a series of alternating dark and light domains, known as Williams' domains that run perpendicular to the original homogeneous alignment. The wavelength of the periodic deformation is determined by the thickness of the cell.[64] The birefringence of the nematic fluid and the periodicity of the instability combine to form periodic cylindrical lenses in the fluid.[65] The domain lines are the result of light focused by the periodic array of lenses. This periodic lens array also acts as a transmission phase grating. Consequently, colinear diffraction spots can be observed when a laser beam propagates through the liquid crystal lens array.[61,66,67]

The presence of fluid flow in the seemingly static periodic domains has been observed by the motion of dust particles in the fluid.[65] Below the threshold voltage, the dust particles are stationary. At threshold, they move in an oscillatory pattern closely related to the domain spacing. The velocity of the particles increases with increasing voltage and with decreasing cell thickness.

Neglecting the frequency response of the electrohydrodynamic flow, Helfrich[62] calculated the threshold voltage for the domain instability. A slightly rewritten form[68] of his expression is

$$V_W^2 = \frac{V_0^2}{\zeta^2 - 1} = \frac{\dfrac{\pi^2}{\Delta\epsilon\epsilon_0}\dfrac{\epsilon_\parallel}{\epsilon_\perp}K_{33}}{\left[1 - \dfrac{\epsilon_\parallel}{\Delta\epsilon}\dfrac{1}{1 + \dfrac{\eta_0}{\gamma_1}}\right]\left[1 - \dfrac{\sigma_\perp}{\sigma_\parallel}\dfrac{\epsilon_\parallel}{\epsilon_\perp}\right] - 1}. \qquad [5]$$

where $\sigma_\perp$ and $\sigma_\parallel$ are the perpendicular and parallel components of the conductivity, and η_0 and γ are viscosity coefficients. Eq. [5] is in good agreement with the experimental data for *p*-azoxyanisole. The

Orsay Liquid Crystal Group[68–70] solved the electrohydrodynamic problem for a variable frequency, sinusoidal voltage source. The fluid instability occurs at the frequency-dependent threshold voltage

$$V_W^2 = \frac{V_0^2(1 + (2\pi f)^2\tau^2)}{\zeta^2 - (1 + (2\pi f)^2\tau^2)} \qquad [6]$$

where f is the frequency, $\tau = \epsilon_\parallel\epsilon_0/\sigma$ is a dielectric relaxation frequency, and V_0 and ζ are the same terms as in Eq. 5. A cutoff frequency

$$f_c = \frac{(\zeta^2 - 1)^{1/2}}{2\pi\tau}$$

results from Eq. [6]. The cutoff frequency is directly proportional to the conductivity.

The theoretical analysis predicts the existence of two regimes with different frequency dependences. For applied frequencies, $f < f_c$, the space charge in the fluid oscillates at the same frequency as the driving signal. This is the region of cellular fluid flow described in the preceding discussion. Because this region exists for frequencies less than the dielectric relaxation frequency, it is called the "conduction regime." The thickness-independent portion of the solid voltage–frequency curve in Fig. 9 is experimental verification of Eq. 6.

When the applied frequency is greater than f_c, the space charge does not oscillate. The fluid interacts with the applied field to result in the "dielectric regime." The threshold field is proportional to $f^{1/2}$ (see Fig. 9). Contrary to the low-frequency situation, the field is thickness independent. Periodic deformations, known as "chevrons," result from the field–fluid interaction. The spatial frequency of the chevron striations is a monotonically increasing function of the drive frequency.

In the calculation of the threshold voltage, Helfrich assumed that the spatial periodicity of the fluid deformation was proportional to the thickness of the cell. Recently, Penz and Ford[72,73] have solved the boundary-value problem associated with the electrohydrodynamic flow process. They have reproduced Helfrich's results and have also shown several other possible solutions that may account for the higher-order instabilities that cause turbulent fluid flow. Meyerhofer[74] has analytically solved the two-dimensional problem by making one simplifying assumption. He has been able to obtain good agreement between the experimental results and the calculated frequency dependence of the domain spacing and the threshold voltage.

Experimentally, it is observed that, when the applied voltage surpasses the threshold voltage by an increasing amount, the rotational velocity of the fluid increases.[75] The fluid gradually becomes more turbulent until the applied voltage exceeds twice the threshold voltage. The intense wide-angle forward scattering accompanying the strong turbulence is the dynamic scattering region.[59,61] Dynamic scattering only happens below the critical frequency and above the threshold voltage (see Fig. 9). As the voltage increases even further (beyond twice the threshold voltage), the scattering likewise increases and the fluid becomes even more turbulent[61] with virtually all traces of the underlying domain structure disappearing. The increasing turbulence as a function of voltage causes the scattering intensity and transient kinetics to be angle dependent.[76–78]

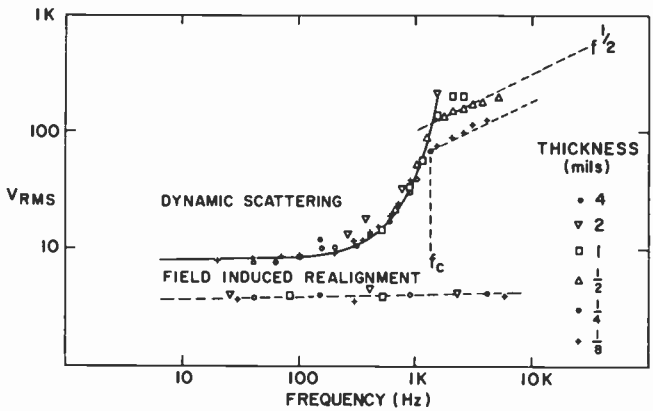


Fig. 9—Various threshold phenomena for nematic fluids with negative dielectric anisotropy and perpendicular alignment. The dashed horizontal line is the threshold voltage for induced birefringence. The curved solid line describes the frequency dependence of the threshold voltage for domains. The sloped dashed lines are the threshold plots for chevron formation. The material is MBBA at 25°C (Ref. [86]).

Though dynamic scattering usually results when the applied voltage exceeds the domain voltage, this need not be true. Dynamic scattering only seems to occur if the fluid is thick enough ($\gtrsim 6\ \mu m$) and has sufficiently low resistivity (less than $1–2 \times 10^{10}$ ohm-cm) and for negative dielectric anisotropy. Domains have been observed when any of the preceding three conditions have been violated.[79,80] With the high resistivity and thin cells, the spatial frequency of the domains is voltage-dependent.[81] At the present time, no theory exists

that defines the exact relationships between the various degrees of fluid instability.

The domain instability and dynamic scattering are also observed for both low-frequency ac and for dc applied voltage. The volume space charge, necessary for hydrodynamic motion, is not produced by the conductivity anisotropy, but by injection of charge from the electrodes.[70,82–84] Meyerhofer and Sussman[85] have measured the voltage–frequency plot for the formation of the domains from very-low-frequency ac to the cutoff frequency f_c. Below a certain frequency, which they relate to the transit time for ions, they have found that the domain threshold voltage decreases from the ac value toward the dc value. Also, the domain spacing changes below the inverse transit-time frequency. This transit-time frequency is of the order of 5 to 10 Hz. At present, the injection mechanism is unknown. It has been hypothesized that the double-layer space charge present at the electrode–liquid interface is responsible for the fluid flow.[86] Vortical fluid flow[87] and laser diffraction patterns[67] have also been observed with dc applied voltage.

Most of the discussion presented above for uniform parallel alignment is still valid with zero-field perpendicular orientation. However, there is one change. The voltage sequence for the production of dynamic scattering has an extra step. For most of the materials used at present, the voltage threshold for induced birefringence is lower than the threshold for domain formation (see Fig. 9), although there are exceptions.[11] As a function of increasing voltage, the fluid progresses from the undeformed state to the induced birefringence texture, then to the presence of domains, and finally to the occurrence of dynamic scattering.

Initial perpendicular alignment provides for greater circular symmetry in the scattering distribution. As shown in Fig. 2, the director for the deformed fluid has only medium range order, approximately 50 μm or less. The projections of the directors in a plane parallel to the fluid are randomly oriented. Because of the circular symmetry about the axis perpendicular to the cell, the laser diffraction pattern consists of a set of circular rings instead of colinear spots.[61,66] The same symmetry is observed for the angular dependence of the dynamic scattering[76] (see Fig. 10).

2.6 Storage Mode

Optical storage effects in mixtures of nematic and cholesteric materials with negative dielectric anisotropy were first observed by Heilmeier and Goldmacher.[88] They reported the following sequence of

events (see Fig. 11). Initially, with no applied voltage, the sample was in a relatively clear state. The application of a dc or low-frequency ac voltage of sufficient magnitude induced the intense scattering known as dynamic scattering. When the voltage was removed, the dynamic scattering disappeared, but a quasi-permanent, forward scattering

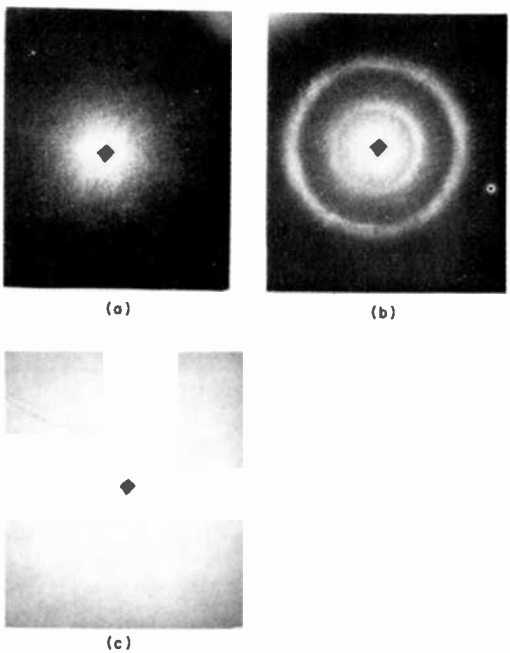


Fig. 10—Diffraction patterns for a perpendicularly oriented cell: (a) $V = 10\ V_{pp}$, very light diffuse scattering; (b) $V = 13\ V_{pp}$, diffraction rings due to domain formation; (c) $V = 16\ V_{pp}$, strong diffuse scattering characteristic of dynamic scattering. The black square in each picture is a piece of tape placed on the screen.

state remained. The reported decay time was on the order of hours at elevated temperature. The scattering texture could be returned to the clear state by the application of an audio frequency signal (greater than 500–1000 Hz). Since the original reports, other investigators have also observed the same effect.[89–91] The off-state has been identified as the Grandjean texture.[89] Due to imperfections in the Grandjean planes, the off-state is slightly scattering.

Rondelez, Gerritsma, and Arnould[54] have reported the presence of two-dimensional deformations at the threshold voltage for scattering. Electrohydrodynamic instabilities were first predicted for negative

cholesteric materials by Helfrich.[50] Hurault[51] has combined Helfrich's theory with the time dependent formalism used by Dubois-Violette, deGennes, and Parodi.[68] His calculations predicted a voltage–frequency relationship similar to that observed for pure nematics of negative dielectric anisotropy (see Fig. 9). Experimental verifica-

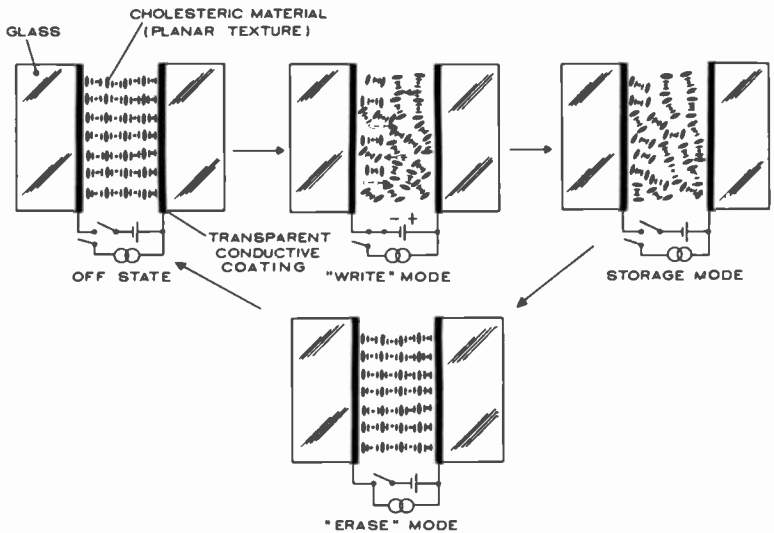


Fig. 11—Schematic illustration of the effect of low- and high-frequency signals on cholesteric fluids with negative dielectric anisotropy.

tion of both "conduction" and "dielectric" regimes has been established.[54,87] The domain periodicity is proportional to $(P_0L)^{1/2}$ in agreement with theory, where P_0 is the zero field value of the pitch and L is the cell thickness. The threshold voltage in the conduction regime is not thickness independent, but is proportional to $(L/P_0)^{1/2}$.

The exact means by which the grid structure becomes distorted so as to form the strongly scattering state is not known. The scattering texture is approximately the same as that found in the cholesteric–nematic transition,[48,60,89] even though the mechanisms for producing the scattering states are probably quite different. As explained, in cholesteric–nematic mixtures of positive dielectric anisotropy, the dielectric forces are sufficient to tilt the fluid into the scattering state. With the materials of negative dielectric anisotropy that are used in the storage effect, the strongly scattering state must arise from the strong fluid deformations associated with dynamic scattering. The light scattered from the storage state is relatively independent of the

concentration of cholesteric material, the direction of the incident light, and the cell thickness.[90]

The restoration of the planar state by the applied field is purely a dielectric interaction. It can only take place when the signal frequency is greater than cutoff frequency f_c, which is proportional to the conductivity of the liquid crystal.

2.7 Transient Response

The theoretical expressions describing the transient response of the fluid deformations are of the same general form for the different effects. A characteristic response time for the director reorientation is given approximately by[91]

$$T = \eta[\Delta\epsilon\epsilon_0 E^2 - Kq^2]^{-1}, \qquad [7]$$

where η is the proper fluid viscosity, E is the applied field, K is the appropriate elastic constant, and q is the wave-vector of the disturbance. For the phenomena occurring in pure nematics, the wave-vector is approximately π/L where L is the cell thickness. Consequently, the rise time and decay time should be of the forms

$$T_{\text{RISE}} = \eta L^2[\Delta\epsilon\epsilon_0 V^2 - K\pi^2]^{-1} \qquad [8]$$

and

$$T_{\text{DECAY}} = \frac{\eta L^2}{\pi^2 K}. \qquad [9]$$

Experimental observations for induced birefringence,[14,92] twisted nematic,[91] and dynamic scattering[60,78,84] are in reasonable agreement with the theory. For a 12-μm-thick fluid, 10 V dc, and 20°C, rise times of the order of 10 msec have been observed with the twisted nematic effect and 200 to 300 msec for dynamic scattering. Decay times are approximately 400 msec for the twisted device and 100 msec for dynamic scattering. For both dynamic scattering and the field effects, the presence of high voltages causes second-order effects to occur and the decay time becomes voltage dependent.[93,94]

The response times for both field- and conduction-induced phenomena can be changed by the presence of a second voltage source whose frequency is above some critical cutoff frequency, while the main source has a frequency less the critical frequency. For dynamic scattering, the critical frequency is f_c, which is inversely proportional to the dielectric relaxation time of the fluid. For an applied frequency $f > f_c$, the conduction torques do not affect the fluid and, through

the negative dielectric anisotropy, the dielectric torque causes a return of the fluid to its nonscattering state. Consequently, the decay time can be significantly shortened as compared to the case when no high-frequency source excites the cell.[95,96]

For field-effect materials, the critical cutoff frequency occurs only in materials that have positive dielectric anisotropy at low frequencies. Above the critical frequency, the dielectric anisotropy is negative. Several materials have been developed with a critical frequency as low as a few kHz at 25°C.[97,98] As with dynamic scattering, the application of a high-frequency source can produce a decay time much shorter than the natural decay time.[98–101] The decay time is inversely proportional to the square of the amplitude of the high-frequency voltage.[98,100]

In the cholesteric–nematic phase change with $L/P_0 > 1$, the wavevector is given by π/P_0 not π/L. The experimental rise and decay times are consistent with the theory for the field-induced phase transition.[91]

The texture change from the highly scattering cholesteric state to the Grandjean texture is described by different kinetic relationships. The erasure time of the scattering state for the storage effect is proportional to V^{-m} where V is the audio frequency signal and m varies between 1 and 3 depending upon the material.[88,102] The natural decay time from the scattering to planar texture is approximately exponentially dependent on the L/P_0 ratio[65] and the inverse of the sample temperature.[102] The inverse temperature dependence suggests that the decay time of the storage state is also directly proportional to the viscosity, which is exponentially dependent on the inverse temperature.[86]

3. Display-Related Parameters

3.1 Display Life

The determination of the operating life is fraught with complications because of the difficulty of defining the conditions that describe the end of useful operation. Subjective evaluations of the steady-state cosmetic appearance and quantitative examination of the variations in response time, nematic–isotropic temperature, and power dissipation are necessary. The changes in cosmetic appearance and response times are usually manifestations of misalignment of the fluid at the liquid–solid interface. The misalignment may be caused either by the application of voltage or by chemical interaction between the fluid and the substrate surface. Time-dependent variations of the current

may also arise from chemical interaction between the fluid and the cell walls. Some of the commonly used liquid-crystal materials are deleteriously affected by the presence of moisture and UV radiation. Proper cell packaging is then necessary to minimize these two unwanted agents.

Sussman[103] has examined the dc electrochemical failure mechanism in the dynamic scattering material *p*-methoxybenzylidene-*p'*-aminophenyl acetate (APAPA). Using the loss of 50% of the cell scattering area as his criterion for the end of life, he showed that the amount of charge passed through the cell determined the operating life. The failure mode was traced to the production of an insulating film at the anode. The utilization of ac drive signals, instead of dc, greatly diminishes the likelihood of failure being caused by electrochemical effects. Consequently, commercial dynamic scattering displays are driven by ac signals. AC operating life is cited as being greater than 5–10,000 hours.[76,104] Equally long operation is to be expected from field-effect displays operating with ac excitation.

3.2 Temperature Dependence

Until the resurgence in liquid-crystal research in the 1960's, most of the mesomorphic materials were solid at room temperature. Today, there are many liquid-crystal systems that exhibit the mesophase over a wide temperature range around 20°C. Some of these materials were cited previously in this article.

The temperature variation of the fluid properties is important for display applications. The rise and decay times are directly proportional to the fluid viscosity as shown in Eqs. [8] and [9]. Since the viscosity is approximately exponentially dependent on the inverse temperature, the response times are strongly temperature dependent.[78,86] At low temperatures, even when the material is still mesomorphic, the viscosity may be so high as to preclude the operation of the liquid-crystal display because of the sluggish transient characteristics.

The threshold voltage for the field effects should be mildly temperature sensitive due to the relatively weak temperature variation of the elastic constants and the dielectric anisotropy. Experimental results on the induced birefringence phenomena confirm this statement so long as the operating temperature is less than 95% of the nematic–isotropic temperature.[92,105] Measurements show that the threshold voltage and contrast ratio in dynamic scattering devices are almost completely insensitive to temperature throughout the nematic range.[78]

The conductivity in nematics is governed by ionic equilibrium[106]

and is inversely proportional to the viscosity and square root of the dissociation constant. Both the viscosity and equilibrium constant are exponentially dependent on temperature as is the conductivity.

4. Addressing Techniques

The presentation of visual information by a display requires a method or methods for exciting multiple positions in the display medium. The process of transmitting signal information throughout the display and exciting the different positions in the display medium is known as addressing. Two general approaches to addressing are useful with liquid crystals. One method involves beam steering, which includes electron-beam addressing (as performed in a cathode ray tube) and light-beam scanning. The alternative approach, which is discussed first, is that of matrix addressing or multiplexing. The two words are equivalent, but, for historical reasons, matrix addressing is used when referring to displays with a large number of elements, and multiplexing is reserved for displays with a relatively small number of elements.

4.1 Matrix Addressing

One of the strong motivating factors in liquid-crystal research is the possibility of constructing two-dimensional displays that dissipate little power. The third dimension, that of the glass–liquid-crystal–glass sandwich, is of the order of $\frac{1}{8}$ inch and is small compared to the other two dimensions. An example of a multi-element liquid-crystal display is the seven-segment, five-digit dynamic scattering display shown in Fig. 12. Each segment of each digit could be individually addressed by a driving signal whose frequency components are lower than the critical cutoff frequency, but this approach is wasteful of both driving circuitry and interconnections between the display and the circuitry. A much more economical approach is one in which the display is rearranged into an X-Y matrix as indicated in Fig. 13. One segment from the first digit is connected to its counterpart on each of the other digits. This interconnection scheme is repeated for the other six segments of the first digit and their counterparts. The seven row lines in the matrix represent the seven coordinated segment leads, while the five column lines are the same as the five digit backplane leads. Consequently, the five-digit seven-segment display with 35 leads can be electrically viewed as a five-by-seven matrix with only 12 leads. In general, a display with $M \times N$ resolution elements can be treated as a matrix array in which only $M + N$ leads are required.

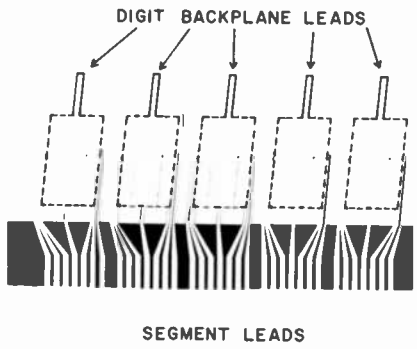


Fig. 12—Top view of a seven-segment five-digit display with electrode leads.

For the proper operation of a matrix array, it is necessary to excite only the desired element in the matrix and no other. As an example, let us discuss the 5 × 7 array. Only the first and third segments of the second digit should be scattering light with all the other segments being nonscattering. For this situation to occur, the voltages $\pm V/2$ must be less (in an absolute magnitude sense) than the threshold voltage required to initiate light scattering. The light scattered versus voltage transfer function for a typical dynamic scattering cell is given in Fig. 14. Since the contrast ratio at $2V_T$ can be in excess of 20:1, it would appear that dynamic scattering displays can be matrix-addressed without any difficulty.

