RCA Review

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

Contents

September 1978 Volume 39 Number 3 ISSN 0033–6831

VideoDisc Optics

Istvan Gorog

RCA Laboratories, Princeton, N. J. 08540

This issue of RCA Review contains a comprehensive description of the optics technology that was developed at RCA Laboratories in support of the RCA VideoDisc program. VideoDiscs are designed to provide consumers with a low cost, prerecorded-video-information distribution medium. A 12 inch diameter VideoDisc contains up to 2 hours of standard color-television-compatible video programs accompanied by high quality two-channel sound. The VideoDiscs are prepared by first recording the appropriately encoded program material onto a master disc and subsequently replicating the masters through a series of plating and vinyl pressing operations that are similar to those employed in the manufacture of standard audio discs. Playback of the final replica is accomplished by either a capacitive or a pressure sensitive stylus, guided by grooves on the disc. The capacitive approach has been selected for the RCA VideoDisc consumer product and a comprehensive review of the technologies employed in the RCA system was published in the March 1978 issue of *RCA Review*.

The primary motivation for undertaking the extensive optical research described in this issue was to develop contact-free playback and high speed measurement techniques suitable for quality and process control evaluation of master discs, replicas, and intermediate parts. A secondary motivation was to explore optical recording as a fully compatible, alternative mastering method to electron-beam and electro-mechanical recording.

The primary technical difficulty that our optical research had to overcome was that the geometric features of interest on the disc surface are substantially smaller than the wavelength of the reading light. The highest frequency signal elements recorded on the disc can be as small as 0.2 micrometers. The understanding and utilization of light diffraction in the vicinity and beyond the well known classical resolution limits of optical systems is the central theme of all papers in this issue.

In addition to the small dimensions of the signal elements, the VideoDisc signal format itself presented a special difficulty with respect to optical resolution. Here the signal elements are cigar shaped surface relief

variations, with their long dimension extending across the entire width of the grooves. The grooves are approximately 2.5 micrometers wide and the edges of adjacent grooves touch each other. The information is contained in the variation of the short dimension of the signal elements along the grooves, i.e., in the tangential disc direction. The optical resolution required to detect the VideoDisc signal elements with a standard optical arrangement, composed of a microscope objective and a light intensity sensor located on the axis of the objective, is exactly twice the resolution that would be required to detect signal elements that have the same information along the track dimension but that are isolated phase relief pits or islands surrounded by a reference surface.

Optical readout with excellent signal quality was successfully achieved using a differential-phase optical detector arrangement. Scalar diffraction theory was used to predict the general performance of this detector system. The experimental data verified the general characteristics and advantages of the differential phase method, but, contrary to the scalar predictions, no decrease in detected signal intensity was found when the signal depth on the reflective disc was increased beyond one-eighth wavelength of the reading light and no signal null was found at the one-quarter-wavelength depth. This gross breakdown of scalar theory necessitated the development of a more accurate vector theory. A better understanding of the VideoDisc diffraction problem was especially needed because we wanted to use the intensity variation in the diffraction patterns for quick and accurate determination of the signal and groove depths on master discs, replicas, and intermediate metal parts. The VideoDisc diffraction problem was solved rigorously for both focused and unfocused incident beams by numerically integrating Maxwell's equations, assuming an infinitely conducting, square-wave phase relief pattern. The results of these exact vector calculations are in agreement with the experimental data; we believe that whatever quantitative differences between theory and experiment that remain are attributable to differences between the assumed and the actual boundary conditions.

The optical resolution available with the shortest laser wavelength suitable for use with high quality optical components is insufficient for producing a frequency independent exposure pattern over the range of interest for VideoDisc recording. We have developed two methods that overcome the optical resolution limitation. One method utilizes the controllable optical standing wave patterns that occur during the exposure of a thin photoresist layer coated on a reflective substrate. The other utilizes the wide range of electronic frequency compensation that is possible when the grooves and the signal elements are simultaneously and linearly recorded onto a thick photoresist layer.

To our knowledge the results described in these papers represent the highest linear density of optical information recording and reading achieved to date.

Optical Readout of the RCA VideoDisc

A. H. Firester, C. B. Carroll, I. Gorog, M. E. Heller, J. P. Russell, and W. C. Stewart

RCA Laboratories, Princeton, N.J. 08540

Abstract—Optical playback of the RCA VideoDisc was once considered at best impracticable if not impossible because the size of the features that encode the information on the disc can be as small as 250 nm. This is less than the wavelength of visible light. Furthermore, conventional optical video disc readout systems require a blank, intertrack optical reference surface. In contrast, the format of the RCA capacitive VideoDisc contains nearly abutting signal tracks. We have developed a phase-sensitive split-detector configuration that enables optical readout of RCA VideoDiscs by providing for this format a resolution twice that of conventional readout systems. Our optical playback techniques are now sufficiently advanced to be of utility in demanding quality control applications, and players based on these techniques are routinely used for evaluating the signal quality of master recordings. This paper details the specific engineering-production role for optical playback of the RCA capacitive VideoDisc and the underlying optical concepts and engineering design of the master reader system.

1. Introduction

The RCA VideoDisc is a mass replicatable disc system from which prerecorded video and audio programming is retrieved by detecting the variations in capacitance between the disc surface and an electroded stylus in intimate contact with the disc surface.[1] The disc is approximately 30 cm in diameter and each side plays for one hour. The stylus follows a spiral groove with a pitch of 9541 grooves/inch while the disc rotates at 450 rpm (7.5 Hz). On each disc rotation, four TV frames are stored in an fm encoded form. On the innermost radius of 7.3 cm, the smallest recorded feature size is about 250 nm.

Although the system is designed for reading out the replicated discs with a physically contacting stylus, it is not practical to monitor the quality of the master recording by a physical-contact readout technique. Contact monitoring might very well be detrimental to the quality of the master recording and thus to the final production replicas. Similar considerations apply to the several generations of metal parts between the master recording and the final production replica. Nevertheless, it is extremely important to verify the quality of the master recording before committing the master to the replication process. Without this capability, a full week of recording production might elapse before it could be determined that a generic defect existed in the master recordings. For this reason it became imperative to develop a nondestructive contactless quality-verifying technique for the masters.

This paper does not describe the results of an isolated development effort. Rather it documents work performed within the atmosphere of the total VideoDisc development effort and, therefore, it includes the direct and indirect contributions of many individuals. In the course of this paper we will distinguish the RCA VideoDisc from its optical counterparts with regard to optical readout, discuss the underlying optical concepts for optical playback of the RCA VideoDisc, review the fundamental and experimental capabilities of lenses operating near the fundamental limits, the design and construction of a VideoDisc master reader, and, the results achieved by it. A summary of this work is included within an overview paper published in the March 1978 issue of *RCA Review*.[2]

2. Optical Principles

The RCA VideoDisc differs from typical optical video disc systems[3–5] in at least three respects important to optical readout. *First*, the disc rotates at 450 rpm—four times slower than typical optical video discs. Thus each groove convolution contains four 33 msec television frames. *Second*, the fm encoding system,[6] while extremely efficient, nonetheless requires a bandwidth extending from several hundred kilohertz to about 6.9 MHz. Together, the rotation speed and bandwidth require signal features as small as 250 nm at the innermost recorded radius. *Third*, the RCA VideoDisc surface is not flat, but contains tracking grooves for guiding a capacitive stylus. The signal encoding elements extend completely across each groove, and the entire two-dimensional area of the disc is used to store the video programming. Fig. 1, which is a scanning electron microscope picture of a VideoDisc surface at high magnification, illustrates the geometry. We shall subsequently explain why such a format, in which there are no blank, intertrack regions, poses a resolution problem for optical playback.

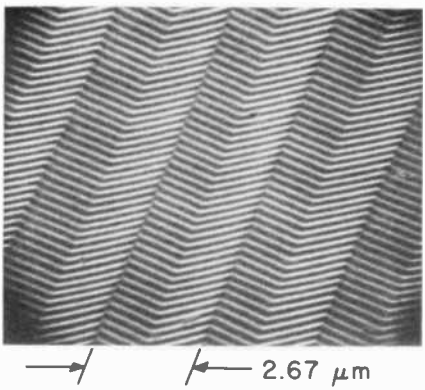


Fig. 1—Scanning electron photomicrograph of the surface of an electromechanically recorded VideoDisc master.

In contrast, optical disc systems operate at a rotational speed of 1800 rpm (30 Hz), or one TV frame per revolution, and have a larger bandwidth.[2,7] They have a minimum signal pit length along the track of approximately 600 nm. Although the tracks are spaced ~1.4 μm apart, the pit width is only ~0.8 μm. Thus, a substantial portion of an optical-format disc comprises a flat reference surface. In typical optical playback systems, the readout signal is derived from the interference of light scattered from the reference surface with the light scattered by the signal relief structures. The significance of this fact and the other differences outlined here will become more apparent after we discuss the principles underlying our optical readout method for the RCA VideoDisc format.

RCA VideoDiscs can be mastered electromechanically, optically, or by means of an electron beam.[8] The finished master disc produced by any of these methods is highly reflective. The electromechanical master is cut directly into the metal substrate, while the optical and electron-beam-exposed masters are overcoated with evaporated gold prior to an electrochemical plating operation. Diffractive scattering results when, upon reflection, the smooth wavefronts of an incident beam are altered in phase by the pattern of signal features on the master-disc surface. Since there is negligible absorption of light by the features, an optical reader must respond to the angular variations of the scattered light. Similarly, the typical optical discs, both reflective and transmissive types, also act as phase objects. Thus both optical video disc readout and RCA VideoDisc master readout share common, underlying, optical principles.

Fig. 2 depicts the principal components of a general optical pickup

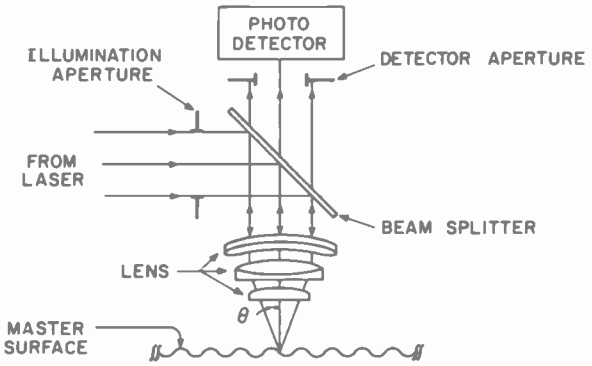


Fig. 2—Basic optical reader system.

system. A nearly diffraction-limited lens, such as a good-quality microscope objective, is used to focus a laser beam of suitable diameter and wavelength λ to a small spot at the record surface. The returning light is collected by the same lens, passes through the beam splitter, and impinges upon a detector which generates a photocurrent proportional to the total radiant flux on its surface. Both the illumination aperture and the detector aperture are in the back focal plane of the lens. In actual lenses, these apertures are often formed by the same physical aperture located within the lens mounting, and the simple configuration depicted in Fig. 2 cannot be realized. Under these circumstances, an auxiliary lens may be used above the beam splitter to project a real image of this aperture onto the detector. The illumination aperture remains as depicted, since the internal physical aperture will simply truncate the incoming beam. For such actual lenses, the maximum detector size is determined by the image of the actual aperture—the effect is as if the detector were situated at that aperture. It should be emphasized that it is *not* an image of the disc surface that is projected onto the photodetector; the photodetector is at an image plane of the lens aperture.

In Fig. 2, the half-angle of the converging cone of light is θ; the numerical aperture, a, of the lens is defined in this case by the relation

$$a = \sin\theta. \qquad [1]$$

This quantity and the wavelength λ fundamentally determine the resolving power of each pickup. If the illumination aperture is uniformly filled, the intensity profile of the focused beam is the Airy diffraction pattern,[9] comprising a central bright area surrounded by alternating dark and bright annular sidelobes whose intensity diminishes rapidly with distance from the center. Approximately half the light flux lies inside a circle of diameter d,

$$d = \lambda/2a. \qquad [2]$$

This circle is also approximately the contour where the intensity of the central area falls to one-half its peak value.

Let us idealize a signal waveform on the VideoDisc master as a periodic relief structure of spatial period L in a flat-bottomed groove. If the wavelength λ and the numerical aperture a are chosen such that d is significantly smaller than the track width, then the signal track appears as a one-dimensional (i.e., edgeless) diffraction grating for the focused beam. The consequences of this situation are sketched in Fig. 3. For illustrative purposes the diffraction pattern produced is more easily sketched for a transmissive phase grating rather than a reflective one. Within the approximations of scalar diffraction theory, upon which all our discussion here is based, transmissive and reflective phase gratings are equivalent when they impart the same spatial phase variation to the readout beam. Three of the diffraction orders (-1, 0, and $+1$) are represented by the three cones of light leaving the record. The central ray (c) of the incident cone is diffracted into the two first orders at an angle β_c from the normal. From the fundamental grating equation, we find

$$\sin\beta_c = \lambda/L. \qquad [3a]$$

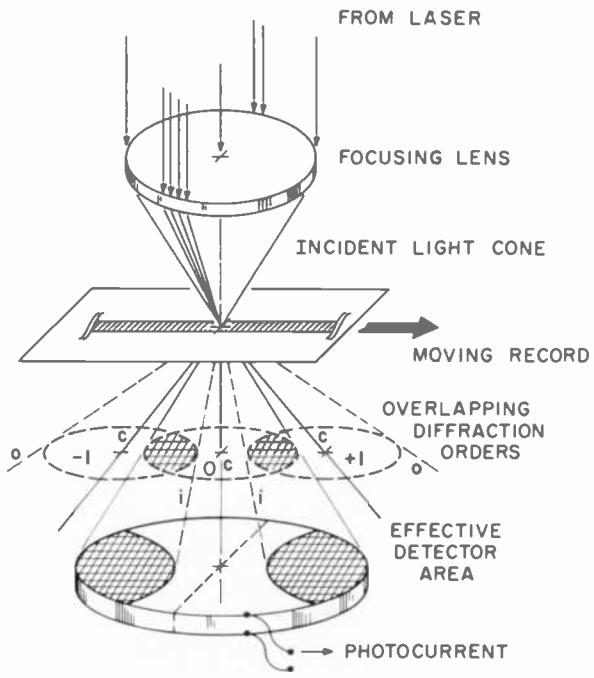


Fig. 3—Illustrating the principles of an optical pickup system. The incident light cone is diffracted by the master waveform into Doppler-shifted orders which overlap on the detector.

Similarly, the innermost (i) and outermost (o) edges of the first-order cones occur at the angles β_i and β_o, respectively, where

$$\sin\beta_i + \sin\theta = \lambda/L. \quad [3b]$$

$$\sin\beta_o - \sin\theta = \lambda/L. \quad [3c]$$

As a consequence of imposing the Abbé sine condition[9] in the design of the lens, rays in the back focal plane of the collecting lens are found at a distance from the optical axis that is proportional to the sine of the angle at which they entered the lens. Thus, the diffracted cones form a series of circles on this plane. The zero-order rays, which are undeviated by the grating, simply form the projected image of the illumination aperture; the higher order rays form offset images of that aperture.

The relative phases of the diffraction orders depend on the shape and depth of the relief grating representing the recorded signal and on the displacement x of the signal track in its plane as the disc rotates below the focused optical beam. The amplitude and phase of the nth order wavefront are described by the complex phasor $\tilde{U}_n$, which for a *reflective, square-wave, relief grating* with a peak-to-peak depth variation of h, is given by

$$\tilde{U}_o = \cos(\phi/2), \text{ and} \quad [4a]$$

$$\tilde{U}_n = \frac{2}{n\pi}\sin(n\pi/2)\sin(\phi/2)\exp\left[-i\pi\left(\frac{2nx}{L} - \frac{1}{2}\right)\right],$$

$$n = \pm 1, \pm 2, \pm 3, \ldots \quad [4b]$$

where

$$\phi = 4\pi h/\lambda \quad [5]$$

is the peak-to-peak phase excursion of the grating profile, and $x = 0$ occurs when the center of the focused spot is midway between two pits. The derivation of Eq. [4] is shown in the Appendix.

Note that the amplitudes and phases given by Eq. [4] are relative to those that would be obtained from a hypothetical smooth reflecting surface located halfway between the upper and lower level of the square-wave relief. Furthermore, only the zero order and odd-numbered orders are present for square-wave relief gratings, and the magnitude of each order in Eq. [4] is independent of the grating displacement. Only the phases change as the record rotates. The phase of an order varies linearly with the product of the displacement and the order number. The real value for $\tilde{U}_o$ means that this order appears to be reflected from the hypothetical reference plane.

When these diffraction orders impinge upon the photodetector, it is the light intensity that generates the photocurrent from which the

grating displacement can be determined. Accordingly we must consider the absolute squared magnitude of the sum of all the diffraction order amplitudes impinging upon the photodetector. In regions where light from at least two overlapping diffraction orders exists, as shown by cross-hatched areas in Fig. 3, the variations in phase caused by continuous motion of the grating will provide cyclic instances of constructive and destructive interference. The mathematical expression for the intensity variation due to two overlapping orders m and n, where $n > m$, is

$$I_{n,m}(x) = |\tilde{U}_m|^2 + |\tilde{U}_n|^2 + |\tilde{U}_m|\,|U_n| \cos[2(n-m)(\pi x/L) + \psi_n - \psi_m], \quad [6]$$

where the ψ's are the phases of the two diffraction orders for $x = 0$. Note that $\psi_n = \pi/2$ for all $n \neq 0$; and $\psi_o = 0$ for square-wave relief gratings, whereas for more general relief structures the ψ's will take on other values. Thus, there are $(n - m)$ cycles of destructive and constructive interference each time the grating translates one period. In regions where there is light flux from only one diffraction order, there is no change in light intensity as the record moves. Where more than two orders overlap, the intensity variations contain a variety of higher harmonic terms as well.

These effects are illustrated in Fig. 4 by photos taken in an experimental arrangement analogous to Fig. 3. A one-dimensional transmission grating having a nearly square phase profile of 4 μm period was used. With a low numerical aperture illumination lens, diffraction outside the zero-order cone was observable. The three overlapping diffraction orders in each photo are -1, 0, and $+1$ from left to right. A pair of extremely weak second orders can be just discerned in plates (b) and (d) of Fig. 4 by their interference with the first orders. In the sequence a-b-c-d-a, the grating is successively translated a distance of $L/4$ each step through one complete period. The successive constructive and destructive interference effects in the overlapping areas of the zero and first orders are clearly evident, as is the constancy of the intensity in the nonoverlapping regions.

If the grating used for the photos of Fig. 4 were of longer period, there would be less separation between diffraction orders, and more orders could overlap in a given region. We have noted above that the photodetector in a practical reader can be no larger than the zero-order area. Each unique combination of diffraction orders occupies a unique region on the photodetector surface and produces a characteristic intensity variation with grating displacement. For a given grating period, summing the products of area and intensity over all interfering sets of diffraction orders on the photodetector gives a quantity proportional to the net

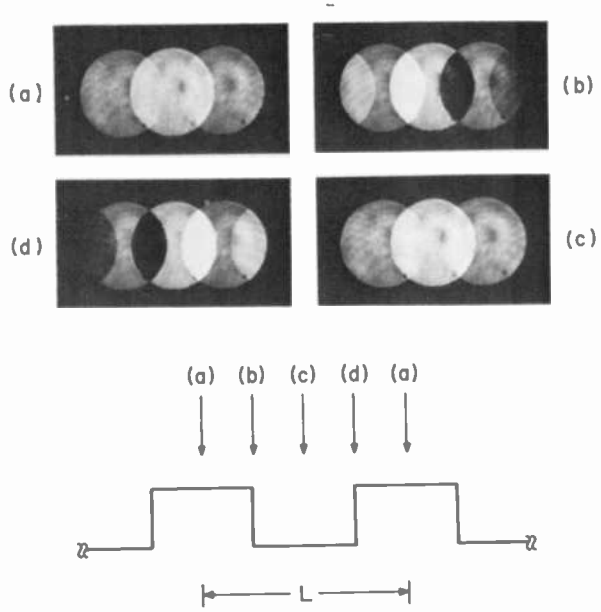


Fig. 4—Interference of diffraction orders as phase grating translates.

photocurrent. Repeating this process for a variety of displacements and grating periods yields waveforms from which the frequency response characteristic of the reader may be calculated. The Appendix outlines the procedure for this calculation.

Without calculating exact frequency response characteristics, we can demonstrate that the typical optical video disc pickup system has severe frequency limitations when reading out the wide signal track that we have been modeling. Let us look more closely at the photocurrents produced when the +1, 0, and −1 orders impinge upon a single, centered, symmetrical photodetector as is used in typical optical video disc readers. From Eqs. [4] and [6], the order overlap intensities are expressed by

$$I_{+1,0}(x) = \cos^2(\phi/2) + (4/\pi^2)\sin^2(\phi/2) - (2/\pi)\sin\phi\,\sin(2\pi x/L) \quad [7a]$$

$$I_{0,-1}(x) = \cos^2(\phi/2) + (4/\pi^2)\sin^2(\phi/2) + (2/\pi)\sin\phi\,\sin(2\pi x/L). \quad [7b]$$

The fact that the x-dependent intensity variations in these two regions are completely out of phase with each other is apparent in Fig. 4, and is a *general* characteristic of *one-dimensional* phase gratings whose profiles, when expressed as a Fourier series, possess no harmonic components of even order. When a single, centered, symmetrical photodetector

is used, such a detector intercepts equal areas of the (+1, 0) and (0, −1) overlap regions. By equating β_c with θ in Eq. [3a] and combining with Eq. [1], we see that for grating periods shorter than λ/a, the two first orders do not overlap and the *net* photocurrent from the detector does not vary with x. When $L \geq \lambda/a$, there is a region in the center of the zero order which also contains overlapping portions of the −1 and +1 orders. The intensity here is obtained by summing $\tilde{U}_{-1}$, $\tilde{U}_0$, and $\tilde{U}_{+1}$, and squaring the modulus.

$$I_{+1,0,-1}(x) = \cos^2(\phi/2) + (2/\pi^2)\sin^2(\phi/2) + (8/\pi^2)\sin^2(\phi/2)\cos(4\pi x/L). \quad [8]$$

This intensity variation has a period $L/2$. For any grating period longer than λ/a, the *net* photocurrent generated by a single centered photodetector consists of a constant term, a fundamental component of period $L/2$, and higher harmonics of the $L/2$ period. The double-frequency signal means that such a reader detects the leading and trailing edges of the signal pits passing through the center of the focused spot as a dip in photocurrent—but there is no distinction between the directions of the edge transitions.

Curve A in Fig. 5 shows the calculated peak-to-peak photocurrent as a function of normalized spatial frequency $\lambda/(2aL)$ for a single circular photodetector that fills the zero order. A pit depth of $\lambda/8$ is assumed here; the readout signal is normalized to the dc photocurrent that would be generated by a featureless record. At a given spatial frequency, the signal strength is proportional to $\sin^2(\phi/2)$, which is maximum for a pit depth

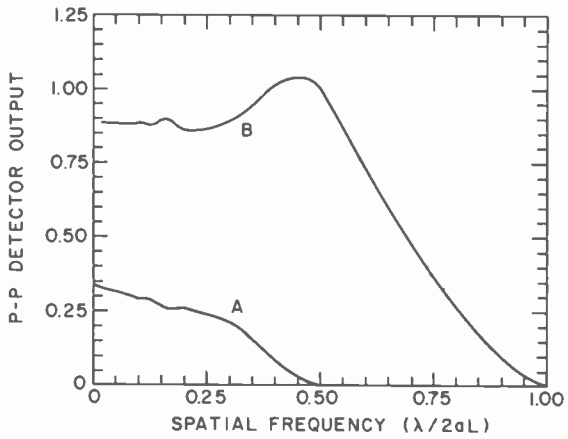


Fig. 5—Peak-to-peak photocurrents calculated for square-wave phase gratings versus spatial frequency. Curve A is for a single circular photodetector and Curve B for a differential photodetector.

of $\lambda/4$. Although microscope objectives are available with numerical apertures approaching unity, in actual performance they are typically no better than $a = 0.8$. With such a lens, a single centered photodetector system would be limited to a shortest grating period of 1000 nm with a laser wavelength of 820 nm (AlGaAs), 790 nm period with a laser wavelength of 633 nm (HeNe), and 552 nm period with a laser wavelength of 442 nm (HeCd). Thus the effective resolution limit is just larger than the smallest signal period on the RCA VideoDisc, and therefore the conventional detector configuration with visible and near infrared laser sources cannot be used for optically reading such discs.

A solution to the resolution problem is to divide the photodetector into two halves, along the dotted line shown in Fig. 3, and subtract the photocurrent induced in one half from that induced in the other half. The two signal terms in Eq. 7 will then reinforce each other, giving an output whose fundamental period is that of the grating. The upper resolution limit for the split-photodetector optical system occurs when the grating period L is so short that the first-order diffraction cones no longer overlap the zero order; the light intensity everywhere in the back focal plane is constant as the grating translates. Equating β_i and θ in Eq. [3b] and combining Eq. [3b] with Eq. [1] gives a relation for the period L_c at which this cutoff occurs:

$$L_c = \lambda/2a. \qquad [9]$$

It is also apparent from these considerations that maximum resolution is obtained by utilizing a detector as large as the projected image of the illumination aperture in order to sense the slightest possible overlap between diffraction orders. However, no additional resolution is to be gained by increasing the detector size further. Curve B in Fig. 5 shows the calculated frequency response for the split photodetector system using the same normalizing factor as curve A. Note that the resolution has been increased twofold to the fundamental limit compared with the single centered photodetector of typical optical video disc pickup systems. With this split photodetector signal, not only is optical readout of RCA VideoDisc masters now practical with both the HeCd and HeNe gas lasers, but it is just barely possible with the 820 nm AlGaAs semiconductor junction laser as well.

When the grating period is substantially longer than L_c and many overlapping diffraction orders contribute to the signal, the differential photodetector response peaks sharply as the steps of the square-wave pass through the focused spot. The peaks, however, alternate in polarity for leading and trailing edges of the signal pits. This type of signal has been shown[10,11] to be related to the derivative of the surface profile. Additionally, the differential output varies with grating depth as $\sin\phi$,

and is *considerably larger* than the output from a single centered detector for $h \leq \lambda/8$.

The tolerance of an optical reader to variations in the spacing between the lens and the grating surface can also be understood in terms of the overlapping diffraction orders. The photos in Fig. 6 use the same phase grating and apparatus as those used to produce Fig. 4. The solid rays leaving the grating in the sketches outline the zero-order cone; the dashed rays are for the +1 order. With perfect focus, the phase difference between two overlapping wavefronts is constant because the incident wave has a constant phase at the grating surface; the diffracted wavefronts appear to originate from a common point at the grating, as suggested by the ray diagram. The degree of constructive or destructive interference is therefore constant wherever the wavefronts coincide.

In the case of a focussing error, as indicated in Fig. 6(b), the incident rays no longer meet the grating at a point—there is a phase variation in the incident wave at the grating surface. The sketch shows both a zero-order ray and a +1 order ray for each of the outermost rays of the incident cone. Extending these rays above the surface of the grating shows that the zero-order and first-order diffraction cones appear to stem from two virtual, laterally displaced points. Thus, the wavefronts have different curvatures, and their phase difference is no longer constant where they coincide. The focus error illustrated here is a very severe one and clearly shows the pattern of interference within a single overlap region. As the focus error decreases, the interference fringes caused by the linearly varying phase difference between the overlapping wavefronts become more widely spaced; equivalently, there is less difference in wavefront curvature.

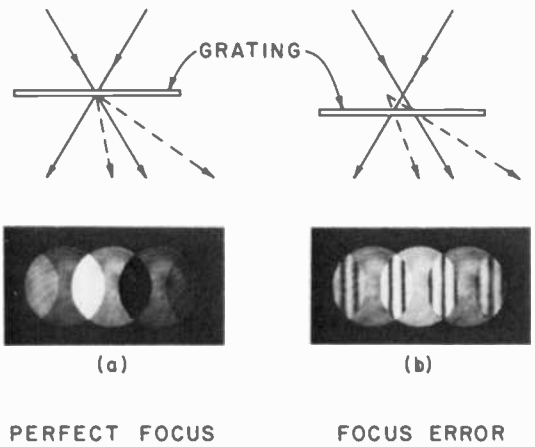


Fig. 6—Effects of defocus on the interference of the overlapping diffraction orders.

When the grating is in motion, the defocus fringe pattern is observed to also translate at the rate of one fringe period per period L of grating motion. When one or more complete fringe periods occupies the overlap area of adjacent diffraction orders, there is little change in the total light flux over each individual area as the grating moves. Consequently, the reader response drops to zero for severe focus errors. Note that if the numerical aperture in Fig. 6(b) were reduced so that the grating is just resolved, there would be fewer fringes in the overlap area and a larger detector response. At $L = 2L_c$, the overlap area between pairs of orders that contribute to the split-detector output is a maximum, and thus at this periodicity the system is extremely sensitive to focus errors. Alternatively, when many diffraction orders overlap, as in the case of a low spatial frequency, the fundamental signal contribution comes from the region where only the zero and first orders overlap and the two first orders have no mutual overlap. This is a narrow crescent shaped region near the edge of the pupil. Since this region is narrow, the defocus fringe pattern must be fine in order that several fringes might exist within this region. Thus, small focus errors produce little change in signal strength for grating periods that are either very long or near the resolution limit. Fig. 7 shows the calculated frequency response characteristics of the differential detector arrangement for various defocussing distances z. The effect of defocussing is clearly most evident for grating periods around twice the fundamental resolution limit.

Let us return for a moment to the question of resolution. We have

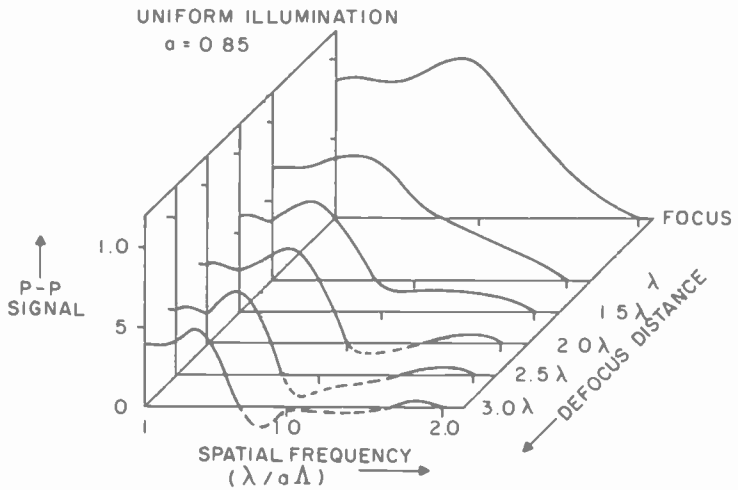


Fig. 7—Calculated peak-to-peak differential detector output versus spatial frequency and defocusing.

shown that the intensity fluctuations at the photodetector result from the interference between diffracted and undiffracted, or zero-order, light. For the RCA VideoDisc format, the apparent reference surface for the zero order is a plane midway between the top and bottom of the signal pits. If the phase of the zero order wavefront only were suitably shifted, the resulting $I_{1,0}(x)$ and $I_{0,-1}(x)$ intensity waveforms would be largely in phase with each other, and a single centered photodetector could be used to resolve the grating structure to the fundamental limit. Although such a zero-order phase shift can be achieved with an interference lens system, such a solution imposes impractical tolerances on the lens-to-disc spacing.

In the case of the disc systems designed specifically for optical readout, there is diffraction not only from the leading and trailing edges of the signal pits, but from the sides of the pits as well. The comparatively wide flat areas between tracks on these discs influences the zero diffraction order so that the effective reference surface is quite close to the actual surface. There is thus a strong in-phase component in the cyclic interference between the zero order and the two first orders having direction components along the track, a component easily sensed with a single centered detector. In effect, the flat surface of the optical disc can be viewed as providing a stable interferometric reference mirror. Resolution of the surface features up to the fundamental limit specified by Eq. [9] is obtained directly with a single centered detector for these optical video discs.

Note that it is only the RCA capacitive disc format, which acts as an edgeless diffraction grating in an optical reader, that requires the use of special phase-contrast methods for full resolution. The differential detector provides a form of differential phase contrast that is particularly simple to implement. Although differential or off-centered detectors can also be successfully used for typical optical video disc formats, the reasons for such a choice involve considerations such as pit depth and surface noise, rather than resolution. Typical readers for optical video discs use a numerical aperture around 0.4 and a He-Ne laser with $\lambda = 0.633\ \mu$m to detect minimum signal periods near 1.8 μm. For the RCA VideoDisc masters, choosing $a = 0.8$ and $\lambda = 0.442\ \mu$m provides resolution of equal quality for signal periods more than three times smaller.

In concluding this discussion of the underlying optical principles, we should also comment briefly on the applicability of scalar diffraction theory, from which we obtained Eq. [4], to surface features comparable to the wavelength of light. When rigorous diffraction theory is used, only the magnitudes of the diffraction orders and the phase terms denoted by ψ in Eq. [6] are changed. The diffraction angles, along with the linear dependence of the phase of the diffracted light on the order number and

the grating displacement, remain the same as predicted by scalar calculations. Thus the basic conceptual description, as well as the quantitative resolution limits that we have presented here are soundly based. The quantitative calculations of signal strength, which depend on the magnitudes of the diffraction orders, become suspect when the scalar theory predicts that a significant fraction of the incident energy is diffracted at angles whose sine is greater than one. In the case of signal profiles much shallower than $\lambda/8$, the incident energy remains predominantly in the zero order; the scalar theory then provides remarkably accurate predictions near the resolution limit of the highest numerical apertures. For VideoDisc signal depths actually encountered, optical output signal-to-noise ratio is not completely explained by the scalar theory.

3. High-Numerical-Aperture Lenses

The previous discussion of optical principles has shown the dependence of the fundamental resolution limit and readout frequency response upon the numerical aperture of the focusing lens. Although microscope objectives with numerical apertures approaching unity are available, their actual performance is equivalent to substantially smaller numerical apertures.

The implementation of optical readout requires an optical lens capable of forming a focused spot whose length along the track is less than the spatial wavelength of the signal to be readout. We have developed a knife-edge scanning technique[12] that measures the focused spot size in one-dimension. This is particularly appropriate for RCA VideoDisc applications since the signal is essentially also one-dimensional. We have used this technique, at a number of wavelengths, to measure several commercially-available, high-quality, high numerical aperture, microscope objectives. Fig. 8 show the effective one-dimensional focal intensity distribution at 633 nm wavelength for a 0.95 numerical aperture lens that is typical of those we have used in optical readout systems. Fig. 9 presents this same data in the form of a one-dimensional modulation transfer function (MTF). This of course is the MTF of an incoherent optical system, and as we have shown in the Appendix, it is similar in its high frequency region to the frequency response of split-photodetector optical readout systems. Curve A is the modulation transfer function of the lens as derived from the data presented in Fig. 8. Curve B is the modulation transfer of a uniformly illuminated, perfect, diffraction-limited lens with a 0.8 numerical aperture. These measurements are indicative of the effective numerical aperture of the lens, and they are an extremely useful tool for checking lens quality. In the actual reader the lens is not uni-

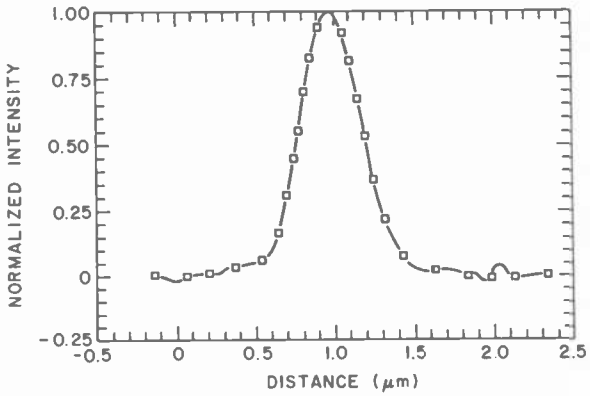


Fig. 8—Measured effective one-dimensional focal intensity distribution of microscope objective with $a = 0.95$ at a wavelength of 633 nm.

formly illuminated; however, measurements under actual conditions are rather difficult.

4. Optical Reader Design and Implementation

In the previous sections, we have discussed the underlying optical principles of RCA VideoDisc optical readout and the experimental capabilities of available high numerical aperture lenses. However, to go from optical readout in principle to optical readout in practice requires the solution to a series of design problems—the high numerical aperture

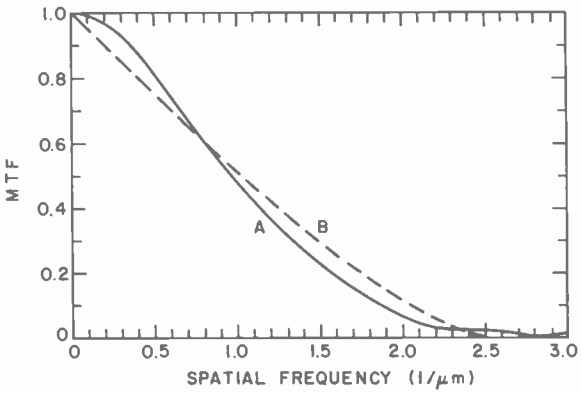


Fig. 9—Modulation transfer function of lens at 633 nm. Curve A shows the MTF of lens data in Fig. 8; curve B the MFT of a uniformly illuminated diffraction-limited lens with $a = 0.8$.

lens must maintain focus to better than 300 nm on a spinning master VideoDisc; even with reasonably small disc eccentricity, the focused spot must be maintained within the spiral track with a servo system; the signal quality should be limited by the master disc and not the optical reader; and finally the reader should be safe, reliable, and easily operated by pilot-production personnel. In this section we will discuss the design and implementation of the VideoDisc master reader.