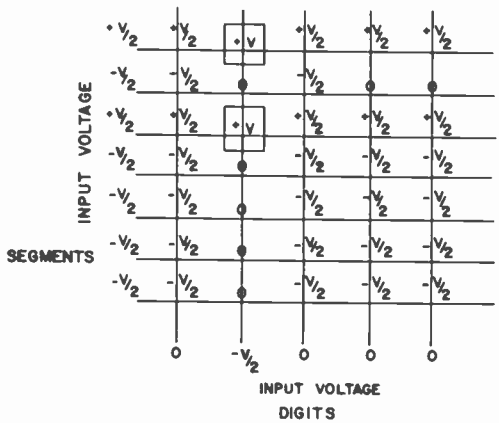


Fig. 13—Representation of the seven-segment five-digit display with leads rearranged in matrix fashion. Half-voltage selection pulses are applied.

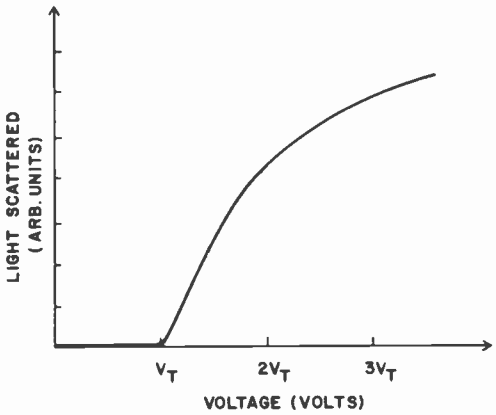


Fig. 14—Typical scattered light versus voltage curve for dynamic scattering.

So far, the matrix array has been treated in a purely static fashion. In actuality, the segment data enters in parallel from the drive circuitry and each digit is selected sequentially in time. After all of the digits have been addressed, the cycle is repeated. The temporal dependence of an arbitrary segment is presented in Fig. 15. The segment is only excited for a time T/N where T is the frame time of the data (typically 30 to 60 Hz) and N is the number of digits in the dis-

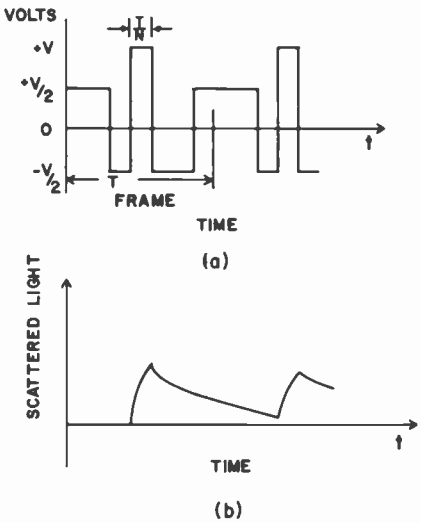


Fig. 15—(a) Applied voltage versus time for an arbitrary segment and (b) time dependence of light scattering for the same segment.

play. The scattered light as a function of time is modified from the voltage waveform by the finite response times of the phenomenon (see Fig. 15B).

Several devices and circuit parameters must be properly controlled to maximize the contrast ratio when operating in a scanning mode. One important consideration is that the rise time should be as short as possible. In practice, fast rise times are only achieved by using voltages that are far greater than twice the threshold voltage, and therefore elements that should be off turn on. Consequently, the contradictory requirements of short rise times and sufficient half-select capability restrict the number of digits that can be addressed in a multiplexing mode.

In order to improve the applied-voltage discrimination ratio for matrix arrays, the so-called one-third-voltage selection method shown in Fig. 16 can be used. This method clearly offers the advan-

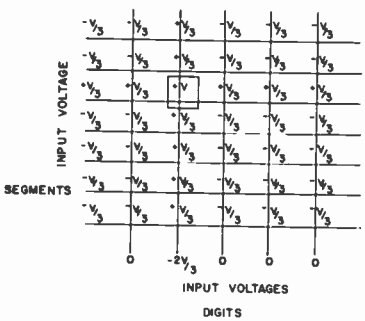


Fig. 16—Applied voltages for the one-third selection method.

tage over the half-select method of applying more voltage to the "on" elements without exceeding the threshold voltage for the elements that are supposed to be nonscattering.

Second, when the decay time is longer than the frame time T, the cell integrates the successive series of input signals that occur every frame. So long as the integration property is obeyed, it has been found that both field-effect[107,108] and dynamic-scattering[109] devices respond to the driving signal in a root-mean-square fashion. As a consequence of the rms behavior, the contrast ratio for a scanned multiplexed display can be no higher than the contrast ratio obtained with a continuous-drive waveform whose rms amplitude is the same as the rms content of the scanned waveform. Kmetz[107] has calculated the

ratio of the rms value for an on element to that for an off element in a $V:V/3$ scheme to be

$$\frac{V_{ON}}{V_{OFF}} = \sqrt{\frac{8}{N} + 1}, \qquad [10]$$

where N is the number of digits being scanned.

Because of the rms behavior, the contrast ratio is reduced at a fixed viewing angle. In addition the angular dependence of the contrast ratio, which is most noticeable when the applied voltage is only somewhat greater than the threshold voltage during continuous drive conditions, becomes more important. For example, in Fig. 5, the level of transmission at 4 V is very asymmetric in its angular dependence, whereas, at 10 V, it is not. In a multiplexed mode, as the number of digits increases, the rms behavior dictates that the angular dependence of the transmitted light becomes closer to that for the 4-V curve than for the 10-V curve. Consequently, the number of digits that can be multiplexed with good contrast is not only a function of the driving signal amplitudes, but also is dependent upon the viewing angle.

Another limitation on the applied waveforms is the requirement that the main frequency components all be much lower in frequency than the cutoff frequency both for dynamic-scattering and for field-effect devices. Frequency components just under the cutoff frequency are not as effective as lower-frequency components at producing the excited state, as can be seen for dynamic scattering in Fig. 9. As explained previously, signals whose frequency content is higher than the cutoff frequency cause a return to the unexcited state.

Alt and Pleshko[109] have extended Kmetz's analysis of rms-responding liquid-crystal devices. They have shown that the voltage-drive configuration that optimizes the number of scanned digits is not the $V:V/3$ scheme, but one in which the ratio of peak voltage to bias voltage is greater than 3:1. They also explicitly demonstrate that the more nonlinear the transmitted light versus voltage curve, the greater the multiplexing capability for a given set of drive-signal amplitudes.

Varying degrees of success have been reported for matrix addressing liquid-crystal displays. For displays using dynamic scattering, anywhere from 3 or 4 digits to 7 digits have been reported as the maximum addressing capability.[107,110] Twisted nematic devices possess approximately the same matrix addressing limitations as dynamic scattering.[107] Hareng, Assouline, and Leiba[111] have reported matrix addressing a 50 × 50 element array, while Schiekel and Fahrens-

chon[112] have successfully operated a 100 × 100 array. Both of these induced birefringence devices are two-color displays with rise times on the order of 1 second.[108,111] Due to the nature of the electro-optic process, the display appearance is a sensitive function of viewing angle, fluid thickness, voltage, and temperature. The small field of view probably limits matrix-addressed birefringence devices to projection display applications. Takata et al[110] have reported the operation of a 260 × 260 liquid-crystal display using the storage-mode effect. Because of the slow rise times at the voltage levels appropriate to matrix addressing, 10 to 20 seconds are required to adddress the whole display. The long decay times associated with the storage effect permit the maintenance of the displayed information for hours or more and the rms behavior does not occur.

The cholesteric-to-nematic phase transition effects have also been utilized in matrix-addressed displays.[55,113] Up to 28 lines have been scanned in the $V{:}V/3$ mode with a bias voltage of 35 V_{rms} and a contrast ratio of 15:1. The relatively large multiplexing capability is due to the long decay time produced by the bias voltage.[56]

The discussion until now has centered on the utilization of driving signals whose frequency components are much lower than the cutoff frequency. However, it is possible to implement useful multiplexing schemes with drive signals whose frequency components are both below and above the cutoff frequency. This approach will first be analyzed for dynamic-scattering devices.

Imagine a cell containing a liquid capable of dynamic scattering. Let us drive this cell with two sinusoidal voltage sources in series, one with applied frequency $f_1 \ll f_c$ and the other with frequency $f_2 \gg f_c$. The high-frequency signal retards the occurrence of dynamic scattering, and it can be shown that the low-frequency threshold voltage for the formation of dynamic scattering is related to the high-frequency signal by the following equation[114,115]

$$V_1^2 = V_0^2 + \gamma V_2^2, \qquad [11]$$

where V_o is the threshold voltage derived by Helfrich[62] for dc and very low frequency ac signals, and γ is a parameter dependent upon several material properties such as dielectric constant, shear torque, viscosity, and conductivity. For the common liquid crystal MBBA, $\gamma = 0.5$ at 32°C.[115]

The increase in threshold voltage for dynamic scattering induced by the high frequency can be readily utilized in a matrix display as shown in Fig. 17. $V_{L/2}$ and V_H are the zero-to-peak amplitudes of the

low- and high-frequency signals. With the simultaneous application of the low- and high-frequency voltages to those elements that are not supposed to be scattering, the low-frequency signal can be increased in amplitude so that the light intensity from the "on" elements is greater than it would have been if a single low-frequency signal had been applied. With pure nematic materials, approximately 16 lines can be addressed with reasonable contrast.[116] The addition of cholesteric material to the nematic should lengthen the decay time and thereby increase the upper limit on the number of lines that can be matrix-addressed.

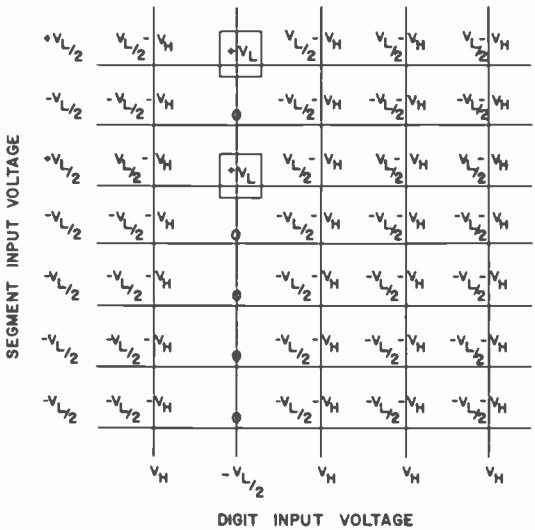


Fig. 17—Dual-frequency addressing of a matrix array.

The two-frequency approach can also be used with field-effect materials. Bücher, Klingbiel, and Van Meter[98] have shown that the low-frequency threshold voltage for the twisted nematic effect is increased by the superposition of a signal whose drive frequency is greater than the critical frequency where the dielectric anisotropy becomes negative. They claim that

$$V_{LF}^2 = V_0^2 + \left|\frac{\Delta\epsilon_{HF}}{\Delta\epsilon_{LF}}\right| V_{HF}^2 \qquad [12]$$

where V_{LF} is the amplitude of the low-frequency signal, V_{HF} is the

amplitude of the high-frequency signal, V_o is the threshold in the absence of the high-frequency signal, $\Delta\epsilon_{HF}$ is the dielectric anisotropy at the high-frequency, and $\Delta\epsilon_{LF}$ is the same quantity at the low frequency. The material they used has a crossover frequency of 2.5 kHz at 25°C, whereas other materials published in the literature have higher critical frequencies.[97,117]

Because of the combined threshold-voltage and rise-time requirements, none of the approaches described here are capable of matrix addressing a high-resolution, high-speed display. Lechner, Marlowe, Nester, and Tults[96] have investigated the application of liquid-crystal matrix displays to television and have concluded that a nonlinear threshold or isolation device, such as a diode or transistor, must be inserted in series with the liquid-crystal element at each matrix intersection to obtain the required speed and legibility for line-at-a-time addressing.

A television-rate line-at-a-time display operates in a manner similar to the small 5 × 7 matrix described previously, with the main differences being size and speed. The frame time is 30 msec and the line time is $30 \times 10^{-3}/500$ sec, or 60 μsec. A small section of a much larger matrix with diodes serving as the isolation devices is shown in Fig. 18.

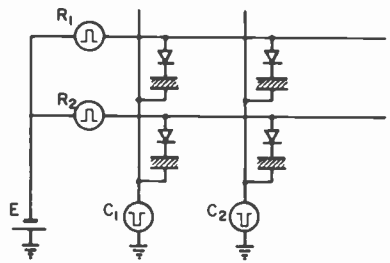


Fig. 18—A small section of a liquid-crystal–diode matrix display.

Assuming that the liquid crystal is exhibiting dynamic scattering, the response of a single diode–liquid-crystal combination is given in Fig. 19. Due to the inherent dielectric relaxation time, ϵ/σ, the addressing voltage across the cell is stored for a time (1 to 10 msec) sufficient to cause the excitation of the fluid deformation. The same basic description of the time response applies to all the electro-optic phenomena. Of course, ϵ/σ varies from material to material.

The design of a color-television display panel that uses the twisted nematic phenomena and polycrystalline thin-film transistors (TFT's)

as the isolation devices has been described by Fischer, Brody, and Escott.[118] In the TFT's, CdSe serves as the semiconductor and Al_2O_3 as the gate insulator. The display has been constructed and is 6 × 6 inches in size with 14,400 TFT's in a 120 × 120 array.[119] Operation of the entire panel has recently been demonstrated, but defects were still present in the panel and the display was only black and white.[120] Good reliability is claimed for the display.

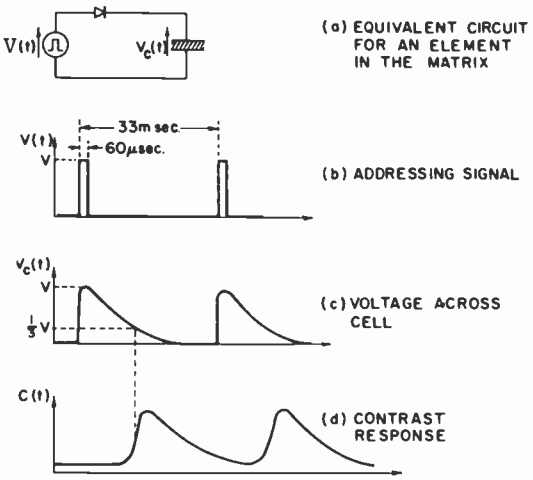


Fig. 19—Time response of a single intersection in the liquid-crystal–diode array.

Lipton and Koda[21] have also recently presented results on a CdSe TFT–liquid-crystal panel. They used relatively high-conductivity dynamic-scattering material; consequently, they added a capacitor in parallel with each TFT and display element to obtain the required electrical decay time. Brody[118,120] and co-workers were able to utilize the long dielectric relaxation time associated with the twisted nematic fluid and did not have to add the supplemental capacitance.

Liquid-crystal displays have also been constructed on a matrix of single-crystal silicon MOS FET's.[122] Pictorial-gray-scale images have been created on a 1 inch display although line defects were present. Though large arrays of TFT's should be much more economical than silicon MOS FET's, TFT's have suffered in the past from stability and reliability problems. More experimentation is necessary to prove their capabilities as the threshold devices in a liquid-crystal panel.

4.2 Beam Scanning

Images are produced on a cathode-ray tube by scanning a high-voltage electron beam across the surface of the cathodoluminescent material. Each position on the phosphor is excited sequentially as the beam is scanned by the deflection electron optics. This is an example of element-at-a-time addressing. Beam scanning in an element-at-a-time mode can be performed using either an electron beam or a light beam. Both techniques have been implemented with liquid crystals.

Van Raalte[123] was the first to describe the results of an electron-beam-scanned dynamic-scattering display. A schematic diagram of his demountable cathode-ray tube and some typical images obtained from the liquid-crystal display are presented in Fig. 20. The liquid

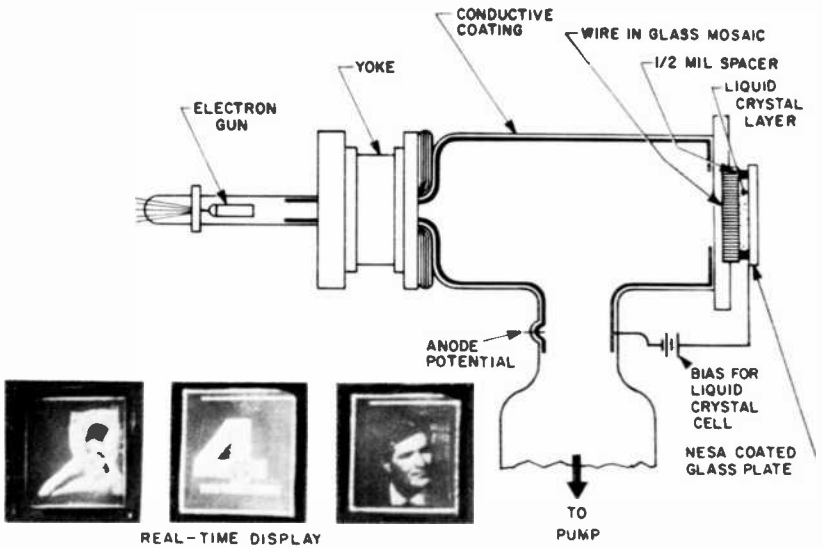


Fig. 20—Illustration of the electron-beam-scanned liquid-crystal display with some images created in the display (Ref. [123]).

crystal was sandwiched between a tin-oxide-coated glass slide and a mosaic feed-through plate constructed with fine wires inserted in glass. A segmented mirror was evaporated on the wire pin mosaic array that provided electrical contact between the electron beam and the liquid crystal. With the electron beam scanning at video rates, a maximum contrast ratio of 7.5 to 1 was obtained. Though very interesting, the display suffered from two deficiencies; (1) the liquid crys-

tal was addressed by unipolar voltage pulses and (2) the pin mosaic was extremely difficult to manufacture at the necessary resolution. Gooch[124] and co-workers have presented a solution to the first problem. They have devised an electron-beam-scanned liquid-crystal display using bistable secondary-electron emission techniques combined with a dielectric layer to drive the liquid crystal with an ac potential. The economical construction of a hermetically sealed high-resolution wire mosaic is still very difficult, although progress has been made.[125]

The complexity of the feed-through mosaic is one of the motivating factors in the development of the light-beam-addressing approaches. In the most common variation of this second method, light illuminates a thin-film photoconductor–liquid-crystal sandwich (see Fig. 21). Margerum, Nimoy, and Wong[126] presented the operation of a

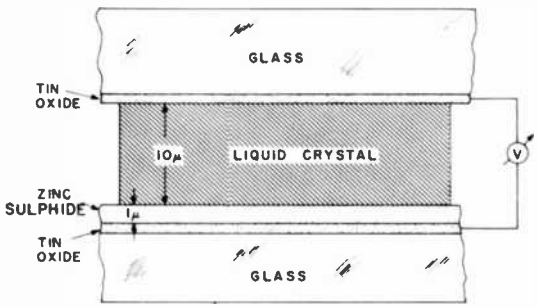


Fig. 21—Schematic diagram of a liquid-crystal–photoconductor structure.

photoconductor–liquid-crystal device. Both dynamic-scattering and storage effects were produced with light irradiating the photoconductor through the glass plate. Unilluminated, the resistance of the ZnS layer was higher than that of the liquid crystal. With UV radiation of sufficient intensity, the resistance of the photoconductor dropped below that of the liquid crystal and the voltage activated the liquid crystal. The ZnS was insensitive to the visible radiation that was used to observe the image in the liquid crystal. A sensitivity of 0.1 mJ/cm^2 was obtained.

White and Feldman[127] improved the sensitivity of the display to light by using evaporated selenium as the photoconductor. An opaque, highly reflecting light barrier was placed between the photoconductor and the liquid crystal to prevent the excitation of the photoconductor by the viewing light. Assouline, Hareng, and Leiba[128] used CdS as the photoconductor, and measured a sensitivity of 5 ×

10^{-6} J/cm^2. Similar results with CdS have been obtained by Jacobson.[29] Haas, Adams, Dir, and Mitchell[130] obtained a sensitivity of 2.5 $\times$ 10^{-6} J/cm^2 using an unspecified photoconductor and storage-effect liquid crystals.

All of the liquid-crystal–photoconductor structures discussed so far were excited by a dc voltage source. Under dc operation, various electrochemical life-degrading interactions have been observed at the liquid-crystal–photoconductor interface.[130] Beard, Bleha, and Wong[131] have reported on a photoconductor–liquid-crystal valve for projection applications that is driven by an ac voltage source (see Fig. 22). In addition, the display is operated in the reflection mode with a

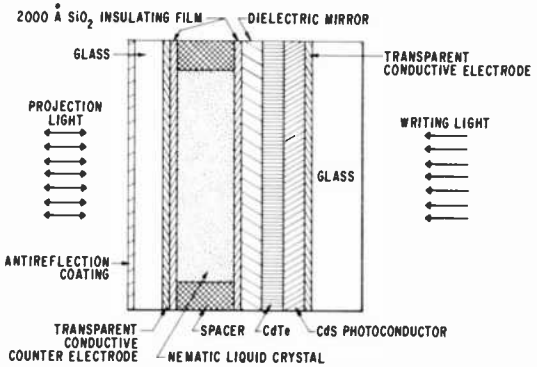


Fig. 22—Side view of an ac addressed liquid-crystal–photoconductor light valve (Ref. [131]).

multilayer combination of dielectric mirror and absorbing layer to separate the CdS photoconductor from the projection light. With 200 lumens/cm^2 of white projection light irradiating the valve, no noticeable interaction between the projection light and the photoconductor has been observed. The light valve has been excited with the projected image from a CRT[132] as well as a static slide image. At the present time, the contrast ratio and speed are still below television standards. The above structure with the induced birefringence mode has been used as the light valve in a color projection display.[133]

The photoconductor–liquid-crystal sandwiches incorporate either a dc driving source or a somewhat complicated multilayer structure. Maydan, Melchior, and Kahn[134] have circumvented both of these aspects by utilizing the thermo-optic properties of nematic–cholesteric mixtures reported by Soref.[135] He showed that nematic–cholesteric mixtures can be converted from the clear planar texture to the highly

scattering state by heating the material from the mesophase into the isotropic phase and then letting it cool. In their laser-beam-scanned display, Maydan and co-workers use the heat absorbed from the laser beam by the $In_{2-x}Sn_xO_{3-y}$ coatings to change the nematic-cholesteric storage material to the scattering state (see Fig. 23). The stored information is removed by exciting the liquid crystal with a high-frequency erase signal.

An improvement of this device has been described in which the nematic-cholesteric mixture is replaced by a smectic material.[136] Thermal writing induces the change of the smectic from the perpendicular to a scattering texture. Unlike the nematic–cholesteric mate-

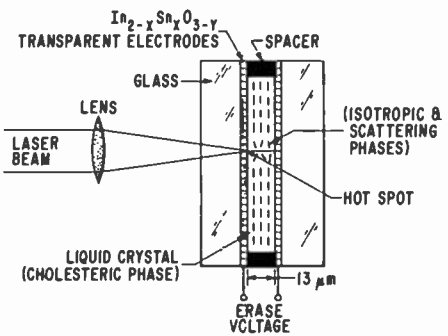


Fig. 23—A laser-beam-addressed thermo-optic liquid-crystal light valve (Ref. [134]).

rials, selective erasure is possible with the smectic device. The thermal writing is too slow for television-rate applications because of the thermal inertia of the glass–liquid-crystal system. With a laser-beam power of 20 mW, addressing speed is approximately 10^4 elements/sec for the smectic device. In the projection mode, the resolution is 50 lines/mm at a contrast ratio of approximately 10:1.

5. Summary

The different electro-optic phenomena have been classified into those that involve only dielectric forces and those that depend upon the interaction of conduction and dielectric torques. The field-effect phenomena possess several common properties. The resistivity of the materials may be as high as chemically practical, i.e., $\rho \gtrsim 10^{11}$ ohm-cm. For the induced birefringence, twisted nematic, and guest–host color switching effects, the threshold voltages are less than 3 or 4

volts, with an observed minimum of approximately one volt. The first two properties imply a very low power dissipation of less than 1 μW/cm^2. In ambient lighting, a diffuse reflector can be used with both the guest–host and twisted nematic effects to obtain a high-contrast and glare-free display.

Dynamic scattering and the storage effect are characterized by forward scattering, turbulent fluid motion during the presence of a field, and fluid resistivities that range between 10^8 and 10^{10} ohm-cm. Both effects can only occur with the applied frequency less than the dielectric relaxation frequency of the liquid crystal. The power dissipation is between 0.1 and 1 mW/cm^2 at an operating voltage of 15 V_{rms}. Neither a polarizer nor an analyzer are required. The static scattering can be stored in the nematic–cholesteric mixtures for anywhere from seconds to months.

Electrical addressing of liquid-crystal displays is accomplished either by X-Y matrix addressing or a scanning electron or light beam. The electron-beam approach is analogous to a cathode ray tube with the phosphor screen replaced by a liquid-crystal–pin wire mosaic feedthrough. Similarly, an amplitude-modulated scanning light beam can activate a photoconductor–liquid-crystal cell to provide the desired spatial pattern in the liquid crystal.

Small and/or slow matrix displays have been fabricated using the inherent threshold characteristics of liquid crystals. Both single-frequency and dual-frequency schemes have been described. The latter approach possesses greater multiplexing capability than the former, but at the price of higher applied voltage and larger power dissipation. However, due to speed of response and contrast ratio limitations, the construction of a 500 × 500 element liquid-crystal display operating at video rates does not appear feasible without the addition of a nonlinear device at each X-Y intersection. Thin-film transistor arrays are being fabricated to serve as the nonlinear devices. Their performance is being evaluated.

References

[1] C. W. Oseen, "The Theory of Liquid Crystals," *Trans. Faraday Soc.*, **2**, p. 833 (1933).

[2] F. C. Frank, "On the Theory of Liquid Crystals," *Disc. Faraday Soc.*, **25**, p. 19 (1958).

[3] P. Sheng, "Introduction to the Elastic Continuum Theory of Liquid Crystals," *RCA Review*, **35**, p. 408 (1974).

[4] P. deGennes, *The Physics of Liquid Crystals*, Oxford University Press, London (1974).

[5] W. Helfrich, "Electric Alignment of Liquid Crystals," *Mol. Cryst. and Liq. Cryst.*, **21**, p. 187 (1973).

[6] H. Gruler, T. J. Scheffer, and G. Meier, "Elastic Constants of Nematic Liquid Crystals I. Theory of the Normal Deformation," *Z. Naturforsch. A*, **27a**, p. 966 (1972).

[7] M. F. Shiekel and K. Fahrenschon, "Deformation of Nematic Liquid Crystals with Vertical Orientation in Electrical Fields," *Appl. Phys. Lett.*, **19**, p. 393 (1971).

[8] F. J. Kahn, "Electric-Field Induced Orientational Deformation of Nematic Liquid Crystals: Tunable Birefringence," *Appl. Phys. Lett.*, **20**, p. 199 (1972).

[9] R. A. Soref and M. J. Rafuse, "Electrically Controlled Birefringence of Thin Nematic Films," *J. Appl. Phys.*, **43**, p. 2029 (1972).

[10] H. Mailer, K. L. Likins, T. R. Tayloer, and J. L. Fergason, "Effect of Ultrasound on a Nematic Liquid Crystal," *Appl. Phys. Lett.*, **18**, p. 105 (1971).

[11] H. Gruler and G. Meier, "Electric Field Induced Deformations in Oriented Liquid Crystals of the Nematic Type," *Mol. Cryst. and Liq. Cryst.*, **16**, p. 299 (1972).

[12] H. Deuling, "Deformation of Nematic Liquid Crystals in an Electric Field," *Mol. Cryst. and Liq. Cryst.* **19**, p. 123 (1972).

[13] L. T. Creagh and A. R. Kmetz, "Mechanism of Surface Alignment in Nematic Liquid Crystals," *Mol. Cryst. and Liq. Cryst.*, **24**, p. 68 (1973).

[14] J. Robert and G. Labrunie, "Transient Behavior of the Electrically Controlled Birefringence in a Nematic Liquid Crystal," *J. Appl. Phys.*, **44**, p. 4689 (1973).

[15] G. Heilmeier and J. Goldmacher, U.S. Patent No. 3,499,702 (1970).

[16] M. Schadt and W. Helfrich, "Voltage Dependent Optical Activity of a Twisted Nematic Liquid Crystal," *Appl. Phys. Lett.* **18**, p. 127 (1971).

[17] F. M. Leslie, "Distortion of Twisted Orientation Patterns in Liquid Crystals by Magnetic Fields," *Mol. Cryst. and Liq. Cryst.*, **12**, p. 57 (1970).

[18] C. J. Alder and E. P. Raynes, "Room Temperature Nematic Liquid Crystal Mixtures with Positive Dielectric Anisotropy," *J. Phys.*, **D6**, p. L33 (1973).

[19] A. Boller, H. Scherrer, and M. Schadt, "Low Electro-Optic Threshold in New Liquid Crystals," *Proc. IEEE*, **60**, p. 1002 (1972).

[20] A. Ashford, J. Constant, J. Kirton, and E. P. Raynes, "Electro-Optic Performance of a New Room Temperature Nematic Liquid Crystal," *Elec. Lett.*, **9**, p. 118 (1973).

[21] R. R. Reynolds, C. Maze and E. P. Oppenheim, "Design Considerations for Positive Dielectric Nematic Mixtures Suitable for Display Applications," Abstracts of Fifth International Liquid Crystal Conf., Stockholm, Sweden, pg 236.

[22] C. H. Gooch and H. A. Tarry, "Optical Characteristics of Twisted Nematic Liquid Crystal Films," *Elec. Lett.*, **10**, p. 2 (1974).

[23] J. Robert and F. Gharadjedaghi, "Rotation du Plan de Polarisation de la Lumière dans une Structure Nématique en Helice," *C. R. Acad. Sc. Paris*, **278B**, p. 73 (1974).

[24] C. J. Gerritsma, W. H. DeJeu and P. VanZanten, "Distortion of a Twisted Nematic Liquid Crystal by a Magnetic Field," *Phys. Lett.*, **36A**, p. 389 (1971).

[25] C. Z. VanDoorn, "On the Magnetic Threshold for the Alignment of a Twisted Nematic Crystal," *Phys. Lett.*, **42A**, p. 537 (1973).

[26] A. I. Baise and M. M. Labes, "Effect of Dielectric Anisotropy on Twisted Nematics," *Appl. Phys. Lett.*, **24**, p. 298 (1974).

[27] D. Meyerhofer, "Electro-optic Properties of Twisted Field Effect Cells," Abstracts of Fifth International Liquid Crystal Conf., Stockholm, Sweden, p. 220.

[28] D. W. Berreman, "Optics in Smoothly Varying Anisotropic Planar Structures: Applicated to Liquid Crystal Twist Cells," *J. Opt. Soc. Amer.*, **63**, p. 1374 (1973).

[29] S. Kobayashi and F. Takeuchi, "Multicolor Field-Effect Display Devices with Twisted Nematic Liquid Crystals," *Proc. of S.I.D.*, **14**, p. 115 (1973).

[30] C. Z. VanDoorn and J. L. A. M. Heldens, "Angular Dependent Optical Transmission of Twisted Nematic Liquid Crystal Layers," *Phys. Lett.*, **47A**, p. 135 (1974).

[31] F. Brochard, "Backflow Effects in Nematic Liquid Crystals," *Mol. Cryst. and Liq. Cryst.*, **23**, p. 51 (1973).

[32] C. J. Gerritsma, J. A. Geurst and A. M. J. Spruijt, "Magnetic-Field-Induced Motion of Disclinations in a Twisted Nematic Layer," *Phys. Lett.*, **43A**, p. 356 (1973).

[33] E. P. Raynes, "Twisted Nematic Liquid Crystal Electro-Optic Devices with Areas of Reverse Twist," *Elec. Lett.*, **9**, p. 101 (1973).

[34] E. P. Raynes, "Improved Contrast Uniformity in Twisted Nematic Liquid Crystal Electro-Optic Display Devices," *Elec. Lett.*, **10**, p. 141 (1974).

[35] P. J. Wild, "Twisted Nematic Liquid Crystal Displays with Low Threshold Voltage," *Comptes Rendus des Journées d'Electronique, EPFL*, p. 102 (1973).

[36] G. H. Heilmeier and L. A. Zanoni, "Guest-Host Interactions in Nematic Liquid Crystals—A New Electro-Optic Effect," *Appl. Phys. Lett.*, **13**, p. 91 (1968).

[37] G. H. Heilmeier, J. A. Castellano, and L. A. Zanoni, "Guest-Host Interactions in Nematic Liquid Crystals," *Mol. Cryst. and Liq. Cryst.*, **8**, p. 293 (1969).

[38] J. J. Wysocki, J. Adams, and W. Haas, "Electric-Field-Induced Phase Change in Cholesteric Liquid Crystals," *Phys. Rev. Lett.*, **20**, p. 1024 (1968).

[39] E. Sackmann, S. Meiboom and L. C. Snyder, "On the Relation of Nematic to Cholesteric Mesophases," *J. Am. Chem. Soc.*, **89**, p. 5981 (1967).

[40] P. G. deGennes, "Calcul de la Distortion D'une Structure Cholesterique Par un Champ Magnetique," *Sol. St. Commun.*, **6**, p. 163 (1968).

[41] R. B. Meyer, "Effects of Electric and Magnetic Fields on the Structures of Cholesteric Liquid Crystals," *Appl. Phys. Lett.*, **12**, p. 281 (1968).

[42] F. Rondelez and J. P. Hulin, "Distortions of a Planar Cholesteric Structure Induced by a Magnetic Field," *Sol. St. Commun.*, **10**, p. 1009 (1972).

[43] F. J. Kahn, "Electric-Field-Induced Color Changes and Pitch Dilation in Cholesteric Liquid Crystals," *Phys. Rev. Lett.*, **24**, p. 209 (1969).

[44] R. B. Meyer, "Distortion of a Cholesteric Structure by a Magnetic Field," *Appl. Phys. Lett.*, **14**, p. 208 (1969).

[45] G. Durand, L. Leger, F. Rondelez and M. Veyssie, "Magnetically Induced Cholesteric-to-Nematic Phase Transition in Liquid Crystals," *Phys. Rev. Lett.*, **22**, p. 227 (1969).

[46] G. H. Heilmeier and J. E. Goldmacher, "Electric-Field-Induced Cholesteric-Nematic Phase Change in Liquid Crystals," *J. Chem. Phys.*, **51**, p. 1258 (1969).

[47] G. W. Gray, K. J. Harrison and J. A. Nash, "New Family of Nematic Liquid Crystals for Displays," *Elec. Lett.*, **9**, p. 130 (1973).

[48] J. P. Hulin, "Parametric Study of the Optical Storage Effect in Mixed Liquid Crystal Systems," *Appl. Phys. Lett.*, **21**, p. 455 (1972).

[49] W. Helfrich, "Deformation of Cholesteric Liquid Crystals with Low Threshold Voltage," *Appl. Phys. Lett.*, **17**, p. 531 (1970).

[50] W. Helfrich, "Electrohydrodynamic and Dielectric Instabilities of Cholesteric Liquid Crystals," *J. Chem. Phys.* **55**, p. 839 (1971).

[51] J. Hurault, "Static Distortions of a Cholesteric Planar Structure Induced by Magnetic or A.C. Electric Fields," Fourth International Liquid Crystal Conf., Kent, Ohio, Aug. 1972.

[52] C. J. Gerritsma and P. VanZanten, "Periodic Perturbations in the Cholesteric Plane Texture," *Phys. Lett.*, **37A**, p. 47 (1971).

[53] T. J. Scheffer, "Electric and Magnetic Field Investigations of the Periodic Gridlike Deformation of a Cholesteric Liquid Crystal," *Phys. Rev. Lett.* **28**, p. 598 (1972).

[54] F. Rondelez, H. Arnould and C. J. Gerritsma, "Electrohydrodynamic Effects in Cholesteric Liquid Crystals Under AC Electric Fields," *Phys. Rev. Lett.*, **28**, p. 735 (1972).

[55] J. J. Wysocki et al., "Cholesteric-Nematic Phase Transition Displays," *Proc. SID*, **13**, p. 115 (1972).

[56] T. Ohtsuka and M. Tsukamoto, "AC Electric-Field-Induced Cholesteric-Nematic Phase Transition in Mixed Liquid Crystal Films," *Jap. J. Appl. Phys.*, **12**, p. 22 (1973).

[57] W. Greubel, "Bistability Behavior of Texture in Cholesteric Liquid Crystals in an Electric Field," *Appl. Phys. Lett.*, **25**, p. 5 (1974).

[58] D. L. White and G. N. Taylor, "A New Absorptive Mode Reflective Liquid Crystal Display Device," *J. Appl. Phys.*, Nov. 1974.

[59] G. H. Heilmeier, L. A. Zanoni and L. A. Barton, "Dynamic Scattering in Nematic Liquid Crystals," *Appl. Phys. Lett.*, **13**, p. 46 (1968).

[60] G. H. Heilmeier, L. A. Zanoni and L. A. Barton, "Dynamic Scattering: A New Electro-Optic Effect in Certain Classes of Nematic Liquid Crystals," *Proc. IEEE*, **56**, p. 1162 (1968).

[61] C. Deutsch and P. N. Keating, "Scattering of Coherent Light from Nematic Liquid Crystals in the Dynamic Scattering Mode," *J. Appl. Phys.*, **40**, p. 4049 (1969).

[62] W. Helfrich, "Conduction-Induced Alignment of Nematic Liquid Crystals: Basic Model and Stability Considerations," *J. Chem. Phys.*, **51**, p. 4092 (1969).

[63] E. F. Carr, "Ordering in Liquid Crystals Owing to Electric and Magnetic Fields," *Advan. Chem. Ser.*, **63**, p. 76 (1967).

[64] R. Williams, "Domains in Liquid Crystals," *J. Chem. Phys.*, **39**, p. 384 (1963).

[65] P. A. Penz, "Voltage-Induced Vorticity and Optical Focusing in Liquid Crystals," *Phys. Rev. Lett.*, **24**, p. 1405 (1970).

[66] T. O. Carroll, "Liquid Crystal Diffraction Grating," *J. Appl. Phys.*, **43**, p 767 (1972).

[67] G. Assouline, A. Dmitrieff, M. Hareng, and E. Leiba, "Diffraction d'un Faisceau Laser par un Cristal Liquide Nématique Soumis à un Champ Électrique," *C. R. Acad. Sci. Paris*, **271B**, p. 857 (1970).

[68] E. Dubois-Vilette, P. G. deGennes, and O. Parodi, "Hydrodynamic Instabilities of Nematic Liquid Crystals Under AC Electric Fields," *J. Physique*, **32**, p. 305 (1971).

[69] Orsay Liquid Crystal Group, "Hydrodynamic Instabilities in Nematic Liquids Under AC Electric Fields," *Phys. Rev. Lett.*, **25**, p. 1642 (1970).

[70] P. G. deGennes, "Electrohydrodynamic Effects in Nematics," *Comments Sol. St. Phys.*, **3**, p. 148 (1971).

[71] R. A. Kashnow and H. S. Cole, "Electrohydrodynamic Instabilities in a High-Purity Nematic Liquid Crystal," *J. Appl. Phys.*, **42**, p. 2134 (1971).

[72] P. A. Penz and G. W. Ford, "Electrohydrodynamic Solutions for Nematic Liquid Crystals," *Appl. Phys. Lett.*, **20**, p. 415 (1972).

[73] P. A. Penz and G. W. Ford, "Electromagnetic Hydrodynamics of Liquid Crystals," *Phys. Rev.*, **6A**, p. 414 (1972).

[74] D. Meyerhofer, "Electro Hydrodynamic Instabilities in Nematic Liquid Crystals," *RCA Review*, **35**, p. 433 (1974).

[75] T. O. Carroll, "Dependence of Conduction-Induced Alignment of Nematic Liquid Crystals Upon Voltage Above Threshold," *J. Appl. Phys.*, **43**, p. 1342 (1972).

[76] L. Goodman, "Light Scattering in Electric-Field Driven Nematic Liquid Crystals," *Proc. SID*, **13**, p. 121 (1972).

[77] L. Cosentino, "On the Transient Scattering of Light by Pulsed Liquid Crystal Cells," *IEEE Trans. Electron Devices*, **ED-1**, p. 1192 (1971).

[78] L. Creagh, A. Kmetz and R. Reynolds, "Performance Characteristics of Nematic Liquid Crystal Display Devices," *IEEE Trans. Electron Devices*, **ED-18**, p. 672 (1971).

[79] W. Greubel and U. Wolff, "Electrically Controllable Domains in Nematic Liquid Crystals," *Appl. Phys. Lett.*, **19**, p. 213 (1971).

[80] W. H. DeJeu, C. J. Gerritsma, and A. M. VanBoxtel, "Electrohydrodynamic Instabilities in Nematic Liquid Crystals," *Phys. Lett.*, **34A**, p. 203 (1971).

[81] L. K. Vistin, "Electrostructural Effect and Optical Properties of a Certain Class of Liquid Crystals and Their Binary Mixtures," *Sov. Phys. Crystallogr.*, **15**, p. 514 (1970).

[82] N. Felici, "Phénomènes Hydro et Aérodynamiques dans la Conduction des Dielectrique Fluide," *Rev. Gen. Elec.*, **78**, p. 717 (1969).

[83] Orsay Liquid Crystal Group, "AC and DC Regimes of the Electrohydrodynamic Instabilities in Nematic Liquid Crystals," *Mol. Cryst. and Liq. Cryst.*, **12**, p. 251 (1971).

[84] H. Koelmans and A. M. VanBoxtel, "Electrohydrodynamic Flow in Nematic Liquid Crystals," *Mol. Cryst. and Liq. Cryst.*, **12**, p. 185 (1971).

[85] D. Meyerhofer and A. Sussman, "The Electrohydrodynamic Threshold in Nematic Liquid Crystals in Low Frequency Fields," *Appl. Phys. Lett.*, **20**, p. 337 (1972).

[86] A. Sussman, "Electro-Optic Liquid Crystal Devices: Principles and Applications," *IEEE Trans. Parts, Hybrids and Packaging*, **PHP-8**, p. 28 (1972).

[87] G. Durand, M. Veyssie, F. Rondelez and L. Leger, "Effect Électrohydrodynamique dans un Cristal Liquide Nématique," *C. R. Acad. Sc. Paris*, **270B**, p. 97 (1970).

[88] G. H. Heilmeier and J. E. Goldmacher, "A New Electric Field Controlled Reflective Optical Storage Effect in Mixed Liquid Crystal Systems," *Proc. IEEE* **57**, p. 34 (1969).

[89] G. Dir et al., "Cholesteric Liquid Crystal Texture Change Displays," *Proc. SID*, **13**, p. 105 (1972).

[90] D. Meyerhofer and E. F. Pasierb, "Light Scattering Characteristics in Liquid Crystal Storage Materials," *Mol. Cryst. and Liq. Cryst.*, **20**, p. 279 (1973).

[91] E. Jakeman and E. P. Raynes, "Electro-Optic Response Times in Liquid Crystals," *Phys. Lett.*, **39A**, p. 69 (1972).

[92] J. Robert, G. Labrunie and J. Borel, "Static and Transient AC Electric Field Effect on Homeotropic Thin Layers," Fourth International Liquid Crystal Conf., Kent, Ohio; Aug. 1972, Paper No 105.

[93] A. Sussman, "Secondary Hydrodynamic Structure in Dynamic Scattering," *Appl. Phys. Lett.*, **21**, p. 269 (1972).

[94] C. J. Gerritsma, C. Z. VanDoorn and P. VanZanten, "Transient Effects in the Electrically Controlled Light Transmission of a Twisted Nematic Layer," *Phys. Lett.*, **48A**, p. 263 (1974).

[95] C. H. Gooch and H. A. Tarry, "Dynamic Scattering in the Homeotropic and Homogeneous Textures of a Nematic Liquid Crystal," *J. of Phys. D. Appl. Phys.*, **5**, p. L25 (1972).

[96] B. J. Lechner, F. Marlowe, E. Nester, and J. Tults, "Liquid Crystal Displays," *Proc. IEEE*, **59**, p. 1566 (1971).