The basic optical reader system consists of five major subsystems; (1) optics to suitably shape the laser beam for illumination of the high numerical aperture lens, (2) a groove tracking subsystem to keep the spot within the signal track despite eccentricity or pitch variations of the VideoDisc master, (3) a focus subsystem to maintain the spot in focus on the master surface, (4) a precise overall structure to reliably maintain accurate alignment and operation, and (5) a photodetector subsystem to convert the optical signals into electrical signals.

4.1 Optics

To make optimal use of the wide signal format of the RCA VideoDisc, the interrogating optical beam should extend across the width of the signal features. If we assume that one noise component is an isotropic surface roughness on the master VideoDisc, then the signal-to-noise ratio of the readout will increase as the square root of the interrogating beam width. We will see later however, that if the beam is made too wide, the tracking signals will be unreliable. To form an interrogating beam that is wide transverse to the grooves and yet is sufficiently narrow along the groove to read out the smallest encoding signal features, it is necessary that the optics for illuminating the readout lens be anamorphic. The readout lens is thus not uniformly illuminated.

Fig. 10 outlines the optical system we have used. The laser is a 2-mW polarized helium-neon laser. Its output beam is expanded by a ×2 Galilean telescope. This beam is then about 2 mm in diameter, and it is further expanded by a commercial eight-to-one anamorphic beam expander utilizing a sequence of prisms.[13] This wide beam is reflected into the focusing lens pupil by a 50% reflecting dielectric mirror mounted on a scanning galvanometer.

In its wide direction, this beam overfills the lens, thereby forming a highly focused rectangular spot on the VideoDisc surface. Note that the beam is sharply focused only in the narrow dimension along the track. While our system is in general astigmatic, we do not pay particular attention to the beam curvature in the wide dimension across the groove.

The spot-forming lens is a 0.95 numerical aperture metallurgical mi-

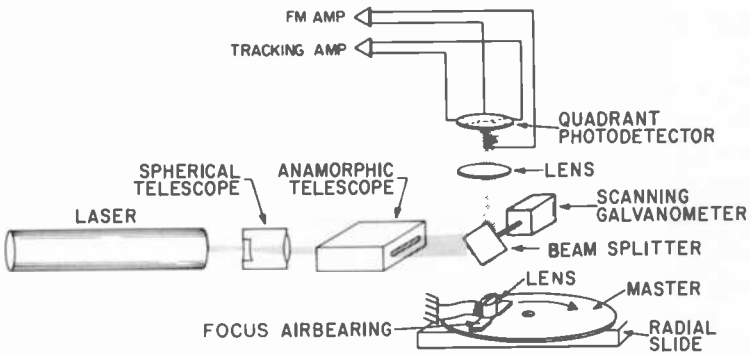


Fig. 10—Optical-reader-system outline.

croscope objective modified to mount in a structure whose function is to keep the lens at a constant distance from the master surface. The light reflected and diffracted by the VideoDisc surface is collected by the same spot-forming lens and again encounters the partially reflecting galvanometer mirror. The light transmitted by this mirror is focused by the photodetector lens to form a reduced image of the spot-forming lens pupil on the photodetector. We use a single quadrant photodetector for both the fm signal detector and the optical tracking sensor.

4.2 Tracking

In the mastering process, masters are routinely centered on the recorder turntable to within a few microns. Similarly masters are centered on the optical reader just as readily to the same precision. All masters contain a closed groove outside of the recorded area. By observing this groove through a microscope while slowly rotating the master, it is simple to quickly center the master relative to the turntable axis. Nonetheless it is necessary for the reader system to incorporate a tracking subsystem. The initial centering accuracy is still typically more than one groove eccentric. Furthermore during the one-hour-per-side play duration, master discs tend to shift on the turntable and change their eccentricity.

Radial motion of the focused spot on the master is accomplished by means of a partially reflecting mirror mounted on the scanning galvanometer. Rotation of the galvanometer changes the angle of incidence in the radial plane of the light illuminating the focusing-lens entrance pupil. This moves the focused spot across the grooves. It provides beam tracking and keeps the interrogating spot within the groove independent of disc eccentricity. Two diametrically opposed quadrants of the photodetector drive a differential amplifier to provide a tracking-error signal.

Fig. 11 shows the detector and its orientation with respect to the grooves and signal elements.

The stylus guiding groove facilitates straightforward optical tracking on the RCA VideoDisc. This is most simply understood from the viewpoint of geometric optics. When the focused spot rides off to one side of track center, as illustrated in Fig. 11, the reflected beam is deviated to the opposite side. Thus, by using a detector with four segments, it is possible to derive both the video signals and the tracking signals from the reflected light.

It should be noted that in our optical system the scanning galvanometer is not imaged onto the photodetector plane; motion of the galvanometer moves the light pattern on the photodetector surface. Normally this would result in reduced tracking range and slight tracking errors; however, because we have so little eccentricity to deal with, and because we prefer to operate close to the field center of our high numerical aperture lens, this is not a problem. The difference of the photocurrents generated by segments C and D provides optical-tracking-error signals. This error signal is amplified to drive the scanning galvanometer. The amplifier electrically compensates for the mechanical resonances of the galvanometer–mirror mechanism to provide a stable servo system with a gain of approximately 55 dB at 7.5 Hz. On a master there are no track distortions, and the only tracking error is caused by the eccentricity, which occurs at the disc rotation rate.

In the design of the reader, it is convenient not to attempt to drive the optical beam inward to follow precisely the spiral track. Rather, the speed of the radial motor drive is set to approximately follow the spiral, and a servo system is used to correct this average speed. This is a narrow-bandwidth servo system, which uses the low-frequency component of the drive current to the scanning galvanometer as its error input.

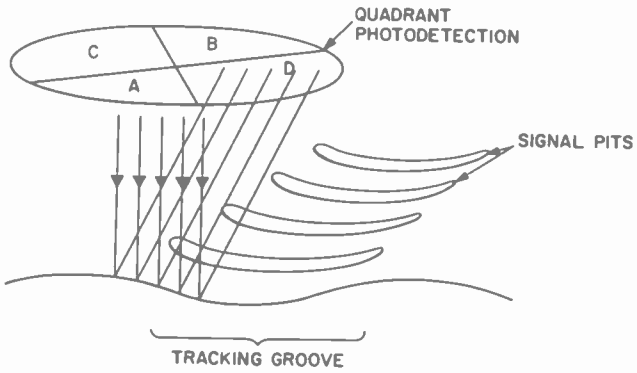


Fig. 11—Tracking sensor geometry.

Although the master reader does compensate for eccentricity by tracking the groove guiding spiral, it does not compensate for the time base errors caused by this eccentricity. For example, a 7 micrometer peak-to-peak eccentricity results in a 0.01% peak deviation of the horizontal sync and color burst reference frequencies. However, at the disc rotation rate of 7.5 Hz, this causes no visible degradation on an average TV set.

4.3 Focus Control

The most critical subsystem in the master reader design is the focus control system. A master on the turntable rests on three balls which can be raised and lowered individually by fine pitched screws. It fits loosely over a center pin. To level a master, an arm is rotated to a precise detent position to place a commercial, capacitive, separation-measuring device over the edge of the master. The three leveling balls are then adjusted to vertically position the master's surface. Masters are of varying thickness, but all masters are adjusted so that the recorded surface is raised to the same absolute height to within 500 nm. However substrates are frequently warped and tend to sag between the supports, resulting in a total peak-to-peak vertical runout of typically 5 μm. Focus has to be maintained on this master to better than 300 nm; this is the function of the structure in which the focusing lens is mounted. This structure is basically a hydrostatic air bearing that supports the focusing lens just above the surface of the master. A detailed drawing is shown in Fig. 12. There are two air pads, one just before and one just after the lens. These are pressurized chambers with porous graphite faces. The faces are lapped smooth with respect to a leveled master surface. Precision machining provides a thin-walled lightweight aluminum assembly. The lens itself is remounted in a light tube that can be screwed up or down with respect to the pads. Including the lens, the total moving mass is approximately 45 g.

When the air bearing is lowered for readout it is pressed toward the master by a parallelogram spring arrangement. The pads are supplied with air at a pressure of about 60 psi which prevents contact as the structure is lowered toward the master surface. Adjustment of the spring pressure to a preload force of about 70 g places the air bearing in an optimum "flying" mode that requires only occasional fine adjustment since each master is leveled to the same height. Coarse optical focusing is accomplished by screwing the lens up or down within the structure. It is essentially a one-time procedure. The air bearing's operation is based on the fact that the "air spring" under the pads is much stiffer than the parallelogram "preload" spring. Thus, as the master surface rises and

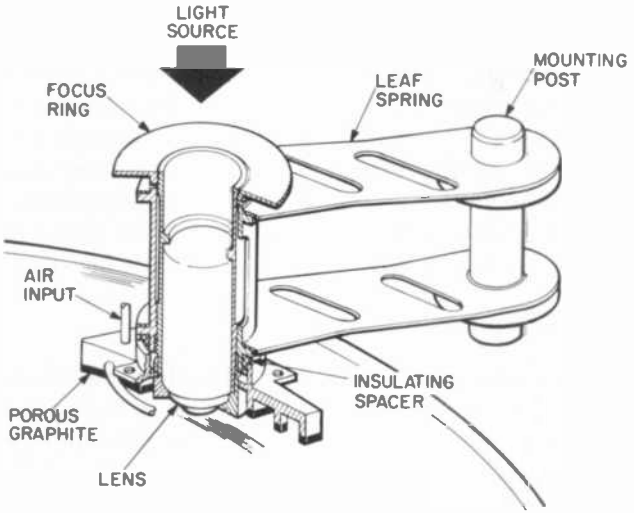


Fig. 12—The focusing-lens air bearing.

falls under the air bearing due to warp and sag, the softer spring does most of the flexing and the "air spring" closely maintains its length. If we model the structure as a mass suspended by two springs, as diagrammed in Fig. 13, then at frequencies well below the natural resonance of the system, the system gain A_o is simply the ratio of the two spring constants. The significance of this gain, A_o, is that the disc-to-lens distance varies by only $1/A_o$ of the vertical variations of the disc surface. In actuality, the spring formed by the hydrostatic air bearing is highly nonlinear and under some conditions the system will be very lightly damped or perhaps even oscillatory. In this regard the two-pad structure

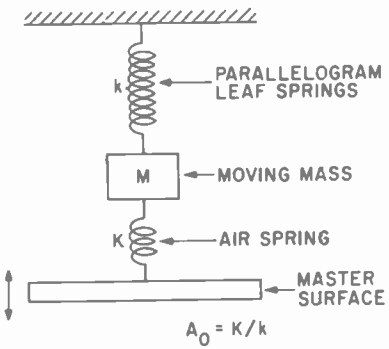


Fig. 13—Air-bearing mechanical model.

is highly tolerant and reliable. We have found this design to be superior to other circularly symmetric designs with which we have experimented. These air bearings typically have gains of 30 to 50. Thus the 5 μm peak-to-peak vertical variations of the disc surface are reduced to 130 to 100 nm variations in focus. Although this is a relatively low gain system, it is adequate for the task and is extremely reliable and tolerant. In the actual reader system, to preclude the possibility of damage to the master under test, there is an automatic device to lift the bearing in the event of air failure. Also, a dashpot assures a gentle lowering.

4.4 Overall Mechanical Structure

The overall mechanical design of the optical-reader structure emphasizes stability in the fixed components and precision in the moving components. The approach was to assure the best fabrication that was possible in the Laboratories' model shop. A photograph of the optical master reader and its accompanying electronics is shown in Fig. 14. The machine is built on a 2 × 4 foot flat slab with a honeycomb core. This provides a rigid base for the optical bench, turntable slide, and leveling-centering

Fig. 14—Optical reader for VideoDisc masters.

station. The base is supported by shock mounts resting on a frame that contains the hydraulic system for the turntable bearing. The frame also contains an electrical wiring panel and connections for all power, control, and air to the reader. A double rack of electronics stands next to the machine and contains a color TV monitor, diagnostic displays, as well as the necessary control and decoding circuitry.

The optical bench is constructed of aluminum jig plate and has a cantilevered section that allows the turntable to move under the focused beam for readout. The laser and beam expanding optics have been described above. Immediately following the laser is an electrically operated shutter that opens only when the focusing lens mounted in the air bearing is lowered for readout. All adjustable movements of the galvanometer that operates the scanning mirror are centered about the optical-beam axis rather than the galvanometer itself. Focus of the reflected light, with its signal and tracking information, onto the quadrant photodetector is assured by providing x-y, and axial adjustments.

We translate the turntable under the optics to provide radial scanning of the recording. A variable speed dc motor drives the lead-screw of a dovetail slide carrying the turntable. Cams on the slide actuate a system of interlocks to prevent damage to the focus air bearing or drive. When the radial drive is clear of the optics it stops in the load/unload position. Here the master is centered and leveled to the correct height.

The turntable bearing is a hydrostatic oil bearing developed for electron beam recording[8] of VideoDiscs and runs smoothly with essentially no error. It is a glass sleeve that floats on the oil pads provided by a steel post. The rotor of a hysteresis synchronous motor was bored out and slipped over the glass. The field windings were then mounted around the rotor. In this way a direct drive to the turntable was obtained. The turntable weighs approximately 35 pounds. A hole in the turntable rim is detected optically to show the period of revolution, which is maintained to be constant within ± 10 microseconds.

In addition to the normal dust covers, there are interlocked light baffles to prevent the escape of any laser light. Basically they shield the light during readout and keep the shutter closed during loading and set-up.

4.5 Photodetection System

The photodetection system is based upon a quadrant, silicon, PIN photodetector oriented as shown in Fig. 11. Two diametrically opposed quadrants (C and D) provide the tracking signal, while the remaining two (A and B) provide the fm signal. This physical separation of the two functions facilitates the design of a low noise, wideband fm signal

preamplifier. The geometry of the fm photodetector elements affects the spatial frequency response of the reader system. All spatial frequencies higher than $\sqrt{2}\, a/\lambda$ will be unaffected, whereas all lower spatial frequencies will be attenuated. This is advantageous since it tends to flatten the system's spatial frequency response. For example, for a uniformly illuminated lens, the signal response is reduced by 0.64 at one half the cut off frequency (λ/a). In actual practice, however, not only do the detector and the lens illumination geometries affect the system frequency response, but so does the actual shape of the track bottom.

The differential preamplifier has a balanced FET front end. The bandwidth of the system extends from essentially dc to 10 MHz; the frequency response is flat to within 2 dB from 2 MHz to 8 MHz. The differential input insures that common-mode laser noise is not present in the detected signal. The noise equivalent power of the amplifier–photodetector unit at 633 nm is typically 25 picowatts/$(\text{Hz})^{1/2}$. A two milliwatt laser has always been sufficient to insure that the system is laser-shot-noise limited and not amplifier-noise limited. In general, the detected optical signal power is about 20 μW, and the system signal-to-noise is limited not by the reader but by the recorded master.

5. Operating Results

For experimental purposes we have constructed VideoDisc optical readers using 442-nm HeCd laser, the 633-nm HeNe laser, and 820-nm and 765-nm[14,15] AlGaAs lasers. As expected, the spatial frequency response at 442-nm laser wavelength is better than at the longer wavelengths. Optical readout with the AlGaAs lasers was rather sensitive to focus errors at the inner radii of the disc. With the 820-nm laser, there was picture degradation due to inadequate frequency response, whereas with the 765-nm laser the defocus tolerance was better and a stable 52-dB carrier-to-noise ratio (30 KHz bandwidth) was achieved on the inside radius.

For constructional and operating convenience we have concentrated primarily on reader systems incorporating the 633-nm HeNe laser. The performance and operation of both laboratory and pilot-production 633-nm optical readers are substantially similar. On the outside radii, at about 14 cm, the system response is flat to within several decibels from 1 to 10 MHz. On an inside radius, at about 7.3 cm, the system response is down about 5 db at 6.7 MHz, which is the peak instantaneous recorded frequency. These measurements are deduced from frequency-response measurements of the optical noise "floor" on a recorded master disc. Two typical results are shown in the two noise-floor measurements depicted in Fig. 15. Also shown are two corresponding frequency sweep readouts,

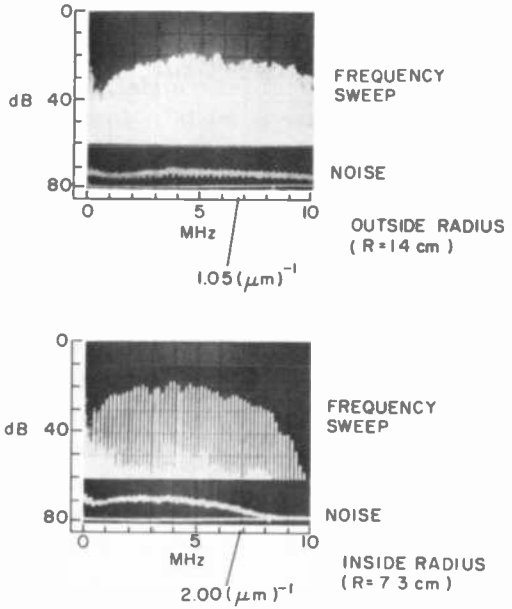


Fig. 15—System frequency response at disc radius of 14 cm (upper) and at disc radius of 7.3 cm (lower).

made by electromechanically recording 0–10 MHz reduced amplitude speed test signals. Under all circumstances the signal-to-noise ratio of the master-reader output is limited by the recorded master. At VideoDisc recording levels, carrier-to-noise ratios are typically in the range of 58–63 dB in a 30-KHz bandwidth.

Despite the fact that the optical system response varies with the disc radius and differs from the standard VideoDisc stylus, no modifications are necessary in the VideoDisc fm decoder system developed for standard players. Optically read masters play well from the outermost to the innermost radius. Nonetheless there are certain unique system characteristics observable when a master designed for capacitive stylus readout is read optically. The two aspects most notable in this context are the system behavior with regard to signal depth variations and signal track drop variations.

Over the signal depth range of interest, 70 to 150 nm, the signal output of the capacitive stylus is approximately proportional to the signal depth. Recall however that scalar optical theory suggests that the optical signal will be a maximum for a signal depth of ~80 nm ($\lambda/8$) and be reduced to zero for a signal depth of ~160 nm ($\lambda/4$). Experimentally this is not the case. Over this depth range, the signal-to-noise ratio of the reader output remains substantially constant. This serious experimental dis-

crepancy from the predictions of the scalar optical theory has necessitated the development of another optical technique for assessing signal depth[2,16] and the reexamination of the underlying optical principles within the context of an exact vector model.[17] In general, we have found a qualitative but not quantitative agreement with this vector model with regard to signal strength and polarization dependence versus signal depth.

Track drop is the term used to describe the peak height of the signal relief relative to the surface of an unrecorded stylus guiding groove. If the recorded master contains frequency dependent track drop variations across the fm encoding bandwidth, then, due to the capacitive stylus nonlinearity and physical asymmetry, spurious intermodulation signals will appear in the video playback. The optical probing beam is inherently symmetrical in the signal direction and thus track drop variations do not cause intermodulation signals in optical playback. This is an intrinsic difference between the optical reader and the capacitive stylus, and in this respect, optical master readout of video signals is not fully indicative of the eventual stylus readout of the replicas. Test signals have been developed, therefore, to provide an optical measure of track drop variations on the master disc.

Optical reading of VideoDisc masters has been an extremely useful experimental tool and pilot-production, quality-control tool. The advantage of rapid assessment of masters has permitted experimental mastering techniques to progress more rapidly. In pilot production it has served as a quality-control monitor for processing and signaling parameters. A VideoDisc master reader that has been in continuous use in the pilot production facility for over two years is shown in Fig. 14. In the development of an optimized electromechanical mastering process, one critical element is the mechanical cutter head and its frequency equalization. In addition to the type of reader illustrated in Fig. 14, another more compact reader has been also constructed to read the signal immediately after it is cut by an electromechanical master recorder. This unit is shown in Fig. 16. This reader has been designed as an add-on to an electromechanical master recorder. Since the electromechanical recorder turntable is not translated radially, the compact reader translates its optical bench over the master for radial tracking. The purpose of this reader is to provide a real-time method for electromechanical-cutter-head evaluation* and equalization and for quality control of the masters.

* Another optical technique for cutter head development and evaluation is described in Ref. [2].

Fig. 16—Compact optical master reader.

Acknowledgements

The development of optical reading of VideoDisc masters could not have taken place without the full cooperation of many others within RCA Laboratories. In addition to the members of the Optical Techniques Team, we gratefully acknowledge the assistance of many others in the VideoDisc development effort. In particular we acknowledge the contributions of Z. Andrevski, J. K. Clemens, E. C. Fox, D. L. Jose, E. O. Keizer, J. H. Martin, R. C. Palmer, J. Reisner, W. R. Roach, M. D. Ross, P. Sheng, and R. L. Truesdell.

Appendix

This appendix outlines the analysis underlying our calculations of output signal waveforms from an optical VideoDisc reader. For heuristic purposes we unfold the optical system about the plane of the disc surface. We have also simplified the situation by ignoring the finite thickness of the phase relief structure on the disc and replacing it with a zero thickness phase transforming interface. Fig. 17 identifies the coordinate variables for each of the three planes of interest in the equivalent transmissive optical system. The two thin lenses of effective focal length f_o each represent the microscope objective. Lens I focusses the spot onto the disc, and lens II collects the scattered light. We treat the optical system as space-invariant by placing the aperture stops in focal planes. The entrance pupil for lens I is formed by the aperture in its front focal plane; this aperture is illuminated from the left by a monochromatic wave. The exit pupil for lens II is located in its back focal plane. This second aperture is referred to as the detector aperture, since it defines

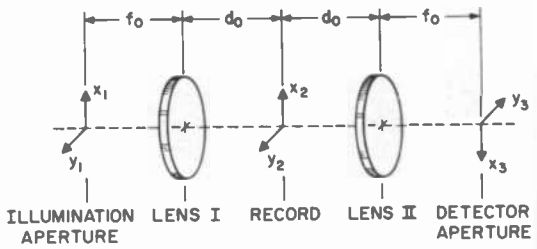


Fig. 17—Diagram of the equivalent transmissive optical system of a reflective reader.

the maximum scattering angle through which the detector can receive scattered light from the disc surface.

We begin by assuming that the microscope objective requires an image conjugate distance of infinity, and that no focus error exists. Thus d_o equals f_o, and a planar (equiphase) wave of unit amplitude fully illuminates the entrance pupil in the (x_1,y_1) plane. The function $g(x_2,y_2; x)$ represents the spatial pattern of optical path-length variations imposed on the incident light by the disc; x_2 and y_2 are fixed with respect to the optical axis, while x measures the disc displacement along the track as a simple translation. The complex transmittance (reflectance) $\underset{\sim}{t}$, representing the signal encoding at the disc surface plane, is

$$\underset{\sim}{t}(x_2,y_2; x) = \exp[i2\pi g(x_2,y_2; x)]. \qquad [10]$$

Let $\underset{\sim}{w}(x_2,y_2)$ represent the complex amplitude of the focussed spot produced by lens I at the disc surface. The departing wave is given by the product $\underset{\sim}{w}\underset{\sim}{t}$. Our ultimate interest is in the wave amplitude at the detector plane caused by $\underset{\sim}{w}\underset{\sim}{t}$.

We proceed by decomposing $\underset{\sim}{w}\underset{\sim}{t}$ into a set of uniform plane waves having various amplitudes and propagation directions. The amplitude $\underset{\sim}{V}'$ of the plane wave component whose propagation vector has direction cosines α and β with respect to the x_2 and y_2 axes is given by[18]

$$\underset{\sim}{V}'(\alpha,\beta; x) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \underset{\sim}{w}(x_2,y_2)\underset{\sim}{t}(x_2,y_2; x) \times \exp\left[-i2\pi\left(\frac{x_2\alpha + y_2\beta}{\lambda}\right)\right] dx_2 dy_2; \qquad [11]$$

this function is called the angular spectrum of $\underset{\sim}{w}\underset{\sim}{t}$. Each plane wave component is focussed by lens II to a point in the (x_3,y_3) plane. Since the Abbé sine condition is fulfilled by well-corrected imaging lenses, a plane wave with direction cosines (α,β) is focused at the point $x_3 = -(\alpha c/a), y_3 = -(\beta c/a)$, where c is the radius of the entrance (exit) pupil of lens I (II), and a is the effective numerical aperture of lenses I and II as given by[8]

$$a = c/f_o. \qquad [12]$$

The negative signs result from our choice of positive coordinate directions. The wave amplitude $\underset{\sim}{V}(x_3,y_3)$ at each point in the detector plane is thus uniquely specified by the amplitude of the corresponding plane wave component in the angular spectrum of $\underset{\sim}{w}\underset{\sim}{t}$. Components whose propagation direction with respect to the optical axis exceeds the angle $\sin^{-1}(a)$ are truncated by the detector aperture. Substituting the variables in Eq. [11] gives

$$\underset{\sim}{V}(x_3,y_3;x) = \text{circ}\left(\frac{x_3^2 + y_3^2}{c^2}\right) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \underset{\sim}{w}(x_2,y_2)$$
$$\times \underset{\sim}{t}(x_2,y_2;x) \exp\left[i2\pi \frac{a(x_2x_3 + y_2y_3)}{\lambda c}\right] dx_2,dy_2, \qquad [13]$$

where the circ function equals zero when its argument is greater than 1, is 1 when its argument is between zero and 1, and is undefined for negative arguments. Constant factors which cancel upon later normalization have been omitted. The integral in Eq. [13] is analogous to the familiar Fourier optics treatment of the relation between wave amplitudes in the front and back focal plane of a lens. We have stressed the angular spectrum approach and the use of the sine condition, however, to demonstrate that no small-angle approximations need be made here. The approximation that *has* been tacitly made is to ignore the field of view of the lenses in the (x_2,y_2) plane. We shall return to this point shortly.

Because Eq. [13] is a Fourier integral, the wave amplitude $\underset{\sim}{V}$ at the detector may also be expressed as the convolution of the individual angular spectra of $\underset{\sim}{w}$ and $\underset{\sim}{t}$:

$$\underset{\sim}{V}(x_3,y_3;x) = \text{circ}\left(\frac{x_3^2 + y_3^2}{c^2}\right)$$
$$\times \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \underset{\sim}{W}(x_3 - \theta, y_3 - \eta)\underset{\sim}{T}(\theta,\eta;x)\, d\theta d\eta, \qquad [14]$$

where

$$\underset{\sim}{W}(x_3,y_3) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(x_2,y_2)$$
$$\times \exp\left[i2\pi \frac{a(x_2x_3 + y_2y_3)}{\lambda c}\right] dx_2dy_2 \qquad [15]$$

and

$$\underset{\sim}{T}(x_3,y_3;x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \underset{\sim}{t}(x_2,y_2;x)$$
$$\times \exp\left[i2\pi \frac{a(x_2x_3 + y_2y_3)}{\lambda c}\right] dx_2dy_2. \qquad [16]$$

By the same argument that leads from Eq. [11] to Eq. [13], the amplitude $\underset{\sim}{H}(x_1,y_1)$ at the illumination aperture is proportional to the angular spectrum of the focussed spot:

$$\underset{\sim}{H}(x_1,y_1) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \underset{\sim}{w}(x_2,y_2) \times \exp\left[i2\pi \frac{a(x_1x_2 + y_1y_2)}{\lambda c}\right] dx_2dy_2. \quad [17]$$

Comparing this relation with Eq. [15] shows that $\underset{\sim}{W}(x_3,y_3)$ is simply an exact replica of the wave amplitude at the illumination aperture. We could have anticipated this result from geometric optics, since the two lenses form a telecentric imaging system between the illumination and detector planes. We shall find Eq. [14] more convenient to use than Eq. [13] because it is possible to calculate the wave amplitude at the detector without explicitly evaluating the expression for the focussed spot. Once $\underset{\sim}{V}$ is known, the photodetector output $S(x)$ is found by integrating the resulting intensity over the detector area.

$$S(x) = \int_{\text{detector}}\int \left|\underset{\sim}{V}(x_3,y_3)\right|^2 dx_3dy_3 \quad [18]$$

Eq. [13] represents an approximation, albeit a rather good one. In order to retain the infinite limits on the integral, we should include an additional factor in the integrand to account for the finite field of view of a real microscope lens in its object plane. The detailed nature of this pupil function is intimately tied to the design and realization of a particular lens. The simplest representation of this restriction is to consider the field of view to be limited by a fictitious clear circular aperture of radius d in the (x_2,y_2) plane. Generally, d will be smaller than c, the radius of the aperture on the long conjugate side of the lens, but will be considerably larger than the characteristic size of the focussed spot at the disc surface.

The new integrand is $\underset{\sim}{w}\underset{\sim}{t}$ circ$[(x_2^2 + y_2^2)/d^2]$ times the exponential factor.

This product has significant amplitude over only a limited spatial extent, since the $\underset{\sim}{w}$ factor represents the focussed spot. Truncation of the product $\underset{\sim}{w}\underset{\sim}{t}$ by a larger aperture therefore has a relatively minor effect. The result of carrying through the mathematics is that the true amplitude in the detector plane is given by the convolution of $\underset{\sim}{V}$ from Eq. [14] with the angular spectrum of the fictitious aperture. Since the

angular spectrum of the aperture is very limited compared to the spectra of $\underline{w}$ and $\underline{t}$, we shall hereafter ignore its small influence on the fine structure of the wave amplitude in the detector plane.

The treatment thus far has been independent of any assumptions about the recorded signal profile, the record format, or the illumination conditions. In the RCA VideoDisc format, the groove width is much larger than the focussed spot size required to resolve the signal features. In this analysis, we assume that the spot is centered in a wide, uncurved groove, and hereafter we ignore the y_2-dependence of g. For a periodic signal-element pattern $g(x_2; x)$ with spatial period L, the transmittance may be written as the Fourier series

$$\underline{t}(x_2; x) = \sum_{n=-\infty}^{+\infty} \underline{b}_n \exp[i2\pi n(x_2 - x)/L],\ n = 0, \pm 1, \pm 2, \ldots \quad [19]$$

The complex coefficients $\underline{b}_n$ of this series

$$\underline{b}_n = L^{-1} \int_0^L \underline{t}(x_2) \exp(-i2\pi n x_2/L)\, dx_2 \quad [20]$$

are defined for $x = 0$, prior to any displacement of the disc. Substituting Eq. [19] into Eq. [16] gives

$$\underline{T}(x_3, y_3; x) = \sum_{n=-\infty}^{+\infty} \underline{U}_n(x)\delta[(x_3 - n\lambda c/aL), y_3] \quad [21]$$

where δ represents the Dirac delta-function, and

$$\underline{U}_n(x) = \underline{b}_n \exp(-i2\pi x/L). \quad [22]$$

If the periodic signal profile on the disc were illuminated by a uniform plane wave and if lens II had an infinite field of view on the (x_2, y_2) plane, Eq. [21] describes how the discrete diffraction orders, having amplitudes $\underline{U}_n$, would be focussed at a series of points along the x_3-axis. Eqs. [4a] and [4b] in the text give the amplitudes obtained for the specific example of a recorded square-wave signal.

The function $\underline{w}(x_3, y_3)$ has been previously identified as a replica of the illumination wave transmitted by the illumination aperture:

$$\underline{W}(x_3, y_3) = \mathrm{circ}\left(\frac{x_3^2 + y_3^2}{c^2}\right). \quad [23]$$

Combining Eqs. [23] and [21] with Eq. [14] yields

$$\underline{V}(x_3, y_3; x) = \mathrm{circ}\left(\frac{x_3^2 + y_3^2}{c^2}\right) \sum_{-\infty}^{+\infty} \underline{U}_n \,\mathrm{circ}\left[\frac{(x_3 - n\lambda c/aL)^2 + y_3^2}{c^2}\right] \quad [24]$$

for the wave amplitude at the detector. Convolving the circ function with a series of laterally displaced delta functions gives a series of displaced circ functions, each associated with a specific diffraction order. The zero-order replica of the illumination aperture coincides with the exit pupil in the detector plane.

Fig. 18 sketches the boundaries of the circ functions in Eq. [24]. The intensity distribution within the exit pupil consists of a mosaic of finite-size areas, within each of which the intensity is uniform and is contributed by a finite number of diffraction orders. If the detector fully fills the exit pupil, then the detector output can be expressed as a finite sum suitable for exact evaluation as follows. The overlap area of the two circles associated with orders n and l, divided by the area of one of the circles, is[19]

$$O(m\mu) = \frac{2}{\pi}\left[\cos^{-1}(m\mu) - m\mu(1 - m^2\mu^2)^{1/2}\right], \quad 0 \le m \le 1$$
$$= 0, \quad m\mu \ge 1, \qquad [25]$$

where

$$m = |n - l|, \qquad [26]$$

and

$$\mu = \lambda/(2aL) \qquad [27]$$

is the normalized fundamental spatial frequency of the periodic signal element pattern on the disc. For example, one of the regions labeled "e" in Fig. 18 has a fractional area given by $O(0.8)$, since $\mu = 0.2$ for this sketch and $m = 4$ for "e." For $\mu = 0.2$, there are only five different shapes involved. Each distinct shape is associated with a distinct value of m,

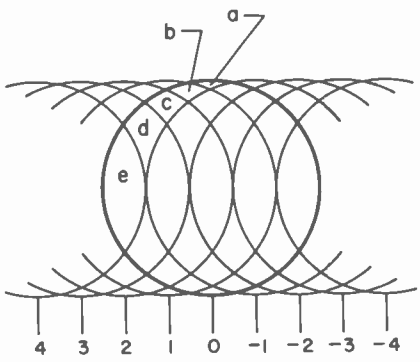


Fig. 18—Diagram showing various regions of the overlapping circles. The numerals specify the order number, n.

and the fractional area of that shape is $O(m\mu) - 2O[(m+1)\mu] + O[(m+2)\mu]$.

A straightforward accounting for all possible overlapping areas and their intensities for any μ eventually yields the exact expression

$$S_{\pm}(x) = \sum_{m=1}^{N} \left\{ \sum_{j=1}^{m} \frac{1}{2} \left(1 \mp \operatorname{sgn}\left[j - \frac{1}{2}(m+1) \right] \right) \times \sum_{k=1}^{m} |\underline{U}_{k-j}|^2 \right\} \{O[(m-1)\mu] - 2O[m\mu] + O[(m+1)\mu]\}, \quad [28]$$

where sgn is defined as -1, 0, or $+1$ when its argument is less than zero, zero, or greater than zero, respectively. Where there is a choice of signs in Eq. [28], use of the uppermost one signifies that the integral in Eq. [18] has been performed only over the semicircular area of the detector aperture for which $x_3 > 0$; conversely, the lower sign is for the semicircular area $x_3 < 0$. The expressions $S_+(x)$ and $S_-(x)$ are normalized to the integrated intensity over the entire zero-order area which would result from a featureless disc of unity transmittance. For a given normalized fundamental spatial frequency μ, the upper limit N of the outer summation may be truncated at the smallest integer greater than or equal to $1/\mu$, because each of the three overlap area functions is zero for larger values of m.

The peak-to-peak output signal versus spatial frequency curve (Fig. 5, curve A) for a single circular detector and square-wave signal elements is obtained by evaluating $S_+(x) - S_-(x)$ at $x = 3L/4$ and at $L/4$. For high spatial frequencies, $\mu > 0.5$, Eq. [27] reduces to

$$S_{\pm}(x) = |\underline{U}_o|^2 \left[\frac{1}{2} - O(\mu) \right] + [(1 \pm 1)|U_o + U_1|^2 + (1 \mp 1)|\underline{U}_o + \underline{U}_{-1}|^2] O(\mu)/2, \quad [29]$$

since only the first-order circles separately overlap the zero-order one. Upon substituting the $\underline{U}_n$ values given by Eq. [4], the high-frequency region of curve B in Fig. 5 is given by the expression $8 \sin(\phi) O(\mu)/\pi$. Although the split-detector arrangement has been shown to be exactly analogous to a differential phase contrast imaging method using partially coherent illumination,[11] the $O(\mu)$ variation at high spatial frequencies has the same frequency dependence displayed by the MTF of an ideal diffraction-limited incoherent imaging system with a circular pupil.[19]

The effect of a focus error may be analyzed by allowing the distance d_o to be different from f_o. The illumination source is not changed. Now the disc surface is no longer in the focal plane of either lens. We will apply the techniques of Fourier optics to separately analyze the effects of defocus on the angular spectrum of the disc, and the effects on the illumination at the system exit pupil.

Let the focus error be z, so that

$$d_o = f_o + z. \qquad [30]$$

The amplitude of the wave departing the (x_2,y_2) plane is again written as the product $\underset{\sim}{w}(x_2,y_2)\underset{\sim}{t}(x_2,y_2; x)$, and has an angular spectrum given by Eq. [11]. However, $\underset{\sim}{w}$ now describes an out-of-focus spot. As the composite wave propagates to the front focal plane of lens II, located a distance z to the right of the disc, its amplitude distribution evolves as the result of diffraction. Rather than expressing the explicit spatial dependence of the wave at this plane, we instead write its angular spectrum as

$$V'_{fII}(\alpha,\beta; x) = \exp\{i2\pi z \sqrt{1-\alpha^2-\beta^2}/\lambda\}$$
$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(x_2,y_2)t(x_2,y_2; x)$$
$$\times \exp\left\{-2\pi i \frac{x_2\alpha + y_2\beta}{\lambda}\right\} dx_2 dy_2 \qquad [31]$$

where we have made use of the propagation properties[18] of angular spectra.