[97] W. H. DeJeu, C. J. Gerristma, P. VanZanten, and W. J. A. Gossens, "Relaxation of the Dielectric Constant Electrohydrodynamic Instabilities in a Liquid Crystal," *Phys. Lett.*, **39A**, p. 355 (1972).

[98] H. K. Bücher, R. T. Klingbiel, and J. P. VanMeter, "Frequency-Addressed Liquid Crystal Field Effect," *Appl. Phys. Lett.*, **25**, p. 186 (1974).

[99] E. P. Raynes and I. A. Shanks, "Fast Switching Twisted Nematic Electro-Optical Shutter and Color-Filter," *Elec. Lett.*, **10**, p. 114 (1974).

[100] T. S. Chang and E. E. Loebner, "Crossover Frequencies and Turn-Off Time Reduction Scheme for Twisted Nematic Liquid Crystal Displays," *Appl. Phys. Lett.*, **25**, p. 1 (1974).

[101] G. Baur, A. Stieb, and G. Meier, "Controlled Decay of Electrically Induced Deformations in Nematic Liquid Crystals," *Appl. Phys.*, **2**, p. 349 (1973).

[102] B. Kellernevich and A. Coche, "Relaxation of Light Scattering in Nematic-Cholesteric Mixtures," *Mol. Cryst. and Liq. Cryst.*, **24**, p. 113 (1973).

[103] A. Sussman, "Dynamic Scattering Life in the Nematic Compound p-Methoxybenzylidene-p-Amino Phenyl Acetate as Influenced by Current Density," *Appl. Phys. Lett.*, **21**, p. 126 (1972).

[104] L. Pohl, R. Steinsträsser, and B. Hampel, "Performance of Nematic Phase V and VA in Liquid Crystal Displays," Fourth Internat. Liq. Cryst. Conf., Kent, Ohio; Aug. 1972, Paper No. 144.

[105] I. Haller, "Elastic Constants of the Nematic Liquid Crystalline Phase of p-Methoxybenzylidene-p-n-Butylaniline (MBBA)," *J. Chem. Phys.*, **57**, p. 1400 (1972).

[106] A. Sussman, "Ionic Equilibrium and Ionic Conductance in the System Tetra-Iso-Pentyl Ammonium Nitrate-p-Azoxyanisole," *Mol. Cryst. and Liq. Cryst.*, **14**, p. 183 (1971).

[107] A. R. Kmetz, "Liquid Crystal Displays Prospects in Perspective," *IEEE Trans. Elec. Dev.*, **ED-20**, p. 954 (1973).

[108] M. Hareng, G. Assouline and E. Leiba, "La Biréfringence Électriquement Contrôlée dans les Cristaux Liquides Nématiques," *Appl. Opt.*, **11** p. 2920 (1972).

[109] P. M. Alt and P. Pleshko, "Scanning Limitations of Liquid Crystal Displays," *IEEE Trans. Elec. Dev.*, **ED-21**, p. 146 (1974).

[110] H. Takata, O. Kogure, and K. Murase, "Matrix-Addressed Liquid Crystal Display." *IEEE Trans. Elec. Dev.*, **ED-20**, p. 990 (1973).

[111] M. Hareng, G. Assouline, and E. Leiba, "Liquid Crystal Matrix Display by Electrically Controlled Birefringence," *Proc. IEEE*, **60**, p. 913 (1972).

[112] M. F. Schiekel and K. Fahrenschon, "Multicolor Matrix Displays Based on Deformation of Vertically Aligned Nematic Liquid Crystal Phases," Digest 1972 Soc. for Information Display International Symp., San Francisco, Calif., p. 98.

[113] T. Ohtsuka, M. Tsukamoto. and M. Tsuchiya, "Liquid Crystal Matrix Display," *Jap. J. Appl. Phys.*, **12**, p. 371 (1973).

[114] C. R. Stein and R. A. Kashnow, "A Two Frequency Coincidence Addressing Scheme for Nematic Liquid Crystal Displays," *Appl. Phys. Lett.*, **19**, p. 343 (1971).

[115] P. J. Wild and J. Nehring, "An Improved Matrix Addressed Liquid Crystal Display," *Appl. Phys. Lett.*, **19**, p. 335 (1971).

[116] C. R. Stein and R. A. Kashnow, "Recent Advances in Frequency Coincidence Matrix Addressing of Liquid Crystal Displays," Digest 1972 Soc. for Information Display International Symp., San Francisco, Calif. p. 64.

[117] M. Schadt, "Dielectric Properties of Some Nematic Liquid Crystals with Strong Positive Dielectric Anisotropy," *J. Chem. Phys.*, **56**, p. 1494 (1972).

[118] A. G. Fischer, T. P. Brody, and W. S. Escott, "Design of a Liquid Crystal Color TV Panel," IEEE Conf. Record 1972 Conf. on Display Devices, New York, NY, p. 64.

[119] T. P. Brody, J. Asars and G. D. Dixon, "A 6 × 6 Inch 20 Lines per Inch Liquid Crystal Display Panel," *IEEE Trans. Elec. Dev.*, **ED-20**, p. 995 (1973).

[120] T. F. Brody, F. C. Luo, D. H. Vavies, and E. W. Greeneich, "Operational Characteristics of a 6 × 6 Inch, TFT Matrix Array, Liquid Crystal Display," Digest 1974 Soc. for Information Display International Symp., San Diego, Calif., p. 166.

[121] L. Lipton and N. Koda, "Liquid Crystal Matrix Display for Video Applications," *Proc. SID*, **14**, p. 127 (1973).

[122] M. Ernstoff, A. M. Leupp, M. J. Little and H. T. Peterson, "Liquid Crystal Pictorial Display," Technical Digest 1973 International Electron Devices Meeting, Washington, D.C., p. 548.

[123] J. A. van Raalte, "Reflective Liquid Crystal Television Display," *Proc. IEEE*, **56**, p. 2146 (1968).

[124] C. H. Gooch et al., "A Storage Cathode-Ray Tube with Liquid Crystal Display," *J. Phys.*, **D6**, p. 1664 (1974).

[125] C. Burrowes, "Electrical Fiber Plates—A New Tool For Storage and Display," *IEEE Conf. Record 1970 IEEE Conf. on Display Devices*, New York, NY, p. 126.

[126] J. D. Margerum, J. Nimoy and S.-Y. Wong, "Reversible Ultraviolet Imaging with Liquid Crystals," *Appl. Phys. Lett.*, **17**, p. 51 (1970).

[127] D. L. White and M. Feldman, "Liquid Crystal Light Valves," *Elec. Lett.*, **6**, p. 837 (1970).

[128] G. Assouline, M. Hareng, and E. Leiba, "Liquid Crystal and Photoconductor Image Converter," *Proc. IEEE*, **59**, p. 1355 (1971).

[129] A. Jacobson et al., "Photoactivated Liquid Crystal Light Valve," Digest 1972 SID International Symp., San Francisco, Calif., p. 70.

[130] W. Haas, J. Adams, G. Dir and C. Mitchell, "Liquid Crystal Memory Panels," *Proc. SID*, **14**, p. 121 (1973).

[131] T. D. Beard, W. P. Bleha and S.-Y. Wong, "Alternating Current Liquid Crystal Light Valve," *Appl. Phys. Lett.*, **22**, p. 90 (1973).

[132] W. P. Bleha, J. Grinberg, and A. D. Jacobson, "AC Driven Photoactivated Liquid Crystal Light Valve," Digest 1973 SID International Symp., New York, NY, p. 42.

[133] J. Grinberg et al., "Photoactivated Liquid Crystal Light Valve for Color Symbology Display," *Conf. Record 1974 IEEE SID Conf. on Display Devices and Systems*, New York, p. 47.

[134] D. Maydan, H. Melchior and F. Kahn, "Thermally Addressed Electrically Erased High-Resolution Liquid Crystal," *Appl. Phys. Lett.*, **21**, p. 392 (1972).

[135] R. A. Soref, "Thermo-Optic Effects in Nematic-Cholesteric Mixtures," *J. Appl. Phys.*, **41**, p. 3022 (1970).

[136] F. J. Kahn, "IR-Laser-Addressed Thermo-Optic Smectic Liquid Crystal Storage Displays," *Appl. Phys. Lett.*, **22**, p. 111 (1973).

Liquid-Crystal Optical Waveguides

D. J. Channin

RCA Laboratories, Princeton, N. J. 08540

Abstract—This paper discusses the use of liquid-crystal layers as optical waveguides. The basic aspects of waveguide propagation are reviewed, with emphasis on mode spectra, phase matching and coupling, and scattering. Experimental techniques appropriate to liquid crystals are presented, along with measurements of attenuation and electro-optic phenomena in nematic waveguides.

Introduction

In recent years a new optical technology has developed in the field of light guiding in thin films and fibers. Liquid-crystal materials have yet to contribute to this technology to the extent that they have to display devices and optical beam processing; nevertheless, examples of liquid-crystal waveguide modulators,[1,2] switches,[3] and deflectors[4] have been demonstrated. Furthermore, optical waveguiding holds potential as a tool for investigating physical and chemical processes in thin films,[5] including liquid crystals.

This paper reviews the basic theory of optical waveguiding in planar structures, with emphasis on the general concepts of mode spectra, energy distribution within the guide, phase matching, and scattering. Experimental techniques appropriate to liquid crystals are discussed, and some experimental results on attenuation measurements and electro-optic effects are described. More detailed discussions of optical waveguide theory and technology (often called integrated optics) are available in recent reviews and books.[6–8]

Guided Optical Waves

The simplest optical waveguide structure is an infinite planar slab of perfectly transparent, isotropic, dielectric material with refractive index n_g bounded on both surfaces by similar material of index $n_o < n_g$ (see Fig. 1). According to Snell's law, light rays in the slab will be totally internally reflected if they have an incidence angle θ with a surface satisfying

$$\sin\theta > \frac{n_0}{n_g}. \qquad [1]$$

Since both surfaces of the slab are parallel, the light will undergo an identical reflection at the opposite boundary and be trapped within

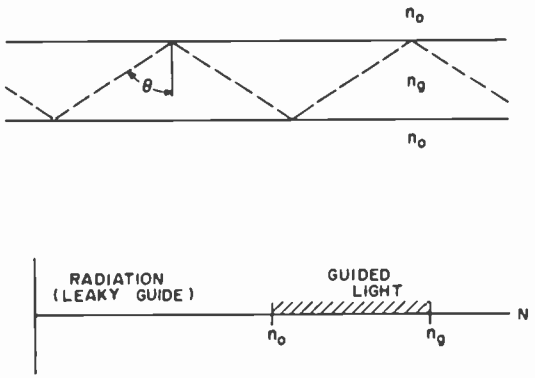


Fig. 1—Total internal reflection of light in slab waveguide. $N = c/v$ is the effective refractive index for phase velocity v.

the slab. The phase velocity v of the light as it travels within the slab is given by

$$v = \frac{c}{n_g \sin\theta}, \qquad [2]$$

where c is the speed of light in free space. From Eqs. [1] and [2] it is apparent that the guided light may have phase velocities within the range

$$\frac{c}{n_0} < v < \frac{c}{n_g}. \qquad [3]$$

Propagation under such conditions is called waveguiding.

If θ is too small to satisfy Eq. [1], some light is transmitted out of the slab at each encounter with the surface. Nevertheless, sufficient reflection may occur that the light in the slab can be characterized as a "leaky" wave, subject to loss by radiation away from the slab. It is convenient to deal with the effective refractive index $N \equiv c/v$ to characterize both leaky and truly guided light.

It has been tacitly assumed that the thickness of the slab is much greater than an optical wavelength, so that geometric optics is applicable. Should the thickness become comparable to the wavelength, destructive interference between the multiple reflections prevents guided propagation except at specific incidence angles for which the interference is constructive. Waveguiding then occurs only for a discrete set of guided modes, with effective indices N_j determined by the guide thickness d as well as the two refractive indices n_g and n_0. This situation, though retaining many features of the geometric optics limit, is properly described by wave optics.

Suppose now that the slab and surroundings are replaced by a medium that is translationally invariant in the y-z plane, but has refractive index $n(x)$ that varies with position in the x direction. The conditions

$$1 \leq n_0 \leq n(x) \leq n_g \qquad [4a]$$

$$n(\pm\infty) = n_0 \qquad [4b]$$

are imposed on $n(x)$.

Consider light propagating parallel to the z axis, translationally invariant along the y axis, and having an electric field ϵ_y. Such light satisfies the wave equation

$$\frac{\partial^2 \epsilon_y}{\partial x^2} + \frac{\partial^2 \epsilon_y}{\partial z^2} + K_0^2 n^2(x) \epsilon_y = 0. \qquad [5]$$

Separating the variables by coordinates yields

$$\epsilon_y = \epsilon(x) \exp\{iKz\} \qquad [6]$$

$$\frac{\partial^2 \epsilon}{\partial x^2} + K_0^2 \left[n^2(x) - \frac{K^2}{K_0^2} \right] \epsilon = 0. \qquad [7]$$

In these equations $K_0 = 2\pi/\lambda_0$, where λ_0 is the free-space optical wavelength. It is convenient to make again the identification of $K/K_0 = c/v = N$, the effective refractive index of the guided modes.

Equations formally identical to Eq. [7] arise in many physical problems, such as the Schrödinger equation for a particle in a potential well. N is the eigenvalue for this equation and, in accordance with Eqs. [4a] and [4b], will take on a discrete and finite spectrum N_o, $N_1, \ldots N_j, \ldots N_{\max}$ within the range $n_o < N_j < n_g$, and a continuous spectrum in the range $1 < N < n_o$. The former characterizes the guided modes and the latter the so-called radiation modes. The two kinds of modes together are associated with a complete orthogonal set of eigenfunctions of Eq. [7]. The leaky modes are solutions of Eq. [7] but are not part of the complete set of eigenfunctions. They have a discrete spectrum within the range of N spanned by the continuous radiation mode spectrum. The ranges spanned by the different modes are shown in Fig. 2.

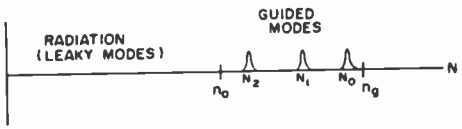


Fig. 2—Ranges of effective indices for guided and radiation modes.

We return now to the particular case of the slab waveguide defined by

$$n(x) = \begin{cases} n_g, |x| < d/2 \\ n_0, |x| > d/2 \end{cases} \qquad [8]$$

The guided modes have effective indices N_j determined by the solutions of the dispersion relation

$$\tan\left(\frac{K_0 d}{2}\sqrt{n_g^2 - N_j^2}\right) = \sqrt{\frac{n_g^2 - N_j^2}{N_j^2 - n_0^2}}\,. \qquad [9]$$

The number of modes increases with the guide thickness d. The lowest-order mode N_o is closest in effective index to n_g, while the highest order mode has effective index $N_{\max}$ closest to n_o.

The spatial distributions of the electric fields for these modes are given by

$$\epsilon_y(x,z) = A_j \exp(iK_0N_jz) \begin{cases} \sin(xK_0\sqrt{n_g^2 - N_j^2}), \\ j = 1,3,\ldots |x| < d/2 \\ \cos(xK_0\sqrt{n_g^2 - N_j^2}), \\ j = 0,2,\ldots \end{cases}$$

$$\epsilon_y(x,z) = B_j \exp(iK_0N_jz) \begin{cases} \exp(-xK_0\sqrt{N_j^2 - n_0^2}), \\ x > d/2 \\ \exp(+xK_0\sqrt{N_j^2 - n_0^2}) \\ x < d/2 \end{cases} \qquad [10]$$

The exponentially decaying field outside the slab is called the evanescent field. Since the electric field ϵ_y lies in the waveguide plane the modes are called transverse electric or TE. A corresponding set of transverse magnetic or TM modes also exist. These satisfy a wave equation for H_y similar to Eq. [7]. The dispersion relationship for TM modes is

$$\tan\left(\frac{K_0d}{2}\sqrt{n_g^2 - N_j^2}\right) = \frac{n_0^2}{n_g^2}\sqrt{\frac{n_g^2 - N_j^2}{N_j^2 - n_0^2}}. \qquad [11]$$

Eq. [11] differs slightly from Eq. [9], indicating differing phase velocities for TE and TM waves despite the assumption of isotropic materials. The velocity differences come from the differing phase shifts on reflection at the interfaces for light polarized parallel to (TE) or perpendicular to (TM) the reflecting surface. Should the waveguide be composed of anisotropic material there will be an additional velocity difference due to the variation of guide index n_g with optical polarization.

In practice, many optical waveguides are asymmetric, comprising thin films with substrate on one side and air on the other. Others have graded continuous index distributions produced by diffusion of atoms into or out of a bulk material. For such waveguides the range of effective indices spanned by the guided modes is

$$(n_0)_{\max} < N_j < n_g, \qquad [12]$$

where $(n_o)_{\max}$ is the *greatest* of the indices surrounding the guide. The spacing of the modes and the light distribution within the guide is determined by the specific refractive index profile, though some features are common to all guides.

In general, the optical energy distribution is concentrated in the

highest-index material for low-order modes, and spreads out into the lower-index material as the mode order increases. Thus, the optical field of a multimode waveguide will sample different regions of the guide as different modes are excited. The techniques of optical waveguiding are therefore potentially useful in studying the structure of thin films and layers.

Phase Matching and Coupling

The recent surge of activity in planar optical waveguides was initiated by the development of practical and efficient ways to couple light between guided modes and beams in free space.[9,10] These techniques are based on the introduction of an additional element to the wave-

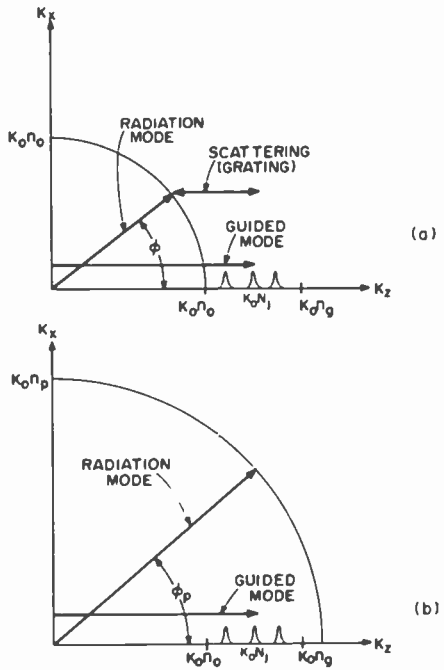


Fig. 3—Wave vector diagrams for (a) grating coupler and (b) prism coupler.

guide to match the phase velocities of the guided modes to those of the radiation modes.

Fig. 3 represents the extension of Figs. 1 and 2 to include the plane perpendicular to the waveguide. The axes now represent wave vec-

tors, the magnitudes of which are related to the effective indices by multiplication by K_o. Fig. 3a shows the basis for a grating coupler. A small periodic modulation of the guide thickness or refractive index phase matches the guided mode with wave vector K_oN_j to radiation in the form of a ray at angle ϕ to the guide surface. The periodic modulation has wavelength λ_s such that

$$K_s = \frac{2\pi}{\lambda_s} = K_0(N_j - n_0 \cos\phi). \quad [13]$$

The basis for a prism coupler is shown in Fig. 3b. A small part of the waveguide is bounded on one side by material of index $n_p > n_g$. The condition

$$n_p \cos\phi_p = N_j \quad [14]$$

establishes coupling between the guided mode N_j and a ray with incidence angle ϕ_p inside the high-index material. In practice the high-index material takes the form of a prism that refracts the internal ray into free space at a convenient angle.

Both kinds of couplers allow selective excitation of particular waveguide modes. TE or TM modes are determined by setting the polarization of the incident beam parallel to or perpendicular to the waveguide plane. The particular effective index N_j is determined by varying the coupling angle after having chosen a prism index n_p or periodic modulation spacing λ_s. In output coupling, a waveguide with one or more modes excited will cause beams of light to be coupled out at angles corresponding to the effective indices of the excited modes.

Scattering

Inhomogeneities in the waveguide material will scatter the guided light and thereby attenuate the waveguide modes. Such scattering may put light into other guided modes or into radiation modes. For waveguides with sharp boundaries such as the uniform slab, a distinction is made between scattering centers located at the material interfaces (surface scattering) and scattering centers distributed throughout the waveguide material (bulk scattering). In both cases the attenuation is a function of N_j because the different spatial distributions of light in the guide have differing overlap of the regions containing scattering centers.[8,11]

For TE modes in a symmetric slab guide with lossless surroundings, the bulk loss α_b is given by

$$\alpha_b = \alpha_{bo} \frac{n_g \sqrt{N_j^2 - n_0^2}}{N_j} \frac{\sqrt{N_j^2 - n_0^2} + K_0 d(n_g^2 - n_0^2)}{(n_g^2 - n_0^2)(1 + K_0 d\sqrt{N_j^2 - n_0^2})}, \quad [15]$$

where α_{bo} is the bulk scattering coefficient of the waveguide material. If the scattering itself is wavelength dependent, α_{bo} will also vary with N_j. For example, if the attenuation is due to Rayleigh scattering, α_{bo} is proportional to $(N_j)^{-4}$. Eq. [15] shows that when α_{bo} is constant, α_b is maximum at the lowest order modes ($j = 0$), and decreases rapidly at the highest modes, where most of the light is in the evanescent field and out of the scattering material.

The surface loss α_s for TE modes of a symmetric slab waveguide is given by

$$\alpha_s = \alpha_{so} \frac{(n_g^2 - N_j^2)\sqrt{N_j^2 - n_0^2}}{N_j(n_g^2 - n_0^2)(1 + Kd\sqrt{N_j^2 - n_0^2})}. \quad [16]$$

If the surface-scattering coefficient α_{so} is constant, α_s is seen to approach zero for the lowest-order mode as $N_{j=0}$ approaches n_g. The increase in loss for higher mode numbers comes from the increasing optical field at the scattering interface. For the highest-order modes the scattering goes again to zero, since the light is distributed over a large volume in the evanescent field.

The dependence of bulk and surface scattering on N in multimode waveguides is sketched in Fig. 4. The total scattering is the sum of the two contributions, and this of course is what is determined when the attenuation of the various modes is measured. In many cases it is possible to resolve the total scattering into the bulk and surface components. This is based on the observation that (1) the surface scattering is zero and the factor multiplying α_{bo} is unity for $N = n_g$ and (2) the bulk scattering is nearly constant except near $N = n_o$, so that the rate of increase in scattering for $N < n_g$ is proportional to the surface scattering. The second condition fails if either α_{bo} or α_{so} depends on N. A more detailed discussion of waveguide scattering has been published elsewhere.[12]

Liquid-Crystal Waveguides

Optical waveguides have been made of nematic liquid-crystal layers. There is no apparent reason why light could not be guided in other mesophases as well, but to date this has not been reported, and the subsequent discussion is restricted to nematics. Since the liquid-crys-

tal layer must be confined on both surfaces and since the bounding material can easily be identical on both sides, the symmetric slab model is appropriate for such waveguides. The high birefringence of nematics makes it necessary that the materials be well aligned and free from domains of differing orientation. The resulting uniaxial layer is of course different from the isotropic materials of the previ-

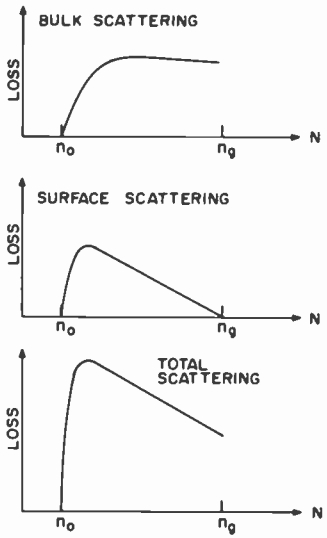


Fig. 4—Dependence of scattering loss on attenuation for multimode slab waveguides.

ous discussion. However, if propagation and polarization directions are restricted to the principle axes, the theory of the isotropic case is applicable if the appropriate value of n_g is chosen for the particular mode polarization (TE or TM) involved.

The technique used for coupling light into the liquid-crystal waveguide must be compatible with the need to confine the layer on both surfaces. We have found it possible to combine the optical coupling prism and one bounding surface into a unitized structure. Alternatively, light may be coupled into a solid-film waveguide that terminates at the liquid crystal, but this technique makes it difficult to selectively excite specific modes in the liquid-crystal layer.

Fig. 5 shows some coupling arrangements we have used. In Fig. 5a the prism with index $n_p = 1.95$ is bonded to low-index glass or fused quartz of index $n_o = 1.48$–1.51. The base of the structure is optically polished. The laser beam is focused just to the left of the prism cor-

ner. Leaky modes excited in the liquid-crystal film under the high-index prism become true guided modes after passing into the film bounded by low-index material. A second prism bonded on the other end of the low-index glass may be used to couple the guided light out again.

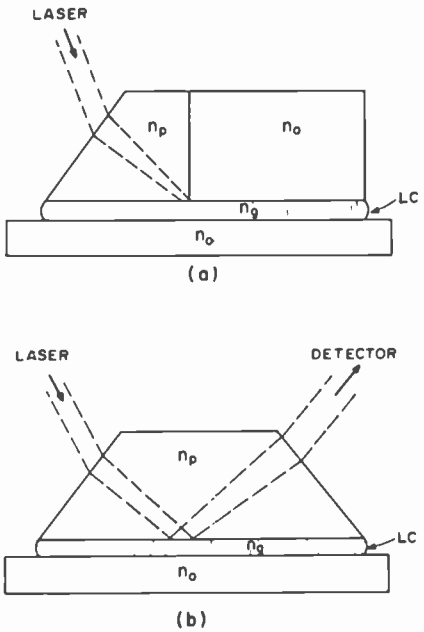


Fig. 5—Prism coupler for exciting (a) guided or (b) leaky modes in liquid-crystal (LC) layers.

A leaky wave coupler is shown in Fig. 5b. Here light coupled into guided modes of the liquid-crystal layer is directly coupled out again. This structure is useful for measuring the mode structure of the liquid-crystal film by measuring the discrete angles for which light rays are coupled into and out of the particular waveguide modes.

The attenuation by scattering of light guided in the arrangement of Fig. 5a is determined by observing the intensity of light scattered from the guided modes into radiation. The streak of scattered light is imaged on a slit mounted in front of a photomultiplier, and this assembly is translated to record the decrement of scattered light intensity as a function of distance from the input coupling point.

Fig. 6 shows attenuation data obtained at room temperature for

TE and TM modes in guides of 6- to 12-μm thickness. The liquid-crystal layers were MBBA aligned with lecithin to have optic axis perpendicular to the bounding surface. For TM modes, $n_g = n_{ext} = 1.75$, while for TE modes, $n_g = n_{ord} = 1.54$. Each data point represents a separate guided mode.

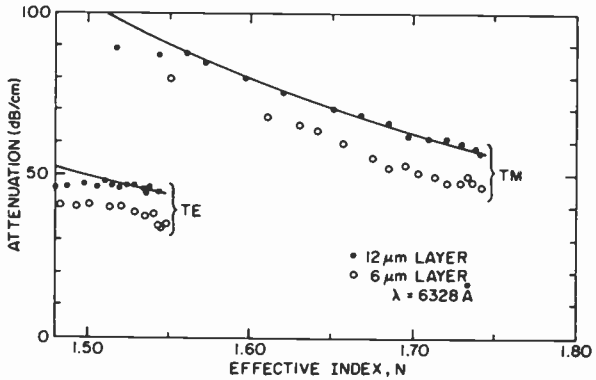


Fig. 6—Attenuation measured for nematic waveguides consisting of MBBA layers with hemeotropic alignment.

That the measured attenuation is due to scattering rather than absorption was established from the absence of significant loss when the liquid crystal was heated above the nematic–isotropic phase transition. The waveguide layers were well aligned, homogeneous, and free from disclinations. The scattering results from thermal fluctuations in molecular order[13] over volumes comparable to optical wavelengths.

As described previously, the scattering of the lowest order modes is equal to the bulk scattering coefficient of the waveguide material. This was in the range 35–45 dB/cm for TE modes and 45–60 dB/cm for TM modes. According to Eq. [15], the bulk scattering strength α_b is independent of waveguide thickness d in the limit $N_j \rightarrow n_g$. The differences in attenuation must be attributed to differences in α_{bo} due, perhaps, to different sample purities or to experimental inaccuracy. The measured attenuation in all samples measured, including samples with molecular alignment parallel to the waveguide plane, was much greater than that reported by Sheridan et al.[2] The very low attenuations they report (less than 1 dB/cm) could not be reproduced in any of our experiments.

The TM mode attenuation increases with increasing mode number. The rate of increase with decreasing N_j is seen to be independent of

sample thickness. Surface scattering strength increases with mode number, but is inversely proportional to the waveguide thickness (see Eq. [16]). Since the thickness dependence is not observed, the mode dependence of the TM scattering is attributed to wavelength dependence in the bulk scattering. The solid curve was fitted to the TM_0 attenuation measured for the 12-μm guide, and then scaled by the factor $(N_j/N_o)^{-4}$ to represent the wavelength dependence of Rayleigh scattering. The fit to the data is obviously very good. Further work is needed, however to understand in detail the scattering mechanisms. It may be particularly rewarding to try and resolve any surface scattering effects resulting from various techniques of molecular alignment.

Electro-optic effects were produced in nematic waveguides by applying pulsed electric fields. In one series of experiments, interdigital electrodes were deposited on one of the bounding surfaces and driven with voltage pulses. The light coupled out of the guide downstream from the electrodes was monitored with a photomultiplier and the response viewed on an oscilloscope.

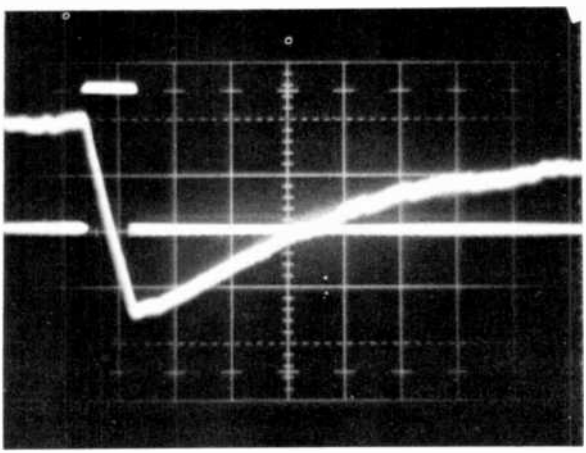


Fig. 7—Transmission of light in nematic waveguide is diminished in response to a pulsed electric field (rectangular pulse) and recovers when field is off (drive voltage of 1-msec duration and 50-volt amplitude).

Fig. 7 shows a voltage pulse (1 msec duration and 50 V amplitude) and its effect on waveguide propagation. Transmission drops during the pulse duration and recovers exponentially afterwards. The light transmission may be completely turned off by this process. Response

times T_{off} and recovery times T_{on} are shown in Fig. 8. It is presumed that the modulation mechanism is refraction of light out of the waveguide by the index modulation caused by molecular rotation in response to the applied electric field.

As in the case of the attenuation, a more detailed investigation of these phenomena is warranted. For example, the variation of the re-

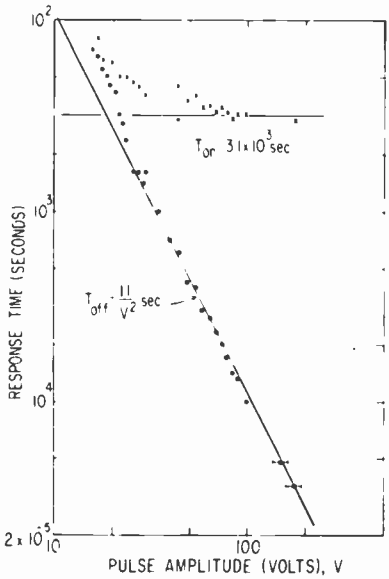


Fig. 8—Response times to application of field T_{off} and recovery times after removal of field T_{on} for waveguide transmission shown in Fig. 7.

covery time T_{on} with pulse voltage is quite unexpected, and suggests that complex patterns of molecular rotation may be excited by the interdigital electrodes. Such effects are relevant to display-device technology, and optical-waveguide techniques appear to be an advantageous way of studying them.

Dynamic scattering has also been used to modulate guided light. In this experiment a liquid-crystal layer was established on the surface of a passive thin-film waveguide. Electrically excited turbulance scattered light in the evanescent field of the thin-film waveguide, which extended into the liquid crystal. The extinction of waveguide light as a function of ac and dc dynamic-scattering excitation voltages is seen in Fig. 9. Very high voltages are necessary, since only a small part of the optical field interacts with the liquid crystal.

Conclusions

Optical-waveguide propagation in liquid-crystal layers has been achieved, and basic electro-optic device phenomena have been demonstrated. The usefulness of such devices has yet to be established, particularly in view of the considerable progress being made with other active waveguides, such as single-crystal electro-optic films.[14] Liquid-crystal materials offer the advantages of high electro-optic index changes and a variety of birefringence and scattering operation

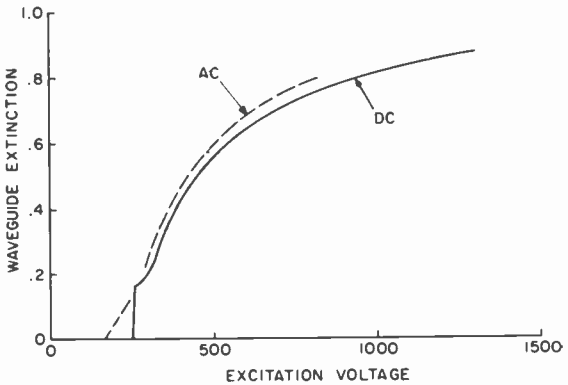


Fig. 9—Transmission reduction in thin-film waveguide when dynamic scattering is excited in nematic overlay.

modes. They are handicapped at present by slow response and recovery speeds, high scattering loss, and the need for confinement in sealed cells. Possibly the most fruitful area of application may be for side-illuminated display panels, should useful materials with low scattering loss be developed.

Optical waveguiding may be most useful as a technique for studies of liquid-crystal phenomena. Accurate measurements of the mode spectrum can be used as a probe of transient changes in the refractive index profile of the liquid-crystal layer. Light scattering into radiation modes and between guided modes determines the spatial frequency spectrum of the alignment inhomogeneities and fluctuations. Scattering between closely spaced modes is a sensitive measure of very-small-angle forward scattering. These same techniques may be useful in studying physical phenomena in other liquid layers as well as liquid crystals.

References:

[1] D. J. Channin, "Optical Waveguide Modulation Using Nematic Liquid Crystal," *Appl. Phys. Lett.* **22,** p. 365 (1973).

[2] J. P. Sheridan, J. M. Schnur, and T. C. Glallorenzi, "Electro-Optic Switching in Low-Loss Liquid Crystal Waveguides," *Appl. Phys. Lett.*, **22,** p. 561 (1973).

[3] J. P. Sheridan, "Liquid Crystals in Integrated Optics," OSA Topical Meeting on Integrated Optics, New Orleans, Jan., 1974.

[4] Chenming Hu, John R. Winnery, and Nabil M. Amer, "Optical Deflection in Thin-Film Nematic-Liquid-Crystal Waveguides," *IEEE J. Quan. Elect.*, **QE-10,** p. 218 (1974).

[5] H. A. Weakliem, D. J. Channin, and A. Bloom, "Determination of Refractive Index Changes in Photosensitive Polymer Films by an Optical Technique," to be published.

[6] P. K. Tien, "Light Waves in Thin Films and Integrated Optics," *Appl. Opt.*, **10,** p. 2395 (1971).

[7] Dietrich Marcuse, ed., *Integrated Optics*, IEEE Press, New York (1973).

[8] Dietrich Marcuse, *Light Transmission Optics*, Van Nostrand Reinhold Co., New York (1972).

[9] P. K. Tien, R. Ulrich, and R. J. Martin, "Modes of Propagating Light Waves in Thin Deposited Semiconductor Films," *Appl. Phys. Lett.*, **14,** p. 291 (1969).

[10] M. L. Dakss, L. Kuhn, P. F. Heidrich, and B. A. Scott, "Grating Coupler for Efficient Excitation of Optical Guided Waves in Thin Films," *Appl. Phys. Lett.*, **16,** p. 523 (1970).

[11] J. Kane and H. Osterberg, "Optical Characteristics of Planar Guided Modes," *J. Opt. Soc. Am.*, **54,** p. 347 (1964).

[12] D. J. Channin, J. M. Hammer, and M. T. Duffy, "Scattering in ZnO-Sapphire Optical Waveguides," to be published in *Applied Optics*.

[13] P. G. deGennes, *The Physics of Liquid Crystals*, Oxford University Press, London (1974).

[14] J. M. Hammer and W. Phillips, "Low-Loss Single-Mode Optical Waveguides and Efficient High-Speed Modulators of $LiNb_xTa_{1-x}O_3$ on $LiTaO_3$," *Applied Phys. Lett.*, **24,** p. 545 (1974).

Lyotropic Liquid Crystals and Biological Membranes: The Crucial Role of Water

Peter J. Wojtowicz

RCA Laboratories, Princeton, N. J. 08540

Abstract—The composition and structure of lyotropic liquid crystals and biological membranes is reviewed. An understanding of the stability of these structures is obtained in terms of the hydrophobic and hydrophilic interactions of the constituent amphiphilic molecules with the aqueous solvent. It is emphasized that it is the unique properties of the water itself that gives rise to the crucial interactions that stabilize lyotropic liquid crystals and biological membranes.

1. Introduction

No collection of papers on liquid crystals would be complete without some discussion of lyotropic liquid crystals and biological membranes. At the present time these two subjects constitute areas of intensive research effort providing a literature of rapidly increasing size. The high current interest mandates a discussion of these topics, but at the same time makes it very difficult to select the material to be presented in the limited space available. The discussion in this paper will therefore be confined to two main topics, (1) the composition and structure of lyotropic liquid crystals and biological membranes and (2) the nature of the principal interactions that give rise to their existence and stability.

Lyotropic liquid crystals and biological membranes are similar to the thermotropic liquid crystals described in previous papers in that

they are fluid phases that possess considerable molecular order. They are quite different, however, in that they are necessarily systems of two or more components being composed of large organic molecules dissolved in a highly polar solvent, most often water. They are also different because it is not the intermolecular interaction between the molecules partaking in the order that is responsible for the formation of the basic structures of lyotropic liquid crystals and biological membranes. Rather it is the interaction of the organic molecules with the aqueous solvent that is most crucial in providing the stability of these ordered phases. This interaction with the water, the so-called *hydrophobic* interaction, was first recognized in the study of the solubility of simple hydrocarbons in water.[1] The principles involved were then successfully applied to the elucidation of the native conformation of the complex proteins.[2] Very recently these principles have also aided in the understanding of the structure and function of biological membranes.[3] The importance of the hydrophobic interaction, however, has not yet been fully appreciated in the case of lyotropic liquid crystals.[4] One of the intentions of this article, therefore, is to help call attention to the decisive role of the hydrophobic interaction in providing the stability of lyotropic liquid crystals.

The following sections of this paper will briefly review the composition and structure of lyotropic liquid crystals and biological membranes. We will then consider the different interactions that are important in determining the structure of these phases. The hydrophobic interaction will be examined in some detail; the properties of the water itself will be shown to be the predominant driving force leading to the existence of lyotropic liquid crystals and biological membranes. Finally, we will discuss an unusual characteristic of these ordered phases. Unlike most of the other ordered phases encountered in physics and chemistry, the lyotropics and the membranes derive their stability not from a competition between the energy and entropy, but from a competition between two different kinds of entropy.

2. Lyotropic Liquid Crystals

The literature on the composition and structure of lyotropic liquid crystals has been extensively reviewed by Winsor.[5] A somewhat shorter review has been given by Brown, Doane and Neff.[6] Our discussions in this section will be based principally on these two papers; both are recommended to the reader seeking additional information or further detail.

2.1 Constituents of Lyotropics

Lyotropic liquid crystals are chemical systems composed of two or more components. Specifically, they are mixtures of amphiphilic compounds and a polar solvent, most frequently water. *Amphiphilic* compounds are characterized by having in the same molecule two groups that differ greatly in their solubility properties. One part of the molecule will be *hydrophilic*, highly soluble in water or other polar solvents, while the other portion will be *lipophilic*, highly soluble in hydrocarbon or nonpolar solvents. In the context of our following discussions of the interactions responsible for the stability of lyotropics it is perhaps more proper to call these latter groups *hydrophobic*, emphasizing their insolubility in water rather than their hydrocarbon solubility. Typical hydrophilic groups are -OH, $-CO_2H$, $-CO_2Na$, $-SO_3K$, $-O(CH_2-CH_2-O)_nH$, $-N(CH_3)_3Br$, and $-PO_4-CH_2CH_2-NH_2$. Typical lipophilic or hydrophobic groups are $-C_nH_{2n+1}$, $-C_6H_4-C_nH_{2n+1}$, and any other radicals containing long hydrocarbon chains, with or without aromatic rings included. An example of an amphiphilic compound that has been studied extensively is sodium laurate, whose molecular structure is displayed in Fig. 1a. The hydrophilic portion is the carboxylic acid group (shown ionized as it is in aqueous solution), while the hydrophobic part is the long straight chain hydrocarbon group. In describing the structures of aggregates of such molecules it is convenient (as shown in Fig. 1a) to represent the hydrophilic "head" by a black dot and the hydrophobic "tail" by the zig-zag line.

Depending on the relative strengths of the hydrophilic and lipophilic tendencies of the two different parts of the molecule, amphiphilic compounds can vary widely in their solubility behavior. They can be predominantly hydrophilic, water soluble and hydrocarbon insoluble (such as $C_nH_{2n+1}-CO_2K$ with $n = 1,2,3$) or predominantly lipophilic, hydrocarbon soluble and water insoluble (such as $C_nH_{2n+1}-OH$ with $n > 12$). The most striking amphiphilic properties (solubilization, formation of micelles, formation of liquid crystals) occurs when the hydrophilic and hydrophobic tendencies are both strong but evenly balanced (such as in $C_nH_{2n+1}-CO_2Na$ with $n = 8$ to 20).

Because of their dual characteristics, amphiphilic compounds are capable of displaying remarkable solubility properties. They are soluble in both water and hydrocarbons, and show strong co-solvent or solubilization effects. Moderately concentrated soap solutions, for example, can dissolve many different kinds or organic compounds that will not dissolve in water alone. Similarly, some amphiphilic compounds (such as Aerosol OT) dissolved in hydrocarbon can solubilize

water or other polar compounds that do not ordinarily dissolve in hydrocarbons. The proper way to describe both situations is to say that the amphiphilic compound acts as co-solvent for both the water and the hydrocarbon.

2.2 Micelles

At extreme dilution, amphiphilic molecules are distributed randomly and uniformly throughout the solution. As the concentration of the

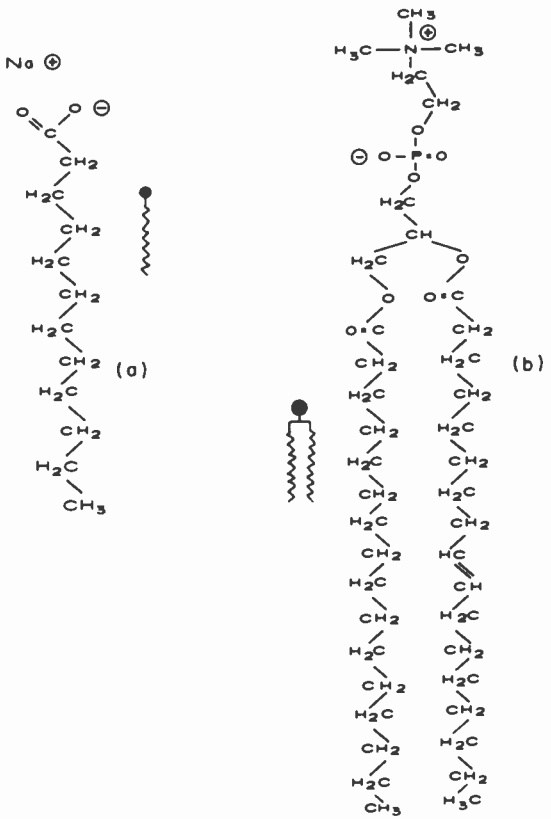


Fig. 1—(a) Molecular structure of a typical amphiphilic molecule, sodium laurate. (b) Molecular structure of a typical lipid found in membranes, phosphatidylcholine. (part b after Ref. [10]).

amphiphilic molecules is increased, however, aggregates of molecules begin to form. Groupings of molecules called micelles arise in which like is associated with like. Fig. 2 displays the structure of spherical

and cylindrical micelles. In both forms, the hydrophilic heads are associated with each other on the outer periphery of the aggregates, while the hydrophobic tails are grouped together in the fluid-like interior of the micelles. Of even greater significance, however, is the observation that the hydrophilic heads are placed in close association with the aqueous solvent, while the hydrophobic tails are sequestered in the interior of the micelles completely out of contact with the water.