Applying the Abbé sine condition as before, we note that again the amplitude in the (x_3,y_3) plane can be expressed as a convolution of the individual angular spectra of $\underset{\sim}{w}$ and $\underset{\sim}{t}$,

$$\underset{\sim}{V}(x_3,y_3; x) = \mathrm{circ}\left(\frac{x_3^2+y_3^3}{c^2}\right) \exp\{i2\pi(z/\lambda)$$
$$\times \sqrt{1-[x_3a/c]^2-[y_3a/c]^2}\}$$
$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underset{\sim}{W}(x_3-\xi,y_3-\eta)\underset{\sim}{T}(\xi,\eta; x)d\xi d\eta. \qquad [32]$$

Both $\underset{\sim}{w}$ and $\underset{\sim}{t}$ are defined as in Eqs. [15] and [16]; however now $\underset{\sim}{w}$ no longer simply represents a replica of the wave amplitude at the illumination aperture in the (x,y) plane but includes a wavefront curvature due to the defocus condition. We can show this by propagating the angular spectrum of the focused spot in the back focal plane of lens I a distance z to the right. Recall that the angular spectrum of the wave in the back focal plane of lens I is $\mathrm{circ}[(\alpha^2+\beta^2)/a^2]$. Then

$$\underset{\sim}{W}(\alpha,\beta) = \mathrm{circ}\left(\frac{\alpha^2+\beta^2}{a^2}\right) \exp\left\{+i2\pi z \frac{\sqrt{1-\alpha^2-\beta^2}}{\lambda}\right\}, \qquad [33]$$

and

$$W(x_3,y_3) = \mathrm{circ}\left(\frac{x_3^2+y_3^2}{c^2}\right)$$
$$\times \exp\left\{+i2\pi z \frac{\sqrt{1-(x_3a/c)^2-(y_3a/c)^2}}{\lambda}\right\} \qquad [34]$$

We note that there are two spherical phase factors in the wave amplitude on the (x_3,y_3) plane. The first one, outside the integral in Eq. [32], is the result of the record not being in the front focal plane of lens II. This spherical phase factor is of no consequence, since it does not affect the *intensity* on the (x_3,y_3) plane. The second, the spherical phase factor in Eq. [34], is the result of the spot defocus. In the paraxial approximation, both spherical phase factors can be shown to have a radius of curvature r, such that

$$zr = f_o^2. \qquad [35]$$

The convolution of $\underset{\sim}{T}$ with $\underset{\sim}{W}$ in the presence of a focussing error once more produces a series of spatially overlapping replicas of the illumination aperture, as in Eq. [24], but with the additional spherical phase factor in each term. As a result, the intensity in the overlap areas is no longer uniform, but forms the spatial pattern of interference fringes visible in Fig. 6. The integral in Eq. [18] representing the detector signal can no longer be reduced to the simple form of Eq. [28].

Analysis of an optical system with microscope objectives designed for finite conjugates proceeds similarly. From the Newtonian imaging relation (Eq. [35]), an axial point source located a distance r to the left of the illumination aperture gives an in-focus spot at the record surface whose distance from the back focal plane of lens I is z. The resulting wave amplitude in the (x_3,y_3) plane contains the same spherical phase factor found in Eq. [32], which does not affect the intensity. The angular spectrum of the spot focussed on the (x_2,y_2) plane depends only on λ and the numerical aperture of the converging light cone, not on the conjugate distances of the source and its image. Thus the detector response for a focussed, finite-conjugate system is the same as the expression (Eq. [28]) that we have calculated in detail for an infinite-conjugate system. In the case of a focus error, the additional spherical phase factor is analogous to that in Eq. [34]; it depends only on the axial distance between the record and the plane of focus for the spot.

Finally, we wish to comment on several practical aspects of the VideoDisc master reader. The optical efficiency of the system is improved by allowing the illumination aperture to truncate the expanded laser beam near the $1/e^2$ contour of its Gaussian intensity profile. Since this nonuniform illumination gives relatively less intensity near the edges of the overlapping circular diffraction orders, the reader response is attenuated at the higher spatial frequencies compared to the middle frequencies. The frequency response characteristic is also affected by the shape of the detector. Choosing a $\pi/2$ sector rather than a semicircle

(π sector) for each half of the differential detector increases the high-frequency output relative to that at low frequencies. Using a focussed spot that is elongated in one direction to more nearly fill the track width changes the overlapping diffraction orders from circles to ellipses. The formalism which we have outlined above is easily adapted to these alternative configurations. The qualitative nature of the altered output waveforms and frequency response characteristics, however, remain similar to those given by the relatively simple result in Eq. [28].

References:

[1] See the March 1978 Special Issue of *RCA Review*.

[2] I. Gorog, "Optical Techniques Developed for the RCA VideoDisc," *RCA Review*, **39**, p. 162, March 1978.

[3] K. Broadbent, "A Review of the MCA Disco-Vision System," *J. SMPTE*, **83**, p. 554 (1974).

[4] G. Bouwhis and P. Burgstede, "The Optical Scanning System of the Philips 'VLP' Record Player," *Philips Tech. Rev.*, **33**, p. 186 (1973). Also, *J. SMPTE*, **83**, p. 572 (1974).

[5] G. Hrbek, "An Experimental Optical Videodisc Playback System," *J. SMPTE*, **83**, p. 580 (1974).

[6] J. K. Clemens, "Capacitive Pickup and the Buried Subcarrier Encoding System for the RCA VideoDisc," *RCA Review*, **39**, p. 33, March 1978.

[7] W. van den Bussche, A. H. Hoogendijk, and J. H. Wessels, "Signal Processing in the Philips 'VLP' System," *Philips Tech. Rev.*, **33**, p. 181 (1973). Also, *J. SMPTE*, **83**, p. 567 (1974).

[8] E. O. Keizer, "VideoDisc Mastering," *RCA Review*, **39**, p. 60, March 1978.

[9] M. Born and E. Wolf, *Principles of Optics*, 3rd ed., Section 4.5.1, Pergamon Press, Oxford (1965).

[10] N. Dekkers and H. deLang, "Differential Phase Contrast in STEM," *Optik*, **41**, p. 452 (1974).

[11] W. C. Stewart, "On Differential Phase Contrast with an Extended Illumination Source," *J. Opt. Soc. Am.*, **66**, p. 813 (1976).

[12] A. H. Firester, M. E. Heller, and P. Sheng, "Knife-Edge Scanning Measurements of Subwavelength Focused Light Beams," *Appl. Optics*, **16**, p. 1971 (1977).

[13] A. H. Firester, T. E. Gayeski, and M. E. Heller, "Efficient Generation of Laser Beams with an Elliptic Cross Section," *Appl. Optics*, **11**, p. 1648 (1972).

[14] H. Kressel and F. Z. Hawrylo, "Red-Light-Emitting Laser Diodes Operating cw at Room Temperature," *Appl. Phys. Lett.*, **28**, p. 598 (1976).

[15] I. Ladany and H. Kressel, "Visible CW (AlGa)As Heterojunction Laser Diodes," Int. El. Dev. Meeting, Wash., D.C., Dec. 6–8, 1976, Paper No. 6.4, Conf. Proc., p. 129.

[16] W. R. Roach, C. B. Carroll, A. H. Firester, I. Gorog and R. W. Wagner, "Diffraction Spectrometry for VideoDisc Quality Control," *RCA Review*, **39**, p. 472, Sept. 1978 (this issue).

[17] P. Sheng, "Theoretical Considerations of Optical Diffraction from RCA VideoDisc Signals," *RCA Review*, **39**, p. 512, Sept. 1978 (this issue).

[18] J. W. Goodman, *Introduction to Fourier Optics*, Section 3-7, McGraw-Hill Book Co., Inc., New York (1968).

[19] See Ref. [18], Section 6-3.

Optical Recording Techniques for the RCA VideoDisc

A. H. Firester, I. Gorog, J. P. Russell, J. J. Gibson, C. B. Carroll, and W. R. Roach

RCA Laboratories, Princeton, N.J. 08540

Abstract—Real-time optical mastering of the RCA VideoDisc has been achieved using materials and techniques that are process and format compatible with other VideoDisc mastering techniques, as well as with the VideoDisc capacitive stylus playback system. The resolution requirements for recording are particularly stringent. Signal encoding features on the RCA VideoDisc can be as small as 200 nm, since each disc plays 1 hour per side and rotates at only 450 rpm. We explored three distinct optical recording methods incorporating differing photoresist exposure-development nonlinearities, optical interference patterns caused by reflections from the substrate, electronic compensation of optical resolution limitations, and methods for forming the stylus guiding groove. This paper details the concepts underlying the three methods, their relative merits, and the achieved results as determined by capacitive stylus playback and scanning-electron-microscope analysis.

1. Introduction

This paper specifically discusses real-time optical mastering of RCA VideoDiscs. Electromechanical and electron-beam mastering techniques have been described elsewhere.[3,5] As discussed in Ref. [3], electron-beam and optical mastering techniques were developed as alternatives to electromechanical mastering, which is the presently preferred method. Optical mastering is fully compatible with the VideoDisc format, replication processes, and capacitive stylus playback of replicas. The principal technical difficulties associated with optical recording of

high-quality video signals in the four-TV-frames-per-revolution RCA VideoDisc format are optical resolution and optical standing-wave patterns within the photosensitive layer. On the innermost radius of one-hour-per-side discs, the signal slot dimensions at 6.3 MHz are approximately 250 nm along the groove and 2.7 μm transverse to the groove. The small dimension along the groove approaches the limit of achievable optical resolution. We have developed three distinct optical recording methods: (1) optical recording on pregrooved substrates with thick sagged photoresist that forms the stylus guiding tracks; (2) resolution-enhanced recording on thin photoresist layers coated on substrates that are pregrooved by a radiused tool to generate the stylus guiding tracks; and (3) simultaneous linear recording of both the signal and the stylus guiding grooves on smooth, thick, photoresist layers by an electronically equalized optical recorder.

This paper describes the basics of optical recording, each of the three recording methods, and the achieved results as demonstrated by playback and scanning-electron-microscopic analysis.

2. The VideoDisc Signal

The VideoDisc signal, which has been described in detail in an earlier paper,[1] consists of the sum of two frequency-modulated sine-wave carriers. One of the carriers is centered around 5.3 MHz and carries the video information; the other is centered around 716 kHz and carries the sound information. The sound carrier is 20 dB weaker than the video carrier. The instantaneous frequency of the video carrier is 4.3 MHz at the video sync tips and 6.3 MHz at the video white level. Due to preemphasis of the video signal and to the presence of a color subcarrier, the instantaneous frequency of the video carrier may swing from 3.9 to 6.9 MHz. Significant frequency components of the modulated carrier range from 1.9 to 9.3 MHz. While frequencies out to 6.9 MHz must be transmitted with an acceptable response to avoid FM system failure due to insufficient carrier-to-noise ratio, loss of response at higher frequencies can be compensated with aperture correction on playback to maintain system performance.[2] The VideoDisc signal is recorded as a "vertical" relief pattern in the grooves of a disc, which is either made out of conductive material or has a conductive coating. The disc rotates at 450 rpm and plays for one hour on each side. The recording starts at an outer radius of about 5.8 inches and proceeds inwards at a groove pitch of about 10,000 grooves per inch to an inner radius of about 2.8 inches. The grooves are about 2.7 micrometers wide and 300 nm deep, and the peak-to-peak amplitude of the recorded signal is about 70 nm. At a radius of R inches the track velocity is $v = 1.2R$ meters/sec. Thus for a frequency of f MHz, the recorded wavelength is $\Lambda = 1.2R/f$ micrometers. The shortest

wavelength of concern is $\Lambda = 400$ nm, which may occur at an inner radius of 2.8 inches for a maximum frequency of 8.3 MHz.

Upon playback, the recorded signal is converted by a capacitive stylus into an electrical signal, which in turn is delivered to two FM demodulators, one for the video and one for the sound.

A somewhat different signal format obtained by hard limiting the sum of the two FM carriers can also be used compatibly with the same player. This ON-OFF format, shown in Fig. 1, consists of a square wave that is frequency modulated by the video information and duty-cycle modulated by the sound carrier, which in turn is frequency modulated by the audio signal. Fig. 1 illustrates how this ON-OFF waveform is obtained by hard limiting the linear sum of a high-frequency video carrier and a relatively weak low-frequency sound carrier. A special analysis of this ON-OFF signals shows that it consists of the sum of the two desired FM carriers plus a number of other components, which in general cause negligible interference in the outputs of the FM demodulators in the player. As a consequence, the linear recording format and the ON-OFF recording format can be used compatibly, each having its advantages and disadvantages depending on the recording technique. The ON-OFF format is used in electron-beam recording,[3] which does not have a linear capability, while the linear format is used in electromechanical recording,[3] which does not have an ON-OFF capability. As shall be discussed in this paper, both formats can be used in optical recording.

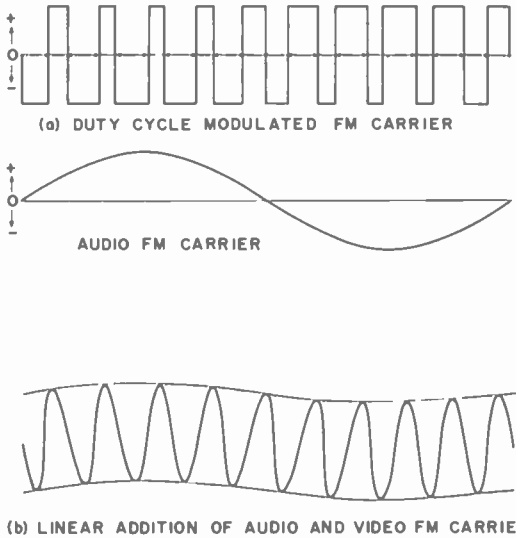


Fig. 1—Methods of combining the audio and video signals: (a) duty-cycle modulation of video FM carrier by audio FM carrier and (b) linear addition of audio and video FM carriers.

The VideoDisc system is remarkably tolerant of variations in signal geometry and stylus characteristics. Capacitive styli are, however, quite nonlinear and read the recorded video carrier with an amplitude and phase response that depends upon the height of the stylus over the surface of the disc. To ensure an acceptable signal-to-noise ratio and an acceptable level of interference between sound and video, specifications for the signal geometry, applicable to the ON-OFF format as well as to the linear format, were developed. These geometric signal specifications, defined in Fig. 2, are:

- *Signal Amplitude* (*SA*): The signal amplitude at 4.3 MHz (peak of sync) shall be larger than 70 nm,

  $SA_{4.3} \geq 70$ nm.

- *Signal Amplitude Ratio* (*R*): The ratio of the signal amplitude at 6.3 MHz (white level) to the signal amplitude at 4.3 MHz (peak of sync) shall be larger than 0.7, i.e.,

  $SA_{6.3}/SA_{4.3} > 0.7$.

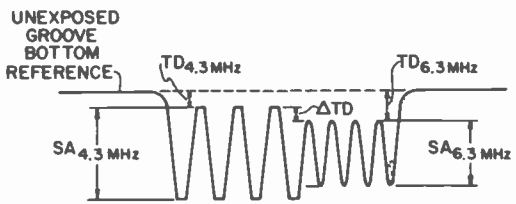
- *Absolute Track Drop* (*TD*): The difference in the level of the bottom of an unexposed groove to the level of the top of a recorded 4.3-MHz carrier shall not exceed 30 nm,

  $TD_{4.3} \leq 30$ nm.

- *Differential Track Drop* (ΔTD): The difference between the track drop at 6.3 MHz and at 4.3 MHz shall not exceed 5 nm,

  $\Delta TD = TD_{6.3} - TD_{4.3} \leq 5$ nm.

Before detailing the various methods we have used to optically master VideoDiscs, we briefly describe the recording apparatus and the optical properties of the recording lens.



TD = TRACK DROP
SA = SIGNAL AMPLITUDE
$\Delta TD = TD_{4.3\,MHz} - TD_{6.3\,MHz}$

Fig. 2—Physical specification and nomenclature of the VideoDisc signal geometry.

3. Recording System

3.1 Recorder

The optical master recorder is very similar in design and construction to the optical reader system.[4] The optical recording system block diagram is shown in Fig. 3. The laser (a) is a 15 mW HeCd laser (RCA LD2186). The polarized output beam is modulated by an electro-optic modulator (b) having a 50-MHz bandwidth. An optical attenuator (c) permits adjustment of the light intensity as a function of recording radius in order to obtain a uniform exposure as the recording progresses toward smaller recording radii. Both the signal modulation of the light beam and the average power level are monitored by a photodetector (i) that senses a fraction of the optical beam. The beam forming optics (d) shape the laser beam to properly fill the spot forming lens (j) so that a chisel-shaped beam as wide as the desired track width is formed on the output of the lens. Typically this beam-shaping system might consist of a spherical telescope in conjunction with an anamorphic telescope. A mirror (e) is mounted on a scanning galvanometer to permit servocontrolled tracking when recording on pregrooved substrates. Two track-

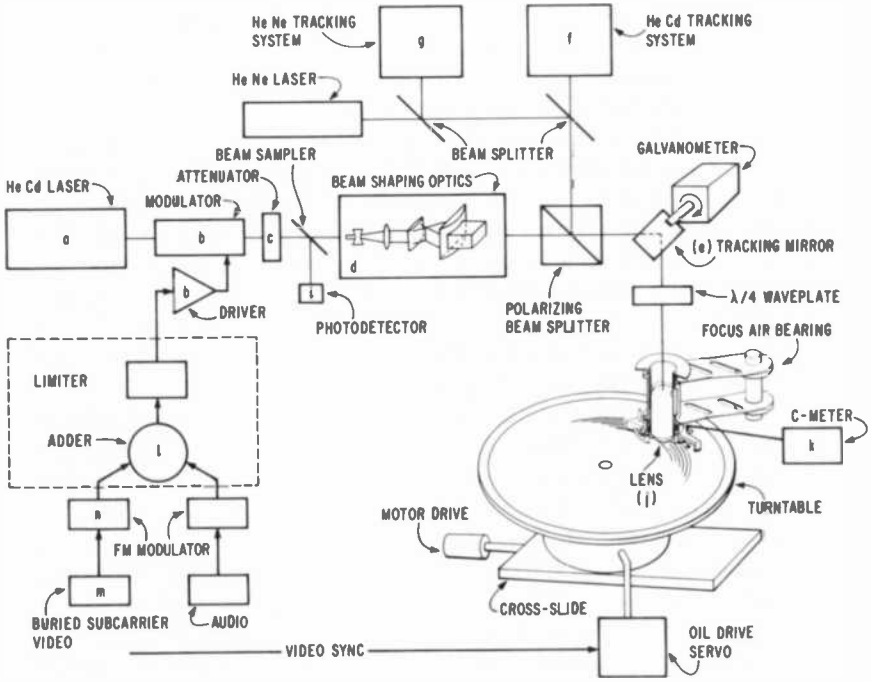


Fig. 3—System outline for optical recorder for VideoDisc masters.

ing-error-signal-detection systems (f) and (g) are shown which are based respectively on either the HeCd recording laser or an auxiliary 1 mW HeNe laser.

We have found that reliable tracking depends upon both the thickness of the photoresist and the optical depth of the tracking groove. For thin photoresist and optically shallow tracking grooves, tracking with the HeCd laser using a transform plane method similar to that employed in the optical reader[4] is satisfactory. However with thick photoresist layers and optically deep tracks, it is necessary to use the auxiliary HeNe laser. The 633-nm radiation readily penetrates the photoresist without inducing any photochemical changes. Under these circumstances we have more successfully used image-plane tracking, i.e., in this case, a split photodetector is located at an image plane of the substrate surface. For the optically deep tracks (see Fig. 6), internal reflections produce multiple images that are symmetric only when the focused beam is centered within the substrate track.

The recording lens is a 0.95 numerical aperture fluorite microscope objective. It is mounted in an air-bearing structure that maintains a constant distance from the recording surface despite small surface warps.[4] The air-bearing performance is monitored by a capacitance meter (k) which monitors the capacitance between the lens mount and the recording substrate. A fixed lens-to-substrate distance is maintained, despite slowly varying height changes of the substrate, by adjustment of the air flow to the air-bearing structure. A coarse focus setting can be obtained by optical readout of the substrate surface microroughness by means of either the HeCd or HeNe detector systems (f) or (g) operating in an optical reader mode.[4] An extremely precise hydrostatic oil bearing forms the heart of the turntable system. Regulated jets of oil provide a smooth drive with a precise speed control;[5] the turntable speed is electrically locked to the vertical-sync signal of the recording signal. The turntable is mounted on a motor driven cross-slide (m) to provide the radial motion necessary to form the recording spiral. Control of this motor drive is either open loop, on substrates without pregrooving, or closed loop with the error signal derived from tracking signals of systems (f) or (g) following a pregrooved substrate.

The signal source for the recorder is also indicated schematically in Fig. 3. For much of our recording, the signal source was similar to that used in the electron-beam disc recorder. The buried-subcarrier-encoded color video signal (m) FM modulates (n) a 5-MHz carrier. To this is added the audio signal which is also impressed on an FM carrier. The resultant sum is clipped and hard-limited, and it is this composite signal that is applied to the electro-optic modulator (b). The portion of the signal-source system enclosed in the dotted box is changed when linear signal recordings are made. This will be discussed in greater detail later.

3.2 Recording Lens

Fundamental to optical recording is the ability of the recording lens (j) to form a tightly focused spot. We will first consider the experimental capabilities of available lenses and then consider the effects of this focussed spot upon the recorder capabilities.

The minimum focused spot size is fundamentally determined by the wavelength of light and the numerical aperture of the lens. For a uniformly illuminated diffraction-limited lens the spot width, w, between half-intensity points is approximately given by

$$w = \frac{\lambda}{2\,\mathrm{NA}}, \qquad [1]$$

where λ is the wavelength of the light, and NA is the lens numerical aperture. In practice because actual high numerical aperture lenses are imperfect, the focused spot width is substantially larger than that predicted by Eq. [1]. In the course of our work, we developed a knife-edge scanning technique[6] for measuring the intensity profile of the focused spot. Fig. 4 illustrates one such intensity profile of a 0.95-numerical-aperture lens measured *in situ* on the optical recorder. As shown in this figure, the full width at half intensity of this spot is ~320 nm, which is appreciably larger than the 233 nm predicted by Eq. [1] for NA = 0.95 and λ = 442 nm.

Now let us consider the effects of the intensity profile of the focused spot on the recorder capabilities. We assume that the recording is made

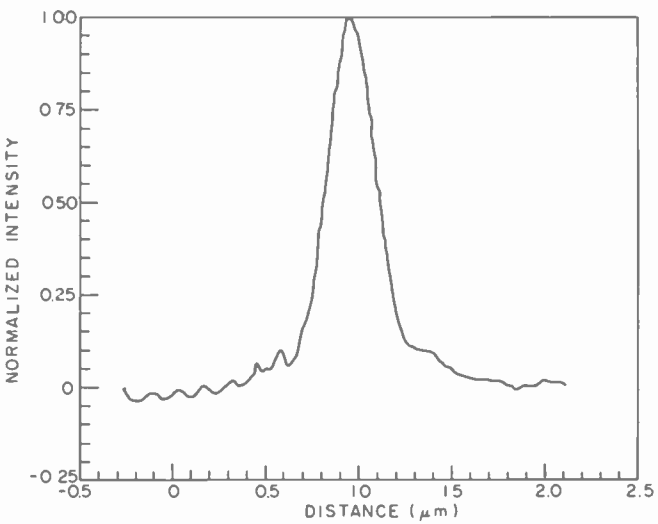


Fig. 4—One-dimensional focused intensity profile for optical recording lens (wavelength = 442 nm, NA = 0.95).

on a strip of unit width that moves with velocity $v = \Omega r$, where Ω is the angular velocity of the master recording turntable and r is the radius at which the recording is taking place. The strip is exposed by a focused spot of intensity profile $\sigma(z)$ along the recording direction and unit width in the disc radial direction. This focused spot is modulated in intensity by the signal, $S(t)$, as a function of time. Thus the time rate of exposure, E, is

$$\frac{dE(z)}{dt} = \sigma(z - vt)S(t). \qquad [2]$$

Consequently the total exposure, $E(z)$, is

$$E(z) = \int_{-\infty}^{\infty} \sigma(z - vt)S(t)dt \qquad [3a]$$

or

$$E(z) = \frac{1}{\Omega r} \int_{-\infty}^{\infty} \sigma(z - u)S(u/\Omega r)du. \qquad [3b]$$

Thus we see that the exposure is the convolution of the focused-spot intensity profile and the temporal recording signal. Since we have assumed a uniform velocity, v, along the recording strip, we can consider both σ and S in terms of spatial variables. If $S(t)$ varies sinusoidally with temporal frequency, f, the recorded sinusoid will have a spatial frequency $\nu = f/\Omega r$. Similarly the temporal signal spectrum, $\tilde{S}(f)$ defined by

$$\tilde{S}(f) = \int_{-\infty}^{\infty} S(t) \exp(-i2\pi ft)dt \qquad [4]$$

can be alternately expressed in terms of spatial frequency,

$$\tilde{S}(\nu\Omega r) = \frac{1}{\Omega r} \int_{-\infty}^{\infty} S(u/\Omega r) \exp(-i2\pi\nu u)du. \qquad [5]$$

Correspondingly, the spatial frequency response or modulation transfer function (MTF) of the focused spot, $\tilde{\sigma}(\nu)$ is

$$\tilde{\sigma}(\nu) = \int_{-\infty}^{\infty} \sigma(z) \exp(-i2\pi\nu z)dz. \qquad [6]$$

Thus the spatial frequency response of the exposure, $\tilde{E}(\nu)$, according to Eq. [3] is proportional to the product of the spot MTF and the signal spectrum. That is,

$$\tilde{E}(\nu) \equiv \int_{-\infty}^{\infty} E(z) \exp(-i2\pi\nu z) dz = \Omega r \tilde{S}(\nu \Omega r) \tilde{o}(\nu). \quad [7]$$

Thus if $\tilde{o}(\nu) = 1$ for all ν, then the response will have the spatial frequency response $\Omega r \tilde{S}(\nu \Omega r)$ corresponding to the temporal modulation $S(z/\Omega r)$. In practice $\tilde{o}(\nu)$ is not everywhere equal to one, but generally decreases and approaches zero at frequency $2NA/\lambda$, the cutoff frequency of an ideal lens. Fig. 5 shows the MTF for the lens data depicted in Fig. 4.

If the FM encoded signals are directly applied to the modulator and the recording process is linear, then the recording transfer function will be given by the MTF of the recording lens. The performance of a typical high-quality lens, as shown in Figs. 4 and 5, is insufficient to meet the VideoDisc signal geometry specifications previously discussed. In this paper we will discuss three optical recording methods we explored and will show how high-quality, VideoDisc optical recordings can be made with a frequency response that is significantly better than that predicted from the recording lens MTF.

4. Direct Recording on Thick Photoresist

4.1 Analytical Estimates

The first method we attempted for optical recording of VideoDisc masters was to directly expose photoresist-covered pregrooved substrates with duty-cycle-modulated ON-OFF FM signals. Both the signal source

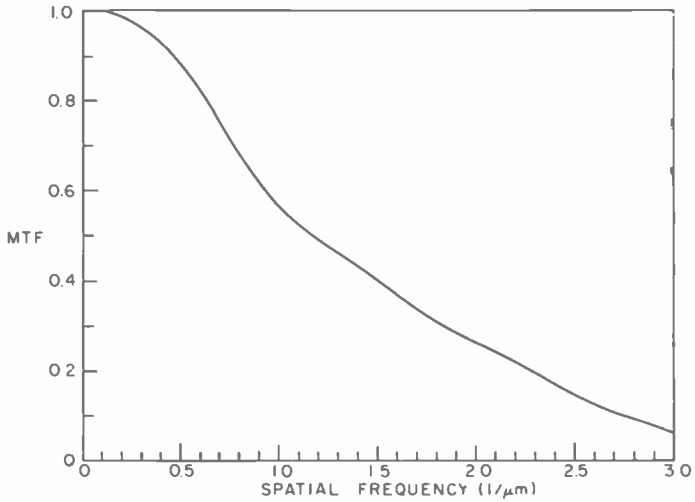


Fig. 5—Modulation transfer function for intensity profile depicted in Fig. 4.

and substrates were identical to those used in electron-beam recording.[3] The recording substrates were trapezoidally pregrooved copper-clad discs onto which a dilute solution of photoresist was spun and the solvent was allowed to evaporate. The photoresist surface sags to form a smooth, sinusoidal-cross-section groove. The cross section of a photoresist-coated pregrooved substrate for a 5555-groove-per inch 1/2-hour-per-side format is shown in Fig. 6. Most of our initial recordings were made on this substrate geometry. As discussed in connection with the physical specifications for the RCA VideoDisc, the peak-to-peak signal dimensions in the recording are the important geometric parameters for a recorded substrate master. For an ON-OFF duty-cycle-modulated FM signal, the ideal recorded depth profile replicates the recording signal. By appropriate choice of exposure–development conditions for the photoresist, it is possible to generate a recording that is substantially better than predicted solely by the lens modulation transfer function. We will first consider the recordings based solely on the lens modulation transfer function and then consider how the recordings can be improved by taking advantage of the photoresist exposure–development nonlinearities that are possible under high-exposure conditions. We will finally illustrate these techniques with experimental results.

Since we are recording with an ON-OFF signal and the disc specifications are based on only peak-to-peak recording dimensions, what is pertinent is not the lens modulation transfer function but rather the contrast transfer function.[7] For a square-wave signal the contrast transfer function, ctf(ν), is given by

$$\mathrm{ctf}(\nu) = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\tilde{o}(n\nu)}{(2n-1)}, \qquad [8]$$

where ν is the spatial frequency of the square wave. Thus unless the sinusoidal MTF diminishes to less than (π/4), the contrast for ON-OFF signals remains at about 100%.

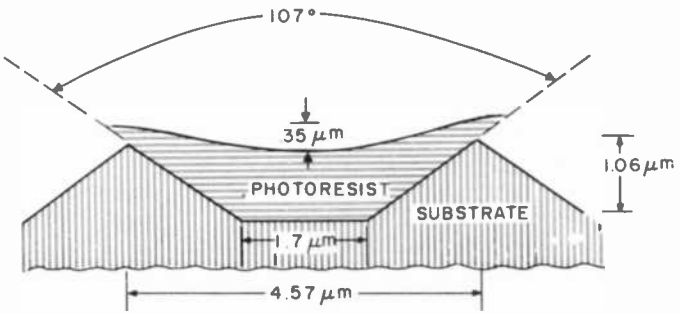


Fig. 6—Cross section of photoresist-covered pregrooved substrate.

In order to estimate the recorded signal geometry as a function of the optical-system modulation transfer function we need to model the behavior of the photoresist. We assume that, in the development of the exposed photoresist, the rate of material removal is proportional to the local exposure raised to the power γ, that is,

$$\frac{dD(z)}{dt} = KE(z)^{\gamma}, \qquad [9]$$

where D is depth removed by the developing solution, E is the local exposure, and K is a constant of proportionality that depends upon both the photoresist and the developer type. As before, z measures the tangential distance along the spiral signal track. Note that we also assume that the exposure is uniform throughout the depth of the photoresist. That is, we ignore possible attenuation of the exposing beam as it penetrates the resist layer as well as the possible interference structure that may be caused by reflections from the copper substrate. If we allow the development to proceed for a time τ, then the total depth removed will be simply,

$$D(z) = (K\tau)E(z)^{\gamma}. \qquad [10]$$

Applying this model to exposure with a square-wave signal of spatial frequency ν, we see that the exposure $E_v(\nu)$ in the signal valleys is given by

$$E_v(\nu) = E_{max}[1 + \mathrm{ctf}(\nu)]/2, \qquad [11]$$

and the exposure $E_p(\nu)$ in the signal peaks is given by

$$E_p(\nu) = E_{max}[1 - \mathrm{ctf}(\nu)]/2, \qquad [12]$$

where E_{max} is the maximum exposure for unmodulated light and $\mathrm{ctf}(\nu)$ is the system contrast transfer function for spatial frequency ν.

By combining Eqs. [10], [11], and [12] for the spatial frequencies corresponding to 4.3 and 6.3 MHz at various disc radii, one can calculate at what radii a particular optical system is capable of meeting the VideoDisc geometric signal specifications. For example, if we assume for the optical system a Gaussian point-spread function with a full width at half maximum of 275 nm, a *linear* (i.e., $\gamma = 1$) photoresist exposure model will meet specifications from the outermost disc radius to a radius of ~4.7 inches at which time the differential track drop (ΔTD) will exceed 5 nm.

If however, we assume that the photoresist development–exposure is nonlinear, for example with a $\gamma = 2$, then the calculated signal geometry will meet specifications up to a radius of 3.25 inches at which point the ratio (R) of the 6.3 and 4.3 MHz signal amplitudes will decrease to

less than 0.7. These results are depicted in Fig. 7, where we have normalized Eq. [10] so that SA(4.3 MHz) at $r = 5.75$ inches would be 100 nm. The fall-off of the signal amplitude at 4.3 MHz (SA), as well as the ratio of the 6.3 and 4.3 MHz signal amplitudes (R), are identical for both $\gamma = 1$ and $\gamma = 2$. However, whereas the differential track drop (ΔTD) exceeds 5 nm at about 4.7 inches radius for $\gamma = 1$, the differential track drop (ΔTD) for $\gamma = 2$ is less than 5 nm for radii greater than 3.0 inches. The effect of the nonlinearity ($\gamma = 2$) is to decrease both the track drop (TD) and the differential track drop (ΔTD). If we increase the nonlinearity to $\gamma = 2.5$, the differential track drop (ΔTD) decreases further; however, the signal amplitude ratio (R) also drops more rapidly and is less than 0.7 for all radii less than ~3.4 inches.

Also shown in Fig. 7 are the geometric parameters plotted on the assumption of another photoresist exposure–development model, which has been empirically established[8] and will be discussed in further detail in the next section of this paper. In this model the development rate versus exposure is given by

$$\frac{dD(z)}{dt} = K' \exp[aE(z)]. \qquad [13]$$

For the calculations with this model, the constants have been chosen so that the signal amplitude, SA (4.3 MHz), at a radius of 5.75 inches would be 100 nm. In Figs. 7 and 8, the parameter ρ is the ratio of the development rate at the peak exposure for a 4.3 MHz signal at radius of 5.75 inches to the development rate at zero exposure, i.e., the erosion rate. As can be seen from Fig. 7, a general property of increasing the nonlinearity is that while track drop (TD) and differential track drop (ΔTD) decrease, so do the signal amplitude (SA) and the ratio of the 6.3 and 4.3 MHz signal amplitudes (R).

Before discussing our experimental results let us relate these calculations to the anticipated results with an actual optical system. Although we have used several different lenses during our experiments, the MTF data depicted in Fig. 5 is typical of our optical recording system MTF. Fig. 8 shows the calculated geometric parameters assuming this experimental MTF. From this figure, the recording models described above predict that satisfactory recordings cannot be made at radii much smaller than about 4.5 inches. Recordings will have either too large a differential track drop (ΔTD) or too small a signal-amplitude ratio (R).

4.2 Results

A number of recordings were made on trapezoidally grooved substrates

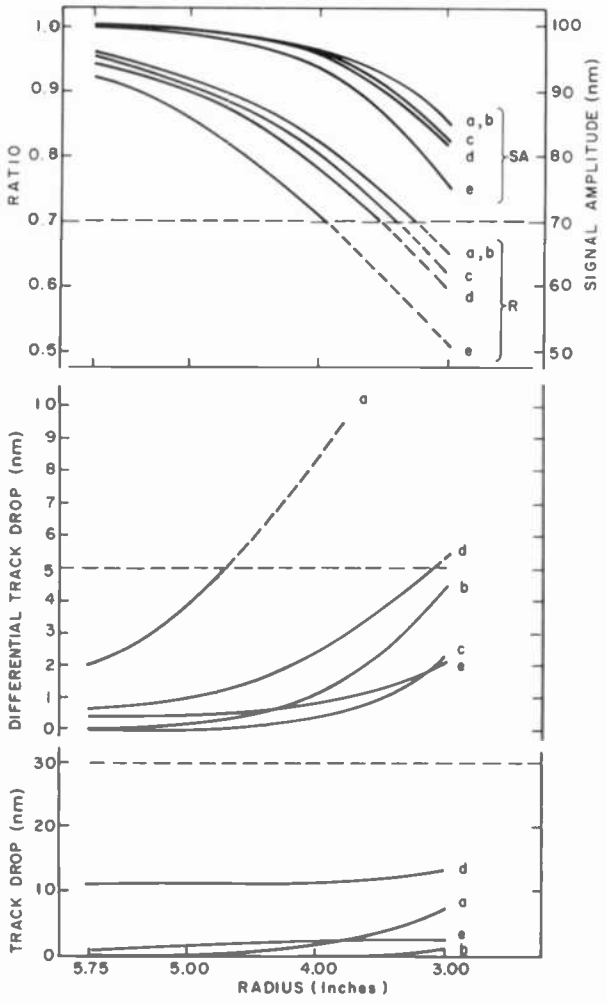


Fig. 7—Calculated geometric specification performance of exposed and developed photoresist as a function of disc radius, assuming a Gaussian beam profile with 275-nm FWHM.

a: Power-law model, $\gamma = 1$ (linear)
b: Power-law model, $\gamma = 2$
c: Power-law model, $\gamma = 2.5$
d: Exponential-law model, $\rho = 10$
e: Exponential-law model, $\rho = 50$
The TD curve for case c is not shown; it falls below that for case b.

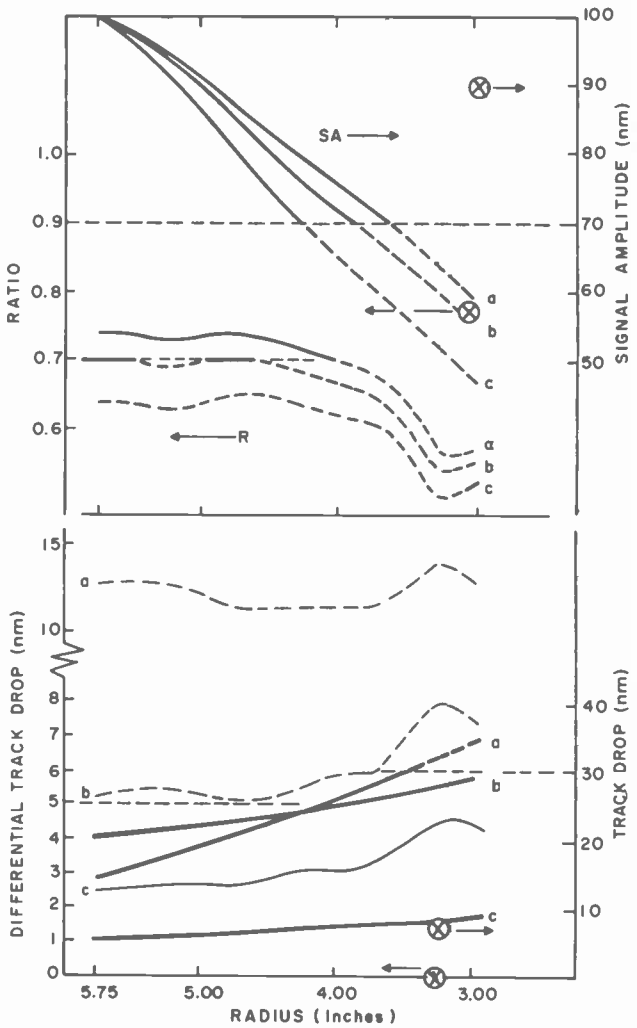


Fig. 8—Calculated geometric specification performance of exposed and developed photoresist as a function of disc radius, assuming the experimentally measured beam profile depicted in Fig. 4.

a: Power-law model, $\gamma = 1$ (linear)
b: Exponential-law model, $\rho = 10$
c: Exponential-law model, $\rho = 50$
Track drop curves are shown in heavy lines, while differential track drop curves are shown in lighter lines. The symbol ⊗ indicates one data point taken from recording $\theta 64$ for each of the geometric signal specifications.

that were coated with either AZ1350B* photoresist or a proprietary electron-beam resist that is also photosensitive. In general, most of these recordings did not meet the VideoDisc specifications; however, several recordings were particularly good. One such case is tabulated below and plotted in Fig. 8.

Recording	SA	R	TD	ΔTD
θ64	90 nm	0.77	7 nm	0 nm

This data was obtained from measurements of SEM photomicrographs. The anomalous absence of measurable differential track drop was noted on other recordings too, although generally either or both the signal amplitude and the signal ratio failed to meet specifications.

These anomalously good results obtained in several recordings we attribute to a fortuitous set of circumstances involving optical interference effects caused by reflections of the recording beam at the photoresist–substrate interface. These effects can be seen by examining carefully the SEM photomicrographs of the recorded signal. Fig. 9a shows the signal geometry for recordinng θ64. Notice the radial structure in the signal elements. Another example is shown in Fig. 9b of a recording in which the differential track drop was also zero (Recording θ79). Fig. 10 shows a striking composite of these effects on two other recordings.

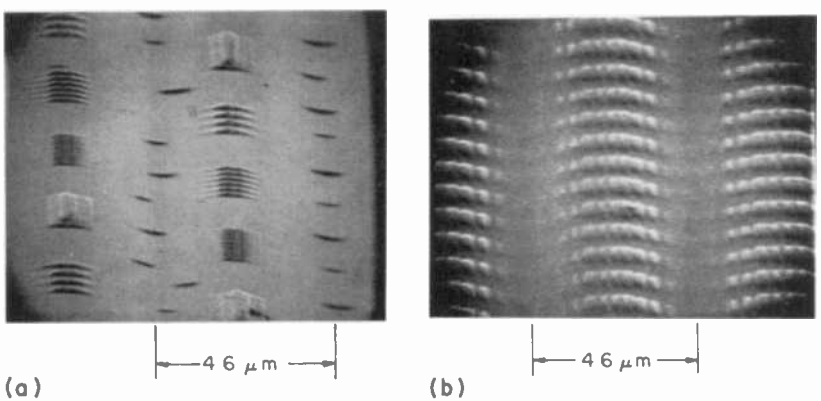


Fig. 9—Scanning electron photomicrograph of surface of VideoDisc master for (a) recording θ64 and (b) recording θ79.

* Shipley Chemical Company.

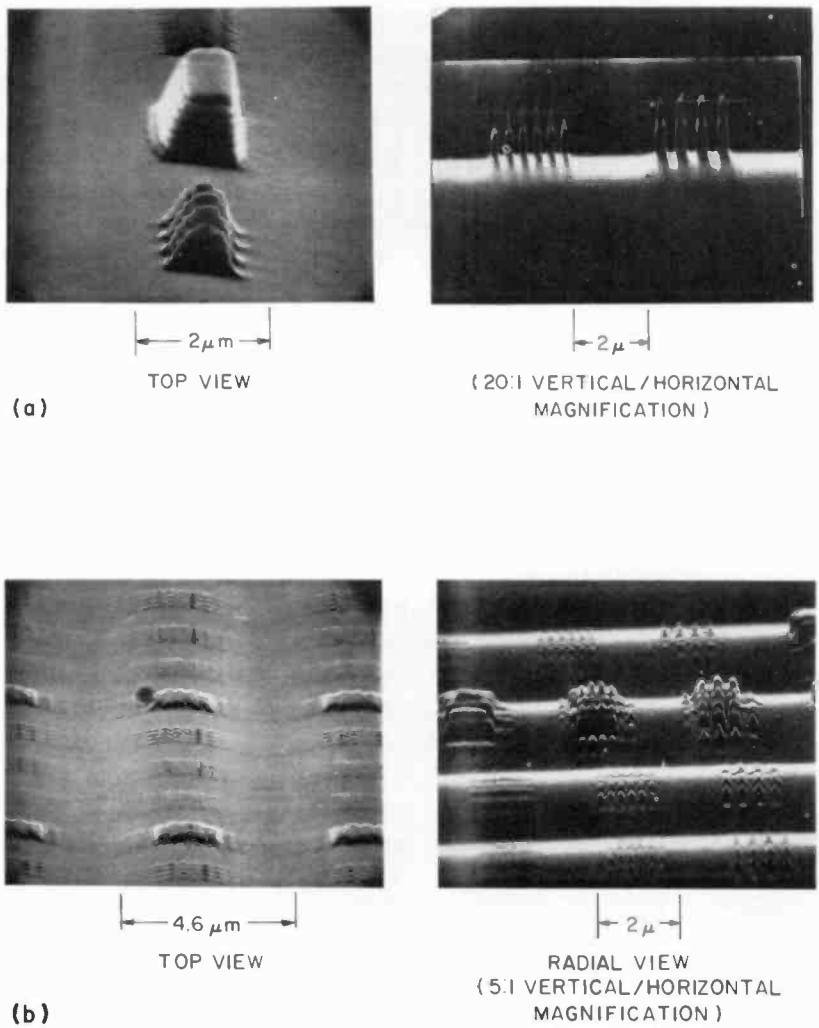


Fig. 10—Scanning electron photomicrographs of VideoDisc master showing top and radial view for (a) recording $\theta 51$ and (b) recording $\theta 74$.

These figures illustrate the strong role that optical interference plays. In recording on thick photoresist coated onto trapezoidally grooved substrates, these interference effects are largely uncontrollable. In the next section, we will examine a second recording technique in which the effects of interference are controlled and utilized.

5. Thin-Resist Recording

5.1 General Comments

Optical recording of VideoDisc signals into thick photoresist coated on trapezoidally pregrooved substrates resulted in microscopic signal element nonuniformities. As discussed above, these nonuniformities were caused by uncontrolled reflections of the recording beam from the bottom and the side walls of the groove cut into the substrate. The simplest solution to the problem is to eliminate substrate reflections altogether. Experimentally, this is very difficult to achieve, because simple black coatings are not sufficiently absorbing and the preparation of high quality anti-reflection coatings on top of the microgrooves would be difficult and probably costly. Furthermore, properly utilized interference effects are beneficial.

A possible approach to eliminating the undesirable nonuniformities is to change the groove geometry and substrate structure in such a manner that the substrate reflections produce a predictable and experimentally controllable interference pattern. One way to do this is to design a resist-coated reflective groove structure in such a way that the phase of the interfering reflected wave is constant across the groove. This constant phase can only be maintained if the groove walls in the substrate are not too steep. Simple geometric considerations suggest that a shallow, cusp-shaped substrate, coated conformally with a thin layer of resist with precisely controlled thickness, satisfies the constant-phase-interference requirement. We have successfully tried such a structure. We found that these thin-resist structures eliminated the undesirable signal-element nonuniformities and enhanced the relative recording response at high frequencies, where the finite size of the recording beam limits the achievable signal modulation depth.

This enhanced frequency response is the result of the combined effect of optical standing waves in the resist and photoresist nonlinearities. To realize this enhancement for improving the signal quality of VideoDisc recordings, the photoresist thickness and development must be carefully chosen. For a given combination of photoresist and recording-beam characteristics, photoresist thickness and development time form a two-dimensional space. Within this space a closed region exists where all geometric requirements set by the VideoDisc signal standards can be met. In the following, we first describe a set of results obtained by numerically evaluating the combined interference and nonlinear effects in terms of the simple power-law nonlinearity model. In order to test the sensitivity of our calculated results to changes in the model, we subsequently carry out a second set of calculations where the resist non-

linearities are treated in terms of the exponential-law model. We find that even though the exact range of the predicted resist thickness, development time, and beam resolution combinations where the VideoDisc standards can be met depends on the model chosen, both models predict a range of parameters where significant signal enhancement is possible. We also present experimental results obtained by recording onto thin-resist-coated cusp substrates. The data thus obtained is in good agreement with our analytical predictions.

5.2 Thin-Resist-Recording Models

Let us consider a thin layer of weakly absorbing photoresist coated onto a flat substrate. As a result of reflection at the photoresist–substrate boundary, an optical standing wave will be set up in the resist. In order to keep our model simple, we assume that the absorption in the resist has a negligible effect on the energy distribution in this standing wave. Also, we neglect the reflections that occur at the resist–air interface. Fig. 11a shows schematically the exposure pattern that results from the presence of the standing waves. Here $E(x)$ denotes the optical energy density that exposes the resist as a function of the depth coordinate x. The solid-line portion of the $E(x)$ curve is the actual standing-wave pattern and the dotted portion is an analytical continuation of this pattern inside the substrate. The distance $x_2 - x_1 = \lambda/2$, where λ is the light wavelength in the resist. Throughout the region of interest to us, the resist thickness x_s is less than $\lambda/2$. The standing-wave energy reaches its maximum value at x_m. The energy density is normalized so that in the absence of standing waves the total exposure is unity at the point of signal maximum for signal wavelengths that are much longer than the characteristic dimension of the recording beam. At shorter signal wavelengths, because of the imperfect beam response, the exposure characteristic of the recording is as schematically shown in Fig. 11b. Here z is the coordinate along the signal track, the signal wavelength is Λ, and $I(z)$ is the incident exposure variation along the track. For a perfect beam or for $\Lambda \rightarrow \infty$, and for a perfectly reflecting substrate $x_2 = x_s$, the maximum exposing energy density is $E(x_m) = 2$ at $z = 0$. Also, for a perfectly reflecting substrate $E(x_s) = 0$ for all z.

The rate of photoresist removal during development is, in general, a nonlinear function of exposure. The simplest way to model this nonlinearity is to assume that the removal rate is the sum of an exposure-independent erosion term and of an exposure-dependent term that is proportional to $E(x)^\gamma$, where the exponent γ is an empirically determined nonlinearity coefficient.

For a general periodic signal, the incident time-integrated beam intensity as function of distance along the track is

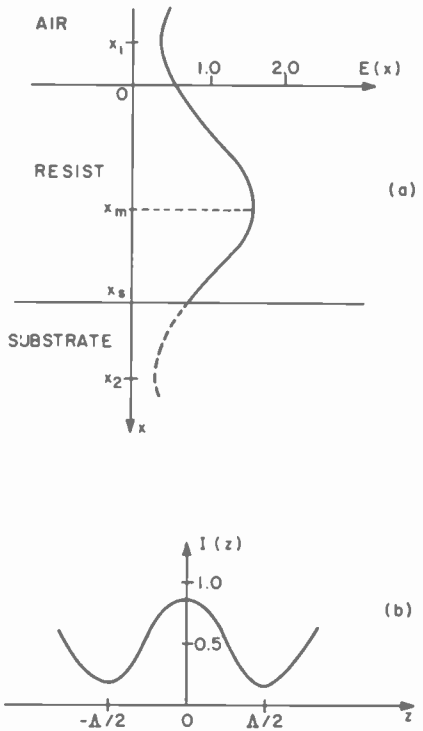


Fig. 11—(a) Exposure variation with depth for a thin resist layer coated onto a partially reflective substrate. (b) Incident exposure pattern produced by an imperfect (finite size) recording beam 100% modulated by a sinusoidal signal that produces a signal wavelength Λ on the moving disc.

$$I(z) = 0.5 + \sum_n M\left(\frac{n}{\Lambda}\right) \cos\left(\frac{2\pi z}{\Lambda}\right), \qquad [13]$$

where $M(n/\Lambda)$ is the product of the beam modulation transfer function evaluated at the nth signal harmonic, $\tilde{\sigma}(n/\Lambda)$, and the corresponding Fourier expansion coefficient of the temporal intensity modulation. The use of the pure cosine expansion is justified here because in practice the recording beams can be assumed to be symmetrical and the input signal waveshapes assumed to be either sinusoidal or square wave.

Based on the foregoing discussion we write the rate of resist development as

$$\frac{dD}{dt} = R_1 \left[I(z) \left\{ 1 + A^2 - 2A \cos\left(\frac{4\pi x}{\lambda} + \phi\right) \right\} \right]^\gamma + R_2, \qquad [14]$$

where D is the resist removed during the development time t, R_1 is the development rate at unity exposure, A is the absolute value of the am-

plitude reflection coefficient at the resist–substrate boundary, $\phi = -4\pi x_1/\lambda$ is the phase of the standing-wave pattern at the air-resist interface, and R_2 is the erosion rate.

The phase coefficient ϕ is determined by the complex amplitude reflection coefficient at the resist–substrate boundary, and by the resist thickness x_s. For any given substrate reflectivity, ϕ can be set by changing the resist thickness. Thus ϕ can be considered as a thickness parameter and, apart from the development time, it plays the most significant role in establishing the developed signal surface relief contours. The rate coefficients R_1 and R_2 are both dependent on the resist composition and on the developer constituents and concentration. Also, R_1 depends on the actual exposure level. For the purposes of numerical calculations we assumed $R_1 = 100$ nm/min and $R_2 = 10$ nm/min. Unless otherwise stated, all results presented here were obtained with these rate values; they approximately correspond to typical VideoDisc recording conditions. Any proportional change in the values of R_1 and R_2 such that the ratio R_1/R_2 is constant simply changes the development time-scale factor. Changes in the ratio R_1/R_2 will affect the quantitive details of our results, but over a rather wide range the most important general predictions of the model remain unaffected. All calculations were carried out with $\lambda = 268$ nm, corresponding to the wavelength inside typical photoresists for the 442 nm HeCd laser.

The surface contour that results from a given exposure and subsequent development can be obtained from straightforward numerical integration of Eq. [14]. Typical results that correspond to recording a 6-MHz square wave signal at a 3.25-inch radius on a disc rotating at 450 rpm, with a 300 nm full-width-at-half-intensity (FWHI) Gaussian beam, are shown in Figs. 12 and 13 for $\phi = 0$ and $\phi = 1$ rad, respectively. In calculating the curves shown in these figures an amplitude reflection coefficient of $A = 0.6$ was assumed; however, again, the results are relatively insensitive to changes in the value of A. When $\phi = 0$ radians, the air–resist interface coincides with a standing-wave minimum. In this case, as shown in Fig. 12, only erosion occurs during the initial stages and the signal shape is not discernible until about 1.5 minutes of development. When $\phi = 1$ the signal is completely developed in less than 1 minute (see Fig. 13). The difference in initial resist thickness between the $\phi = 0$ rad and $\phi = 1$ rad cases is only approximately 20 nm.

The presence of standing waves and nonlinearities also has a strong influence on the frequency response of the recording system. The response curves shown in Fig. 14 were again obtained by integrating Eq. [14]. Here we define the frequency response as the frequency-dependent relative signal pit depth obtained after development of photoresist that was exposed by a modulated beam. For Fig. 14 we again assume the same beam and resist parameters as used above for Fig. 13. For reference we

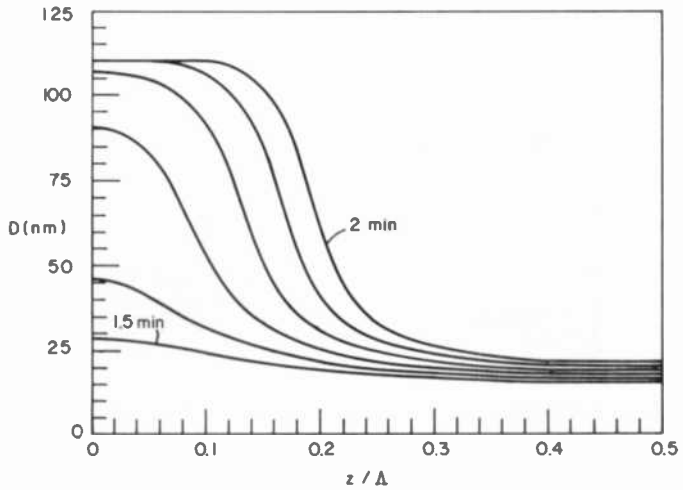


Fig. 12—Evolution of the signal pattern during development when a standing-wave null occurs at the resist–air interface. The recording parameters are: 300-nm FWHI Gaussian beam, 6-MHz square-wave modulated signal at 3.25-inch radius, $\lambda = 268$ nm, $A = 0.6$, $\phi = 0$ rad, $R_1 = 100$ nm/min, $R_1/R_2 = 10$, and $\gamma = 2$. The signal profiles are plotted in 0.1-minute intervals between 1.5 and 2 minutes development time.

include in Fig. 14 the response curve for an interference-free linear recording medium. As Fig. 14 indicates, the improvement of the high-frequency response as a result of the combined nonlinearity and interference effects is substantial. For example, at 0.5 μm signal wavelength the relative response can be increased from approximately 35% to 85%. The significance of the above frequency-response improvement can be further appreciated by considering the recording specification that the ratio (R) of the depth of the 6.3-MHz signal elements to those at 4.3 MHz is to be greater than 0.7. In the absence of standing waves and nonlinearities, 760 nm is the shortest signal wavelength recordable with a 300-nm full-width-at-half-intensity Gaussian beam. This corresponds to 6.3 MHz at a 4-inch recording radius. However, with a $\gamma = 3$ nonlinearity, this short wavelength limit can be extended down to 400 nm, corresponding to 6.3 MHz at a radius of 2.1 inches.

This improvement in frequency response can be understood as follows. The presence of nonlinearities increases the differential rate of development between high and low exposure regions of the resist. This differential rate can then be utilized to flatten the frequency response by choosing a combination of resist thickness and development time such that in the regions of maximum signal ($z = 0$ in Fig. 11b), the resist is developed down to the substrate before significant development can

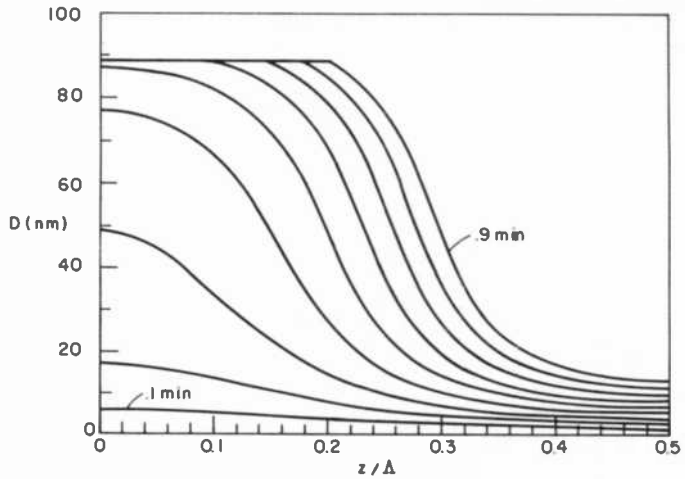


Fig. 13—Evolution of the signal pattern during development. All recording parameters are the same as for Fig. 2 except that here the resist thickness has been reduced so that now $\phi = 1$ rad. The signal profiles are plotted in 0.1-minute intervals between 0.1 and 0.9 minute development time.

occur in regions of minimum signal ($z = \pm\Lambda/2$ in Fig. 11b). Furthermore, during development the signal pit will reach the substrate first in the regions of low-frequency-signal maxima. With further development, the regions of high-frequency-signal maxima will eventually also reach the substrate surface, and at that time the pit depths at the low- and high-frequency-signal maxima are equal.

The foregoing discussion assumed that the initial resist thickness is less than $\lambda/2$. If the resist thickness is greater than $\lambda/2$, then standing-wave minima may occur inside the resist. With sufficient substrate reflectivity, the exposure at this standing-wave minimum is sufficiently small to inhibit development beyond the first minimum encountered by the growing signal pit during development. Thus the signal depth will be effectively clipped at the plane of this minimum. This mode of operation has the advantage that the substrate remains protected throughout the development process by a layer of undeveloped resist. Frequently, when the resist is fully developed to the substrate, contamination or etching of the substrate surface may render it useless for subsequent reuse.

The presence of nonlinearities results in significant distortion of the recorded signal waveshape. These distortions are clearly visible in Figs. 12 and 13, which show the signal contours at various stages of development for a symmetric-square-wave modulation signal. A properly designed FM system is insensitive to signal shape distortions provided the

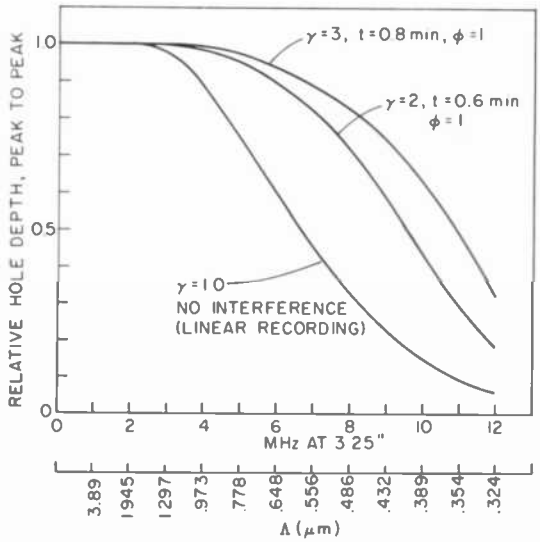


Fig. 14—Relative signal pit depth as a function of frequency. With the exception of the parameters explicitly shown here, all recording parameters are the same as those in Figs. 2 and 3. The parameter t denotes the development time.

positions of the zero-crossings are unperturbed. A useful operational definition for the zero-crossings is that they are the locations where the signal crosses its average value. For a symmetric signal waveshape, this always occurs halfway between adjacent minimum and maximum signal points. As indicated by the curves in Figs. 12 and 13, nonlinear thin-resist recording may produce significant zero-crossing errors. However, as also indicated by these curves, for any signal frequency there are resist thickness and development time combinations such that this error is zero. In general this optimum combination may be different for different frequencies. Detailed numerical calculations were carried out to investigate this question, and we found that in the vicinity of zero error the frequency dependence of this error is negligible. The results of a sample calculation are shown in Fig. 15. Here we again used the same resist and beam parameters as before (square-wave-modulated, 300-nm Gaussian beam, $R_1 = 100$ nm/min, $R_1/R_2 = 10$, $\gamma = 2$, and $A = 0.6$), and we show the zero-crossing error as a function ϕ for a constant 0.6 min development time for 4- and 6-MHz signals recorded at 3.25-inch radius. The zero-crossing error is given in degrees; 45° corresponds to a crossing position error of $\Lambda/8$. As shown in Fig. 15, with the assumed parameters both the 4- and 6-MHz signals can be recorded with no zero-crossing error provided that the resist thickness is so chosen that $\phi = 1.1$ rad. In the vicinity of this error-free operating point, there exist some finite ranges of

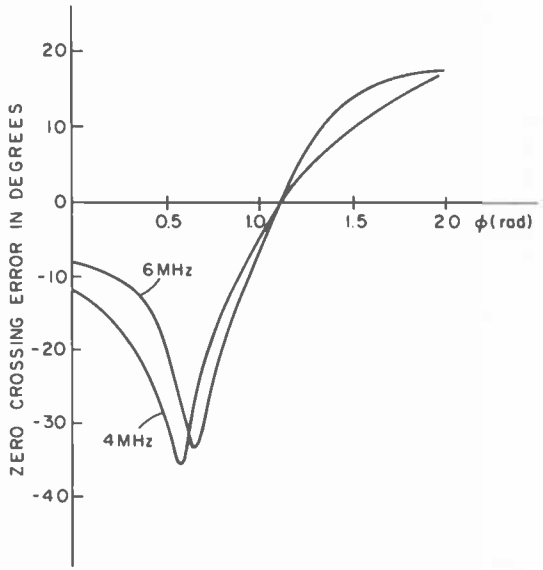


Fig. 15—Zero-crossing error of the recorded wave shape as a function of the resist thickness parameter ϕ for 4-MHz and 6-MHz square-wave-modulated signals at 3.25-inch radius. The development time is $t = 0.6$ min. All other recording parameters are the same as those in Fig. 2 and 3.

thickness and development-time combinations where, even though the zero-crossing error is finite, the FM signal distortions are acceptably small.

In the course of several years of VideoDisc research, it has been determined that, given the wide variety of signal distortions associated with the various mastering techniques, the most convenient practical way of specifying the signal geometry that will produce acceptable playback is in terms of a set of signal amplitude ratios and signal minimum and maximum values. These signal standards were described and summarized in an earlier section of this paper.

In the foregoing discussion of thin-resist recording we have explored the key features of the signal evolution process. Now we shall use the geometric signal standards as our acceptance criteria and explore the range of recording parameters where all of these standards can be met simultaneously.

The previous thin-resist development calculations were based on a simple model of resist behavior that assumed that the development rate is equal to the sum of an exposure-independent erosion term plus another term that is a nonlinear function of exposure. This simple model was found to be very convenient and useful for exploring the effects of

nonlinearities; however, its formation is purely heuristic and the physical foundations of its details are suspect. For example, it is not possible to fit the experimental exposure-versus-development-rate curve over the entire range of interest with a single γ. For calculating the range of recording parameters where the VideoDisc signal standards can be met, therefore, we have also used another model. Even though this alternative model also cannot be rigorously derived from the first principles of the physical chemistry of photoresists, its key assumptions are well founded and in certain experimental situations it was found to fit the data over a wide range of parameters. For clarity in the forthcoming discussions, we shall refer to our original model as the "power-law" model and to this alternative model as the "exponential-law" model.

The exponential-law model is based on the assumption that, for given development conditions, the development rate of a positive resist is uniquely determined by the inhibitor concentration.[8] The change in inhibitor concentration during exposure can be calculated from the well-known absorption law

$$\frac{dC}{d\tau} = -K_1 C(x,z,\tau) P(x,z,\tau), \qquad [15]$$

where C is the inhibitor concentration, P is the optical power, τ is the exposure time variable, and K_1 is a proportionality constant. The change in concentration during exposure is

$$\Delta C = \int_0^{\tau_0} \frac{dC}{d\tau} d\tau, \qquad [16]$$

where τ_0 is the exposure time. If $C \gg \Delta C$ for all τ, then

$$\Delta C = -K_1 C_0 E(x,z). \qquad [17]$$

Here C_0 denotes the uniform inhibitor concentration prior to exposure and E is the exposure energy density.

Based on an experimentally obtained curve that describes the relationship between development rate and inhibitor concentration for certain resists,[8] we assume that over the range of parameters of interest to us

$$\log\left(\frac{dD}{dt}\right) = K_2 - K_3 \Delta C, \qquad [18]$$

where K_2 and K_3 are constants of the development process. Combining Eqs. [17] and [18], we obtain

$$\frac{dD}{dt} = K_4 e^{K_5 E(x,z)}, \qquad [19]$$

where K_4 and K_5 are again constants that characterize the resist and the

development process. In the absence of exposure, $E(x,z) = 0$ and thus K_4 describes the erosion rate. In the presence of substrate reflections, Eq. [19] can be written as

$$\frac{dD}{dt} = K_4 \exp\left\{K_5 I(z)\left[1 + A^2 - 2A\cos\left(\frac{4\pi x}{\lambda}\right)\right]\right\}. \qquad [20]$$

Eq. [20] specifies the signal contours for the exponential-law model in the same manner as Eq. [14] does for the power-law model.

The range of resist-thickness and development-time combinations where the VideoDisc signal standards can be met was calculated using both the power-law and the exponential-law models, and representative results are presented, respectively, in Figs. 16 and 17. Here D_0 is the photoresist thickness prior to development and t is the development time. The curves drawn on these figures are the boundaries within which, for a given set of parameters, the signal standards are met. These curves were generated by first numerically integrating Eqs. [14] and [20] for the 4.3 MHz and 6.3 MHz carriers at 2.9-inch disc radius and, subsequently, testing whether the signal contours thus obtained met the standards.

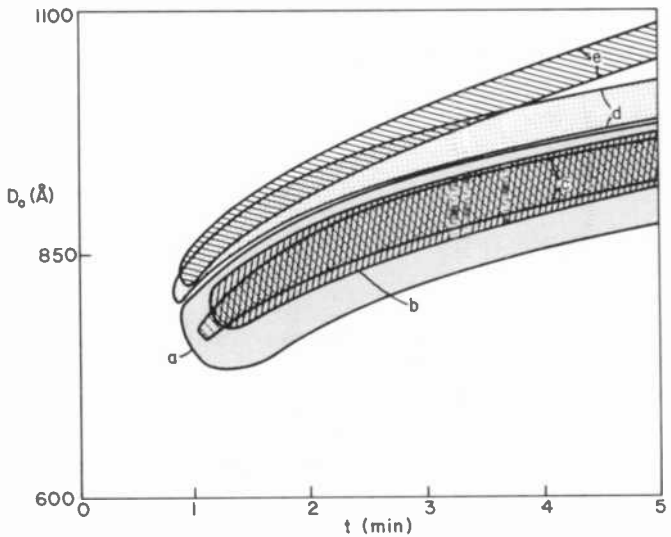


Fig. 16—Power-law model prediction of the range of resist-thickness and development-time combinations where the VideoDisc recording standards can be met with a square-wave-modulated signal. For all curves $\lambda = 268$ nm, $A = 0.7$, and $R_1^\circ = 200$ nm/min. The other parameters are as follows:

a: $w = 300$ nm, $\gamma = 2.5$, $R_2 = 2$ nm/min;
b: $w =$ ELD, $\gamma = 2.5$, $R_2 = 2$ nm/min;
c: $w = 350$ nm, $\gamma = 2.5$, $R_2 = 2$ nm/min;
d: $w =$ ELD, $\gamma = 2.0$, $R_2 = 2$ nm/min;
e: $w =$ ELD, $\gamma = 2.0$, $R_2 = 4$ nm/min.

Experimentally, the development time is adjustable by appropriate dilution of the developer. Therefore, the exact values of the development rate coefficients are insignificant. They were so chosen that the resulting development times fall within the typical experimental range.

All curves in Fig. 16 were calculated for a constant peak development rate $R_1^* = 200$ nm/min, where $R_1^* = R_1 I(0)(1 + A)^2$ is the maximum local signal development rate that occurs at the 4.3-MHz signal peak at the standing-wave maximum. Likewise, all curves in Fig. 17 were calculated for a constant maximum development rate $R_{max}(4.3) = 500$ nm/min, where $R_{max}(4.3) = K_4 \exp[K_5 I(0)(1 + A)^2]$. Note that the power-law model allows independent specification of the signal development and erosion rates, while in the exponential-law model these two rates are coupled. The degree of nonlinearity is specified for the power-law model by γ and for the exponential-law model by $\rho = R_{max}(4.3)/K_4$. The influence of recording-beam resolution was tested by examining two Gaussian beams with full widths at half intensity $w = 300$ and 350 nm, respectively. The curves shown with w = ELD were generated by using the experimental lens data (ELD) shown in Fig. 4. The substrate–resist interface reflection value used was $A = 0.7$, which is typical for copper substrates and AZ1350B photoresist. The optical wavelength in the resist was assumed to be 268 nm, corresponding to the

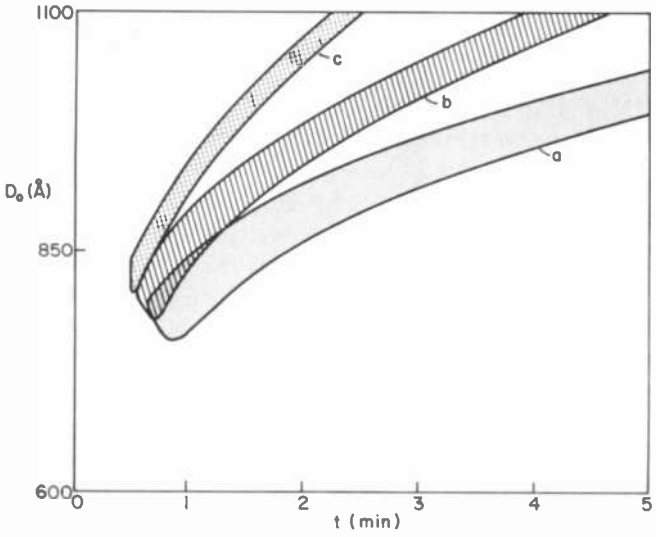


Fig. 17—Exponential-law-model prediction of the range of resist-thickness and development-time combinations where the VideoDisc recording standards can be met with a square-wave-modulated signal. The parameters are as follows: for all curves $\lambda = 268$ nm, $A = 0.7$, w = ELD and $R_{max}(4.3) = 500$ nm/min. The other parameters are as follows: a: $\rho = 200$; b: $\rho = 100$; c: $\rho = 50$.

442 nm HeCd laser. The beam was again assumed to be 100% modulated by square-wave signals.

Figs. 16 and 17 indicate that for a given resist thickness the development time must be carefully adjusted. Conversely, for a given development time, the thickness variation must be maintained within a narrow range, typically less than 5 nm. In general, increasing the nonlinearity (by increasing the exposure) and reducing the spot size will increase the tolerable thickness variation and the allowed development timing error. Outside the closed boundary of the acceptance regions, signal geometry fails for the following reasons. For a given development time, if the thickness is increased beyond the acceptance boundary, then first the high-frequency signal becomes shallow and the signal-amplitude-ratio specification (R) can not be met. As the thickness is further increased, eventually the depth of the 4.3 MHz signal (SA) becomes less than the required 70 nm. If for a constant development time the thickness is reduced below the acceptance boundary, first the differential track drop (ΔTD) increases and then, with further thickness reduction, the absolute track drop (TD) also increases beyond its allowed range. Furthermore, if the thickness is sufficiently reduced, the signal-depth requirement (SA) can not be met either. All of these conclusions are independent of the exact value of the parameters and of the model chosen; they are fundamental consequences of the combined effects of the standing waves and of the nonlinearities present in the thin-resist exposure and development process.

5.3 Experimental Considerations

A number of thin-resist recordings were made using copper substrates into which grooves were cut with a smooth-radius-tipped tool. The groove bottoms thus produced were cylindrical and the cross section of the closely spaced grooves had the appearance of a series of cusps. These substrates were then coated with a thin layer of photoresist. Most experiments were performed with AZ1350B positive photoresist. The resist-coated groove cross section is shown in Fig. 18. Here g denotes the groove width; d is the resist thickness at the groove center; R_s and R_r denote, respectively, the radii of curvature of the groove bottom and of the resist surface; and s is the signal width in the radial disc direction.

All thin-resist experiments were carried out in an experimental 5555-grooves-per-inch disc format. The typical dimensions were $R_s = 7\ \mu$m, $R_r = 8\ \mu$m, $s = 3.5\ \mu$m, and d was in the range of 80 to 100 nm. The radii of curvature of the groove bottom and resist surface were determined from diffraction data.[3,9,10] The developed resist thickness was estimated from the diffraction spectra produced by the signal elements.

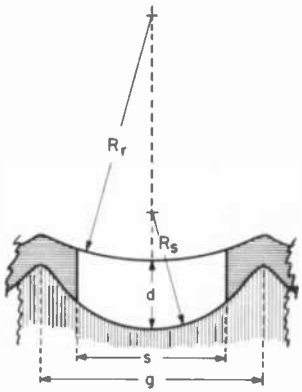


Fig. 18—Thin-resist-coated cusp substrate geometry.

The thin-resist-model calculations indicated that the resist must be maintained uniform within a narrow tolerance range in order to meet the signal standards. The groove geometry described above indicates that the resist does not coat the substrate conformally. Simple geometric calculations suggest that the difference in resist thickness between that at groove center and at locations ± 1.75 μm off center is approximately 30 nm. This differential thickness is clearly outside the range of allowed tolerance indicated by Figs. 16 and 17.