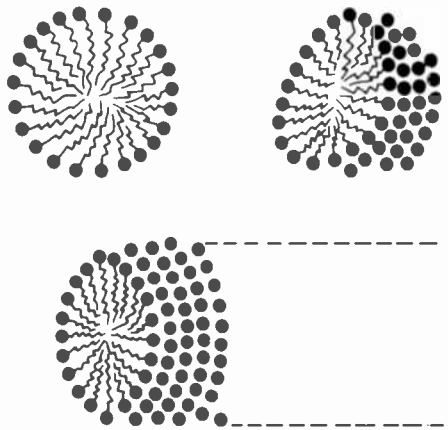


Fig. 2—Schematic representation of the structures of spherical (upper drawings) and cylindrical (lower drawing) micelles.

Micelles are not entities composed of fixed numbers of molecules having a fixed geometrical shape. They must be regarded as statistical in nature, in equilibrium with the surrounding amphiphilic molecules, and fluctuating constantly in size and shape in response to temperature. On dilution of the mixture, micelles dissociate rapidly, while on concentrating the solution, more extended micellar structures appear, eventually forming the many different lyotropic liquid-crystal phases.

The observed structure of micelles permits the rationalization of the solubilization property of aqueous solutions of amphiphilic compounds. The interior of the micelles can be thought of as small pockets of essentially pure liquid hydrocarbon. The interiors of the micelles, the pockets of hydrocarbon, are then capable of dissolving other lipophilic or hydrophobic molecules added to the solution. The action of soap in cleansing materials of oily or greasy dirt or soil is precisely of this nature.

2.3. Structures of Lyotropics

As the proportion of amphiphilic compound to water increases, an impressive variety of different lyotropic liquid crystal phases are observed. A systematic classification of the different types and their structure is presented in References [5] and [6] to which the reader is referred for the extensive details. In this paper we shall only attempt to give the flavor of the situation.

One type of lyotropic structure is the so-called "isotropic" phase. In this phase spherical micelles form the basic unit of the liquid crystal structure. These are then deployed in either a face-centered cubic or body-centered cubic arrangement within the fluid aqueous medium. Although the micelles are essentially arranged on a lattice, they do not touch and the structure is not rigid as in a true crystal. The presence of the intervening aqueous solvent provides sufficient fluidity so that this structure (as well as the others to be described below), though highly ordered, is still properly classified as a liquid. Another broad class of structures is the so-called "middle" phase. In this phase the basic structural unit is the rod-like cylindrical micelle of essentially infinite length. These are then disposed in a hexagonal arrangement within the aqueous medium. Related to this category are the phases composed of rod-like micelles of rectangular cross section arranged in square or rectangular packings. In all of these structures the common feature is, of course, the sequestering of the hydrophobic tails away from contact with the water while allowing the hydrophilic heads to reside within the aqueous solvent.

When the proportion of water to amphiphilic compound becomes low, structures of the "reversed" type occur. The basic units here are spherical or cylindrical reversed micelles in which the hydrophobic tails are on the outside, while the hydrophilic heads line the interior which contains the water. Liquid crystalline arrangements of the reversed micelles then occur in the several ways outlined above. In all cases the hydrophobic tails on the exterior of the reversed micelles are in contact only with each other; the water is sequestered in the polar interiors.

Among the most interesting of the lyotropic phases are the "lamellar" structures. The most important of these is the so-called "neat" phase, whose structure is depicted in Fig. 3. In this structure we find the amphiphilic molecules arranged in double layers of essentially infinite extent in two dimensions. The internal structure of the double layer is such that the hydrophobic tails occupy the interior out of contact with water, while the hydrophilic heads line the exterior in

contact with the fluid aqueous medium (regions denoted by A in Fig. 3). The double layers then stack periodically along the third dimension alternating with layers of the aqueous solvent. The thickness of the double layer is somewhat less than twice the length of the amphiphilic molecules (layer thickness is thus about 30 to 40 Å). The thickness of the intervening aqueous layers is about 20 Å. The hydrocarbon region within the double layers is essentially fluid. Some experiments, however, show a gradual transition from a fluid-like property of the hydrocarbon chains to a more rigid type of behavior at lower temperatures.[7] The overall structure of the neat phase is rather analogous to that of the smectic liquid crystals.

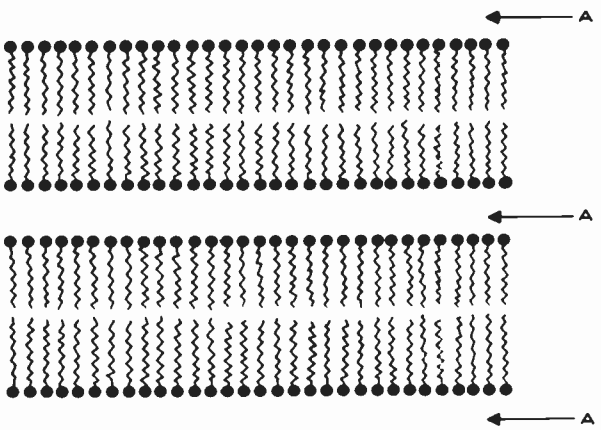


Fig. 3—Schematic representation of the structure of a portion of the neat phase. The regions containing the aqueous solvent are denoted by the A's.

Transitions from one kind of phase to another occur with changes in both the temperature and the concentration of the amphiphilic compound in water. A typical phase diagram is that for sodium laurate shown in Fig. 4. The T_c line gives the temperatures of transition from solid to liquid or liquid crystalline phases. It can be thought of as representing the depression of the melting point of sodium laurate by the water. The T_i line gives the temperatures of transition from liquid crystalline to normal liquid phases. The regions of stability of the "neat" and "middle" phases are shown. The cross-hatched areas are the regions of stability of the "isotropic" structures. The intermediate regions labelled *I* are most probably conjugate mixtures of the adjacent stable phases.

3. Biological Membranes

Studies of the structure and functions of biological membranes currently constitute an area of very active and intensive research. The literature on this subject is extensive and continually growing. The material selected for this section is primarily taken from articles by Rothfield[8] and by Singer.[3] The reader interested in further information is encouraged to read these articles as well as others appearing in the same volume. Also recommended are the more popular articles by Singer[9] and by Capaldi,[10] plus the recent review by Singer.[11]

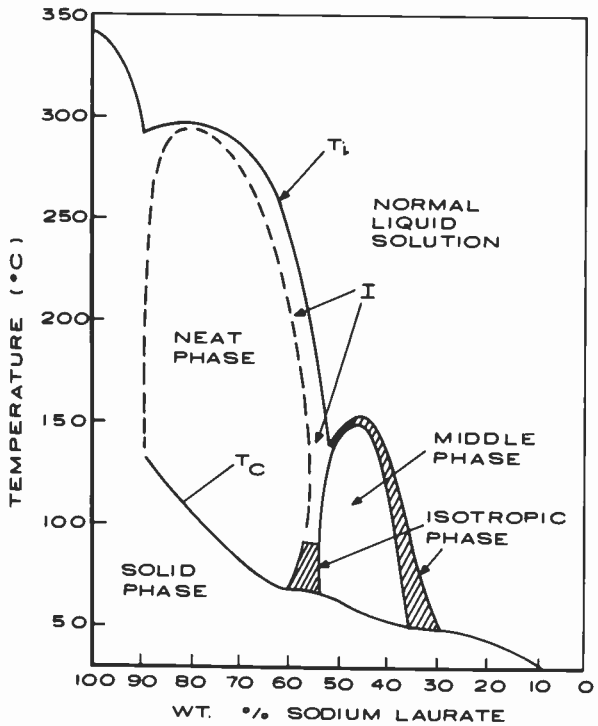


Fig. 4—Temperature-composition phase diagram of the sodium-laurate–water system (after Refs. [5] and [6]).

In spite of the extensive knowledge that has been obtained, it is still difficult to define a biological membrane. Rothfield,[8] however, has provided a particularly succinct description. Reproduced verbatim it reads: "Biological membranes are continuous structures separating two aqueous phases. They are relatively impermeable to water-

soluble compounds, show a characteristic trilaminar appearance when fixed sections are examined by electron microscopy, and contain significant amounts of lipids and proteins." We shall enlarge on this description in the following two sections.

3.1 Constituents of Membranes

The principal constituents of all biological membranes are lipids, proteins, and oligosaccharides. The aqueous environment surrounding the membrane should properly also be considered one of the major components of membranes. The constituents that give the membranes their primary structure (and account for approximately half their mass) are the lipids. A bewildering diversity of lipids is found in membranes, and any one membrane will contain several different lipids. The lipids observed to be present belong to a variety of classes including phosphatidylcholine, phosphatidylethanolamine, sphingomyelin, glycolipids, cholesterol, etc.

The molecular structure of one example of lipid, phosphatidylcholine, is shown in Fig. 1b. The molecule is obviously recognized to be amphiphilic. At the top of the drawing we find the hydrophilic "head." In the aqueous environment of biological systems, the head is ionized as shown; the zwitterion (hybrid ion charged both positively and negatively) in this case is composed of phosphate and trimethylamine groups. In some lipids the hydrophilic head may, however, be an un-ionized or neutral group. Below the head is the glycerol group (sometimes called the "backbone"). Attached to the backbone are the twin hydrophobic "tails." The tails consist of long hydrocarbon chains. In any one class of lipids the chains will appear in many different lengths and several degrees of saturation (number of carbon-carbon double bonds). Their means of attachment can consist of a variety of covalent linkages to the glyceryl phosphate moiety. In discussing the structure of biological membranes it is convenient (as shown in Fig. 1b) to represent the hydrophilic head as a large dark dot and the twin hydrophobic tails as a pair of zig-zag lines.

The second major constituent of biological membranes (accounting for approximately half their mass) is the proteins. As with the lipids, a perplexing variety of different proteins are found. Membranes from different sources are observed to contain large numbers of proteins of different molecular weights, ranging from less than 15,000 to over 100,000. No single type of protein dominates. While the complete role of the proteins in the total function of the membranes is still not completely understood, it does seem clear that they are of lesser consequence in determining the primary structure to be described below.

We will, therefore, not give the proteins any further consideration here; the same holds true for the oligosaccharides.

3.2 Structure of Membranes

In spite of the significant heterogeneity in the molecular structure of lipids occurring in biological membranes, one important property is common: all membrane lipids are strongly amphiphilic. Because of this, one should expect that in the aqueous environment of biological systems these molecules will aggregate into structures similar to the lyotropic liquid crystals described above. This is indeed the case. Aqueous solutions of phospholipids and synthetic mixtures of naturally occurring lipids dissolved in water display various forms of liquid-crystal behavior.

The most common structure observed is the so-called "bimolecular leaflet" or phospholipid bilayer. The structure is analogous to the neat phase of the lyotropic liquid crystals and is shown schematically in Fig. 5. The most prominent feature of this structure is the arrange-

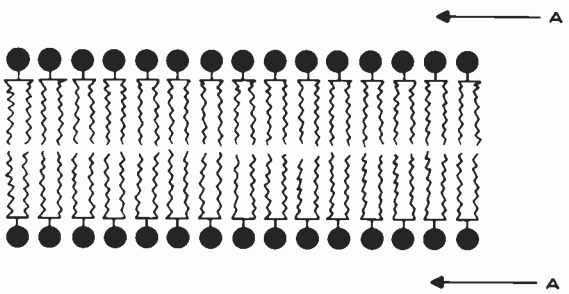


Fig. 5—Schematic representation of the structure of a portion of the lipid bilayer component of biological membranes. The regions containing the aqueous environment are denoted by the A's.

ment of the lipids into two contiguous layers of essentially infinite extent, such that the hydrophobic tails occupy the interior of the bilayers out of contact with the aqueous environment (regions denoted by A in Fig. 5), while the hydrophilic heads line the exterior in contact with the water. The thickness of the bilayer is approximately twice the length of the individual lipid molecules (40–50 Å).

The bimolecular leaflet is not a rigid structural entity. In keeping with its liquid-crystalline nature, the hydrocarbon interior of the bi-

layer is quite fluid. The fluidity or mobility can be viewed simply as random movements of the individual hydrocarbon chains of the molecules in the bilayer. In general, mobility is favored at higher temperatures, by greater degrees of unsaturation of the hydrocarbon chains, and by shorter chain lengths. Gradual transitions from more-fluid states to less-fluid states are seen in several experiments when the temperature is lowered. The implications of this fluidity (and its changes) to the biological function of membranes are many but cannot be examined here.

The structure of actual biological membranes cannot be as simple as the phospholipid bilayer. Membranes do, after all, contain large numbers of proteins of varying weights, shapes, and sizes. One of the major problems of membrane biology is the question of the detailed incorporation of both the lipids and the proteins into the overall membrane structure. The experimental situation is very difficult and no clearly observed membrane structure is available. Many models, however, have been proposed. The basic philosophy is that a general pattern of organization exists and that the heterogeneity and distinctiveness of different membranes can be understood as variations on a common structural theme. A particularly appealing model of this kind has been presented by Singer and others.[3,9,10]

The model is variously called the Lipid-Globular Protein Mosaic model or the Fluid Mosaic model. The basic structure of the membrane is assumed to be the bimolecular leaflet of lipid molecules. The leaflet is not considered to be continuous, however. The globular integral proteins and patches of the lipid bilayer are assumed to be arranged in an alternating mosaic pattern throughout the membrane. The hydrophobic portion of the lipids and a large fraction of the nonpolar amino acid residues of the proteins are sequestered from contact with water, mainly in the interior of the membrane. The hydrophilic groups of the lipids and the ionic residues of the proteins are in direct contact with the aqueous environment on the exterior of the membrane. Some of the proteins lie predominantly near the surface of the membrane, others penetrate the interior. Some of the latter may penetrate a short distance into the inside; others may extend clear through the entire thickness of the membrane. The saccharide components, being hydrophilic, presumably reside on the surface in direct contact with the aqueous environment.

The Lipid-Globular Protein Mosaic model is based on experimental data on the conformation of proteins in intact membranes and on general thermodynamic considerations (maximization of hydrophilic interactions and minimization of hydrophobic interactions of the lip-

ids and proteins with the water). At the present time, there do not seem to be any data or experimental results that are clearly inconsistent with this model. While there is not yet any direct experimental evidence for this structural pattern, its consistency and thermodynamic feasibility recommend this model as a working hypothesis for further investigation.

4. Interaction of Amphiphilic Compounds with Water

Lyotropic liquid crystals and biological membranes are ordered assemblies of amphiphilic molecules situated in an aqueous environment. By analogy with the other ordered or condensed phases encountered in physics and chemistry, one might suspect that the stability of lyotropics and membranes derive from favorable attractive interactions between constituent amphiphilic molecules, and that the water only serves to provide the medium in which the ordered aggregates can reside. This notion is, however, incorrect. The participation of the water is far from passive. The role of the water is, in fact, crucial to the formation and stability of lyotropics and membranes. The chief mechanism by which the water acts to promote the various ordered structures is the *hydrophobic effect*.

Simplistically stated, the hydrophobic effect may be defined as the tendency of water to reject any contact with substances of a nonpolar or hydrocarbon nature. The existence of this effect was first recognized in the study of the extremely low solubility of hydrocarbons in water.[1] The principles involved were later successfully applied to the elucidation of the native conformation of protein molecules by Kauzman.[2] The application of these ideas to the study of membrane structures has been advanced by Singer.[3] Recently, Tanford[4] published an entire book on the hydrophobic effect, including the influence of this interaction on the formation of micelles, lipid bilayers, membranes and other ordered structures. Aside from Singer's[3] and Tanford's[4] statements on the decisive role of the hydrophobic effect on lyotropics, the lyotropic liquid-crystal literature seems peculiarly unaware of this phenomenon. Winsor's[5] extensive review with its systematic analysis (R-theory) of the many lyotropic phases does not take the hydrophobic effect into account. More recent reviews[6,7] of lyotropic liquid crystals do not mention the phenomenon. We hope that the present discussion will help to advance the realization of the importance of the hydrophobic effect to lyotropics. The material of the following sections is taken chiefly from Ref. [3] with some assistance from Refs. [2] and [4].

4.1 Solubility of Hydrocarbons in Water

Most of the thermodynamic data concerning the hydrophobic interaction comes from studies of the solubility of hydrocarbons in water. An examination of these data provides an understanding of the nature and magnitude of the hydrophobic effect. The knowledge gained here can then be qualitatively applied to the more complex systems of interest.

The thermodynamics of solute–solvent interactions is most conveniently described in terms of *unitary* quantities.[12] The unitary free energy and unitary entropy changes accompanying some process (such as the transfer of hydrocarbon from nonpolar solvent to water or the transfer of hydrocarbon from pure hydrocarbon to water) are the standard free-energy and entropy changes corrected for any translational entropy terms (the *cratic* entropy) that are not intrinsic to the interaction under consideration. The cratic entropy is simply the entropy of mixing the solute and solvent into an ideal solution. With the cratic contribution removed, the unitary free energy and entropy contain only contributions to the thermodynamics of the process that come from the interaction of the individual solute molecules with the solvent.

Table I—Thermodynamic Changes in the Transfer of Hydrocarbons from Nonpolar Solvents to Water at 25°C (After Ref. [7])

Process	ΔF_u (cal/mole)	ΔH (cal/mole)	ΔS_u (cal/mole °K)
CH_4 in benzene → CH_4 in H_2O	+2600	−2800	−18
CH_4 in ether → CH_4 in H_2O	+3300	−2400	−19
CH_4 in CCl_4 → CH_4 in H_2O	+2900	−2500	−18
C_2H_6 in benzene → C_2H_6 in H_2O	+3800	−2200	−20
C_2H_6 in CCl_4 → C_2H_6 in H_2O	+3700	−1700	−18
C_2H_4 in benzene → C_2H_4 in H_2O	+2920	−1610	−15
C_2H_2 in benzene → C_2H_2 in H_2O	+1870	−190	−7
Liq. propane → C_3H_8 in H_2O	+5050	−1800	−23
Liq. n-butane → C_4H_{10} in H_2O	+5850	−1000	−23

Table 1 contains a compilation of thermodynamic data for the transfer of simple hydrocarbons from nonpolar solvents to water. The unitary free energy, enthalpy, and unitary entropy changes are denoted by ΔF_u, ΔH and ΔS_u, respectively; the temperature is 25°C in all cases. ΔF_u is unfavorably positive in all examples. This is in keeping with the empirical observation that hydrocarbons do not dissolve in

water to any appreciable extent. The ΔH are, however, exothermic, demonstrating that the intermolecular interaction energies favor the solution of hydrocarbon molecules in water. The unfavorable positive ΔF_u can therefore only arise because of a strongly unfavorable unitary entropy ($\Delta F_u = \Delta H - T\Delta S_u$). Table 1 shows that this is indeed the case. For all examples, ΔS_u is large and negative.

The low solubility of hydrocarbons in water is therefore a consequence of the large decrease in unitary entropy accompanying the introduction of such molecules into an aqueous environment. In their classic investigation, Frank and Evans[1] concluded that this large entropy decrease must be due to some kind of ordering of the water molecules around the hydrocarbon molecules dissolved in the water. The water molecules at the surface of the cavity created by the introduction of the hydrocarbon molecule must be capable of rearranging themselves in order to regenerate broken hydrogen bonds. In doing so, however, they create a higher degree of order than existed in the undisturbed water; the result of this ordering is the decrease in the unitary entropy. It is important to realize that it is the property of the water alone (its highly structured nature resulting from the considerable degree of hydrogen bonding) that is responsible for the hydrophobic interaction. The hydrophobic effect is relatively insensitive to the precise nature of the nonpolar solutes involved, and, furthermore, is not nearly so pronounced in the case of other polar solvents. Expressing the description of the hydrophobic effect in simple terms, water rejects contact with hydrocarbon and other nonpolar groups because not to do so would require the water to increase the local order of its structure, thereby reducing the entropy of the system.

4.2 Solubility of Ionized Species in Water

The interaction of the hydrophilic portions of amphiphilic molecules with the aqueous solvent is another important factor in the stabilization of the structures of lyotropics and membranes. Information on the hydrophilic interaction may be obtained from thermodynamic studies of the solubility of simple ionized molecules in water and other polar solvents.

Simple electrostatic arguments suggest that the free energy of ionized species is inversely proportional to the dielectric constant of the medium in which the ions reside. This should be so because ions interact more strongly with the molecules of polar solvents than with those of nonpolar solvents. This trend is indeed observed in thermodynamic data. Table 2 contains a compilation of the solubilities of a

typical charged molecule, the zwitterion of glycine, $^{+}H_3N\text{-}CH_2\text{-}CO_2^{-}$, in various solvents of decreasing polar character. Also included is the unitary free energy of transfer of this species from water to the other solvents. ΔF_u is seen to be strongly positive for all nonaqueous solvents. It is clear that this ionized molecule is at a much lower free energy in contact with water than with any other solvents.

Table 2—Solubility and Free Energy of Transfer of Glycine in Various Solvents at 25°C (After Ref. [3]).

Solvent	Solubility (mole/liter)	ΔF_u (cal/mole)
Water	2.886	—
Formamide	0.0838	1680
Methanol	0.00426	3430
Ethanol	0.00039	4630
Butanol	0.0000959	5190
Acetone	0.0000305	6000

The data in Table 2 reveal another of the unique properties of the solvent water. The interaction of the water with the ions is not simply attributable to the high dielectric constant; formamide has almost the same dielectric constant as water, yet it requires 1680 cal/mole of free energy to transfer glycine from water to formamide. This suggests that, as with the nonpolar solutes, ionic species induce significant changes in the local order of the water structure.

Another estimate of the importance of the hydrophilic interaction may be obtained from a consideration of the free-energy difference between the ionized species in water versus the uncharged species in nonpolar solvent. The thermodynamic data show that it is much more favorable for the molecules to exist as ionized species in water. For example, in the case of the carboxyl group, $-CO_2^{-}$, it takes a unitary free energy of 3300 cal/mole at 25°C just to protonate it (change it into the neutral group, $-CO_2H$) in water at pH 7. These considerations demonstrate that it is a thermodynamic necessity for the hydrophilic groups of amphiphilic molecules to be ionized and in direct contact with the aqueous environment.

4.3 Aggregation of Amphipnilic Compounds

The two fundamental thermodynamic principles described in the previous sections may now be applied to the question of the stability of lyotropic liquid crystals and biological membranes.

A very large cratic entropy is associated with the uniform dispersal

of amphiphilic molecules throughout the aqueous medium; at extreme dilution this contribution to the free energy will stabilize isolated molecules. But, as the concentration of amphiphilic species is increased, large amounts of negative unitary entropy are produced by the hydrophobic effect. The free energy of the system can then be more effectively minimized by having the amphiphilic molecules aggregate. The cratic entropy is lost, but far larger amounts of unitary entropy are gained by forming micelles, bilayers, or other lyotropic structures in which the hydrophobic groups are completely sequestered from contact with the water. We noted in previous sections that all the structures observed did indeed sequester the hydrophobic groups. These structures, moreover, also arranged all the hydrophilic groups onto the exterior of the aggregates in direct contact with the water and away from association with the nonpolar hydrophobic parts. Thus, the observed structures simultaneously satisfy both of the important thermodynamic requirements: (1) minimization of the hydrophobic interaction and (2) maximization of the hydrophilic interaction with the aqueous solvent.

This interpretation is consistent with experiments involving the addition of nonaqueous solvents to aqueous solutions of amphiphilic compounds. The addition of 20 to 30 mole % of ethanol to solutions of amphiphilic compounds significantly reduces the stability of micelles; isolated molecules become much more favored. The primary effect of the ethanol is to reduce the hydrophobic effect, thereby diminishing the stability of aggregates. In addition, conductance measurements show that the net charge on the amphiphilic molecules is decreased in 30% ethanol because of formation of ion pairs (with the small cations) in the medium of lower dielectric constant. Thus, the hydrophilic interaction is also reduced on addition of the less polar solvent, and this also contributes to the destabilization of micelles.

Several other interactions play an important but secondary role in determining the stability of aggregates of amphiphilic molecules. These include van der Waals attraction among the hydrocarbon tails, electrostatic repulsion between similarly charged hydrophilic heads, electrostatic attraction between zwitterionic heads, and possible hydrogen bonding among the polar portions of the molecules. At any given concentration and temperature, the interplay of these forces plus the major hydrophobic and hydrophilic interactions determines the precise form of aggregation of the amphiphilic molecules. A quantitative theory of the composition and temperature dependence of the various structures of the lyotropic phases that includes all of the above considerations is not yet available. A reconstruction of the R-

theory[5] to explicitly include the hydrophobic effect along the lines given by Kauzman,[2] Singer,[3] and Tanford[4] would be a highly desirable beginning.

5. Conclusion

The composition and structure of lyotropic liquid crystals and biological membranes have been examined. An understanding of the varied structures was obtained in terms of the hydrophobic and hydrophilic interactions. That is, aggregates of amphiphilic molecules in aqueous solution always form in such a manner as to minimize the hydrophobic interaction between the hydrocarbon tails and the water, while simultaneously maximizing the hydrophilic interaction of the polar heads with the aqueous solvent. In this way, we saw that the water was not merely the medium in which these phenomena take place. Instead, it became clear that it is the unique properties of the water itself that give rise to the crucial interactions that stabilize lyotropics and membranes.

The ordered structures discussed here are also unusual from another point of view. While most of the ordered phases encountered in physics and chemistry owe their existence to a successful competition between the energy and the entropy, the ordered structures discussed here derive their stability from a competition between two different kinds of entropy. In magnetism, superconductivity, ferroelectricity, thermotropic liquid crystals, solid-liquid-vapor equilibrium, multicomponent phase separation, etc., the ordered phase is always one of low energy and low entropy, while the disordered phase has high energy and high entropy. At low temperatures the free energy (given by $F = E - TS$) will be minimized by the low energy ordered state. At higher temperatures the TS term gains in importance and eventually the free energy will be minimized by the high entropy disordered state.

In the case of micelles, lyotropics and membranes, however, the ordered state is characterized by having a low cratic entropy and a high unitary entropy. The disordered state with the molecules dispersed throughout the solvent, on the other hand, is a state of high cratic entropy and low unitary entropy (because of the hydrophobic effect). Since the cratic and unitary entropies have different dependencies on composition and temperature, the minimization of the free energy will sometimes be accomplished by maximizing the cratic entropy, and sometimes by maximizing the unitary entropy, depending on the exact conditions.

The outcome of the competition between cratic and unitary entropies, as the temperature and concentration are varied, thus gives rise to the broad features of the phase diagram shown in Fig. 4. This situation is somewhat analogous to existence of an ordered nematic state in the hard rod model.[13] In this case, the two kinds of entropy in competition are the cratic and orientational entropies.

References:

[1] H. S. Frank and M. W. Evans, "Free Volume and Entropy in Condensed Systems III," *J. Chem. Phys.*, **13**, p. 507 (1945).

[2] W. Kauzmann, "Some Factors in the Interpretation of Protein Denaturation," *Adv. Protein Chem.*, **14**, p. 1 (1959).

[3] S. J. Singer, "The Molecular Organization of Biological Membranes," in *Structure and Function of Biological Membranes*, ed. by L. I. Rothfield, Academic Press, N. Y. (1971).

[4] C. Tanford, *The Hydrophobic Effect*, John Wiley and Sons, N. Y. (1973).

[5] P. A. Winsor, "Binary and Multicomponent Solutions of Amphiphilic Compounds," *Chem. Reviews*, **68**, p. 1 (1968).

[6] G. H. Brown, J. W. Doane and V. D. Neff, *A Review of the Structure and Physical Properties of Liquid Crystals*, CRC Press, Cleveland, Ohio (1971).

[7] A. Saupe, "Liquid Crystals," in *Annual Reviews of Phys. Chem.*, ed. by H. Eyring, Annual Reviews, Inc., Palo Alto, Vol. 24 (1973).

[8] L. I. Rothfield, "Biological Membranes: An Overview at the Molecular Level," in *Structure and Function of Biological Membranes*, ed. by L. I. Rothfield, Academic Press, N. Y. (1971).

[9] S. J. Singer, "The Fluid Mosaic Model of the Structure of Cell Membranes," Science, **175**, p. 720 (1972).

[10] R. A. Capaldi, "A Dyanmic Model of Cell Membranes," *Scientific American*, p. 27, March 1974.

[11] S. J. Singer, "The Molecular Organization of Membranes," in *Annual Review of Biochemistry*, ed. by E. E. Snell, Annual Reviews, Inc., Palo Alto, Vol. 43, 1974.

[12] R. W. Gurney, *Ionic Processes in Solution*, Chapter 5, McGraw-Hill Book Co., Inc., N. Y. (1953).

[13] P. Sheng, "Hard Rod Model of the Nematic-Isotropic Phase Transition," *RCA Review* **35**, p. 132, March 1974.

Recent Papers by RCA Authors

Listing is alphabetical by name of primary author. For copies of reprints, the reader should contact the publication directly.

B. Abeles and P. Sheng, "Electron Localization in Granular Metals," **Low Temp. Physics,** Vol. 3, p. 578.

M. S. Abrahams, J. Blanc, and C. J. Buiocchi, "Interdependence of Strain, Precipitation, and Dislocation Formation in Epitaxial Se-Doped GaAs," **J. Appl. Phys.,** Vol. 45, No. 8, p. 3277, Aug. 1974.

M. S. Abrahams and C. J. Buiocchi, "Cross-Sectional Specimens for Transmission Electron Microscopy," **J. Appl. Phys.,** Vol. 45, No. 8, p. 3315, Aug. 1974.

K. Ametani, "Atomic Absorption Spectrophotometric Determination of Rare Earths (Y, Eu, Gd, Dy, Ho, and Er) in Single Crystals of Magnetic Garnets and Sulfides," **Bull. Chem. Soc. of Japan,** Vol. 47, No. 9, p. 2238, Sept. 1974.

A. J. Aukstikalnis, "A Survey of Spacecraft Computers," **Astronics & Aeronautics,** Aug. 1974.

D. E. Carlson, K. W. Hang, and G. F. Stockdale, "Ion Depletion of Glass at A Blocking Anode: II, Properties of Ion-Depleted Glasses," **J. Amer. Ceramic Soc.,** Vol. 57, No. 7, p 295, July 1974.

D. E. Carlson, "Ion Depletion of Glass At a Blocking Anode: I, Theory and Experimental Results for Alkali Silicate Glasses," **J. Amer. Ceramic Soc.,** Vol. 57, No. 7, p. 291, July 1974.

D. J. Channin, "Liquid-Crystal Technique for Observing Integrated-Circuit Operation," **IEEE Trans. Electron Devices,** Vol. 21, No. 1, p. 650, Oct. 1974.

R. S. Crandall, "Lifetime of Surface-State Electrons on Liquid ^{4}He: II. Electron Lattice," **Phys. Rev. A,** Vol. 10, p. 1370, Oct. '74.

P. A. Crossley and W. E. Ham, "Use of Test Structures and Results of Electrical Tests for Silicon-on-Sapphire Integrated Circuit Processes," **J. Electronic Materials,** Vol. 2, No. 4, p. 466, 1973.

G. W. Cullen and J. F. Corboy, "A Comparison of the Semiconducting Properties of Thin Films of Silicon on Sapphire and Spinel," **J. Electrochem. Soc.,** Vol. 121, No. 10, p. 1345, Oct. 1974.

E. J. Denlinger, "Design of Partial Height Ferrite Waveguide Circulators," **IEEE Trans Microwave Theory & Techn.,** p. 810, Aug. 1974.

E. C. Douglas and A. G. F. Dingwall, "Ion Implantation for Threshold Control in COS/MOS Circuits," **IEEE Trans. Electron Devices,** Vol. 21, No. 6, p. 324, June 1974.

J. Dresner, "Mobility in Epitaxial GaAs Under 1 Mev Electron Irradiation," **J. Appl. Phys.,** Vol. 45, No. 9, p. 4118, Sept. 1974.

R. S. Engelbrecht, "Electrostatic Recording on Insulators by Nonwetting Conducting Liquids," **J. Appl. Phys.,** Vol. 45, No. 8, p. 3421, Aug. 1974.

J. G. Endriz, "Surface Waves and Grating-Tuned Photocathodes," **Appl. Phys. Lett.,** Vol. 25, No. 5, p. 261, Sept. 1974.

J. G. Endriz, "The Role of Plasmons in Photoemission from Metals," **Proc. Conf. on Polaritons,** p. 275.

R. E. Enstrom, C. J. Nuese, J. R. Appert, and J. J. Gannon, "Influence of Gas-Phase Stoichiometry on the Defect Morphology, Impurity Doping, and Electroluminescence Efficiency of Vapor-Grown GaAs P-N Junctions," **J. Electrochem. Soc.,** Vol. 121, No. 11, p. 1516, Nov. 1974.

D. G. Fisher, "The Effect of Cs-O Activation Temperature of the Surface Escape Probability of NEA (In,Ga)As Photocathodes," **IEEE Trans. Electron Devices,** Vol. 21, No. 8, p. 541, Aug 1974.

D. G. Fisher, R. E. Enstrom, J. S. Escher, H. F. Gossenberger, and J. R. Appert, "Photoemission Characteristics of Transmission-Mode Negative Electron Affinity GaAs and (In,Ga)As Vapor-Grown Structures," **IEEE Trans. Electron Devices,** Vol. 21, No. 10, p. 641, Oct. 1974.

J. J. Gannon and C. J. Nuese, "A Chemical Etchant for the Selective Removal of GaAs Through SiO_2 Masks," **J. Electrochem. Soc.,** Vol. 121, No. 9, p. 1215, Sept. 1974.

A. K. Ghosh, "Review on High-Power-Laser Damage to Materials II," **RCA Review** Vol. 35, No. 2, p. 279, June 1974.

G. B. Herzog, "The Impact of LSI Technology on Computer System," **Information Processing** (Book) 1974.

P. T. Ho and W. R. Curtice, "A Microstrip Trapatt Amplifier for X-Band Operation," **Proc. IEEE,** Vol. 62, No. 7, p. 1029, July 1974.

H. Kawamoto, S. G. Liu, H. J. Prager, and E. L. Allen, "S-Band Trapatt Amplifiers with Four-Layer Diode Structures," **RCA Review,** Vol. 35, No. 3, p. 372, Sept. 1974.

H. P. Kleinknecht and A. E. Widmer, "(GaAl)P Optical-Waveguide Modulators Fabricated by Liquid-Phase Epitaxy," **J. Appl. Phys.,** Vol. 45, No. 8, p. 3453, Aug. 1974.

H. Kressel, P. Robinson, R. V. D'Aiello, and S. H. McFarlane, "Properties of High-Voltage Silicon Epitaxial Diodes," **J. Appl. Phys.,** Vol. 45, No. 9, p. 3930, Sept. 1974.

H. Kressel, "Gallium Arsenide and (AlGa)As Devices Prepared by Liquid-Phase Epitaxy," **J. Electronic Materials,** Vol. 3, No. 4, p. 747, 1974.

H. Kressel, H. F. Lockwood, I. Ladany, and M. Ettenberg, "Heterojunction Laser Diodes for Room Temperature Operation," **Opt. Eng.,** Vol. 13, No. 5, p. 416, Sept./Oct. 1974.

A. W. Levine, M. Kaplan, E. S. Poliniak, "The Interaction of 5-KeV Electrons with Polymers of Methyl Isopropenyl Ketone," **Polymer Eng. & Sci.,** Vol. 14, No. 7, p. 518, July 1974.

J. D. Levine, "Power Law Reverse Current-Voltage Characteristic in Schottky Barriers," **Solid State Elect.,** Vol. 17, p. 1083, 1974.

R. U. Martinelli and M. Ettenberg, "Electron Transport and Emission Characteristics of Negative Electron Affinity $Al_xGa_{1-x}As$ Alloys ($0 \leq x \leq 0.3$)," **J. Appl. Phys.,** Vol. 45, No. 9, p. 3896, Sept. 1974.

R. U. Martinelli, J. I. Pankove, "Secondary Electron Emission from the GaN:Cs-O Surface," **Appl. Phys. Lett.,** Vol. 25, No. 10, p. 549, Nov. 1974.

R. U. Martinelli and D. G. Fisher, "The Application of Semiconductors with Negative Electron Affinity Surfaces to Electron Emission Devices," **Proc. IEEE,** Vol. 62, No. 10, p. 1339, Oct. 1974.

S. H. McFarlane and E. M. Ettenberg, "The Effect of Substrate Preparation on Surface Appearance of Liquid-Phase Epitaxial (Al,Ga)As on GaP," **J. Crystal Growth,** Vol. 23, p. 233, 1974.

C. J. Nuese, M. Ettenberg, and G. H. Olsen, "Room-Temperature Heterojunction Laser Diodes from Vapor-Grown $In_{1-x}Ga_xP$/GaAs Structures," **Appl. Phys. Lett.,** Vol. 25, No. 10, p. 612, Nov. 1974.

J. O'Brien, E. Fischer, and R. Glicksman, "A Large Optical Cavity Laser Diode for Communications Applications," **Conf. Proc. 6th DOD Conf. on Laser Tech.**

H. F. Olson, "Comment on an Analysis of Design Conditions for a Phase-Inverter Speaker System with a Drone Cone," **IEEE Trans. Acoustics, Speech, & Signal Process,** Vol. 22, No. 5, p. 389, Oct. 1974.

L. S. Onyshkevych, R. Shahbender, S. Tomkiel, and F. Putzrath, "Design, Construction, and Testing of a Magnetic Bubble Memory Chip," **RCA Review,** Vol. 35, No. 2, p. 216, June 1974.

J. I. Pankove and M. A. Lampert, "Model for Electroluminescence in GaN," **Phys. Rev. Lett.,** Vol. 33, No. 6, p. 361, Aug. 1974.

J. I. Pankove and J. E. Berkeyheiser, "Properties of Zn-Doped GaN: II Photoconductivity," **J. Appl. Phys,** Vol. 45, No. 9, p. 3892, Sept. 1974.

R. J. Powell, J. R. Ligenza, and M. S. Schneider, "Selective Oxidation of Silicon in Low-Temperature High-Pressure Stream," **IEEE Trans Electron Devices,** Vol. 121, No. 10, p. 636, Oct. 1974.

D. Redfield and J. W. Burke, "Optical Absorption Edge of $LiNbO_3$," **J. Appl. Phys.,** Vol. 45, No. 10, p. 4566, Oct. 1974.

W. R. Roach, "Resolution of Electro-Optic Light Valves," **IEEE Trans. Electron Devices,** Vol. 21. No. 8, p. 453, Aug. 1974.

D. P. Schnon, "Printed Wiring Processing Materials Techniques," **Electronic Packaging and Product,** Vol. 14, No. 7, July 1974.

B. R. Schwartz, "Interconnections," **Electromechanical Design,** Aug. 1974.

R. Shahbender, "Novel Magneto-Optic Systems," **IEEE Trans. Magnetics,** p. 575, Sept. 1974.

F. Sterzer, "An Electronic License Plate for Motor Vehicles," **RCA Review,** Vol. 35, No. 2, p. 167, June 1974.

R. A. Sunshine, "Infrared Observation of Current Distributions in Large Area Power Transistors," **IEEE Trans. Industrial Elect.,** Vol. 21, No. 3, p. 116, Aug. 1974.

R. A. Sunshine, "Multi-Dimensional Current Flow in Silicon Power Transistors Operating in the Saturation Mode," **Power Electronics Specialists Conf. Record,** p. 154, 1974.

A. Sussman, "The Electro-Optic Transfer Function in Nematic Liquids," **RCA Review,** Vol. 35, No. 2, p. 176, June 1974.

K. Suzuki and M. Nishikawa, " 'Sing-Around' Method for Acoustic Surface-Wave Velocity Measurements Using an Optical Probe," **Japan. J. Appl. Phys.,** Vol. 13, No. 8, p. 1216, Aug. 1974.

T. Takahashi, K. Ametani, and O. Yamada, "Synthesis and Some Properties of RCr_3S_6 Crystals (R = Y, Gd, Dy, Ho and Er)" **J. Cryst. Growth,** Vol. 24, No. 25, p. 151, 1974.

L. J. Vieland and A. W. Wicklund, "High T_c Nb_3Ga and Nb_3Ge by CVD," **Phys Lett.**, Vol. 49, No. 5, p. 407, Oct. 1974.
J. L. Vossen, "A Sputtering Technique for Coating the Inside Walls of Through-Holes in Substrates," **J. Vacuum Sci. & Tech.**, Vol. 11, No. 5, p. 875, Sept./Oct. 1974.
C. C. Wang, S. H. Ladany, S. H. McFarlane, and F. C. Dougherty, "Two-Stage Epitaxial Growth of GaP on Spinel," **J. Cryst. Growth**, Vol. 24, No. 25, p. 239, 1974.
P. P. Webb, R. J. McIntyre, and J. Conradi, "Properties of Avalanche Photodiodes," **RCA Review**, Vol. 35, No. 3, p. 234, June 1974.
J. Weisbecker, "A Practical, Low-Cost Home/School Microprocessor System," **Computer Mag.**, Vol. 7, No. 8, p. 20, Aug. 1974.
A. E. Widmer and R. Fehlmann, "A Millivolt Source for Temperature Programming of Laboratory Furnaces," **J. Phys. E: Scientific Instruments**, Vol. 7, p. 610, 1974.
R. Williams and A. M. Goodman, "Wetting of Thin Layers of SiO_2 by Water," **Appl. Phys. Lett.**, Vol. 25, No. 10, p. 531, Nov. 1974.
J. Wittke, "Dispersion-Limited Modulation Bandwidths of Optical Fibers," **RCA Review**, Vol. 35, No. 2, p. 198, June 1974.