The systematic thickness variation across the grooves can be compensated for by appropriately shaping the recording-beam profile. In the thickness range of interest for constant exposure, the thin regions develop faster than do thick regions. Thus, as discussed above in connection with Figs. 16 and 17, if the development time is correctly chosen for the thick region at groove center, then the thin off-center regions will show excessive track drop. By reducing the off-track-center beam intensity with respect to its on-center value, this difficulty can be overcome. Numerical calculations using the power-law model with $\gamma = 2$ indicate that, with the substrate geometry described above, the signal profile standards can be met everywhere over a 3.5-μm-wide region provided the beam intensity is smoothly reduced from its maximum value at track center to approximately 60% of this maximum at ± 1.75 μm off center.

The signal structure achievable with thin-resist-coated cusp substrates is shown in Fig. 19. This test recording was made at a rotational speed of 450 rpm, and the SEM picture shown was taken at approximately a 5.7-inch disc radius. The recording was made into an RCA proprietary positive photoresist. The nominal resist thickness at groove center was 80 nm. The smooth signal structure across the grooves is a clear indication of a well-controlled standing-wave pattern. The spatial frequency

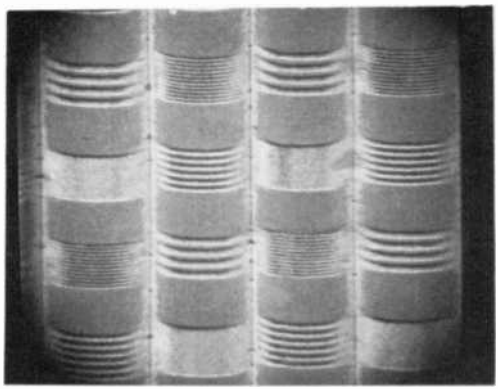


Fig. 19—Thin-resist VideoDisc recording observed with a scanning electron microscope. The test signal shown here consisted of a continuously repeated sequence of a 1-μsec-long pulse followed by 4-, 6-, and 12-MHz bursts. The features are separated by 1-μsec-long signal-off intervals.

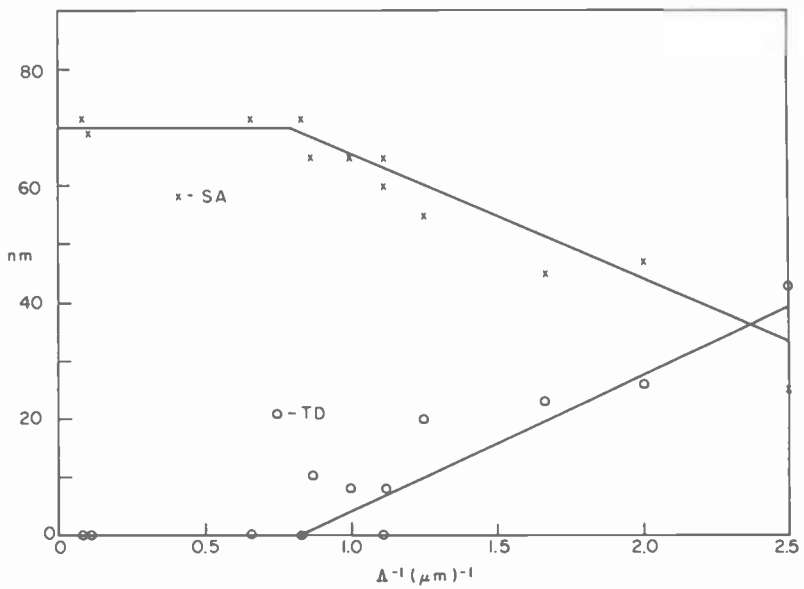


Fig. 20—Signal amplitude response (SA) and track drop (TD) as a function of spatial frequency, measured with a scanning electron microscope at track center. The straight lines are drawn for the purpose of qualitative trend indication.

response of this recording is shown in Fig. 20. The data points shown in Fig. 20 were measured with a scanning electron microscope on samples taken at various locations on the disc. The scatter in the data points is a characteristic consequence of slight coating-thickness nonuniformities in thin-resist recordings. The carrier-to-noise ratio measured at 5 MHz in a 30-kHz bandwidth was 52 dB and 50 dB, respectively, at approximately 5.5- and 3.2-inch disc radii.

Comparison of the signal-amplitude data points in Fig. 20 with the frequency response of the recording lens, shown in Fig. 5, clearly indicates the frequency-response enhancement in thin-resist recordings. The flat low-frequency response at $\Lambda^{-1} \lesssim 1\ (\mu m)^{-1}$ is the result of "bottoming" of the signal elements and the substrate. The slope improvement is the result of the combined standing-wave and nonlinear effects.

6. Linear Recording

6.1 Principle of Linear Recording

In the nonlinear recording process discussed so far and illustrated in Fig. 3, the modulator turns the beam on and off as fast as possible in response to the ON-OFF signal that emerges from the limiter. The locations of these ON-OFF transitions on the disc convey the information. When the beam width is much smaller than the wavelength of the recorded signal, a reference "land" level on the disc is established when the beam is turned OFF, and a constant valley level is established when the beam is turned ON. Under these ideal conditions, the nonlinear response of the photoresist causes no errors in the location of the transitions. However, when the width of the beam is not negligible compared with the signal wavelength, distortion in the system causes errors in the apparent location of the ON-OFF transitions as they are seen by the readout device. These errors are difficult to control or to compensate with predistortion of the input signal. In the outputs of the FM demodulators in the player, the effect of these errors is distortion of the information, resulting in a variety of undesirable beats.

Whenever the beam width is not negligible compared with the wavelength of the recorded carrier, linear recording becomes attractive. In linear recording, the intensity of the recording beam varies continuously in linear response to the applied input signal.

Fig. 21 is a simplified block diagram of a linear recording and playback system. The system shown in Fig. 21 differs from the more detailed system shown in Fig. 3 only by the blocks called "linear aperture corrector", and "nonlinear corrector," and by the fact that the modulator is assumed to be linear. In Fig. 21 the input signal at interface X is shown to be the sum of frequency-modulated sine-wave carriers, but it could

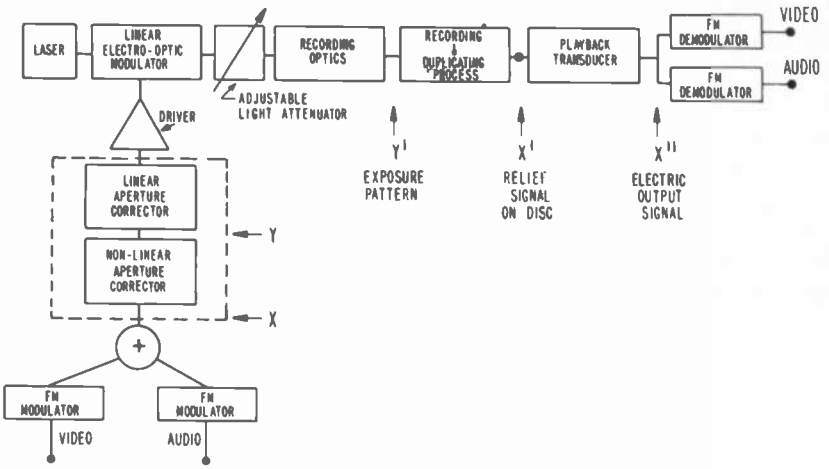


Fig. 21—Block diagram of a linear recording and playback system.

also be any other signal to be recorded on the disc. The purpose of a linear recording system is to reproduce a faithful linear replica of the input signal as a relief pattern on the disc, shown as interface X′ in Fig. 21. Ideally the system is to be completed by a player with a linear transducer, so that the electric signal at the output of the transducer at interface X″ is a linear replica of the input signal X. In Fig. 21 the output at interface X″ is shown to be delivered to FM demodulators.

To obtain a distortion-free reproduction of the input on the disc (interface X′), it is necessary to compensate for the linear distortion caused by the aperture of the recording beam and the possible nonlinear distortion caused by the development characteristic of the photoresist as well as by the disc replication process. The linear aperture corrector, which precedes the modulator driver, ideally corrects for the MTF characteristic of the beam so that the transfer characteristic from the input of the aperture corrector (interface Y) to the light exposure of the photoresist (interface Y′) has flat amplitude and group delay responses over the frequency band of interest. In a system with an ideal aperture corrector, it is also theoretically possible to correct for a nonlinear relation between depth of the relief pattern and the exposure. Such nonlinear correction must precede the aperture corrector, as shown in Fig. 21. Nonlinear correction was not included in the experimental system.

6.2 Principles and Limitations of Aperture Correction

Fig. 22 shows the transfer characteristic of the electro-optic modulator. The horizontal axis is the input voltage V volts from the driver, and the vertical axis is the output power P watts of the optical beam. The transfer characteristic is linear only over a limited range centered around a dc bias point, at which the input voltage is $\overline{V}$ and the output light power is $\overline{P}$. The output from the modulator is unavoidably attenuated by the optical system and intentionally by the adjustable attenuator (c in Fig. 3) with a gain g_0. The average exposure of the disc in Joules per meter of track length is consequently

$$\overline{E} = g_0\overline{P}/v \text{ Joules/m}, \qquad [21]$$

where

v = track velocity m/s and

g_0 = beam power gain from modulator to disc surface.

The average depth of the recorded relief signal is established by the average exposure $\overline{E}$, which is kept constant by keeping the average modulator output power $\overline{P}$ constant and by varying the gain g_0 in proportion to the track velocity v.

A sine-wave carrier of frequency f emerging from the modulator with an amplitude $p\overline{P}$ will generate an exposure pattern with a wavelength $\Lambda = v/f$ which has an exposure amplitude $e\overline{E}$, where

$$e = \breve{\sigma}(v)p. \qquad [22]$$

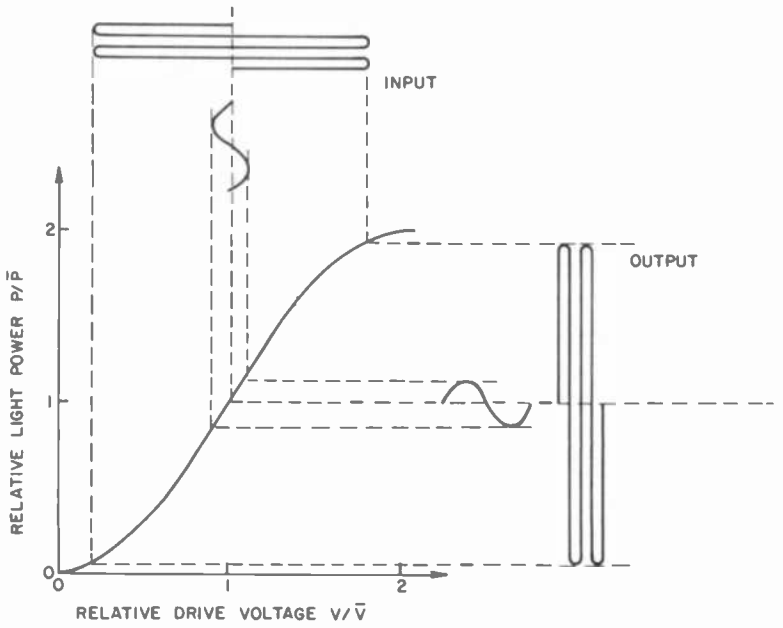


Fig. 22—Transfer characteristic of electro-optical modulator.

Here,

p = amplitude of power variation of beam emerging from modulator relative to the average power $\overline{P}$ of the beam emerging from the modulator.

$\tilde{\sigma}(\nu)$ = response of the recording beam normalized to make $\tilde{\sigma}(0) = 1$ (i.e., MTF),

$\nu = 1/\Lambda = f/v$ = spatial frequency in cycles/meter, and

e = amplitude of exposure variation relative to average exposure $\overline{E}$.

Fig. 5 shows the measured MTF response, $\tilde{\sigma}(\nu)$, of the recording beam used in the experimental system.

To keep e constant over a band of spatial frequencies ranging from zero to ν_{max}, and assuming that the MTF response of the beam $\tilde{\sigma}(\nu)$ decreases monotonically with ν, the modulator must deliver a maximum output p_{max} when recording the spatial frequency ν_{max} and a much smaller amplitude $p_{min} = e$ at low frequencies where $\tilde{\sigma}(\nu) \approx 1$. With an ideal linear modulator the largest linear output is $p_{max} = 1$, which is obtained when the modulator is driven from cutoff to a peak output of $2\overline{P}$. Even if the modulator is driven hard into a nonlinear mode of operation, the maximum output can only be slightly larger than $p_{max} = 1$. Assuming $p_{max} \simeq 1$, then

$$e \cong \tilde{\sigma}(\nu_{max}). \qquad [23]$$

Thus, it is seen that to achieve a significant aperture correction, the average exposure $\overline{E}$ must be much larger than the exposure amplitude $e\overline{E}$ of the recorded signal. This suggests a recording technique in which the grooves in the disc are recorded simultaneously with the signal. The average exposure $\overline{E}$ establishes the average groove depth while $2e\overline{E}$ determines the peak-to-peak amplitude of the recorded relief pattern. As an illustration, assume that the development characteristic of the photoresist is linear, i.e., that the depth is proportional to the exposure, and that the groove depth is to be 320 nm while the peak-to-peak signal is to be 80 nm. Then $e = 1/8$ and a flat frequency response can be obtained up to a frequency ν_{max} at which the MTF response is $20\log 8 = 18$ dB below the low frequency response.

In practice the band that can be equalized is further limited by the nonlinear characteristic of the modulator. As the relative output level p of the signal from the modulator is increased, the distortion increases. The modulator which was used in the experimental system may serve as an illustration. The light output power P of this modulator varies with the applied voltage V as $\sin^2(\text{const} \times V)$. If the modulator is biased at the point of symmetry, the dynamic transfer characteristic is, as illustrated in Fig. 22,

$$\Delta P/\overline{P} = \sin((\pi/2)(\Delta V/\overline{V})), \qquad [24]$$

where ΔP and ΔV are the deviations from the bias point $\overline{P}$, $\overline{V}$. If the applied input voltage is a sine wave $\Delta V = V\sin\phi$, where $\phi = 2\pi ft$ is the instantaneous phase, then

$$\Delta P/\overline{P} = 2\sum_{n=0}^{\infty} J_{2n+1}(x)\sin((2n+1)\phi), \qquad [25]$$

where $x = (\pi/2)(V/\overline{V})$ and $J_n(x)$ are Bessel functions. Expressed in terms of the fundamental component p of the output signal from the modulator, the relative output signal is

$$\Delta P/\overline{P} = p(\sin\phi + d_3(p)\sin 3\phi + \ldots), \qquad [26]$$

where

$$p = g_m(p)x \qquad [27]$$

$g_m(p) = p/J_1^{-1}(p/2)$ = gain of modulator normalized to be unity when the signal is small

$d_3(p) = 2J_3(J_1^{-1}(p/2))/p$ = third harmonic distortion

and where J_1^{-1} is the inverse of the Bessel function J_1 for $p < 1.8$.

Fig. 23 shows the modulator gain 20log g_m and the third-harmonic distortion d_3 in dB as a function of the relative output level of the fundamental 20log p in dB. Although the third harmonic of a high-frequency carrier is harmless *per se*, the presence of a third-order harmonic is an

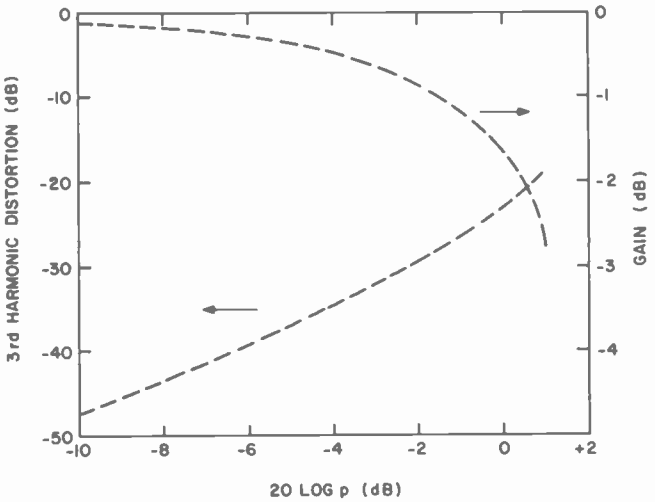


Fig. 23—Gain and 3rd-harmonic distortion of electro-optical modulator as a function of the output amplitude of the fundamental component. The modulator has a characteristic as shown in Fig. 22, is biased at the point of symmetry, and is fed with a sine-wave input signal.

indicator that more harmful third-order intermodulation products may be generated when a composite signal is applied to the input of the demodulator. In the experimental system, third-harmonic distortion for the largest signal was chosen to be about −30 dB, with the consequence that the output of the modulator must be kept 2 dB below the maximum level $p_{max} = 1$ of an ideal linear modulator, as can be seen from Fig. 23. For example, for a ratio of the peak-to-peak exposure to the average exposure of $2e = 1/4$, a linear modulator can correct for an MTF roll-off of $20\log e = 18$ dB, while this modulator can only correct for a 16-dB MTF roll-off if the third harmonic is to be less than −30 dB. Fig. 23 also shows that at the maximum level $20\log p = -2$ dB, the gain $g_m(p)$ of the modulator is about 1 dB less than at low signal levels, and, consequently, at high signal levels the input signal must be increased even further. More precisely, for a given e the frequency response of the aperture corrector must be

$$G(f/v) = \frac{1}{\tilde{\sigma}(f/v)g_m[e/\tilde{\sigma}(f/v)]}. \qquad [28]$$

Since $G(f/v)$ is a function of the track velocity v, the required frequency response of the aperture corrector is different at different recording radii when the disc rotates at a constant rotational frequency. The aperture corrector must consequently be adaptive. The design of such an adaptive equalizer is greatly simplified if the only requirement on its phase characteristic is that it is linear. This implies that the beam is symmetrical, i.e., is not skewed along the track.

6.3 A Practical Aperture Corrector for the VideoDisc Signal

Fig. 5 shows that the MTF response of the recording beam is −16.5 dB (15%) below the low-frequency response at a spatial frequency of 2.5 cycles/μm ($\Lambda = 0.4$ μm), which is the highest spatial frequency of concern in the VideoDisc signal as specified in Sect. 2. Thus, according to the discussion in Sec. 6.2, perfect aperture correction can theoretically be achieved for the entire VideoDisc spectrum if the recorded peak-to-peak signal is less than 1/4 the average groove depth ($e = 1/8$).

For the VideoDisc signal, however, only the band occupied by the video carrier and its sidebands needs aperture correction, which furthermore does not have to be perfect. The sound carrier is at such a low frequency, 716 kHz, that it does not need any aperture correction and can, therefore, be added at a constant level to the aperture-corrected video carrier.

A simple aperture corrector was designed with the following objectives: (a) to keep the peak-to-peak exposure of the video carrier at 5.3 MHz constant and independent of the recording radius; (b) to keep the signal

amplitude at 6.3 MHz equal to the signal amplitude at 4.3 MHz and independent of radius; and (c) to keep a linear phase-versus-frequency characteristic.

A simplified block diagram of the aperture corrector is shown in Fig. 24. It consists of a terminated delay line with three nonloading taps—one at the input, one at the center, and one at the termination. The total delay of the line is $1/(2f_0)$ where $f_0 = 5.3$ MHz. The input and output taps are added and weighted with a factor of $a/2$ to form a network with a frequency response

$$\frac{a}{2}\left(e^{+i2\pi f/4f_0} + e^{-i2\pi f/4f_0}\right) = a\cos[(\pi/2)(f/f_0)].$$

The output of this network is subtracted from the output from the center tap to form a phase-linear network with a response

$$H(f) = b[1 - a\cos\{(\pi/2)(f/f_0)\}]. \qquad [29]$$

This transfer characteristic is illustrated in Fig. 25. The parameter b is varied with the radius to keep the peak-to-peak exposure at 5.3 MHz constant, while the parameter a, which controls slope compensation, is varied with the radius to keep the peak-to-peak exposure at 6.3 MHz equal to the peak-to-peak exposure at 4.3 MHz.

Figs. 26a, b, and c show the calculated aperture-corrected frequency response of the system outlined in Fig. 24, assuming the MTF characteristic shown in Fig. 5. The recording radii are 6, 4.5, and 3 inches. Also shown in Fig. 26 are the uncorrected frequency responses. It is seen that at a 3-inch radius the response is flat to within one dB between 3.5 and 7 MHz. At the larger radii the bandwidth is wider.

The VideoDisc geometric signal specifications can all be met. With an average groove depth of 350 nm, the peak-to-peak video carrier at 5.3 MHz is 350/4 = 87.5 nm. The ratio of signal amplitudes at 4.3 and 6.3 MHz is kept at unity by the aperture corrector. The absolute track drop

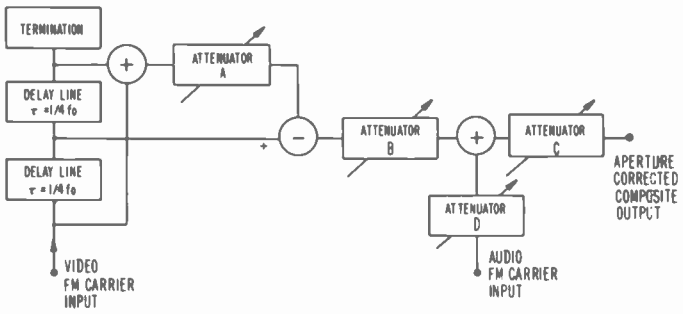


Fig. 24—Block diagram of an adjustable, phase-linear aperture corrector.

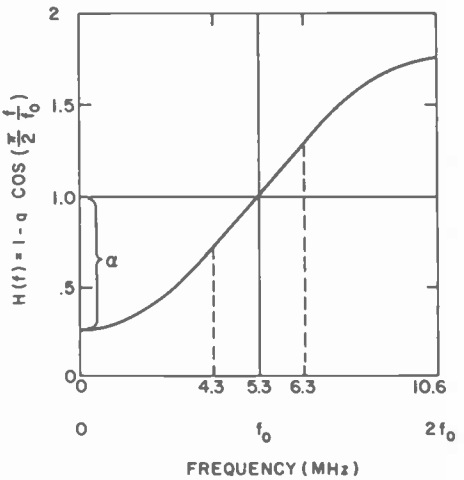


Fig. 25—Amplitude versus frequency characteristic of the aperture corrector shown in Fig. 24.

is zero since the signal and the groove are recorded simultaneously. The differential track drop is also zero, since the signal amplitudes at 4.3 and 6.3 MHz are identical.

The aperture corrector was constructed with broadband operational amplifiers. Care was taken to equalize the delays in the different paths to assure a linear phase characteristic. The actual characteristics of the

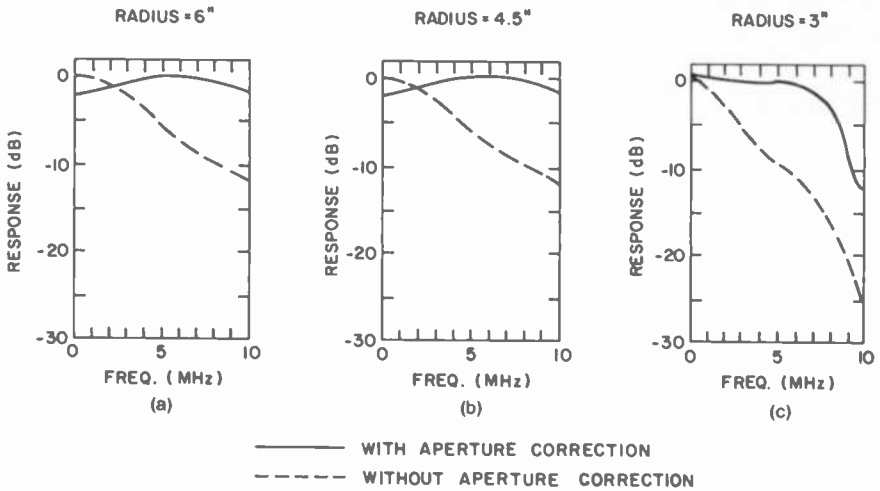


Fig. 26—Calculated frequency response of the linear recording system with and without aperture correction for a beam with an MTF characteristic as shown in Fig. 5.

aperture corrector departed very little from the theoretical characteristics illustrated in Fig. 25. A computer program was developed to determine the attenuator settings as a function of the recording radius, given a measured MTF characteristic and the measured characteristics of the aperture corrector.

6.4 Experimental Results

Figs. 27a and 27b are microphotographs of a groove containing a linearly recorded signal. The groove and the signal were recorded simultaneously. Fig. 27c shows the signal profile along a center cross section of the groove. The signal amplitude is seen to be somewhat less than 100 nm.

Video programs and a variety of test signals were recorded with aperture correction as described in Sec. 6.3. The masters could be read optically[4] and the replicas could be read with capacitive styli. Television pictures were of good quality, but somewhat noisy at the inner radius.

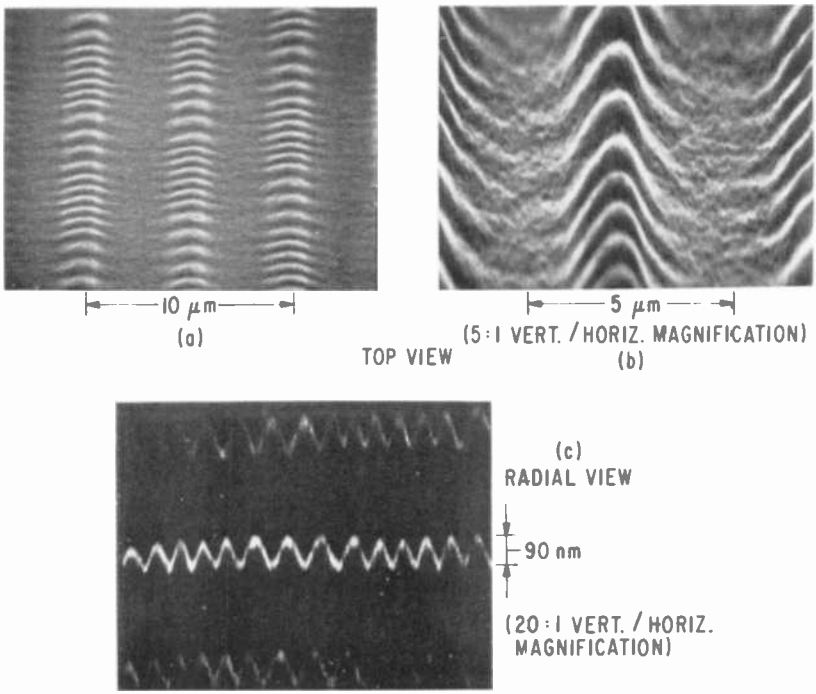
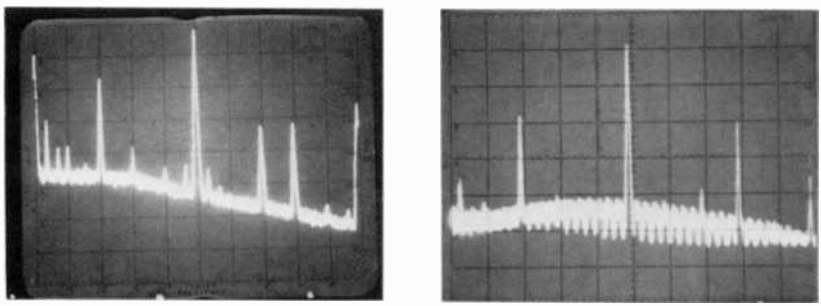


Fig. 27—(a) Microphotograph of a linearly recorded groove plus signal; (b) microphotograph at higher magnification and 5:1 vertical to horizontal distortion; and (c) microphotograph of a cross section along the center of the groove showing the signal.

The quality of the recording is difficult to assess, since all measurements strongly depend on the variable performance of the playback transducer. Figs. 28a and b show the same test signal, a 5-MHz carrier frequency modulated with a deviation of 1 MHz by a 3-MHz tone, read optically from the master and with a capacitive stylus from a replica. The noise bandwidth of the spectrum analyzer is 30 kHz. The capacitive readout shows a carrier-to-noise ratio of about 50 dB, a noise floor that tilts because of the frequency response of the stylus, and components (beats) at 7 MHz and 10 MHz that are caused by the nonlinear properties of the stylus. The optical readout shows a carrier-to-noise ratio of 48 dB and much weaker distortion components. Both readouts show, however, that the FM sidebands at 2 MHz and 8 MHz are equal, indicating that the aperture correction worked well. The upper part of Fig. 29 shows a playback of a recorded frequency sweep from 0 to 10 MHz, and the lower part shows the noise floor when the sweep signal is removed. Fig. 29 thus yields an estimate of the carrier-to-noise ratio as a function of frequency. A summary of the carrier-to-noise ratio as a function of frequency obtained with a capacitive stylus at different radii is shown in Fig. 30. If it is assumed that the noise is all caused by "white" surface roughness noise, then the level of the noise power density should be independent of frequency and inversely proportional to the radius. With this assumption, Fig. 30 suggests that aperture correction has worked quite well, in that at each radius the frequency response is quite flat within the frequency band (2–8 MHz) of concern and that at the center of the band the signal level stays constant within ±1 dB, independent of the radius.

The signal-to-noise ratio of the linearly mastered records did not quite meet the VideoDisc signal playback specifications of 49 dB. A possible explanation is that the removal of photoresist to a depth equal to the



(a) (b)

Fig. 28—Spectrum of a linearly recorded 5-MHz carrier frequency modulated with a 3-MHz tone with a deviation of 1 MHz: (a) playback with capacitive stylus from replica and (b) playback with optical reader from master.

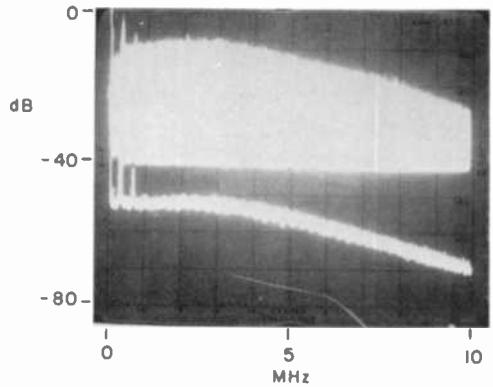


Fig. 29—Spectrum of a recorded frequency sweep signal played back with capacitive stylus. The lower trace shows the noise floor.

groove depth (350 nm) causes significant roughness due to the statistical removal of photoresist particles in the development process. A number of test records designed to explore the noise problem, however, did not confirm this simple explanation and did not produce a basis for drawing any general explanation for the cause of the excess noise.

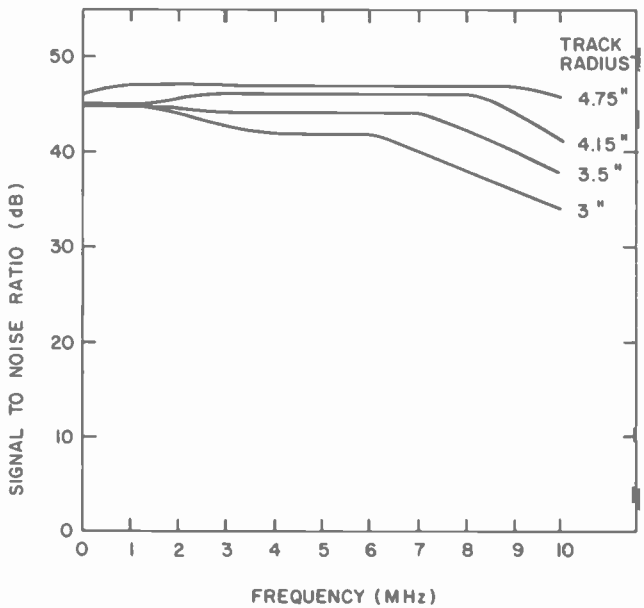


Fig. 30—Signal-to-noise ratio as a function of frequency, with radius as parameter, for linear recording and playback from replica with capacitive stylus.

7. Conclusions

We have discussed three approaches to optical mastering of RCA VideoDiscs; (1) optical recording on pregrooved substrates with thick sagged photoresist; (2) resolution-enhanced recordings on thin photoresist layers; and (3) simultaneous linear recording of both the signal and stylus guiding grooves by an electronically equalized optical recorder. All three methods are capable of producing capacitive-stylus-playable VideoDisc replicas; however some methods are clearly superior in some respects and each has its own particular advantages and disadvantages. Fig. 31 shows spectrum analyzer measurements of stylus playback of a test signal from replicas that were mastered by the three optical methods. The test signal is a 5-MHz carrier frequency modulated with a 1-MHz deviation by a 3-MHz tone.

Fig. 31a shows playback at a radius of 3.4 inches from a thick, sagged, photoresist recording. The carrier-to-noise ratio in a 30-kHz bandwidth is 46 dB, while the 2/8 MHz aperture tilt (adjusted for the stylus system response, as measured by the noise floor) is −10 dB. Both the carrier-to-noise ratio and the aperture tilt values are below the minimum signal playback specifications, although the record passed the geometric specifications. This performance, coupled with the difficulty of even meeting the geometric specifications makes the thick sagged photoresist mastering technique the least desirable with existing technology. Higher quality recording lenses and shorter wavelength lasers may make this technique practicable, but at present it is not.

Fig. 31b shows stylus playback from a thin photoresist mastered recording at approximately 3-inch radius. The carrier-to-noise ratio is 50 dB (30 kHz bandwidth) and the 2/8 MHz adjusted aperture tilt is −2 dB. Fig. 31c (a duplicate of Fig. 28a) shows stylus playback from a linear recording of the signal and groove. The carrier-to-noise ratio is 52 dB (30 kHz bandwidth) while the 2/8 MHz adjusted aperture tilt is about −3 dB. The high carrier-to-noise ratio of this replica is not typical; more often linear recordings are in the range of 46–48 dB. Both the thin-resist and the linear data shown here meet the signal specifications for VideoDisc playback.

Thus the choice of optical recording method is between the thin-photoresist mastering and the simultaneous linear recording of signals and grooves. The problems associated with both are primarily process related. For the thin-photoresist recording, it is necessary to conformally coat the substrate with a thickness tolerance of about 5 nm. Ideally even the intergroove cusps should be coated. If the coating is not conformal, then the exposing beam profile in the radial direction must be configured to compensate for the photoresist thickness profile. This becomes more and more difficult to accomplish as the coating becomes progressively thinner. The processing problem associated with the simultaneous linear

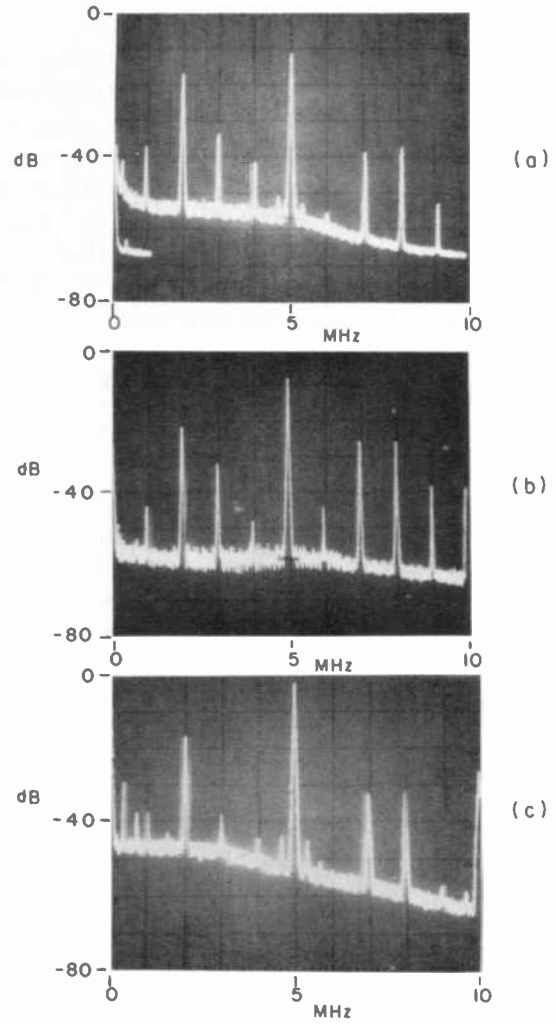


Fig. 31—Stylus playback on a recorded 5-MHz carrier, frequency modulated with a 3-MHz tone, with a deviation of 1 MHz. Optically mastered by (a) recording on thick photoresist on pregrooved substrated, (b) recording on thin photoresist on a cusped substrate, and (c) simultaneous recording of groove and linearly compensated signal on thick photoresist.

recording of signals and grooves into thick photoresist is primarily the control of the photoresist constituents, treatment, and development so that excessive roughness, which limits the playback signal-to-noise ratio, does not occur following the deep developments required to form the groove. An ancillary optical problem is that the radial beam profile de-

termines both the groove shape and the radial signal profile. Thus it is impossible to achieve uniformly deep signals across the entire groove cross section. Nonetheless, linear simultaneous recording of the signal and grooves is the *only* method that is capable in principle of generating a master with a flat, well-controlled frequency characteristic despite the resolution limitations of the optical recording beam. This capability leads us to prefer linear recording for optical mastering of VideoDiscs. Furthermore, in comparison to all known recording methods, linear optical recording offers the most convenient means for making master recordings with flat amplitude and linear phase response. Thus, we believe that simultaneous linear recording of grooves and signals offers the best approach to optically mastering RCA VideoDiscs, provided that photoresist process control can be improved.