"Patents Issued to RCA Inventors Third Quarter, 1974

July

A. A. A. Ahmed Temperature-Sensitive Control Circuit (3,825,778)
J. J. Amodei Apparatus for Providing an Optical System Using Adaptive Holographic Components (3,825,316)
E. F. Belohoubek and D. M. Stevenson Transistor Carrier for Microwave Stripline Circuit (3,825,805)
W. M. Binder Device for Mounting and Orientating a Workpiece (3,826,047)
H. L. Blust and N. L. Lindburg Plug Sealing of Hermetic Enclosures (3,826,634)
B. F. Bogner and D. F. Bowman Lens Fed Antenna Array System (3,827,055)
R. M. Carrell Rotatable Paraboloidal Reservoir Useful in an Ink Jet Printer (3,823,409)
R. M. Christensen, J. J. Gibson, and A. L. Linberg Omnidirectional Sound Field Reproducing System (3,824,342)
C. F. Coleman Method of Soldering Circuit Components to a Substrate (3,825,994)
E. B. Davidson Unsaturated Polyester Resin Photoresist Composition (3,825,428)
W. F. W. Dietz Raster Correction Circuit Utilizing a Parabolically Varying Load Circuit (3,825,793)
N. Feldstein and J. A. Weiner Method of Making a Semitransparent Photomask (3,822,155)
R. Feryszka and J. O. Preisig MOS FET Reference Voltage Supply (3,823,332)
M. B. Finkelstein Recording Web Tension Control (3,823,896)
P. Foldes Adjustable Polarization Antenna System (3,827,051)
S. V. Forgue Targets for Television Pickup Tubes (3,818,262)
N. S. Freedman, C. W. Horsting, W. F. Lawrence, and J. J. Corrona Thermo-Electric Molular Structure and Method of Making Same (3,787,958)
J. B. George Television Automatic Gain Control Circuitry Providing for Compatible Control of VHF Tuner and UHF Tuner (3,823,379)
P. E. Haferl Cascode Video Output Feedback Amplifier (3,823,264)
D. Hampel Multi-Function Logic Gate (3,825,770)
B. A. Hegarty and L. H. Trevail High Heat Dissipation Solder-Reflow Flip Chip Transistor (3,823,469)
M. C. Johnson and D. P. Clock Power Supply Keep Alive System (3,824450)
M. A. Leedom and M. E. Miller Information Playback System Stylus (3,826,877)
D. P. Marinelli and T. E. Stockton Method of Depositing Epitaxial Layers on a Substrate from the Liquid Phase (3,825,449)
P. C. Phillips Balanced Line Driver, Line Receiver System (3,825,682)
N. E. Prysiak Thermoelectric Generator (3,794,526)
G. D. Pyles Delayed Alarm and Drowse for Clock Receivers (3,825,836)
G. D. Pyles Cartridge Changer with Cartridge Sensing Means (3,825,949)
W. E. Riggle Method of Aligning a Two-Capillary Tube Gas Discharge Device (3,824,089)
T. M. Shrader Method of Making Cathode Ray Tube Internal Shields (3,822,453)
W. F. Speer and L. E. Reed Parity of Tuning Apparatus (3,824,507)
L. J. Vieland and A. W. Wicklund Process for Preparing Superconducting Niobium-Gallium Alloy (3,824,082)
J. H. Wharton Pulse Delay Circuit (3,824,411)

August

A. A. A. Ahmed Relaxation Oscillator (3,831,113)
L. J. Bazin Beam Current Stabilization and Blanking Apparatus (3,831,056)
B. W. Beyers, Jr. Video Disc Playback Eddy Current Speed Control System (3,829,612)
J. J. Brandinger, D. H. Pritchard, G. L. Fredendall, and A. C. Schroeder Color Signal Producing System Utilizing Spatial Color Encoding (3,828,121)
J. W. Daniel, Jr. Parametric Amplifier Having an Idler Circuit Reducing Spurious Idler Signal Magnitude (3,831,037)
W. F. W. Dietz Horizontal Deflection System with Boosted B Plus (3,832,595)
R. A. Dischert and J. F. Monahan Automatic Centering Control System for Television Apparatus (3,830,959)
D. P. Dorsey and W. E. Rodda Storage Tube Erase Control (3,831,054)
D. D. Freedman High-Speed Analog-to-Digital Converter (3,829,853)
E. A. Goldberg Motor Speed Control System (3,828,234)
R. B. Goyer Wide Range. Monostable Multivibrator Circuit having a Constant Current Source (3,829,716)
V. E. Hills and L. Wu Signal Duration Sensing Circuit (3,828,258)
N. Hovagimyan and M. Rosenblatt Information Transfer System for a PBX (3,832,495)
S. Y. Narayan Dynamic Dividing Circuit for Dividing an Input Frequency By Two (3,832,651)
J. M. S. Neilson Thyristor Having Capacitively Coupled Control Electrode (3,831,187)
R. W. Paglione Encapsulated Microstrip Circulator with Mode Elimination Means (3,831,114)
K. J. Phillips Active Nutation Damping in Dual-Spin Spacecraft (3,830,447)
P. J. Smaiser Rotator System including a remote drive Motor and a Local Indicator-Control Motor (3,831,074)
C. L. Upadhyayula and S. Y. Narayan Dynamic Dividing Circuit for Dividing an Input Frequency by at least Three (3,832,652)
H. R. Warren Record Web Control and Drive Apparatus (3,828,996)
B. Zuk High Speed Driving Circuit for Producing Two In-Phase and Two Out-of-Phase Signals (3,828,206)

September

D. H. Block Pulse Width Sensing Circuit (3,835,336)
D. S. Bond Television Communication System with Time Delay Compensation (3,835,253)
D. S. Bond and J. M. L. Holman Geo-Synchronous Satellites in Quasi-Equatorial Orbits (3,836,969)
R. E. Debrecht and L. S. Napoli Microwave Transmission Line and Devices Using Multiple Coplanar Conductors (3,835,421)
J. A. Dodd, Jr. and R. F. Okomoto Method of making Elliptically or Rectangularly Graded Photoprinting Masters '3,834,905)
D. P. Dorsey and W. E. Rodda Storage Tube Focus Control (3,838,311)
W. G. Einthoven, W. H. Schilp, Jr., and A. A. Todd Semiconductor Darlington Circuit (3,836,997)
W. B. Hall and J. A. Koskulitz Heat Dissipation for Power Integrated Circuit Devices (3,836,825)
J. R. Harford Keyed AGC Circuit (3,835,248)
A. F. McDonie Method of Making a Bialkali Photocathode with Improved Sensitivity and High Temperature Operating Characteristics (3,838,304)
D. L. Matthies Video Discs Having a Methyl Alkyl Silicone Coating (3,833,408)
J. J. Moscony and R. L. Kennard Method of Repairing an Imperfect Pattern of Metalized Portions on a Substrate (3,833,375)
M. I. Payne and J. D. Mazgy Apparatus for Testing the Linearity of a Circuit by Using Ratio Determining Means (3,836,845)
H. Perkel Closed Loop Roll and Yaw Control for Satellites (3,834,653)
M. Pradervand Chip Removal in the Mastering of Fine-Grooved Discs (3,837,656)
J. A. Rajchman Holographic Memory Including Corner Reflectors (3,833,893)
R. S. Ronen Method of Simultaneously Making a SIGFET and a MOSFET (3,837,071)
R. J. Ryan and L. A. Dimarco Holographic Recording Medium (3,833,383)
R. L. Schelhorn Bonding Tool and Method of Bonding Therewith (3,838,240)
W. H. Schilp, Jr. and A. A. Todd Semicondcutor Darlington Circuit (3,836,996)
C. F. Wheatley, Jr. and W. G. Einthoven Semiconductor Darlington Circuit (3,836,995)
H. A. Wittlinger Current Amplifier (3,835,410)

AUTHORS

D. J. Channin received the BS in physics from Case Institute of Technology in 1964 and the PhD in physics from Cornell University in 1970. His dissertation and post doctoral work at Cornell involved thermal and microwave phonon pulse experiments at cryogenic temperatures. From 1970 to the present he has been at RCA Laboratories, Princeton, N.J. He has worked on acoustooptic laser beam modulators and deflectors, active and passive phenomena in optical waveguides, and the application of liquid crystals to optical waveguides and to integrated-circuit testing. Dr. Channin is a member of the American Physical Society, the Optical Society of America, Tau Beta Pi, and Sigma Xi.

T. N. Chin graduated with the B.S. degree from National Tsing Hua University, Kunming, Yunnan in 1942. He received the M.S. and Ph.D. degrees from the University of Illinois, Urbana in 1949 and 1952, respectively. From 1942 to 1945 he worked at the Central Radio Manufacturing Works in Kunming, and from 1945 to 1947 he was an instructor at Tsing Hua University in Peking. In 1952 he became a research associate in the E.E. Department, University of Illinois, carrying out studies on the transport properties of the electron flows and electron emission from a hollow cathode. Since 1955 he has been employed by RCA where he worked first on several types of vidicons in Lancaster, Pennsylvania. In 1957 he was transferred to the Laboratories at Princeton, N.J., where he has been engaged in investigations related to display devices and color kinescopes. From 1963 to 1964 he was an associate professor in the E.E. Department at Rutgers, N.J. In that summer, he participated in M.I.T.'s program on Experimental Solid State Physics under the support of the National Science Foundation.

Dr. Chin is a member of Sigma Xi and the IEEE.

Roger W. Cohen joined RCA Laboratories in 1960 after receiving the B.S. degree in Physics from M. I. T. He participated in the Graduate Study Program, which provided for study leading to a Masters Degree, while maintaining research activities at the Laboratories. In 1966 he received the Ph.D. degree in Physics from Rutgers. He is presently head of the Physics and Chemistry of Solids Research group at RCA Laboratories, 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 He^3. He has been very active in the field of superconductive tunneling, the enhancement of the transition temperature in superconductors,

the normal state properties of beta-tungsten compounds, and the relationship of lattice instabilities to high-temperature superconductivity. He has contributed to the theory of the optical and magnetic properties of dispersions of fine metallic particles. Recently, he has studied fundamental problems in the theory of imaging in incoherent optical display systems. He has collaborated in the establishment of new image quality descriptors that include the effects of both the display and the human visual system.

Dr. Cohen is a member of the American Physical Society and Sigma Xi.

Louis S. Cosentino received the B.E.E. degree from City College of New York, N.Y., in 1960 and the M.S. degree from Princeton University, Princeton, N.J. in 1962. Since 1960 he has been a Member of the Technical Staff at RCA Laboratories, Princeton, N.J. His past endeavors have included the design and development of tunnel-diode circuits for high-speed computers and several years of research in cryoelectrics, including switching devices, memories, and associated electronics. More recently he has worked on the optics and components for read-write optical memory systems utilizing holography.

Mr. Consentino was co-recipient of an RCA Laboratories Outstanding Achievement Award in 1968 and 1971. He is a senior member of the IEEE, and a member of Eta Kappa Nu and AAAS.

Maurice D. Coutts received a B.Sc. with 1st Class Honors from Glasgow University in 1954, and was awarded a Department of Scientific and Industrial research fellowship from 1954 to 1957 for research in electron microscopy. Since joining the staff of RCA Laboratories in 1961 he has worked with electron microscopy and diffraction on a variety of solid state problems. He has been the recipient of two RCA Laboratories Achievement awards for outstanding work in electron microscopy and diffraction. He has also been a member of the Committee on Structure for the National Academy of Sciences and National Academy of Engineering. He is a member of the Electron Microscope Society and of the American Chemical Society.

K. F. Etzold received his B.S. in Physics from the City College of New York in 1963, the M.S. from New York University in 1965, and the Ph.D. in Physics in 1973, also from New York University. He has taught undergraduate courses at N.Y.U. in Electricity and Magnetism and Electronic Instrumentation. He recently joined the Information Sciences group at RCA Laboratories. In addition to his background in physics, Dr. Etzold has extensive experience in scientific instrumentation. He has done consulting work for the New York Psychiatric Institute where he designed instrumentation for behavioral studies and equipment for large-tip and microelectrode recording of evoked responses from the hypothalamus. He also has extensive experience with circuitry and concepts in the audio range through work at firms specializing in professional equipment and independent consulting. He is a member of the American Physical Society.

Lawrence A. Goodman received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from M.I.T. He joined RCA Laboratories in 1970 after completing his doctoral dissertation on the optical properties of cadmium sulfide. Since coming to RCA, Dr. Goodman has been investigating various aspects of liquid-crystal displays. These studies have included conductive-electrode preparation, surface-alignment phenomena, bulk electro-optic effects, and various approaches to electrically addressing liquid-crystal displays. He is a member of IEEE, SID, Tau Beta Pi, and Eta Kappa Nu.

Ho-Chung Huang received his B.S. degree from National Taiwan University, Taiwan, Republic of China, in 1959, and his M.S. and Ph.D. degrees from Cornell University in 1965 and 1967, respectively. As a graduate student at Cornell, Dr. Huang discovered the hybrid mode of transferred electron devices. From 1967 to 1969 he was a Senior Research Physicist with Monsanto Company in St. Louis, Missouri, where he was engaged in the research and development of the transferred electron oscillators operating in the hybrid mode and LSA mode. Since joining RCA Laboratories in 1969, Dr. Huang has worked on the processing of microwave devices such as transferred electron devices, Impatts, electron-beam semiconductor targets, bulk acoustic delay lines, and GaAs field-effect transistors. Dr. Huang is a member of the IEEE.

S. T. Jolly received his B.Sc. degree in Metallurgy from the University of Wales in 1939. From 1940 to 1946, he served in the British Army. From 1946 to 1952, he was employed by EMI Ltd. supervising the Materials and Components Test and Evaluation Laboratory. Since 1953, he has been employed by RCA, initially as engineering leader responsible for the development of packaging hardware, power supplies, and magnetic head design used in the production of the Bizmac 301 and 501 computers. As manager of the Magnetic Head Design department from 1959 to 1963, he supervised the design and development of magnetic recording heads for professional audio, video, and digital recording, and assisted in the set up of associated production facilities. From 1963 to 1968, Mr. Jolly was an engineering group leader with RCA Defense Electronic Products, responsible for development work in computer memories, automated test equipment studies, and setting up a facility for the production of high resolution photomasks to be used in the manufacture of integrated-circuits. Mr. Jolly was the senior engineer responsible for initial design and process development for liquid-crystal displays. His present assignment is the development of epitaxial-growth processes and production of material for solid-state-microwave devices with the Microwave Technology Center, RCA Laboratories, Princeton, N.J.

Achilles G. Kokkas received the B.S.E.E. degree with honors from Robert College, Istanbul, Turkey, in 1963, and the M.S., E.E., and Sc.D. degrees from the Massachusetts Institute of Technology, in 1966, 1970, and 1972, respectively. From 1963 to 1965 and 1968 to 1972 he was at the M.I.T. Center for Materials Science and Engineering, working primarily in the areas of anomalous diffusion in silicon and electrical-thermal interactions in integrated circuits. During the years 1968–1971 his work was supported by an RCA Doctoral Study Award.

He joined the technical staff of RCA Laboratories in 1965. At RCA he has been involved with the electrical characterization of silicon films grown on sapphire and spinel, the development of lateral Gunn-effect devices and silicon PIN diodes, and has investigated design trade-offs in CMOS circuits. He is presently working on high performance integrated circuits. His interests include radiation effects in MOS/SOS circuits and thermal effects in discrete devices and integrated circuits.

Dr. Kokkas is a member of the Society of Sigma Xi and the IEEE.

B. J. Levin received the BSEE degree from Drexel University in 1964, a MSEE degree from MIT in 1965, and a PhD in Electrical Engineering from the University of Pennsylvania in 1969. He worked on the design of digital circuits on the Early Bird Satellite program at Bell Telephone Laboratories in 1964. In the summer of 1966, he worked on the development of GaAs microwave diode avalanche oscillators as a member of the technical staff of the RCA Laboratories. In the summer of 1967 he worked on frequency locking of microwave diode avalanche oscillators at American Electronics Laboratories. In September 1967 he became a Research Associate at the Moore School of Electrical Engineering. He was concerned with various theoretical and experimental problems of millimeter wave imaging. His dissertation was a detailed analysis of the reflection mode-photoconductive readout millimeter wave image converter. In July 1969 Dr. Levin joined the Applied Physics Laboratory of RCA's Advanced Technology Laboratories in Camden, N.J. At ATL he has continued his investigation of millimeter wave imaging techniques. His work has included extending the imaging system to higher frequencies, the study of various readout techniques, the experimental imaging of complex targets, and various quasi-optical approaches to improving the target recognition ability of the imaging system. Since July 1970, he has also made a major contribution to the development of an electronically controlled low-loss phase shifter operating in the 100 GHz region. Recently, Dr. Levin has worked on the development of high power, high efficiency GaAs Impatt diodes at RCA Laboratories in Princeton, N.J. and on the Combat Theater Ground Communications (CTGC) study program conducted by RCA Government Communications Systems.

Dr. Levin is a member of the IEEE, Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi.

Reuben S. Mezrich received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from the Polytechnic Institute of Brooklyn. In 1961, while studying for the B.S. he received a National Science Foundation grant for research on materials capable of laser action. In 1963 he joined RCA Laboratories and was associated with the cryoelectric devices group working on cryogenic logic elements. At the same time he was on the RCA Graduate Study program, which enabled him to fulfill the requirements for the master's degree. Dr. Mezrich received a RCA Laboratories Graduate Study award in 1964, allowing him to fulfill the course requirements for the Ph.D. in Electrical Engineering. Since his return to the Labora-

tories in 1965, Dr. Mezrich has been involved in research in holography and its applications to computer storage. These studies have included read-only memories, image storage and retrieval systems, and reversible memories. In 1968 he was co-recipient of an IR 100 award for the development of a holographic read-only memory. In 1969 he was the recipient of an IR 100 award for the development of holographic storage on magnetic films. He received an RCA Laboratories Achievement Award in 1970 for his work on magnetic holography and in 1972 for a team effort on read-write optical memories.

Dr. Mezrich is a member of Tau Beta Pi, Eta Kappa Nu and IEEE.

E. B. Priestley received his B.Sc. degree in Chemistry from the University of Alberta in 1965 and his Ph.D. degree in Chemical Physics from the California Institute of Technology in 1969. He spent two years as a Research Fellow in the Department of Engineering and Applied Physics at Harvard University, and in 1971 joined the technical staff of the Physical Electronics Laboratory at the David Sarnoff Research Center, Princeton, N.J. Dr. Priestley has been studying long-range orientational molecular ordering in nematic and smectic liquid crystals using a Raman scattering technique he developed while at Harvard University. At present he is using Raman scattering and inelastic electron tunneling spectroscopy to investigate chemisorption of liquid crystal and surfactant molecules on a variety of substrates in an effort to understand the forces responsible for surface alignment of liquid crystals.

Dr. Priestley is a member of the American Physical Society and the American Association for the Advancement of Science.

Wilber C. Stewart received the B.S., M.S., and Ph.D. degrees in electrical engineering at Duke University. Upon joining RCA Laboratories in 1964, his interests in superconducting devices included studies of field-induced switching phenomena, flux coupling in films, general inductance relations, signal propagation in large device arrays, and Josephson tunneling phenomena. His more recent activities in optical holography have included research on electrical and optical switching phenomena in ferroelectric ceramics, analyses of random phase data masks, and studies of the optical system requirements for holographic memories.

Dr. Stewart was the recipient of RCA Laboratories Outstanding Achievement Awards in 1967 and 1970. He is a member of the IEEE, Optical Society of America, Phi Beta Kappa, Tau Beta Pi, Eta Kappa Nu, Pi Mu Epsilon, and Sigma Xi.

Alan Sussman received the bachelors degree in chemistry from Rensselear Polytechnic Institute, Troy, N.Y., and the Ph.D. degree in physical chemistry from Yale University, New Haven, Conn. He then joined the David Sarnoff Research Laboratory (RCA), working on solid state light amplification, organic thin film rectifiers, and liquid crystal display devices. For the latter two projects, he received RCA Achievement Awards. After serving as a technical consultant for Hewlett-Packard, N.E.L.C. San Diego, and Optel Corporation, he joined the RCA Solid State Division in Somerville, N.J., continuing his involvement in liquid-crystal displays.

Chainulu L. Upadhyayula received the B.Sc. degree in Physics and the M.Sc. in Applied Physics from Andhra University, India, in 1955 and 1958, respectively, and the Ph.D. in Engineering from Brown University in 1968. During 1958–59, he was a trainee in the Atomic Energy Establishment, Bombay, India. From 1959 to 1964, he worked in the Electronics Division of the Atomic Energy Establishment. His work was in the area of nuclear electronic instrumentation. From 1964 to 1967, Dr. Upadhyayula was a research assistant and teaching assistant in engineering at Brown University where he was engaged in the study of electrical transport properties of semiconductors. During 1968–69, he was a Post Doctoral Fellow in Engineering at Brown University studying tunneling through superconducting metal-insulator-metal structures at cryogenic temperatures. Dr. Upadhyayula joined the RCA Microwave Technology Center at the David Sarnoff Research Center in Princeton, N.J., in 1969 and is presently concerned with semiconductor devices and device physics.

In 1970, Dr. Upadhyayula was a co-recipient of an RCA Laboratories Outstanding Achievement Award for a team effort in the development of GaAs transferred electron amplifiers. He is a member of Sigma Xi and the IEEE.

David Vilkomerson received the BSEE from the Massachusetts Institute of Technology in 1962, where he had been a National Merit Scholar. In 1962, he joined the RCA Missile and Surface Radar Division, working on radar data processing. He transferred to RCA Laboratories in 1963 and did research in alloy semiconductor lasers and solid-state cryogenic devices, developing a new superconducting amplifying device. This work provided the thesis for the M.S. in Engineering he was awarded in 1964 by the University of Pennsylvania. Dr. Vilkomerson received an RCA Research Award for Doctoral Study in 1964, allowing him to pursue full-time study toward the Ph.D. at Columbia University. After his return to RCA Laboratories in 1965, he was engaged in research on holograms and optical memories. He received the Ph.D. in 1969. He was a co-recipient of an IR 100 Award for the develop met of a holographic read-only memory. Dr. Vilkomerson was awarded a Post-Doctoral Fellowship to the Hebrew University of Jerusalem (1969–1970). While there he studied work on ultrasonic holography for medical diagnosis. He is continuing that research presently.

He is a member of Sigma Xi, the IEEE, and the Optical Society of America.

Peter J. Wojtowicz received his B.Sc. degree in chemistry (with highest honors) from Rutgers University in 1953. He received his M.S. in chemistry in 1954 and his Ph.D. in physical chemistry in 1956 from Yale University. He was a National Science Foundation Predoctoral fellow while at Yale, 1953–1956. His thesis was concerned with the statistical mechanics of the order-disorder phenomenon in solids and the theory of molten salts and fluid mixtures. Dr. Wojtowicz joined RCA Laboratories in 1956. He is a Member of the Technical Staff of the Physics and Chemistry of Solids Group of the Physical Electronics Research Laboratory. During 1966–67 he was acting head of the General Research Group. His research

effort while at RCA has been directed chiefly toward the theory of magnetic materials including quantum and statistical mechanics of the thermal, structural, and magnetic properties of these substances. This work included the theory of the cooperative Jahn-Teller effect, the statistical mechanics of magnetic interactions and phase transitions, the theory of magnetic semiconductors, the theory of granular ferromagnetic metals, and the theory of the application of ferrites and garnets to various magnetic devices. He is currently engaged in the theory of liquid crystals and liquid-crystal phase transformations. Dr. Wojtowicz is the recipient of two RCA Laboratories Achievement Awards for the years 1962 and 1966. He is a fellow of the American Physical Society and a member of Sigma Xi and Phi Beta Kappa.

RCA Review

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

Index Volume 35, 1974

March 1974 Volume 35 Number 1

June 1974 Volume 35 Number 2

September 1974 Volume 35 Number 3

December 1974 Volume 35 Number 4

Index to Authors, Volume 35, 1974