Acknowledgements

We acknowledge the cooperation and support of the other members of the VideoDisc team and in particular we thank L. A. Barton for reduction of our SEM data; E. P. Cecelski for his contribution in the development of the linear recording aperture corrector; J. K. Clemens, R. Drake, R. C. Palmer, M. D. Ross, and their co-workers for their assistance with systems and recording signals; R. L. Crane and R. W. Klopfenstein for their help with the computer modeling; M. Kaplan and D. Meyerhofer for their contributions to the understanding of photoresist behavior; D. L. Jose for his evaluation of the signals on replicated recordings; E. Pasierb for thin photoresist coating; J. H. Reisner and co-workers for SEM preparation and analysis; H. G. Scheible and co-workers for photoresist and replica processing; P. V. Valembois for fabrication of master substrates; M. E. Heller, J. H. Martin, and W. J. Mitchell for their assistance in the design and fabrication of the recording system; and L. P. Fox, E. O. Keizer, and D. L. Ross for their advice and continuous support.

References:

[1] J. K. Clemens, "Capacitive Pick-up and the Buried SubCarrier Encoding System for the RCA VideoDisc," *RCA Review*, **39**, No. 1, p. 33 (March 1978).

[2] M. Ross, R. Palmer, and J. Clemens, "The Influence of Carrier-to-Noise Ratio and Stylus Life on the RCA VideoDisc System Parameters," to be published in a forthcoming issue of the RCA Review.

[3] E. O. Keizer, "VideoDisc Mastering" *RCA Review*, **39**, No. 1, p. 60 (March 1960).

[4] A. H. Firester, C. B. Carroll, I. Gorog, M. E. Heller, J. P. Russell, and W. C. Stewart, "Optical Readout of the RCA VideoDisc," *RCA Review*, **39**, No. 3, p. 392, Sept. 1978 (this issue).

[5] This system was originally developed for use in an electron-beam disc recorder. For further details see for example Ref. (3) or "Electron Beam Recording of Masters for the VideoDisc," by J. H. Reisner, G. H. N. Riddle, S. M. Zollers, L. H. Lin, E. D. Simshauser, W. P. Morewood, R. R. Demers, J. Guarracini, *Proc. 8th Intl Conf. on Electron and Ion Beam Science and Technology*, R. Bakish, ed. (Electrochem. Soc., Princeton, N.J., 1978) (to be published).

[6] A. H. Firester, M. E. Heller, and P. Sheng, "Knife-Edge Scanning Measurements of Subwavelength Focused Light Beams," *Applied Optics*, **16**, No. 7, p. 1971 (July 1977).

[7] J. C. Richmond, "Image Quality of Photoelectric Imaging Systems and its Evaluation," *Advances in Electronics and Electron Physics,* **40B**, p. 519.

[8] F. H. Dill, "Optical Lithography," *IEEE Trans. on Electron Devices,* **ED-22**, p. 440, (July 1975) and F. H. Dill, et al, "Characterization of Positive Resist," *IEEE Trans. on Electron Devices,* **ED-22**, p. 445 (July 1975).

[9] I. Gorog, "Optical techniques developed for the RCA VideoDisc," *RCA Review,* **39**, No. 1, p. 162 (March 1978).

[10] W. R. Roach, C. B. Carroll, A. H. Firester, I. Gorog, and R. W. Wagner, "Diffraction Spectrometry for VideoDisc Quality Control," *RCA Review,* **39**, No. 3, p. 472, Sept. 1978 (this issue).

Diffraction Spectrometry for VideoDisc Quality Control

W. R. Roach, C. B. Carroll, A. H. Firester, I. Gorog, and R. W. Wagner

RCA Laboratories, Princeton, N.J. 08540

Abstract—The diffraction of light from the RCA VideoDisc surface has been used to measure groove and signal dimensions with nanometer accuracy and to detect the presence of localized defects comparable in size to the groove width. Theoretical descriptions of the diffraction process, details of measuring and inspection instruments based on this process, and results of measurements on VideoDisc parts are reported.

1. Introduction

The present RCA VideoDisc is a 12-inch-diameter vinyl disc, on each side of which is recorded the signals necessary to provide a one hour audiovisual playback.[1] The video signal is FM encoded and then embossed as a series of pits spaced along a spiral groove having a pitch of approximately 10,000 grooves per inch. To insure high quality playback from the VideoDisc, frequent testing and evaluation of the various parts used in making the VideoDisc (master, replicas, and the disc itself) are required for process development during research, for process monitoring during manufacture, and for quality control of the final product. To develop the pressing process for vinyl discs, measurements of signal and groove depths are especially useful during the initial phases of compound formulation when the optimum pressing parameters are unknown and when the rapid determination of the disc topography is essential. During the manufacturing process it is necessary to make accurate measurements of the dimensions of both the grooves and the signal elements and

to detect the presence of localized defects on the master recording, the intermediate electrochemically replicated metal parts, and the final pressed vinyl discs.

An optical microscope enables a trained observer to make a qualitative estimate of groove pitch uniformity and perhaps even signal depth uniformity on a single disc, but it is very limited in its usefulness for measuring the dimensions of VideoDisc grooves or signal elements. The scanning electron-microscope (SEM) can provide a detailed picture of the microscopic structure, but sample preparation is time consuming, and the estimation of signal and groove depths from electron micrographs is both tedious and requires expert judgement. Also, since all signal elements are not identical, a single SEM estimate can not be sufficiently representative to characterize the disc surface, and many such measurements would be required to provide a statistically significant sample. This procedure is extremely time consuming. From a capacitive or optical playback of the video signal we can infer some characteristics of the signal geometry, but this method is not suitable for obtaining detailed information on groove shape. Capacitive playback can not be used for playing intermediate parts of the negative polarity since the signals then appear on the tops of the grooves. Optical playback can overcome this tracking problem, but only relatively flat parts can be read out. Finally, for certain inspection tasks playback at real time rates is often too slow to be satisfactory.

To fulfil the need to quickly and accurately determine characteristic groove and signal depths and to quickly and completely inspect the surface of the various parts for defects that would result in playback errors on the final disc, we have developed two optical inspection systems based on the diffraction of a weakly focused laser beam reflected from the VideoDisc surface. A brief summary description of these systems was presented in a previously published review paper.[2] In this paper we present a more detailed description of these systems. Sec. 2 describes the microtopography of the VideoDisc surface; Sec. 3 gives a theoretical treatment of the diffraction by that surface of a laser beam whose diameter is large enough to simultaneously illuminate many signal elements and grooves; Sec. 4 describes a diffraction spectrometer that utilizes this diffraction theory to provide quantitative estimates of groove and signal depths; and Sec. 5 describes a laser defect detector that utilizes these diffraction principles to isolate nonperiodic defects on the VideoDisc surface.

2. The VideoDisc Microtopography

The RCA VideoDisc has been mastered by a variety of techniques.[3,4] These include (a) electron beam exposure of the signal elements into an electron sensitive resist layer that has been sag-coated into a pregrooved substrate, (b) optical exposure of signal elements into a similar sag-coated photoresist surface, (c) optical exposure of a thin photoresist layer conformally coated onto a pregrooved substrate, (d) the simultaneous optical exposure of grooves and signal elements into a thick flat photoresist layer, and (e) electromechanical cutting into a smooth copper substrate. The basic groove shape is determined by the mastering process used. It is typically a sinusoid for the sag-coated resist surfaces and a V-groove for the electromechanically cut master. These two cases are illustrated schematically in Fig. 1a and Fig. 1b. Note that the left hand portion of each figure shows the smooth-groove region and the right hand shows an area recorded with video information. In general, the signal trains on adjacent tracks are not synchronized in phase and both frequency and duty cycle differ, although these aspects are not illustrated in Figs. 1a and 1b.

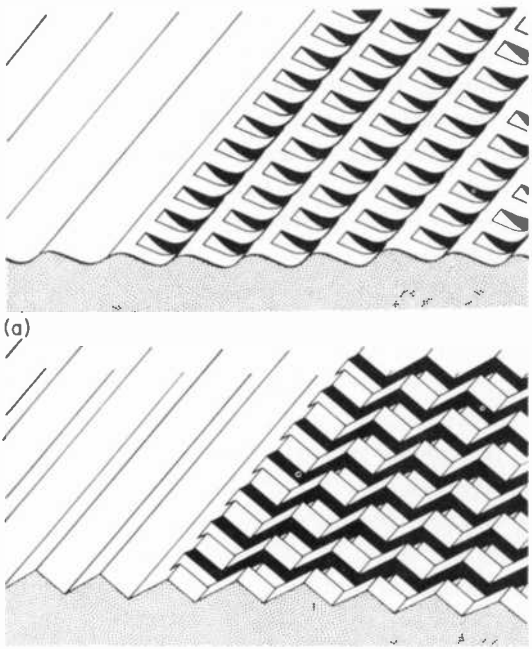


Fig. 1—Surface topography of RCA VideoDisc (vertical scale is expanded): (a) Sinusoidal (sag-coated) groove profile and (b) triangular (electromechanically cut) groove profile.

The most dramatic differences between the two types of discs are immediately evident. The signal elements for the sag-coated discs are typically deepest at the center of the groove and taper to essentially zero depth near the edge of the groove, while the signal elements for the electromechanically cut discs extend across the full width of the groove with uniform depth. The typical signal-element cross section along the track can also be different for electromechanical and for sag-coated discs. The techniques described in this paper are sensitive to changes in groove profile but are not sensitive to subtle differences in signal profile.

The RCA VideoDisc grooves are ~2.7 μm wide and ~0.3–0.5 μm deep. The signal elements are ~0.07 μm deep at the center of the groove. For recording the 4 frames per revolution format, the disc is rotated at a constant angular velocity of 7.5 revolutions per second. The length of the signal elements along the groove depends upon both the instantaneous frequency of the FM encoded video signal and on the radial location of the signal element. The wavelength of the highest recorded frequency at the innermost radius is ~0.5 μm, while the wavelength of the dominant signal frequency, corresponding to the 5 MHz video carrier, lies between ~0.7 μm near the inside and ~1.4 μm at the outer recorded radius.

3. Diffraction Theory

For all VideoDisc parts of interest, the surface that is to be inspected is a reflecting relief surface. The monitoring task is performed by shining a weakly focused laser beam onto this surface and measuring the resultant diffracted-light intensity distribution by suitably placed detectors. One does not observe a direct image of the disc surface but a diffraction pattern that represents a Fourier transform of that surface. Each diffraction direction corresponds to a particular set of spatial frequencies on the disc, and the light flux in a given direction is determined by the corresponding Fourier amplitudes of the surface modulation.

As indicated in Fig. 2, the diffraction pattern produced by normal illumination of the smooth grooved area of an RCA VideoDisc with a He-Ne laser consists of a set of 9 groove orders characteristic of a one-dimensional diffraction grating. When a signal area is similarly illuminated, diffraction by the signal elements occurs in the orthogonal direction. If the signals were all of the same frequency and if all the tracks were in phase, the resultant diffraction pattern would be a two-dimensional array with intensity maxima corresponding to directions of complete constructive interference. The RCA VideoDisc diffraction pattern, shown schematically in Fig. 3, differs from the two-dimensional array characteristic of crossed gratings in one significant respect. Instead

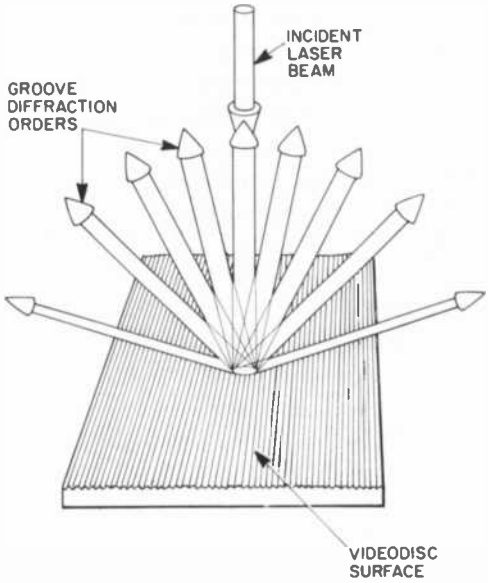


Fig. 2—Diffraction pattern produced by 6328 Å illumination of smooth grooved region of 9541-groove-per-inch RCA VideoDisc.

of discrete orders corresponding to simultaneous signal and groove diffraction, we find that in the signal diffraction direction the orders are well defined but the pattern in the groove diffraction direction is smeared out. (This smearing is not severe in the 0th signal-order groove diffraction pattern.) The discrete signal-diffraction directions can be explained by assuming that the disc is made up of strips of gratings (the signal tracks), and for each strip constructive interference can only occur at the angle determined by the signal frequency. If the diffracted amplitudes from all strips were combined with random phases, partial constructive interference would occur with equal probability at all points along the signal band. On an RCA VideoDisc, the smearing process is more or less complete depending on the number of tracks illuminated and on how completely randomized the signal phases are. The signal diffraction angle can be sharply defined only in the regions where the signal frequencies are relatively constant, as in the case of the illumination of the elements in the vertical retrace interval. When many tracks having varying signal frequencies are simultaneously illuminated, the signal diffraction pattern is spread into broad bands.

In the following analysis, however, the two orthogonal diffraction processes are considered to be operating independently. This means that when the groove shape is being analyzed only the 0th signal order (one-

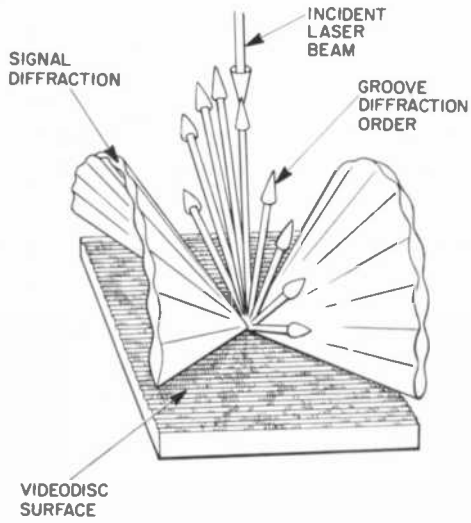


Fig. 3—Diffraction pattern produced by 6328 Å illumination of signalled region of 9541-groove-per-inch RCA VideoDisc.

dimensional discrete groove diffraction array) is considered, and the presence of the signal trains is accounted for solely by its effect on the average profile of the bottom of the groove. When the signal diffraction is being analyzed, although the presence of the grooves causes light to be diffracted in the radial direction, the collection of all the power in the signal band and the comparison of this with the sum of the powers in the discrete groove orders is assumed to be equivalent to monitoring the one-dimensional array of diffracted beams that would result if the signal elements were lined up, in phase, on an ungrooved surface to produce a one-dimensional grating. Experimentally we have found that the independent treatment of groove and signal diffraction processes appears to be satisfactory. Thus we only utilize one-dimensional diffraction theory to derive numerical estimates for both signal and groove depths from experimentally measured spectra.

The scalar theory of diffraction is discussed first in Sec. 3.1. It describes rather well the general structure of the experimentally measured unpolarized diffraction spectrum for both signal elements and grooves, and it can provide an intuitive framework for understanding the diffraction process. The range of validity for the scalar-theory description of signal diffraction and a comparison of this theory with an exact vector treatment of diffraction from square wave elements is discussed in Sec. 3.2.a. We show there that for the RCA VideoDisc signal dimensions the unpolarized diffraction spectrum can be adequately interpreted by scalar diffraction theory. A similar discussion of the limitations of the scalar

theory when applied to groove diffraction and a description of an approximate vector theory for triangular groove diffraction is contained in Sec. 3.2.b. Although the scalar theory is satisfactory for describing the diffraction from shallow sinusoidal grooves, we show that the more rigorous vector theory of diffraction is necessary to accurately describe the unpolarized diffraction spectrum from the relatively deep triangular grooves.

3.1 Scalar Theory

The scalar theory of diffraction[5] is suitable for describing the diffraction of light by shallow long-wavelength gratings. One assumes that an infinite plane wavefront normally incident on a diffraction grating lying in the x-z plane can be characterized by a scalar field, A_i, where

$$A_i = A \exp\{j(-ky - \omega t)\}. \quad [1]$$

A is the amplitude of the optical field, $j = \sqrt{-1}$, k is the light propagation vector ($= 2\pi/\lambda$, where λ is the wavelength of the light), $-y$ is the incident propagation direction, and ω is the light angular frequency. The scalar theory further assumes that the reflection off a relief surface introduces a uniform amplitude attenuation r and a local phase shift $\phi(x)$ due to the differences in optical path traveled by the light wave. At the plane $y = 0$ the wavefront of the reflected scalar field A_r can be characterized by

$$A_r = rA \exp\{j(\phi(x) - \omega t)\}. \quad [2]$$

The infinite wavefront will now propagate away from the surface, and in the far field the net amplitude will consist of the coherent addition of the amplitudes from the complete set of periodic elements of the grating. For the field of an infinite plane wave reflected from the infinite grating this coherent sum is zero unless the contribution from identical regions of adjacent elements are exactly in phase. This occurs when, the relative phase shift $\Delta\phi$, introduced by the differing path lengths traveled by light after reflection from elements separated by the grating period d, is given by

$$\Delta\phi = 2\pi n. \quad [3]$$

Here n is an integer called the order number. This special condition occurs for waves propagating at the angle θ_n with respect to the incident direction defined by

$$n\lambda = d \sin\theta_n. \quad [4]$$

For these special directions, the contribution to the amplitude of the

scalar field is identical from each element of the grating. The resultant scalar field, A_n, traveling in these directions can be found by an integration over a single period of the diffraction grating, and it takes the simple form

$$A_n = \frac{1}{d}\int_0^d rA \exp\{j(\phi(x) - nKx - \omega t)\}dx. \quad [5]$$

Here $K = 2\pi/d$ is the wave vector of the grating and d is the grating wavelength. The term $nKx = 2\pi nx/d$ just accounts for the path difference introduced in that portion of the wavefront emanating from point x as shown in Fig. 4. For convenience we shall define U_n by

$$U_n = \frac{1}{d}\int_0^d \exp\{j(\phi(x) - nKx)\}dx. \quad [6]$$

Eq. [5] can then be rewritten as

$$A_n = rA \exp\{-jwt\}\, U_n. \quad [7]$$

If the power per unit area in the incident wave is denoted by I, the power per unit area propagating in the nth order diffraction wave I_n is

$$I_n = r^2 I\, U_n{}^* U_n. \quad [8]$$

If the beam is of finite diameter and contains power P, the scalar diffraction theory predicts that the power P_n diffracted into the nth order is

$$P_n = r^2 P\, U_n{}^* U_n. \quad [9]$$

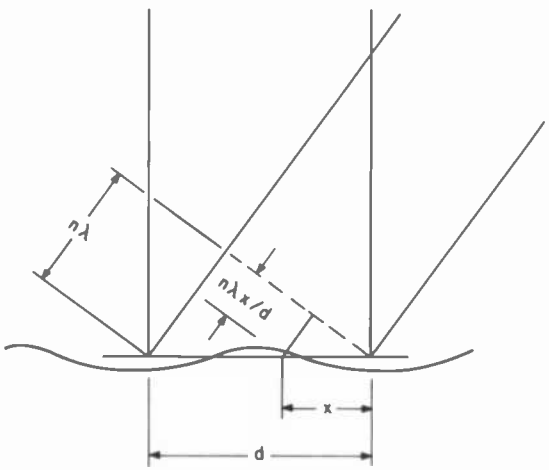


Fig. 4—Construction used to determine local shift in plane-wave phase induced by grating profile.

One check on the validity of any theory of diffraction is that for $P = 1$ and $r = 1$ the sum of the P_n for all real diffracted modes should be precisely unity. By real diffracted modes we mean those modes for which Eq. [4] has a real solution for θ_n. When θ_n is imaginary, the corresponding mode is called evanescent. Evanescent modes must decay exponentially as we move away from the surface, and they cannot carry energy. In general, when the calculated P_n in the evanescent modes is large, the sum of P_n for the real modes calculated from the scalar theory does not equal unity. Thus, when the P_n in the evanescent modes is large, we should be wary of using the scalar theory and resort to more rigorous vector theories of diffraction.

If the beam diameter is made so small that it illuminates only a few grooves, the diffraction peaks are broadened and side lobes are evident. This situation is avoided in practice, because it would also lose the advantage of averaging over an area that is large enough to obtain a statistically significant measure of the surface topography.

3.1.a Scalar Diffraction Spectra

In the course of the RCA VideoDisc development effort, we were interested in sinusoidal, triangular, and square-wave grating profiles. The sag-coated discs have grooves that are very nearly sinusoidal, whereas the electromechanically cut masters have triangular grooves. The signal elements are typically sinusoidal, but near the outside of electron-beam-recorded discs they become almost square wave. The square-wave grating is also of interest since an exact vector diffraction theory is available whose results can be compared with the scalar results. For the three cases of interest, the phase shift introduced by a reflection relief grating of depth Δ is given by

Sinusoidal	$\phi(x) = \alpha \sin(2\pi x/d)$		[10]
Triangular	$\phi(x) = 4\alpha(x/d),$	$0 < x < d/2$	[11a]
	$\phi(x) = 4\alpha(1 - x/d),$	$d/2 < x < d$	[11b]
Square-Wave	$\phi(x) = 0,$	$0 < x < d/2$	[12a]
	$\phi(x) = 2\alpha,$	$d/2 < x < d$	[12b]

where α is one half of the maximum phase shift introduced by the grating and is given by

$$\alpha = 2\pi\Delta/\lambda. \qquad [13]$$

Here Δ is the physical depth of the grating and λ is the wavelength of the laser used. As summarized by Boivin,[6] a direct application of Eq. [6] yields

Sinusoidal	$U_n{}^*U_n = J_n{}^2(\alpha)$		[14]
Trianglar	$U_n{}^*U_n = \sin^2(\alpha)/\alpha^2,$	$n = 0$	[15a]
	$U_n{}^*U_n = \alpha^2 \cos^2(\alpha)/[(n\pi/2)^2 - \alpha^2]^2,$	n odd	[15b]
	$U_n{}^*U_n = \alpha^2 \sin^2(\alpha)/[(n\pi/2)^2 - \alpha^2]^2,$	n even	[15c]
Square-Wave	$U_n{}^*U_n = \cos^2(\alpha),$	$n = 0$	[16a]
	$U_n{}^*U_n = \sin^2(\alpha)/(n\pi/2)^2,$	n odd	[16b]
	$U_n{}^*U_n = 0.$	n even	[16c]

The phase shift introduced by a grating of arbitrary shape can be expressed as

$$\phi(x) = \sum_n \{a_n \cos(2\pi nx/d) + b_n \sin(2\pi nx/d)\}. \quad [17]$$

When the grating is shallow and symmetric, a direct substitution of $\phi(x)$ into Eqs. [6] and [9] yields the following expression for the ratio P_1/P_0 in terms of the Fourier coefficients

$$P_1/P_0 = a_1{}^2/4. \quad [18]$$

For the shallow sine-wave grating this becomes

$$P_1/P_0 = (2\pi\, \Delta_{\text{sine}}/\lambda)^2/4. \quad [19]$$

Therefore, the shallow sine-wave depth Δ_{sine} can be found from the ratio of diffracted power through

$$\Delta_{\text{sine}} = \frac{\lambda}{\pi} \sqrt{P_1/P_0}\,. \quad [20]$$

For the profiles of interest, the peak-to-peak depths Δ_{sq}, Δ_{sine}, and Δ_{tri} corresponding to the fundamental Fourier component of the peak-to-peak depth Δ_{sine}, are

Square-wave	$\Delta_{\text{sq}} = (\pi/4)\Delta_{\text{sine}}$	[21a]
Sinusoidal	$\Delta_{\text{sine}} = \Delta_{\text{sine}}$	[21b]
Triangular	$\Delta_{\text{tri}} = (\pi^2/8)\Delta_{\text{sine}}.$	[21c]

In the shallow-signal limit, the diffraction depends only on the fundamental Fourier coefficient. This is explicitly shown in Table 1, where we give the diffraction spectra calculated from Eq. [8] for the three grating profiles under consideration. Here we assume $r = 1$, $I = 1$, and within each set the depths are chosen so that the fundamental Fourier component is the same for all the gratings. Note that up to $\Delta_{\text{sine}} \sim 1000$ Å the 0th and 1st order powers are nearly identical for all three gratings, and as the depth increases the square-wave grating (which contains large

Table 1—Scalar Diffraction Spectra

	Depth Angstroms	Diffracted Power in Plus (or Minus) Order				
		Zero	One	Two	Three	Four
Square	78.540	.993931	.002460	.0	.000273	.0
Sine	100.000	.995080	.002458	.000002	.000000	.000000
Triangle	123.370	.995009	.002458	.000002	.000030	.000000
Square	157.079	.975872	.009779	.0	.001087	.0
Sine	200.000	.980429	.009761	.000024	.000000	.000000
Triangle	246.740	.980154	.009747	.000037	.000115	.000002
Square	235.619	.946262	.021779	.0	.002420	.0
Sine	300.000	.956370	.021694	.000121	.000000	.000000
Triangle	370.110	.955790	.021620	.000184	.000241	.000011
Square	314.159	.905818	.038170	.0	.004241	.0
Sine	400.000	.923433	.037903	.000379	.000002	.000000
Triangle	493.479	.922495	.037675	.000573	.000387	.000035
Square	392.699	.855523	.058554	.0	.006506	.0
Sine	500.000	.882338	.057914	.000911	.000006	.000000
Triangle	616.849	.881053	.057368	.001375	.000527	.000081
Square	471.238	.796598	.082436	.0	.009160	.0
Sine	600.000	.833979	.081137	.001855	.000019	.000000
Triangle	740.219	.832427	.080038	.002790	.000633	.000160
Square	549.778	.730473	.109235	.0	.012137	.0
Sine	700.000	.779396	.106893	.003363	.000046	.000000
Triangle	863.589	.777732	.104924	.005038	.000683	.000280
Square	628.318	.658754	.138302	.0	.015367	.0
Sine	800.000	.719749	.134429	.005595	.000101	.000001
Triangle	986.959	.718197	.131198	.008344	.000660	.000446
Square	706.858	.583180	.168931	.0	.018770	.0
Sine	900.000	.656287	.162942	.008712	.000200	.000003
Triangle	1110.329	.655129	.157994	.012920	.000562	.000661
Square	785.397	.505588	.200378	.0	.022264	.0
Sine	1000.000	.590316	.191607	.012862	.000367	.000006
Triangle	1233.698	.589873	.184436	.018960	.000402	.000921

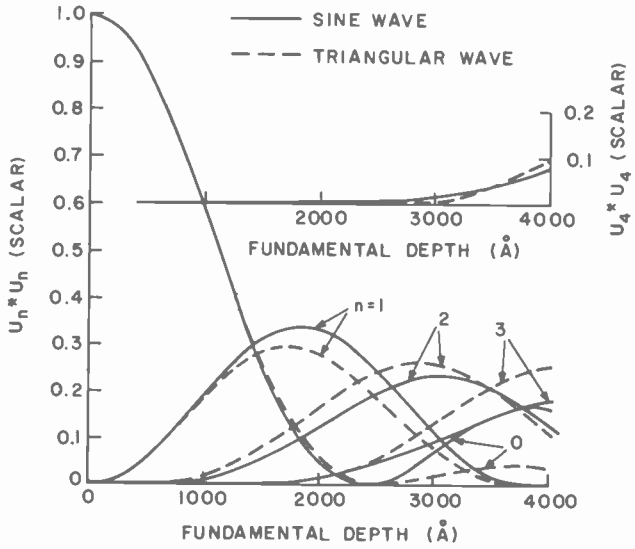


Fig. 5—Scalar diffraction spectrum for sinusoidal and triangular grooves versus the peak-to-peak amplitude of the fundamental Fourier component in the depth profile.

high-frequency Fourier components) diffracts light most strongly into the higher odd orders and the sine-wave grating (which has no higher order Fourier components) diffracts light least strongly into the higher orders.

At larger depths the qualitative behavior of the square-wave grating is very different from the sinusoidal and triangular wave gratings. The square-wave zero order contains all the power when the fundamental depth is an integer number of half wavelengths of light. The other even-integer square-wave orders never contain any power, so that for intermediate depths the light is distributed among the zero and odd orders. The sinusoidal and triangular diffraction spectra are plotted in Fig. 5 versus the peak-to-peak amplitude of the fundamental Fourier component. Note that, in general, for the larger depths the powers in the higher triangular orders are larger than those in the corresponding sine-wave order, which might be predicted from the Fourier decomposition of the groove profiles themselves.

3.2 Vector Theories of Diffraction

Although the scalar theory described above is expected to describe the qualitative behavior of unpolarized laser diffraction by a VideoDisc surface, neither the signal elements (having short wavelengths) nor the grooves (being deep) satisfy the basic geometrical assumptions of this

theory. Thus the quantitative predictions of the scalar theory could not be assumed to be sufficiently accurate to provide the basis for an absolute measurement of grating depth. More rigorous vector theories of diffraction from perfectly conducting gratings were thus used to generate polarized and unpolarized (vector) diffraction spectra. Comparisons of these spectra with the scalar theory spectra and with experimental measurements on finite conductivity grating surfaces were made to establish the ranges of validity for use of the scalar and of an approximate vector theory. These studies are described in the next two sections.

3.2.a Signal Diffraction

The signal depths are typically $\lesssim$1000 Å, so even when orders ± 2 are present their intensities are extremely small. In this shallow signal limit, any measurements of the 0th order and 1st order powers can be fit to an infinite set of signal shapes having the same periodicity and fundamental Fourier amplitude. Thus the measured unpolarized diffraction data provide no internal consistency check on the scalar theory. Furthermore in this range of signal depths, although scanning-electron-microscope measurements are in general agreement with the optically determined signal depths, the accuracy of the SEM measurements is marginal for verifying the optically determined depths.

To estimate the validity of the scalar theory in the case of signal diffraction an extensive theoretical study was made by Ping Sheng.[7] He took as the model problem the vector diffraction from a perfectly conducting square-wave grating and solved Maxwell's equations exactly in both the far-field region and in the square-wave channels, matching boundary conditions everywhere along the surface of the grating. His comparison of the exact vector theory with the scalar solution for the same model problem indicated that when polarized light is used as a probe, for a square-wave depth of 0.125λ (i.e., $\sim$800 Å when using 6328 Å He-Ne lasers), the depth estimate calculated using the scalar theory could be in error by as much as 50% for the shorter wavelength signals encountered on the VideoDisc. However, if unpolarized light is used, the depths predicted by the scalar theory agree closely with those calculated from the exact vector theory for the signal wavelengths of interest. Even near the inside radius, the agreement between the two theories is better than 5%. We expect the same behavior to hold for perfectly conducting sinusoidal signal elements of the same fundamental depth.

The effect of the finite conductivity of the surface is difficult to treat theoretically. Thus the following experiment was performed to observe the effects of surface conductivity on the optically determined depth. Several identical pressings of a disc having sinusoidal signal elements of various wavelengths and depths were made and coated with metals

Table 2—Effective Signal Depths Calculated Using Scalar Theory for Unpolarized Radiation

Signal Wavelength Microns	Effective Signal Depth (Å) for Silver	Inconel	Bismuth
.69	529	514	508
.71	429	408	414
.71	801	780	775
.72	252	239	242
.73	668	652	647
.75	395	379	387
.99	876	855	852
1.01	452	444	441
1.18	1085	1072	1068
1.20	538	547	545
1.56	983	974	963
1.59	474	480	472

of differing conductivity. The particular pressings were made from a sag-coated (sinusoidal groove profile) master for which the signal elements tapered from the center to the edge of the grooves.

Using E, H, and unpolarized laser radiation, the zero-signal-order (groove diffracted) power, P_0, and the signal diffracted power, P_1, were measured, and effective signal depths were calculated for the signal elements from Eq. [20]. Here by $E(H)$ polarization we mean that the $E(H)$ vector is parallel to the grating direction; i.e., it is parallel to the radius in the case of signal elements.* The effective signal depths calculated from the experimental data for the E and H polarization using the scalar theory disagreed with each other in a manner predicted by the exact theory but to an extent that was strongly dependent on the conductivity of the grating surface. For the case of experiments using unpolarized radiation, the effective signal depths calculated using the scalar theory are shown in Table 2. From this table we see that the surface conductivity has remarkably little effect on these optically determined depths. Thus, for the shallow signals encountered on the VideoDisc, unpolarized measurements of the ratio of P_1/P_0 can be interpreted using the scalar theory to yield a unique measure of the effective signal depth that is relatively insensitive to surface conductivity.

3.2.b Groove Diffraction

The grooves on the present RCA VideoDisc have a periodicity $d = 2.667$ μm and, therefore, using a laser wavelength $\lambda = 0.6328$ μm, the maximum integer value for which Eq. [4] yields a real diffraction angle is $n = 4$.

* Since in this paper $E(H)$ always refers to the vector directions measured with respect to the grating under consideration the orientation of the optical electric vector for E polarized groove measurements is parallel to the tangential disc direction.

Thus to monitor the groove diffraction we collected the orders 0, ±1, ±2, ±3, and ±4. The complete measured unpolarized spectrum had the same general behavior as the spectrum predicted by the scalar theory having one adjustable parameter, the groove depth. The grooves were also deep enough to allow relatively accurate, independent measurements of the depth using a scanning electron-microscope. For shallow grooved sag-coated discs the SEM-determined groove depths were in excellent agreement with this optically determined depth. However, although groove depths estimated from the scalar theory for deep triangular grooves were in good agreement with the SEM-determined depths, the unpolarized diffraction spectra for precisely machined V-groove substrates were measured with sufficient accuracy to believe that discrepancies occurring in the fits of the higher orders of the experimental spectra to the scalar theory were significant. A more accurate theory of groove diffraction was thus required for the deep triangular grooves.

An approximate vector solution for the diffraction of light from a triangular grating was obtained by using Rayleigh's method of expanding the diffracted and evanescent modes as a Fourier series of plane and exponentially decaying waves and by matching boundary conditions on the surface of the perfectly conducting grating. The solution presented below was described in detail by Petit.[8] For the E polarization he assumed that the incident plane wave is of the form

$$E^i = \exp\{-jky\}, \tag{22}$$

and that the reflected wave consists of the infinite sum of diffracted waves

$$E^r = \sum_{n=-\infty}^{\infty} B_n \exp\{j(k\sqrt{1 - n^2K^2/k^2}y + nKx)\}. \tag{23}$$

When the square root above is positive, the terms correspond to the real solutions of Eq. [4]. When it is negative the waves are evanescent. The boundary condition that the electric vector vanish on the grating surface $y = f(x)$ requires that the sum of E^i plus E^r be 0 on that surface. Thus

$$-\exp\{-jkf(x)\} = \sum_{n=-\infty}^{\infty} B_n \exp\{j[f(x)\sqrt{k^2 - n^2K^2} + nKx]\}. \tag{25}$$

Both the left hand side and each term on the right-hand side are periodic with period $d = 2\pi/K$, and so they can be expanded in a Fourier series. Thus

$$-\exp\{-jkf(x)\} = \sum_{q=-\infty}^{\infty} D_q \exp\{jqKx\} \tag{26}$$

and

$$\exp\{j[f(x)\sqrt{k^2 - n^2K^2} + nKx]\} = \sum_{q=-\infty}^{\infty} A_{nq} \exp\{jqKx\}. \quad [27]$$

By multiplying both sides of the defining equations by $\exp\{-jqKx\}$ and by integrating over a grating period d the coefficients D_q and A_{nq} can be explicitly calculated. Eq. [25] can thus be rewritten

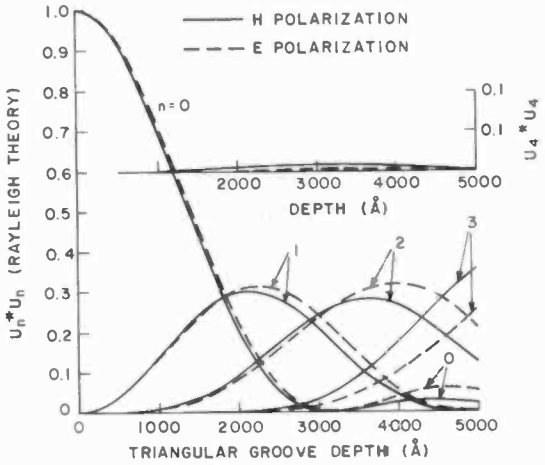
$$\sum_{q=-\infty}^{\infty} D_q \exp\{jqKx\} = \sum_{n=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} B_n A_{np} \exp\{jpKx\}. \quad [28]$$

The only way that this equation can hold everywhere is for the coefficients of corresponding terms to be equal; that is,

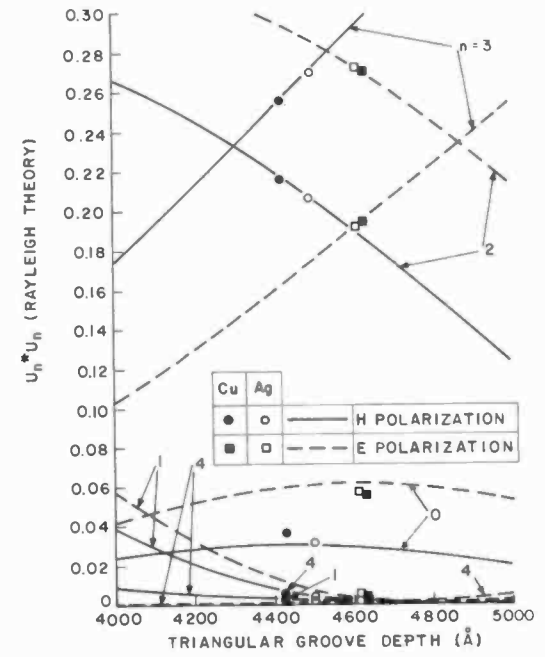
$$D_q = \sum_{n=-\infty}^{\infty} A_{nq} B_n, \qquad q = -\infty \ldots -1, 0, +1, \ldots \infty. \quad [29]$$

Since the D_q and the A_{nq} can be calculated explicitly from the grating profile, the infinite set of Eqs. [29] defines the complex expansion coefficients B_n for the diffracted wave intensities in terms of known quantities. In general, when calculating the expansion coefficients for the real modes, it is sufficient to consider a finite set of equations, each having the same finite number of terms; that number equals the number of real modes plus several evanescent modes. A similar procedure, utilizing the appropriate boundary condition for H polarized reflection, is followed to determine that finite set of equations defining the expansion coefficients for the H polarized modes.

A computer subroutine program that was provided by Ping Sheng was used to solve the finite set of equations for the complex expansion coefficients for each of the two polarizations. We calculated the coefficients for a triangular groove profile and included the real modes plus a varying number, n_e, of evanescent modes. The power in finite diameter diffracted beams was obtained by multiplying the infinite plane wave intensity $B_n{}^*B_n$ by the appropriate[8] area factor $\cos\theta_n$. We found that, for small n_e and a given triangular depth, the total diffracted power oscillated slightly about the exact result of unity as we changed n_e. We further found that a solution that more accurately resulted in a total diffracted power of unity could be obtained by averaging the diffracted power in each order for adjacent values of n_e. For our results, presented below, the spectra are always an average of the case $n_e = 2$ and $n_e = 3$. The diffraction spectra were calculated in this manner for both E and H polarizations for a range of groove depths and these are shown in Figs. 6a and 6b. By averaging intensities for the two polarizations, we obtained the unpolarized spectra shown as Figs. 7a and 7b. The dashed curve in Fig. 7a is the scalar theory, which is shown for reference. Notice that the



(a)



(b)

Fig. 6—Polarized diffraction spectrum calculated from Rayleigh theory versus the depth of triangular grooves: (a) extended range of depth and (b) experimentally determined spectrum. Experimental points always correspond to the nearest theoretical curves except for the *H* polarized 1st and 4th orders for Cu which are labeled specifically.

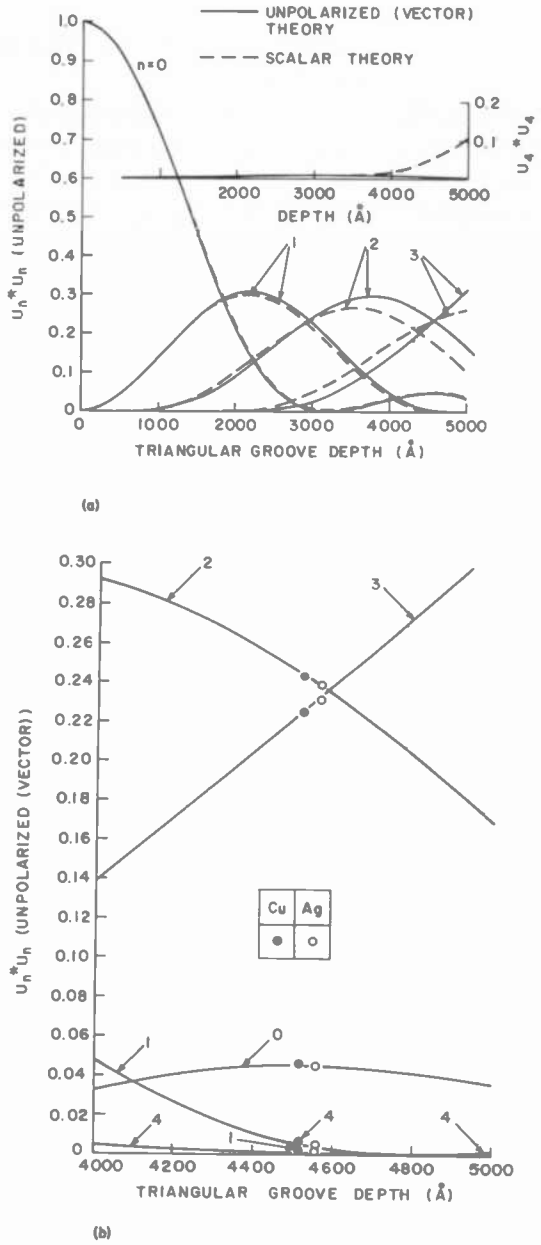


Fig. 7—Unpolarized (vector) diffraction spectrum versus depth of triangular groove: (a) extended range of depth with the scalar theory shown also for reference and (b) experimentally determined spectrum. Experimental points always correspond to the nearest theoretical curve except for the 1st and 4th orders for Cu which are labeled specifically.

0th and 1st orders are in good agreement over the entire range shown; while for the 2nd and 3rd orders, there is marked numerical disagreement between the two theories at the larger depths.

As with the signal diffraction, the effect of surface conductivity on the groove diffraction spectrum was investigated experimentally. A thin silver layer was evaporated onto a portion of a precisely machined V-groove copper substrate, and measurements of polarized spectra for both the copper and the silver grating were made. The experimental data were first normalized such that the sum of all diffraction orders for each polarization equaled unity, and then each set was plotted in Fig. 6b at a depth (that is chosen independently for each of the two polarizations) that gives the best fit to the theoretical spectrum. The procedure for establishing the best-fit groove depth is explained in Sec. 4.3. a. Note that the two measured spectra plotted for the higher conductivity silver grating are in better agreement with the theory than are the copper spectra and that the depths for E and H polarization for the silver grating are in closer agreement with each other than the depths determined by E and H diffraction off of the copper. Moreover, note that if the unpolarized spectra are plotted against the unpolarized (vector) theory in Fig. 7b, there is remarkable agreement between the depths determined for the silver and the copper gratings.

Thus for the triangular depths encountered in the VideoDisc, unpolarized measurements of the diffraction spectrum can be fit to the unpolarized (vector) theory to obtain a depth that is relatively insensitive to surface conductivity and that is believed to be accurate within ~100 Å out of a 4500 Å depth.

4. Diffraction Spectrometer

For observing the detailed structure of groove and signal diffraction spectra of the RCA VideoDisc, it is convenient to utilize a single scanned detector, which is described in Sec. 4.1. From the study of such scanned spectra we can determine groove pitch uniformity and can infer the signal frequency content. For measurements of signal and groove diffraction powers, a fixed detector array that can simultaneously collect all the light from each diffraction order is more suitable. This system is described in detail in Sec. 4.2. Methods of efficiently generating depth estimates from the measured order powers are described in Sec. 4.3. Some results of depth measurements on production parts and a brief description of a diffraction spectrometer that has been built for pilot production use are presented in Sec. 4.4.

4.1 Scanned Diffraction Spectra

Measurements of the diffraction spectra of RCA VideoDiscs have been made by illuminating a horizontally mounted disc with a ~0.5 mm diameter laser beam at near normal incidence. A single small slit-shaped solar cell detector with its active face horizontal and its long axis parallel to the grooves was scanned in the radial plane to measure the energy distribution in the groove diffraction direction. Typical diffraction patterns measured in this manner for sinusoidal and V-groove profiles are shown, respectively, in Figs. 8a and 8b. The widths of the peaks shown are determined by the laser diameter and the detector width. The absence of appreciable spreading indicates a constant groove pitch over the diameter of the laser beam. The slight differences between plus and minus orders are caused by slight groove profile asymmetries. Groove depths have been determined using a scanned detector by increasing the size of the detector so that it could simultaneously collect all of the light from a single diffraction order but not from adjacent orders. When such a large collector was scanned, a peak height in the spectrum was a measure of the total power P_n in a diffraction order n. A horizontally mounted slit-shaped detector having its long axis perpendicular to the grooves has also been scanned in the tangential plane to obtain the signal diffraction spectra shown in Figs. 9a and 9b. Data for both figures were taken for signal elements located at the same radius but at two slightly different angles on the disc surface. Both regions contain vertical retrace signals, but one is predominantly 5.0 MHz and the other predominantly 4.3 MHz.[9]

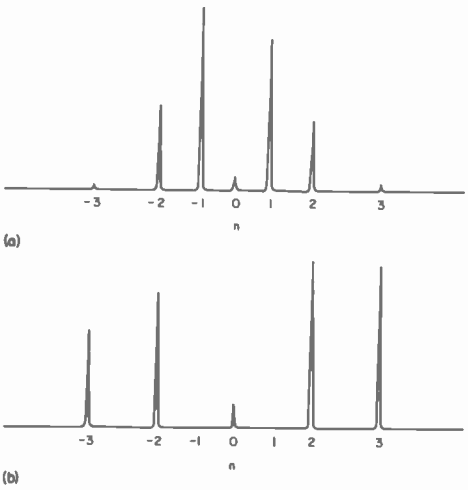


Fig. 8—Measured groove diffraction spectrum in smooth grooved region for (a) sag-coated profile pressed disc, and (b) machined V-groove substrate.

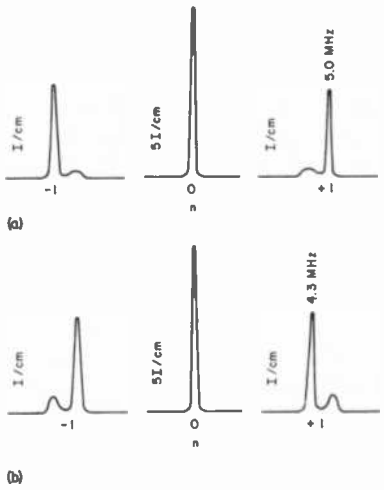


Fig. 9—Measured signal diffraction spectrum in vertical synchronization region of RCA VideoDisc with (a) frequency predominantly 5 MHz and (b) frequency predominantly 4.3 MHz. (Note the change in scale between the 0th order diffraction and the signal diffraction.)

Since the diffraction spectra of a symmetric original part and its exact replica are identical, all generations of VideoDisc parts can be measured using the same instrument. Because the diffraction spectrum of a uniform grating is unaffected by a translation of the grating, measurements of the spectrum can be made while the disc is moving. A convenient method of characterizing the disc surface topography at a particular radius consists of spinning the disc on a turntable and measuring the diffraction spectrum averaged over the annular ring illuminated by the stationary beam. From this averaged spectrum we can calculate groove and signal depths that are characteristic of the disc surface at that radius, although they are not average depths in the usual sense. The depths calculated from the spectra averaged over the relatively large area ring are not sensitive to occasional surface defects. It is also relatively easy to precisely locate the measuring beam at a particular radius on the disc and thereby accurately reproduce measurement conditions, whereas to locate single spot by both radius and angle would require considerably more care and would not be well suited for routine measurements. The ratio of two signal (or two groove) depths determined in this manner for VideoDisc parts of successive generations has been shown to be a reliable estimate of replication efficiency, and it was found to be a useful process control indicator.

4.2 Fixed Detector Arrays

A convenient arrangement for the rapid collection of diffraction data is to position an array of detectors so that each detector intercepts an entire diffraction order. The array must be designed for the laser wavelength used and for the groove pitch of the discs to be measured.

Our current standard system is shown in Fig. 10a. It employs a He-Ne laser incident on the disc at ~2° from the normal (to avoid blocking it by the zero-order detector). The groove detector array consists of 9 solar cell detectors. The detector circuitry is shown in Fig. 10b. The operational amplifier allows the solar cell to operate into a short circuit load while providing an output voltage that is proportional to the light power incident on the active area of the detector over at least 4 decades. The position of the groove detector array is not critical as long as each diffraction order falls entirely on the correct detector. For 1 cm^2 detectors this means that the vertical position of this array relative to the disc

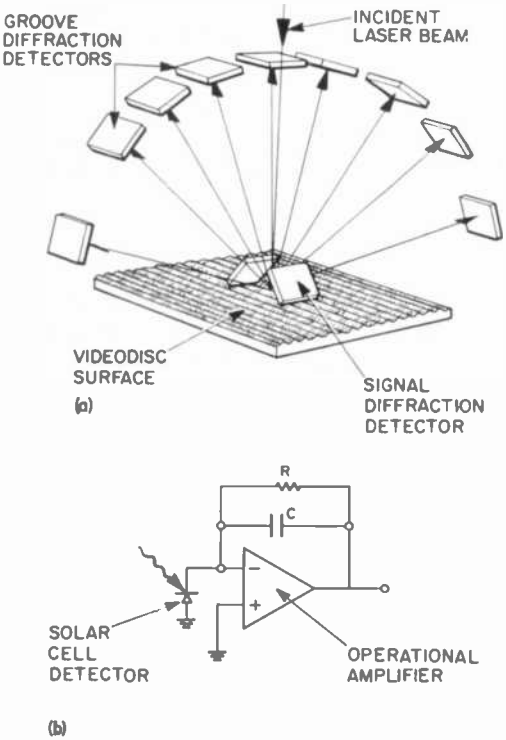


Fig. 10—(a) Fixed detector array used for measuring diffracted power in groove and signal orders and (b) operational amplifier circuit used to convert photocurrent to voltage. *R* is adjusted to provide convenient output level; *C* is chosen to provide proper integration time.

surface needs only to be maintained within a few millimeters. Disc flatness is critical to the extent that gross deviations from flatness can cause the diffracted beams to miss the appropriate detectors. Groove pitch nonuniformities also can cause the diffracted beams to miss the proper detectors and lead to erroneous depth estimates.

For the collection of the signal diffraction two symmetrically positioned 1 cm^2 solar cell detectors are mounted at 45° to the disc surface in the position shown in Fig. 10a. They are located ~1.5 mm above the disc surface and just far enough apart (~3 mm) to allow the incident beam and the zero-order (groove diffracted) pattern to pass between them unimpeded. Oriented in this manner the signal detectors can collect light diffracted by the signal elements through an angle of ~10° to ~80°. The total power P_{+s} (or P_{-s}) collected by one of these detectors is a measure of the sum of all the positive (or negative) signal diffraction orders, although typically for the shallow signals encountered, only the signal powers P_1 and P_{-1} are significant. As we have noted before, the signal diffraction angle varies with the local spatial frequency of the signal elements. For normal incidence the 1st order signal diffraction angle is given by

$$\sin\theta_1 = \lambda/\Lambda \qquad [30]$$

where λ is the light wavelength and Λ is the signal wavelength. For signals recorded at a rotational frequency of 7.5 revolutions per second, the signal wavelength in microns is

$$\Lambda = 1.20r/f. \qquad [31]$$

where r is the radius in inches and f is the signal frequency in MHz. Thus our detector collection angle of 10° to 80° will allow us to measure 1st order diffraction from signals recorded between $r = 2.85$ and 5.72 inches and at frequencies between $f = 1.9$ and 5.3 MHz.

4.3 Treatment of Data

Assuming that the diffraction data have been collected using the fixed array, we can proceed to estimate signal and groove depths. Since the signal diffraction consists of 0th and 1st orders, Eq. [20] can be implemented using either a digital computer or a simple analog circuit to provide an estimate of the effective signal depth. The three aspects of the RCA signal format that insure that the effective signal depth is a reasonably accurate estimate of the signal depth in the case of triangular grooves are: (1) the signal elements extend with uniform depth completely across the grooves; (2) the signal elements are generally shallower than ~1000 Å, which means that only the 0th order and 1st order contain

significant power in them; and (3) the frequency spectrum of the recorded signal is sharply peaked near 5 MHz, resulting in diffraction angles that range from 26° near the outside radius of 5.7 inches to ~70° at the inside radius of 2.6 inches. This angular range fits comfortably within the collection range of the signal detectors.

Estimates for the accuracy of the signal-depth measurements were obtained through the following considerations. The assumption that all light collected by the signal detector is 1st order is valid at the inside, and in fact until r is greater than 5.26 inches no second order diffraction occurs for 5 MHz signals. Thereafter, the inclusion of 2nd order diffraction with the 1st order produces an overestimate of the signal depth. This error is an increasing function of both radius and signal depth. At a radius of 5.72 inches for a 1000 Å deep signal, the signal depth determined by Eq. [20] is in error by ~3%. Near the inside radius, the signal element length is comparable to the wavelength of the light used, and the calculations of P. Sheng[7] showed that the scalar theory for 800 Å deep square-wave grooves overestimates the signal depth by ~2.5% for 0.83 μm long signal elements (5 MHz at r = 3.5 inches) and by ~5% for 0.72 μm long elements (5 MHz at r = 3.0 inches). Similar errors can be expected to occur for a scalar theory depth estimate for ~ 1000 Å deep sinusoidal signal elements. There remains one complicating problem that will not be treated here. This occurs when very short wavelength signals are recorded in very deep V-grooves. The resultant signal diffraction spectrum has a large amount of energy concentrated at each end of the signal bands and thus a large fraction of this energy can slip below the bottom edge of the signal detectors so that the calculated signal depth could be significantly lower than the true depth.

Computations of the groove depth have been performed in two ways. In the off-line scheme, the complete measured spectrum was least squares fit to one of the theoretical spectra already discussed. This required a digital computer. The on-line scheme involved a direct computation of the depth from selected ratios of measured diffraction orders. This calculation was easily implemented with analog computation elements.

4.3.a Least Squares Fitting

The method described below is not limited to off-line computation. However, because it requires a large and/or fast computer, the only way to date that we have implemented this groove-measurement technique is to collect diffraction data from a limited number of regions on the disc and then to perform the calculation using an off-line computer. This computer is used to first normalize the measured groove spectrum by multiplying it by a single scale factor C_M such that

$$C_M \sum_{n=-4}^{4} P_n = 1. \qquad [32]$$

The best-fit groove depth is found by the computer by searching for a depth that minimizes the RMS error, $\epsilon(\Delta)$, where $\epsilon(\Delta)$ is defined by

$$\epsilon(\Delta) = \sqrt{\sum_{n=-4}^{4} (U_n{}^*U_n - C_M P_n)^2}\,. \qquad [33]$$

The best-fit groove depth obtained in this manner depends on the particular theory chosen to calculate the order intensities $U_n{}^*U_n$. The value of $\epsilon(\Delta)$ corresponding to the best fit groove depth, Δ, indicates how well the measured groove shape can be characterized by the assumed theoretical profile.

4.3.b On-Line Calculations

The spectrometer described above utilized a digital computer for fitting the measured diffraction spectrum to the diffraction spectrum of a model surface. The data collection was done point by point and was entered into a computer. The minimization program calculated the groove depth and an error estimate. An alternative scheme that gives rapid estimates of the groove depth as the result of an analog calculation is based on (a) *a priori* knowledge of the grating shape, (b) a knowledge of the approximate grating depth, and (c) the existence of simple analytic relationships between the theoretical power distribution among the various diffraction orders.

In the case of sinusoidal grooves of depth Δ, the ideal scalar diffraction spectrum is characterized by Bessel functions. A recursion relation exists between each three successive orders that can be solved for the argument α. This relationship is

$$\alpha = 2nJ_n(\alpha)/(J_{n-1}(\alpha) + J_{n+1}(\alpha)) \qquad [34]$$

where $\alpha = 2\pi\Delta/\lambda$. From the measured order powers P_n, P_{n-1}, and P_{n+1}, we can derive an estimate Δ_n for the groove depth

$$\Delta_n = \pm \frac{n\lambda}{\pi} \sqrt{P_n}/(\pm\sqrt{P_{n-1}} \pm \sqrt{P_{n+1}}). \qquad [35]$$

For shallow sinusoidal signals, all signs are positive, and when $n = 1$ Eq. [35] reduces to Eq. [20]. For deeper gratings, the $\pm$ signs appear because the $J_n(\alpha)$, $J_{n-1}(\alpha)$, and $J_{n+1}(\alpha)$ are oscillating functions of α, and at the outset we do not know their relative phases. However, the signs can be established from either an *a priori* knowledge of the approximate element depth or by consistency checks concerning (a) the assumed sign and the results of the calculation and (b) a comparison of 2 or more depth estimates. The signs in the expression for Δ_2 and Δ_3 are all positive for groove depths less than 3800 Å, and this covers all cases encountered to date for 9541-grooves/inch, one-hour-per-side discs.[9] Eq. [35] can be

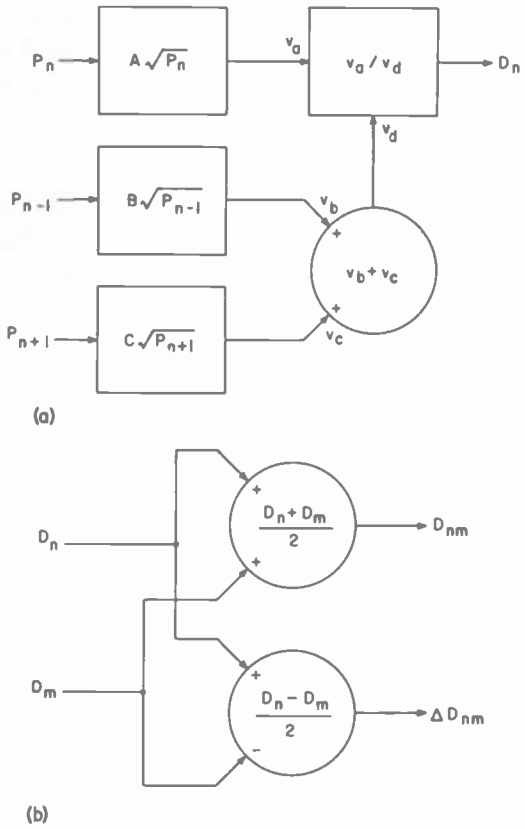


Fig. 11—Analog computer circuit for calculation of (a) scalar sine wave depth and (b) average depth and error estimate.

evaluated using the analog system shown in Fig. 11a. A groove depth estimate and an error estimate are then obtained by taking half of the sum and half of the difference, respectively, of Eq. [35] for two values of n, as shown in Fig. 11b.

For the case of the triangular groove, we notice from the formulas given as Eqs. [15a] and [15b] that by taking the ratio of two successive even orders or of two successive odd orders we can similarly express the grating depth explicitly in terms of the ratio of the power in the two orders

$$\Delta_{02} = (\lambda/2)\sqrt{1/(1 \pm \sqrt{P_0/P_2})}, \qquad [36]$$

and

$$\Delta_{13} = (\lambda/2)\sqrt{\frac{9 \pm \sqrt{P_1/P_3}}{1 \pm \sqrt{P_1/P_3}}}. \qquad [37]$$

Here again the ambiguity as to signs can only be resolved by an *a priori* knowledge of the approximate groove depth. For $\lambda = 6328$ Å, when the groove depths are $\lesssim 3164$ Å, all the signs are positive. Both Eq. [36] and Eq. [37] are easily implemented using analog computational elements, and we can calculate the sum and difference of these two expressions as we did in the sinusoidal groove case.

For a groove that has the exact shape of the assumed profile, the two estimates Δ_2 and Δ_3 (or Δ_{02} and Δ_{13}) would be identical and would equal the digital computer fitted depth. In practice the groove is neither exactly sinusoidal nor exactly triangular and the difference between these two estimates is a measure of the deviation of the groove from the ideal shape.

When we consider deep triangular grooves, the scalar approximation is in error and simple analytic forms for the diffraction spectrum do not exist. We have nonetheless found that we can generate an excellent estimate of the groove depth from diffraction measurements with unpolarized light. This is accomplished by locally expressing the groove depth as a power series expansion in terms of the ratio of the two strongest diffraction orders. For measuring a 140° (included angle) V-groove having a nominal depth of 4852 Å, an expression for the depth Δ_{tri} in terms of the ratio P_3/P_2 is of the form

$$\Delta_{tri} = A + B(P_3/P_2) + C[(P_3/P_2) - D]^2 + E[(P_3/P_2) - D]^4. \quad [38]$$

The coefficients A, B, C, D, and E can be obtained by fitting the above power series expression to a theoretical curve generated by numerical solution of the Rayleigh model of diffraction. These coefficients are shown in Fig. 12. The absolute accuracy of this fit is ±15 Å over the range 3200–5200 Å. The evaluation of Δ_{tri} from measured values of P_3/P_2 can then be easily implemented on an analog computer using the circuit also shown in Fig. 12. This general technique is suitable for calculating any depth shallower than that for which multiple reflections occur. Multiple reflections cause a breakdown of the Rayleigh theory.

A laboratory version of the real-time diffraction spectrometer incorporates the analog depth computers, and the disc to be measured is mounted, on a turntable which is on a roller slide. This allows for simultaneous rotation and translation of the sample along a radius underneath the illuminating laser. A motor driven lead screw provides for uniform, slow translational motion. The radial position is sensed by a rack and pinion driven potentiometer. The outputs of both the signal depth and the groove depth analog computers are typically applied to the two Y drives of an X-Y-Y recorder and the radial position signal is applied to the X drive. With the turntable spinning at ~450 rpm, we can

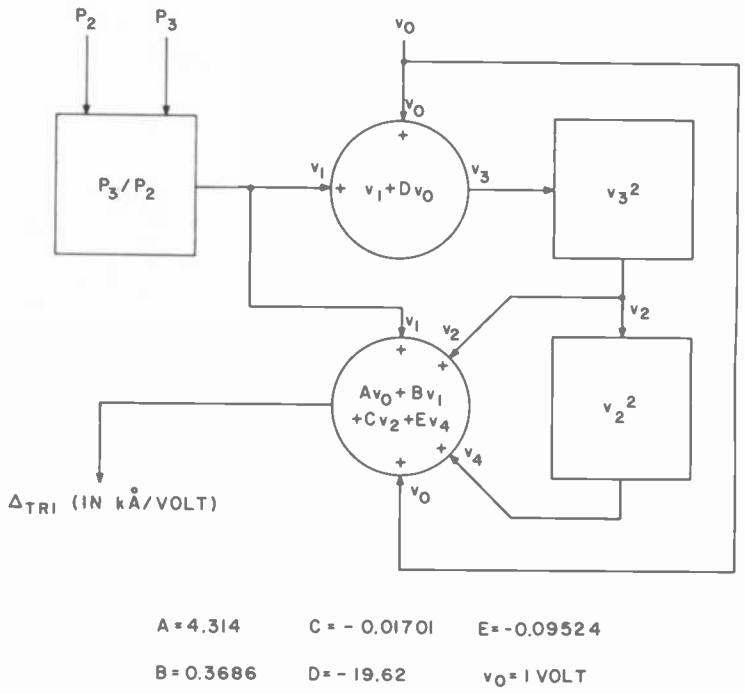


Fig. 12—Analog computer circuit for calculation of triangular groove depth from unpolarized (vector) theory.

obtain a plot of signal depth and groove depth versus radius within ~2 minutes. A typical plot of this type is shown in Fig. 13. By providing a means for slow angular motion of the turntable we can generate a plot of signal depth and groove depth versus angle at a fixed radius, also within a few minutes. A version of the diffraction spectrometer that has been built for the pilot disc-pressing facility is shown in Fig. 14. This system meets all Bureau of Radiological Health safety requirements of a Class I laser system, incorporates a vacuum table for holding metal parts flat, and allows an untrained operator to make plots of signal depth and groove depth versus radius for parts having the 9541 grooves/inch RCA VideoDisc format. After a disc is loaded, the operator initiates the following automatic sequence by pushing one button: (a) transport of the disc into the system, (b) adjustment of the optics and laser power for disc height and surface reflectivity, and (c) scan of the disc to generate a topographic map under preset front-panel-selectable conditions. Thus this system rapidly generates a topographic map of any master, mold, stamper, or pressed vinyl disc of interest.

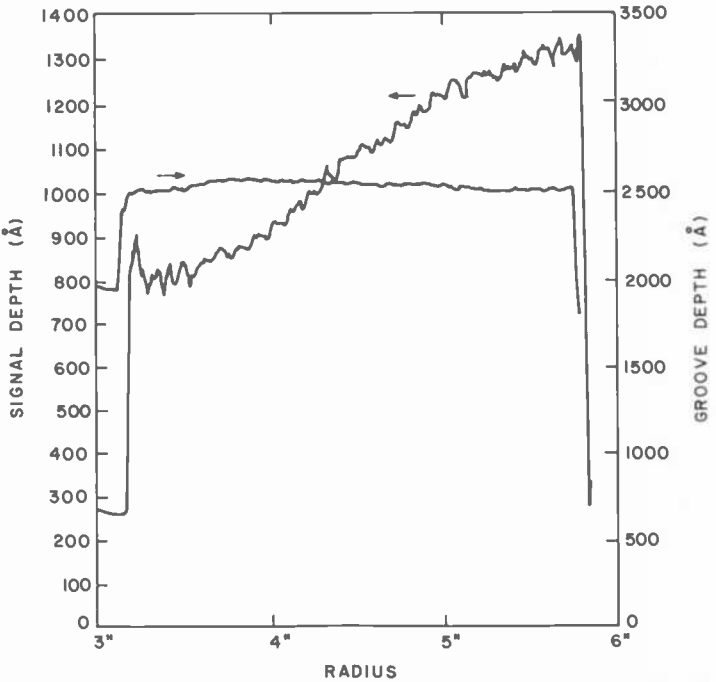


Fig. 13—Plot of effective signal depth and groove depth versus radius for typical electron-beam-recorded (sag-coated groove profile) pressed vinyl disc.

4.4 Results of Depth Measurements on Production Parts

A diffraction spectrometer incorporating the detector array and the off-line computer has been used to measure groove and signal depths on master substrates, successive generations of metal parts (masters, molds, stampers), and on molded vinyl discs. The measured diffraction data was fit to either the sinusoidal or the triangular profile. The reproducability of the best-fit groove-depth estimate was typically better than 1%. The absolute accuracy depends upon how well the actual groove profile could be characterized as sinusoidal or triangular. For the sagged discs, the smooth grooved regions are usually well described by the sinusoidal profile; the presence of signal elements that do not extend the full width of the groove produces an increase in the effective groove depth and increases the 2nd harmonic content. For the mechanically cut triangular grooves, the presence of signal elements does not change the groove profile; thus in this case the groove depth measurement accuracy is independent of whether one measures smooth grooved or signaled areas. The reproducibility of the signal depth on nominally identical triangular grooved discs was ~2%; and, as discussed, above the absolute accuracy was estimated to be ~5%.

Fig. 14—Diffraction spectrometer for use in pilot production facility.

Extensive measurements were performed on both bare metal and metallized sag-coated resist substrates. Bare metal substrates had groove profiles that were either triangular or cusp shaped. Sag-coated resist produced sinusoidal cross sections. The analysis of the triangular and sinusoidal groove data proceeded as discussed above. The cusped grooves (cut with a rounded tool) were analyzed in a different manner. For this case a computer program was written that allowed the fitting of the diffraction data to the scalar spectra of a cylindrical groove shape described by a single parameter, the radius of curvature. The analysis of optical measurements of the groove profiles produced radii of curvature for the tool tips that were in excellent agreement with the SEM determined tool tip radii, typically in the range of 6–7 microns. For the pitch used in cutting the cusp grooves (5555 grooves/inch) the corresponding groove depths were 0.45–0.38 μm which were shallow enough to insure reasonable accuracy from the scalar theory.

By making diffraction measurements of groove depth and signal depth on succeeding generations of replicated parts we could easily check the faithfulness of the replication process. The insensitivity of the diffraction measurement to the polarity of the grooves in the disc allowed us to directly measure all parts using the same instrument, so that small changes in depth could be easily and accurately determined. We have found that the electroformed metal part replication is remarkably faithful. Typi-

cally, the loss in groove depth has been 0.8% in going from one generation of metal parts to the next and it is extremely consistent. The groove depth replication loss for the vinyl pressing process is considerably larger. The vinyl composition, molding pressure, and molding time can all affect the fill of the stamper. To monitor the fill factor (the ratio of groove or signal depth on the pressing to the corresponding depth on the stamper) a diffraction system has been in continuous operation at the pilot pressing facility for 2 years. It is typically used after a metallization step to measure groove and signal depths on representative discs from each press run and to compare these depths with the corresponding data on the stamper. This same instrument has been used also to measure the effects of heat treatment of the vinyl disc on groove and signal profiles. For experimental, sagged-resist, 5555-groove/inch discs the reproducability of measurements has been ±5 Å out of a groove depth of approximately 3500 Å. We have thus been able to measure extremely small changes produced by the processing of discs.

A series of interesting measurements involved the application of the lubricating oil that is used to reduce disc and stylus wear.[10] It was found that the measured groove depth increased upon oiling. This can be explained in the following manner. Light that is reflected off the bottom of the groove and passes through oil of thickness B suffers a phase shift of $2B(n-1)2\pi/\lambda$ in addition to that suffered in reflection from an unoiled groove. Light reflected off of the top of the grooves, where the oil thickness is T, likewise suffers an additional oil-layer-induced phase shift $2T(n-1)2\pi/\lambda$. If the oil is not deposited in a film that uniformly coats the grooves, the difference in these shifts is measured as an apparent change in the groove depth. It was found that for a given groove shape the difference in depth produced by oiling was a universal function of only the average oil thickness and did not depend on the oil composition or on any surface treatment prior to oiling. The empirical relationship between the increase in optical thickness and the absolute oil thickness was established by the simultaneous evaporation of oil onto vinyl discs and silicon wafers. The absolute oil thickness on the silicon wafer, measured using ellipsometry, was assumed to be the same as that deposited on the corresponding area of the disc. The diffraction measurement involving a groove depth determination of a disc before and after oiling is then a fast and accurate method for measurement of oil thickness. The empirical relationship, however, must be redetermined for differing groove profiles.

One additional set of experiments involved the measurement of the signal depth in the smooth groove region. An ideal record illuminated in this region would scatter light only onto the groove detectors. However, when smooth grooved areas on production discs were illuminated, ran-

dom surface roughness having a periodicity in the 0.5–2 μm range scattered light into the signal detectors, and this data was processed to obtain a signal depth. The resultant depth was interpreted as a measure of surface roughness and was used to compare recordings or pressings for surface quality.

5. Laser Defect Detector

The laser defect detector is also based on the detection of the light flux distribution produced when a laser beam simultaneously illuminates many elements and grooves on the VideoDisc surface. However, instead of measuring the peaks in the diffraction spectrum, as was done in the diffraction spectrometer, the laser defect detector is designed to sense the much smaller light component scattered into areas between these high intensity peaks. In this section, we discuss first the concept of Schlieren optics, which forms the basis for the laser defect detector; then, the incorporation of this concept into an easily used inspection system is detailed; and last the results of defect measurements on production parts are described.

5.1 Schlieren Optics

A Schlieren optical system is well suited for detecting the presence of a nonperiodic distortion in the VideoDisc surface topography. Such a distortion causes the wavefront of a diffracted beam emanating from a localized area of the VideoDisc surface (object plane) to differ slightly in phase or in amplitude from the wavefronts diffracted from adjacent periodic elements. When a laser beam that is larger than the periodic elements illuminates an area that contains such a localized defect, the addition of all diffracted amplitudes in the far field (Fourier plane), for directions that usually correspond to nulls in the diffraction pattern, includes a contribution from the defective area that no longer has both the proper phase and amplitude to allow complete cancellation. By monitoring selected null regions in the far field, we are able to sense the presence of defects on the disc surface illuminated by the laser beam. If one wished to localize the defect to an area smaller than the diameter of the illuminating beam, the light from the usually null regions could be focused to form an image of the disc surface. The defect would then appear in the image plane as a bright spot on a dark field. This is a common configuration of a Schlieren system, but the imaging step is unnecessary in our case because on the disc surface we typically focus the laser to a spot that is sufficiently small to provide adequate localization of the defect whose presence is simply "flagged" by an increase

in light intensity in the normally dark background regions of the Fourier plane. The intensity of light scattered into the background regions of the diffraction pattern will depend on (a) the total power in the illuminating beam, (b) the size of the defect relative to the diameter of the illuminating beam, and (c) the precise nature of the defect. When a laser beam illuminates an RCA VideoDisc, the intensity in the far field nulls is generally very low and the presence of small perturbations of the wavefront can be sensed as a large fractional change in the background light intensity. The observation of this increase in background light intensity thus provides the basis for a very sensitive defect detector system.

5.2 Defect Detector System

A defect detector system suitable for process development and control should not only be able to sense the defects as the part is scanned, but it should also provide some means for subsequently inspecting the defects microscopically. A convenient optical system has been devised that allows both the laser system and a high powered inspection microscope to simultaneously access a single small area of the disc surface. This is accomplished by incorporating a laser beam that is incident at 45° to the normal of a disc in a plane containing the radius. The laser beam then produces the diffraction pattern schematically indicated in Fig. 15. Diffraction from an ideal defect-free disc consists of a discrete set of groove diffraction spots along with the signal diffraction bands. A defect scatters light into regions removed from this ideal pattern.

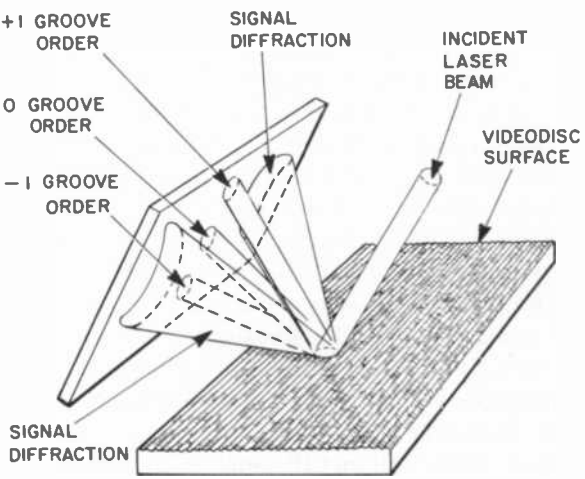


Fig. 15—Diffraction pattern produced by 45° illumination of RCA VideoDisc.

In order to obtain maximum sensitivity, one wishes to detect very small angle scattering, that is light that is very near to the zero-order spot. To accomplish this for an ideal flat record, one would need only to center a large area detector directly in front of the zero order beam, and then use a small diameter stop positioned in the center of this defect detector to interrupt the zero order beam itself. The diameter of the active area of this defect detector should not be so large as to intercept the signal diffraction band. In practice, due to groove pitch nonuniformities, the zero order beam has some spreading in the radial direction, tending to produce a continuous groove diffraction band. In order to reduce the background light collected by the defect detector, the zero order stop must be expanded in the radial direction, thus becoming a strip. In order to generate a map of the defects present on a VideoDisc surface, the disc is mounted on a rapidly spinning turntable that can be slowly translated under the incident beam. This causes the beam to scan the disc surface with an overlapping spiral and hence allows rapid inspection of the entire area of the disc for defects.

The defect detector system also incorporates a second coupled turntable upon which is mounted a disc of electric recording paper. As the VideoDisc is scanned, the presence of defects illuminated by the laser beam produces increases of light at the defect detector which are sensed, amplified, and then recorded as marks on this paper, thus producing a map of the defects. During this mapping process, a low powered microscope is focused on the writing pen. After the map is made, we can reposition the leading edge of the mark under this microscope, thus automatically bringing the corresponding defect back under the laser beam and into the field of view of the high powered microscope. In use, once the map is made, the laser beam is automatically shut off so that the microscopic inspection of the defect can be made under safe and optimum viewing conditions. The system also incorporates a digital readout of both angle and radius that allows us to accurately locate a defect with respect to a polar coordinate system fixed on the disc.

One additional problem involved in measurements of real VideoDisc parts is the result of the warp of the disc surface. This warp typically causes the beam that is reflected off the surface to wander around on the defect detector over distances that are much larger than the width of the zero order stop. Since maximum sensitivity is obtained with the minimum size stop, we cannot afford to increase the width of the stop to allow for this beam wander. Thus a servo controlled beam steering system was developed to stabilize the landing position of the zero order beam on the defect detector. The major features of the resulting optical system are indicated schematically in Fig. 16. Approximately 25% of the zero order beam is reflected by the beam splitter onto the position detector. The

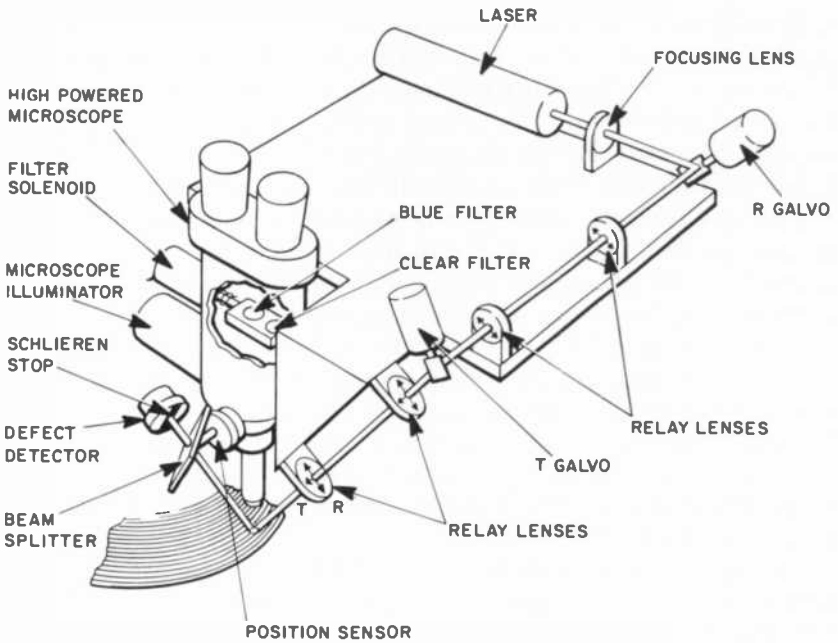


Fig. 16—Principal optical components in laser defect detector.

signals from this detector are amplified and applied to two galvanometers, which are used to move the galvanometer mirrors and thus change the angle of attack of the laser beam on the disc surface in such a manner that a constant landing position of the zero order beam in the center of the position sensor is always maintained. Since the position sensor and the defect detector are in optically equivalent positions, this effectively pins the zero order beam at the center of the defect detector and allows us to use a zero order stop that is designed for maximum static sensitivity.

The defect detector system just described has evolved through successive generations from a laboratory prototype to the system pictured in Fig. 17. This system meets all the Bureau of Radiological Health safety requirements of a Class I laser system. Two nearly identical systems of this type are currently in use in our pilot disc production plant. One monitors the mastering operation, and the second machine is set up in the replication facility where the entire fanout of electroformed metal parts is inspected. These two systems were engineered to allow safe and rapid inspection of VideoDisc parts in production environments. They provide dependable process quality control information for the VideoDisc mastering and replication operations.[11]

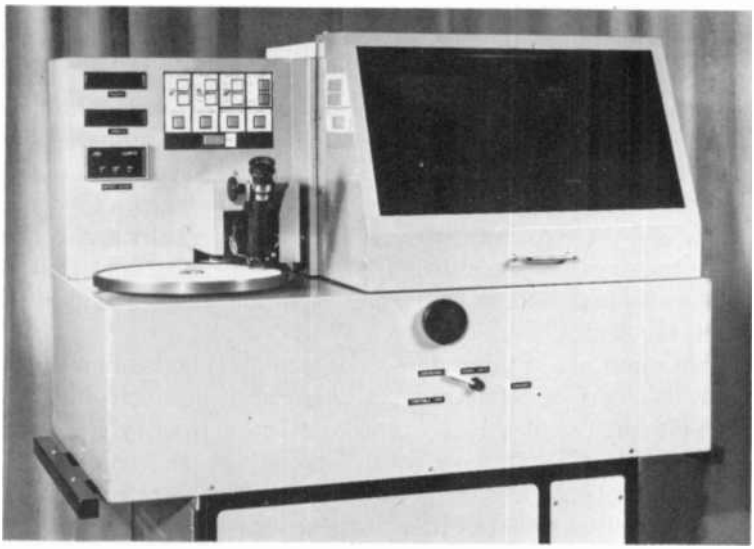


Fig. 17—Laser defect detector in use at pilot production facility.

5.3 Results of Defect Measurements on Production Parts

The result of scanning a VideoDisc part with the laser defect detector system is a map consisting of short tangential lines whose leading ends correspond to areas of increased scattering. Fig. 18 illustrates various map patterns that are occasionally generated when scanning various disc

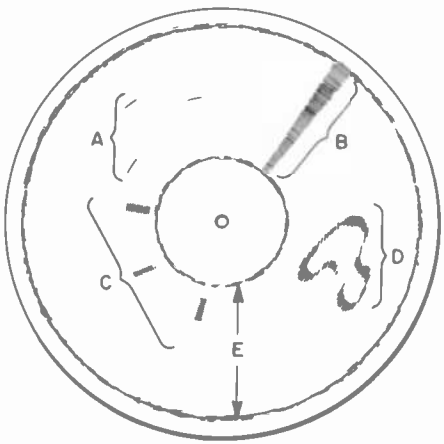
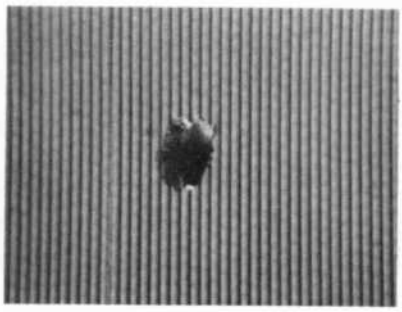


Fig. 18—Schematic representation of a defect map including response of system to point defect (A), radial scratch (B), vertical sync signal (C), surface warp (D), and rough surface outside of recorded region (E).

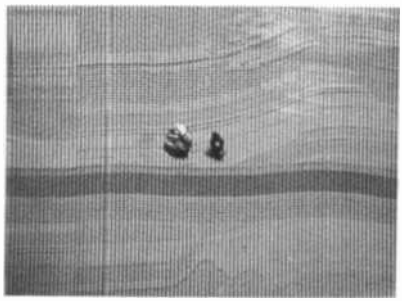
parts. The response of the system to a point defect is a single short tangential line (or a few lines on successive turns of our overlapping spiral scan), and defects of this sort are labeled "A" in Fig. 18. A radial scratch, "B", appears as a narrow wedge whose leading edge locates the scratch itself. If the vertical sync signals are very deep they can cause the pattern indicated by "C". If the surface of the disc is badly warped and the servo gain is insufficient to completely compensate for the warp, a topographic contour is generated as indicated by "D". In general any region that is outside of the precisely machined region of grooves is so rough that there is almost continuous detection of small angle scattering as evidenced by the bands labeled "E".

The minimum size of point defects detected depends upon the laser intensity, disc surface reflectivity, stop size and the electronically adjustable gain and threshold levels of the detector circuitry. For our systems, it can approach one groove width. The highest electronic sensitivity is utilized for inspection of incoming flat copper substrates and the recorded master substrate. In order to reduce the time spent classifying defects after they have been detected, the electronic sensitivity is adjusted for each stage in the process so that only those defects that are likely to adversely affect the final product will be detected. When viewed with a high powered microscope, typical detected defects appear as shown in Figs. 19a and 19b. In Fig. 19a the individual signal elements are also visible. Fig. 19b demonstrates an additional feature of the diffracting RCA VideoDisc surface. By illuminating the surface with a white light beam incident at ~45° angle to the surface in the tangential plane, the light is diffracted upward into the microscope objective by local spatial frequency dependent color dispersion. Thus, different signal frequencies appear as different colors and intensities. This can be seen in this black and white photograph as remarkably contrasting grey levels.

The defect detector system is able to reproducibly scan disc parts for defects larger than a preset adjustable size. By utilizing the digital readout of angle and radius, individual defects can also be followed as they propagate through the replication chain. In the case of electromechanically cut discs, defects on the recorded substrate surface which are very small but detectable by the laser defect detector can result in large defects on the replicated metal master. One mechanism for defect growth occurs when small, subsurface cavities occur on the recorded substrate master. During the separation procedure of the first electroplated metal part replication there is a pullout of the cavity wall. In this case, the defect on the recorded substrate is larger after replication than before, and this also has been observed.[12] The presence of a microscopic subsurface cavity on the substrate can often be sensed by the laser defect



(a)



(b)

Fig. 19—Defects typical of those located by defect detection and subsequently photographed with (a) high magnification and (b) off-axis illumination.

detector even prior to recording, and the elimination of such defects at this early state prevents recordings from being made on "bad" substrates. In the case of resist-coated master substrates the system has been used to check the quality of the resist coating, the uniformity of development process, and the subsequently evaporated gold layer. Once again the early detection of defects saves a tremendous amount of time and effort and prevents the generation of large numbers of useless parts.

As we progress through the replication chain, the number of detectable defects increases. Since for each scan the system can automatically count the number of times that defects have been sensed (large defects cause multiple counts), the total defect count can be used as a measure of surface quality. We can thus quantify any degradation that occurs as we proceed through the chain, can pinpoint those processes responsible for excessive defects, and can tell if changes in those processes result in product improvements.

6. Summary

An analysis of the diffraction spectrum, that is produced when a laser beam illuminates many signal elements and grooves on an RCA VideoDisc surface, provides accurate estimates of signal and groove depths and a sensitive method for detecting the presence of microscopic defects. A complete understanding of the polarization effects in this diffraction spectrum required vector theories for both signal and groove diffraction. These theories assume perfectly conducting grating surfaces. Unpolarized spectra generated from these vector theories were compared with the corresponding square wave and triangular scalar spectra. Theoretical investigations of laser diffraction from ideally shaped perfectly conducting gratings showed that (a) the signal depth determined for shallow ($\lesssim$1000 Å) signal elements using the exact unpolarized (vector) theory and that determined from the corresponding scalar theory are in agreement to within 5%; (b) for deep ($\gtrsim$3000 Å) triangular grooves the scalar theory is not sufficiently accurate to provide reliable depth estimates, so that an unpolarized (vector) theory should be used for calculating these groove depths from the measured spectra. For several finite conductivity RCA VideoDisc surfaces we have shown that an analysis of the unpolarized diffraction data, using the scalar theory for the signal elements and using an approximate unpolarized (vector) theory for the grooves, yields signal and groove depth estimates that are relatively independent of surface conductivity and very reproducible.

We have built diffraction spectrometers that collect the diffraction data and fit the measured spectra to the appropriate theoretical spectra to produce fast and accurate signal and groove depth estimates. Since depth estimates can be made with equal ease and accuracy on both positive and negative parts, all types of RCA VideoDisc parts can be easily compared using the same instrument. One such system has been utilized in a pilot disc-pressing facility for several years to quickly and accurately determine replication efficiency.

We have also built laser defect detectors that are also based on the diffraction of laser light from the RCA VideoDisc surface. Here the presence of a defect is detected by a scattering of light into the region of the Fourier plane that for a perfect disc has zero intensity. Since we are able to detect a small increase in the light intensity above a very small background, this technique is inherently extremely sensitive. Systems based on this concept have been described and results of measurements on production parts have proven this instrument to be a useful and versatile tool for process development and manufacturing quality control.

Acknowledgments

The authors gratefully acknowledge the able assistance of the following members of the RCA staff: J. Y. Avins, S. H. Hagerty, M. E. Heller, M. LaValva, J. H. Martin, R. E. Schneller, J. O. Schroeder, J. P. Walentine, and W. Weiss for assistance in system design and construction; P. V. Valembois for preparation of master substrates; J. H. Reisner for electron-beam recording of test discs; L. J. Levin and F. Tams for evaporative coatings; H. G. Scheible and co-workers for sample preparation; L. Ekstrom and G. Kaganowicz for collaboration in disc oiling experiments; A. E. Bell for ellipsometer oil thickness measurements; L. H. Lin and J. P. Russell for several useful discussions; J. P. Wittke for a critical reading of the manuscript; R. L. Crane for his help in computer program development; P. Sheng for the expert theoretical insight that he provided; and L. P. Fox, E. O. Keizer, and D. L. Ross for their continuous support.

References:

[1] D. S. McCoy, "The RCA 'SelectaVision' VideoDisc System," *RCA Review*, **39**, No. 1, p. 7, March 1978.

[2] I. Gorog, "Optical Techniques Developed for the RCA VideoDisc," *RCA Review*, **39**, No. 1, p. 162, March 1978.

[3] E. O. Keizer, "VideoDisc Mastering," *RCA Review*, **39**, No. 1, p. 60, March 1978.

[4] A. H. Firester, I. Gorog, J. P. Russell, J. J. Gibson, C. B. Carroll and W. R. Roach, "Optical Recording of the RCA VideoDisc," *RCA Review*, **39**, No. 3, p. 427, Sept. 1978 (this issue).

[5] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, New York, N.Y. (1959), pp. 381–386.

[6] L. P. Boivin, "Multiple Imaging Using Various Types of Simple Phase Gratings," *Applied Optics*, **11**, No. 8, p. 1782, Aug. 1972.

[7] P. Sheng, "Theoretical Considerations of Optical Diffraction from RCA VideoDisc Signals," *RCA Review*, **39**, No. 3, p. 512, Sept. 1978 (this issue).

[8] R. Petit, "Contribution à l'Étude de la Diffraction d'une Onde Plane par un Réseau Metallique," *Revue D'Optique*, **42**, No. 6, p. 6, Juin 1963.

[9] J. K. Clemens, "Capacitive Pickup and the Buried Subcarrier Encoding System for the RCA VideoDisc," *RCA Review*, **39**, No. 1, p. 33, March 1978.

[10] D. L. Ross, "Coatings for VideoDiscs," *RCA Review*, **39**, No. 1, p. 133, March 1978.

[11] W. J. Gordon, "VideoDisc Testing Philosophy and Techniques," *RCA Review*, **39**, No. 1, p. 186, March 1978.

[12] P. V. Valembois, private communication.

Theoretical Considerations of Optical Diffraction from RCA VideoDisc Signals

Ping Sheng

RCA Laboratories, Princeton, N. J. 08540

Abstract—Exact model calculations are performed for the diffraction of either a plane wave or a focused spot incident on extremely-short-wavelength signal elements recorded in the RCA VideoDisc. The results not only confirm the experimentally found deviations from the scalar diffraction theory but also suggest general criteria for the optimal design of optical signal readout and quality control systems. It is shown that the H-polarized probing light beam, where the H(E)-polarization denotes the configuration in which the H(E) vector of light is in the radial disc plane, would give higher optical signal as well as better resolution than the E-polarized light when optical readout is achievable by means of a single axial detector. However, for optical playback of the RCA VideoDisc, which uses a differential phase detector, the E-polarization is found to be superior. Furthermore, contrary to the conclusion of the scalar theory, the E-polarization signal for differential phase detector does not vanish at $\lambda/4$ signal element depth, where λ is the wavelength of light. Indeed, the maximum optical signal for the E-polarization occurs at $\lambda/4$ surface relief depth as opposed to $\lambda/8$ predicted by the scalar theory. The cause of this discrepancy is explicitly traced to the breakdown of the scalar theory approximation when either the periodicity of the signal elements or the spot size, or both, are comparable to or smaller than λ. In both the plane wave and the focused spot diffraction calculations, the dependences of the diffraction intensity on signal depth and signal width are examined. Various differences between the E-polarization, the H-polarization, and the scalar theory results are especially pointed out and discussed in relation to the underlying physics.

1. Introduction

In previous papers[1–3] we have seen that a common feature in the optical playback and inspection systems for the RCA VideoDisc is the use of coherent light, in the form of a collimated beam or a focused spot, to probe the disc surface. In each case, it is through diffraction of the incident light that the desired information contained in the recorded surface relief patterns is translated into detectable electro-optical signals. A good theoretical understanding of optical diffraction is therefore important, not only for the correct interpretation of the detected electro-optical signals but also for the optimal design of the optical playback and inspection systems.

The general problem of diffraction theory involves the calculation of the amplitude and the phase of the scattered radiation when an electromagnetic wave is incident on one or more obstacles or apertures on some surfaces. The simplest, and therefore the most widely used, method of calculation is the scalar diffraction theory. The scalar theory ignores the vector nature of light and treats the interaction of light with the scattering object only in an approximate manner. Although satisfactory for diffraction problems where the dimensions of the scattering objects are much larger than the wavelength of light λ, the approximations involved in the scalar theory are known to fail[4] badly when the size or periodicity of the diffracting object becomes comparable to or smaller than λ. In the case of optical diffraction from the RCA VideoDisc signal elements, the HeNe laser has a wavelength of 6328Å. When this wavelength is compared to the repetition distance of the disc signal elements, which range from 0.5 to 1.8 μm, it becomes obvious that the scalar diffraction theory will not be adequate. In fact, experimental results from optical read-out and inspection of RCA VideoDiscs have indicated serious discrepancies with the predictions of the scalar diffraction theory. It therefore seems imperative on both theoretical and the experimental grounds that an exact diffraction calculation be performed.

This paper presents exact model calculations of optical diffraction from RCA VideoDisc signals. Sec. 2 describes the simplified model of VideoDisc signal elements used in the subsequent calculations. Solutions to the plane-wave and the focused-spot diffraction problems are given in Secs. 3 and 4, respectively. In each case the differences between the results of the scalar and the exact theories, and their implications for the optical playback and inspection systems, are pointed out and discussed.

2. Model Description

A rigorous diffraction calculation that takes into account all the realistic details of the signal-element geometry would be impossibly difficult. To facilitate the calculation, we will adopt a simplified model that retains the essential features of the actual diffraction geometry. Fig. 1 gives a schematic picture of the RCA VideoDisc signal elements in a segment of the V-shaped groove. The bottom of the V groove is indicated by the dashed lines. The signal element is narrow in the groove (x) direction and long in the radial (z) direction extending completely across the groove. The cross section of the signal elements has the typical profile shown in the figure.

A plane wave or a focused spot incident on the VideoDisc surface will be diffracted in both the x and the z directions. Experimentally it is found that, to a large degree, the diffractions in the two orthogonal directions are decoupled. That is, if $\mathcal{I}(x,z)$ is the intensity pattern of the diffracted light at some plane $y = y_0$ above the disc surface, then $\mathcal{I}(x,z) \cong \mathcal{I}_x(x)\mathcal{I}_z(z)$, where $\mathcal{I}_x(x)$ is the intensity pattern due to diffraction by signal elements alone, and $\mathcal{I}_z(z)$ is the diffraction pattern caused by the grooves and the relative phase shifts of the signal elements in neighboring grooves. This result suggests that for the purpose of signal diffraction calculation, the geometric structures in the radial direction can be neglected. Therefore, our first simplifying approximation of the model is to regard the signal elements as being recorded in a flat substrate with

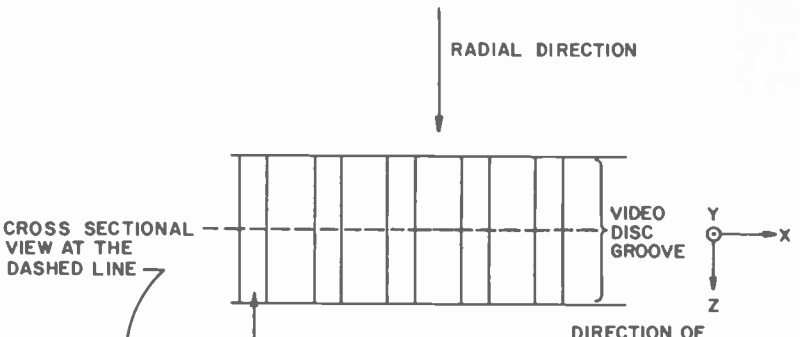


Fig. 1—Schematic representation of RCA VideoDisc signal elements in a section of the groove.

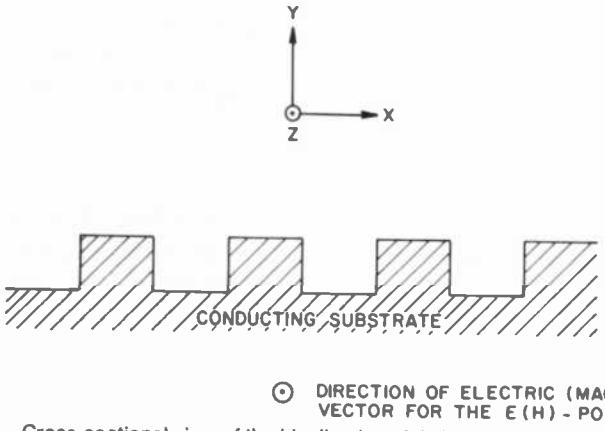


Fig. 2—Cross-sectional view of the idealized model signal elements. For E(H)-polarization of the incident light, the electric (magnetic) vector is perpendicular to the plane of paper.

infinite extent in the z direction. Such idealization in the signal-element geometry allows us to reduce the diffraction calculation to a one-dimensional problem.

The second, perhaps more drastic, approximation of the model is to represent the VideoDisc signal geometry along the x direction by a square-wave surface relief pattern, as shown in Fig. 2, where the disc substrate is assumed to be perfectly conducting. The combination of the two approximations yields an overall picture of the model signal elements as a series of parallel rectangular channels in an otherwise flat, conducting disc substrate. Relative to these signal channels, the incident light can have two distinct polarizations characterized by different orientations of the electric or the magnetic vector. In this paper we will define the E(H)-polarization as the configuration in which the electric (magnetic) vector of the incident light is along the radial direction of the VideoDisc (z-axis).

3. Plane Wave Diffraction

In a previous paper in this issue describing the principle and operation of the diffraction spectrometer,[1] it is seen that plane-wave diffraction by RCA VideoDiscs has been successfully utilized for the measurement of signal depth. However, the data obtained by the diffraction spectrometer also reveal some intriguing features, such as the strong polarization dependence of the diffraction intensity, that are inherently inexplicable by the scalar theory. In this section the problem of plane-wave diffraction is examined by both the scalar approximation and the

exact formulation. It is shown that, while differing significantly in many important respects from the scalar theory predictions, the results of the exact model calculation are consistent with all the experimentally observed diffraction characteristics.

3.1 Scalar Theory

Fig. 3 gives the coordinate system used in the calculation. The surface of the disc is at the plane $y = 0$. The periodicity of the signal elements is denoted by d, the width of the channel by a, and the depth by h. Consider a plane wave of unit amplitude incident on the VideoDisc at an angle θ_i (θ_i is negative as shown in Fig. 3) where $-\pi/2 < \theta_i < \pi/2$. The scalar wave can be written as

$$U_i(x,y) = \exp[-ik(\sin\theta_i x + \cos\theta_i y)], \quad [1]$$

where $k \equiv 2\pi/\lambda$ is the wavevector of light. At the plane $y = 0$, the phase of the scalar wave can be written as a function of x only:

$$V(x) = \exp[-ik \sin\theta_i x]. \quad [2]$$

After interaction with the VideoDisc signals, there will emerge a diffracted scalar wave whose amplitude and phase at the plane $y = 0$ is denoted by $R(x)$. Given $R(x)$, the diffracted light amplitude at any direction θ, $\mathcal{R}(\sin\theta)$, can be calculated by superposition as shown in Fig. 4,

$$\mathcal{R}(\sin\theta) = \lambda^{-1} \int_{-\infty}^{\infty} dx\, R(x) \exp[-ikx \sin\theta], \quad [3]$$

where the normalizing factor λ^{-1} is obtained by requiring that in the case of a flat surface ($R(x) = V(x)$) $\mathcal{R}(\sin\theta)$ should be given by the specular reflection condition $\delta(\sin\theta + \sin\theta_i)$. Here the delta function $\delta(x)$ is defined by

$$\delta(x) = \int_{-\infty}^{\infty} \exp(i2\pi xy)dy. \quad [4]$$

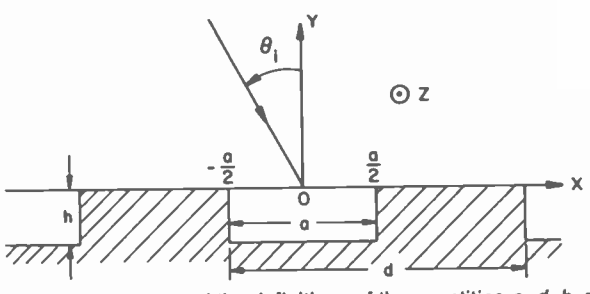


Fig. 3—The coordinate system and the definitions of the quantities a, d, h, and θ_i used in the plane-wave diffraction calculation.

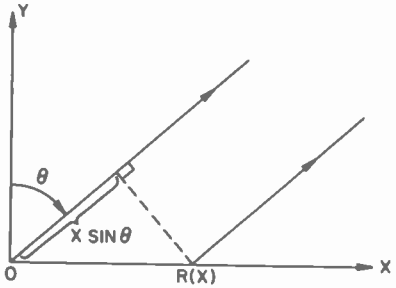


Fig. 4—The (complex) amplitude of light propagating at a given direction θ is obtained by superposing the product of amplitude ($R(x)$) and phase factor ($\exp(-ikx \sin\theta)$) for all values of x.

It is clear that the scalar diffraction calculation is completed if $R(x)$ can be expressed in terms of $V(x)$ and the parameters of the VideoDisc signal elements. The crucial approximation of the scalar theory states that

$$R(x) = V(x) \exp[ikF(x)], \tag{5a}$$

$$F(x) = \begin{cases} 2h; |x - nd| \leq \dfrac{a}{2}, & -\infty < n < \infty \\ 0; \text{otherwise} \end{cases} . \tag{5b}$$

In other words, the net effect of the interaction between the incident light and the signal element is treated as a local phase delay whose magnitude is proportional to the round-trip path length to the bottom of the signal channel. The expression $\exp[ikF(x)]$ can be Fourier-analyzed, giving

$$\exp[ikF(x)] = \sum_{n=\infty}^{\infty} B_n \exp[i2\pi nx/d], \tag{6}$$

$$B_n = \begin{cases} 1 + \dfrac{a}{d}[\exp(i4\pi h/\lambda) - 1]; n = 0 \\ [\exp(i4\pi h/\lambda) - 1][\sin(n\pi a/d)/n\pi]; n \neq 0 \end{cases} \tag{7}$$

Substitution of Eqs. [4] through [7] into Eq. [3] yields

$$\mathcal{R}(\sigma) = \sum_{n=-\infty}^{\infty} \delta(\sigma - \gamma_n) B_n, \tag{8a}$$

$$\gamma_n = \sin\theta_0 + \frac{n\lambda}{d}, \tag{8b}$$

where $\theta_0 \equiv -\theta_i$, and we have replaced $\sin\theta$ of Eq. [3] by the variable σ. Since $\mathcal{R}(\sigma)$ is the amplitude of a particular diffracted wave, the total diffracted wave amplitude at any spatial point can be obtained as[5]

$$U_d(x,y) = \int_{-\infty}^{\infty} d\sigma \mathcal{R}(\sigma) \exp[ik(\sigma x + \sqrt{1-\sigma^2}\, y)]. \qquad [9]$$

Substitution of Eq. [8] into Eq. [9] yields

$$U_d(x,y) = \sum_{n=-\infty}^{\infty} B_n \exp[ik(\gamma_n + \sqrt{1-\gamma_n^2}\, y)]. \qquad [10]$$

Eq. [10] states the expected result that the diffracted wave is composed of different diffraction orders. For the nth diffraction order, the amplitude is given by B_n of Eq. [7] and the angle of propagation is given by $\theta_n = \sin^{-1}\gamma_n$. For $|\gamma_n| > 1$, we have an evanascent wave that attenuates in the y direction as $\exp[-k\sqrt{\gamma_n^2 - 1}\, y]$. It should be noted that the expression for γ_n, Eq. [8b], can be regarded as a statement of conservation of momentum in the x direction and is derived independent of the approximations used to obtain B_n. Therefore, it is expected that the diffracted wave of the exact solution should also be expressible as a superposition of the plane waves $\exp[ik(\gamma_n x + \sqrt{1-\gamma_n^2}\, y)]$.

The normalized energy flux I_n of the nth diffraction order is given by

$$I_n = |B_n|^2 = \begin{cases} 1 - 4\dfrac{a}{d}\left(1 - \dfrac{a}{d}\right)\sin^2 kh; & n = 0 \\ 4\dfrac{\sin^2(na\pi/d)}{n^2\pi^2}\sin^2 kh; & n \neq 0. \end{cases} \qquad [11]$$

It is easily shown that

$$\sum_{n=-\infty}^{\infty} I_n = 1 \qquad [12]$$

as required by conservation of energy. However, this statement is deceptive, because in the summation of Eq. [12] all the evanascent waves have been included, and these in reality can not transport energy. Therefore, if we sum only those $|B_n|^2$ for which $|\gamma_n| < 1$, a net energy deficiency is obtained. This deficiency may be small for $d \gg \lambda$ (so that n is large before $|\gamma_n| > 1$), but for d comparable to λ the discrepancy can be considerable. This defect of the scalar theory points out its obvious inadequacy in the limit of small signal-periodicity.

3.2 Exact Theory

The exact diffraction calculation for a plane wave incident on a metallic square-wave grating has been considered by several previous workers.[6,7] The basic idea is to note that the surface of the square wave relief pattern, i.e., $y = 0$ plane of Fig. 3, naturally divides the space into two half-spaces. In each of the two half-spaces, any arbitrary wave can be expanded in terms of the eigenfunctions of the wave equation for that region. In the free space above $y = 0$, the incident and the diffracted wave can be expressed as superposition of plane waves as shown before. For E-polarization, the z-component of the electric vector E_z can be written as

$$E_z^> = \exp[-ik(\sin\theta_i x + \cos\theta_i y)] + \sum_{n=-\infty}^{\infty} B_n^e \exp[ik(\gamma_n x + \sqrt{1-\gamma_n^2}\, y)]. \quad [13]$$

Here k, θ_i, and γ_n all have the same definitions as before, B_n^e is the (as yet) undetermined amplitude of nth diffraction order for the E-polarization, and $\sqrt{1-\gamma_n^2}$ is understood to take on the value $i\sqrt{\gamma_n^2 - 1}$ when $|\gamma_n| > 1$. For $y < 0$, it is easily verified that

$$F_n^e = \sum_{m=-\infty}^{\infty} \exp[ikmd \sin\theta_0] f_n^e (x - md), \quad [14a]$$

with

$$f_n^e = \begin{cases} \sin\left[\frac{n\pi}{a}\left(x + \frac{a}{2}\right)\right] \sin\left[k\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\,(y+h)\right] \\ \text{for } |x| \le \frac{a}{2},\ -h \le y \le 0; \\ 0, \text{ otherwise} \end{cases} \quad [14b]$$

is a solution of the wave equation

$$\frac{\partial^2 F_n^e}{\partial x^2} + \frac{\partial^2 F_n^e}{\partial y^2} + k^2 F_n^e = 0 \quad [15]$$

and satisfies the tangential E-field boundary conditions of vanishing on the sides ($x = \pm a/2$) and bottom ($y = -h$) of the signal channel. We will call f_n^e's the channel modes of the E-polarization. The presence of $\exp[ikmd \sin\theta_0]$ in Eq. [14a] ensures the conservation of light momentum in the x direction. Since F_n^e's form a complete set of orthogonal functions, it follows that any arbitrary solution of the wave equation which satisfies the same boundary conditions, such as E_z, should be expressible as a linear superposition of F_n^e's. That is,

$$E_z^< = \sum_{n=1}^{\infty} C_n^e F_n^e, \quad [16]$$

where for $n > 2a/\lambda$ it will be understood that $\sin[k\sqrt{1-(n\lambda/2a)^2}\,(y+h)]$ in Eq. [14b] becomes $i\sinh[k\sqrt{(n\lambda/2a)^2-1}\,(y+h)]$. The problem of plane wave diffraction is now centered on the determination of B_n^e's and C_n^e's through boundary conditions fitting at $y = 0$. The condition of tangential component of the electric vector being continuous yields $E_z^> = E_z^<$ at $y = 0$. Since at $y = 0$ both $E_z^>$ and $E_z^<$ can be expressed as the product of $\exp[ikx\sin\theta_0]$ and a periodic function with periodicity d, it suffices to write out the condition for $|x| \le d/2$:

$$\exp(ikx\sin\theta_0) + \sum_{m=-\infty}^{\infty} B_m^e \exp(ik\gamma_m x) =$$

$$\begin{cases} \sum_{n=1}^{\infty} C_n^e \sin\left[kh\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right] \sin\left[\frac{n\pi}{a}\left(x+\frac{a}{2}\right)\right]; & |x| \le \frac{a}{2} \\ 0; & \frac{a}{2} \le |x| \le \frac{d}{2}. \end{cases}$$

[17]

The second boundary condition is the continuity of the tangential component of the magnetic vector H_x. Since from Maxwell's equations $\partial E_z/\partial y = ikH_x$, this condition can be expressed in the interval $|x| \le a/2$ as

$$\cos\theta_0 \exp(ikx\sin\theta_0) - \sum_{m=-\infty}^{\infty} B_m^e \sqrt{1-\gamma_m^2} \exp(ik\gamma_m x)$$

$$= i\sum_{n=1}^{\infty} C_n^e \sqrt{1-\left(\frac{n\lambda}{2a}\right)^2} \cos\left[kh\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right]$$

$$\times \sin\left[\frac{n\pi}{a}\left(x+\frac{a}{2}\right)\right]. \qquad [18]$$

It is important to note that for H_x the condition of continuity is valid only for $|x| \le a/2$. For $a/2 \le |x| \le d/2$ there can be surface currents at the $y = 0$ metallic plane, thereby allowing for a discontinuity of H_x. Multiplying Eq. [17] by $\exp(-ik\gamma_m x)$ and integrating over the interval $|x| \le d/2$ gives

$$B_n^e = -\delta_{n0} + \sum_{m=1}^{\infty} \sin[kh\sqrt{1-(m\lambda/2a)^2}]\, Q_{mn} C_m^e, \qquad [19]$$

$$Q_{mn} = \frac{1}{2\pi}\frac{\lambda}{d}\frac{(m\lambda/2a)}{(m\lambda/2a)^2-\gamma_n^2}[\exp(ik\gamma_n a/2)$$

$$+ (-1)^{m+1}\exp(-ik\gamma_n a/2)]. \qquad [20]$$

Multiplication of Eq. [18] by $\sin[(m\pi/a)(x + (a/2))]$ and integration over $|x| \le a/2$ yields

$$iC_m^e \sqrt{1-\left(\frac{m\lambda}{2a}\right)^2} \cos\left[kh\sqrt{1-\left(\frac{m\lambda}{2a}\right)^2}\right] \frac{a}{2d}$$

$$= \cos\theta_0 Q_{m0}^* - \sum_{n=-\infty}^{\infty} B_n^e \sqrt{1-\gamma_n^2}\, Q_{mn}^*, \qquad [21]$$

where Q_{mn}^* stands for the complex conjugate of Q_{mn}. Eqs. [19] and [21] are two infinite sets of equations that uniquely determine all C_n^e's and B_n^e's.

In the H-polarization case, the channel modes f_n^h are given by

$$f_n^h(x) = \begin{cases} \cos\left\{\frac{n\pi}{a}\left(x+\frac{a}{2}\right)\right\} \cos\left[k\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\,(y+h)\right] \\ \text{for } |x| \le \frac{a}{2},\ -h \le y \le 0; \\ 0, \text{ otherwise} \end{cases} \qquad [22]$$

where $\cos[k\sqrt{1-(n\lambda/2a)^2}\,(y+h)]$ is understood to take the value $\cosh[k\sqrt{(n\lambda/2a)^2-1}\,(y+h)]$ when $n\lambda/2a \ge 1$. The magnetic channel modes, f_n^h's, satisfy the boundary conditions of normal derivatives vanishing on the sides and bottom of the signal channel as required for the z component of the magnetic vector H_z. Given f_n^h, the rest of the formalism is completely analogous to the E-polarization case. The amplitudes of the nth diffraction order and the mth channel mode are denoted by B_n^h and C_m^h, respectively. For completeness we will also present here the equations for their determination:

$$B_n^h = \delta_{n0} + \frac{i}{\sqrt{1-\gamma_n^2}} \sum_{m=0}^{\infty} P_{mn} \sqrt{1-\left(\frac{m\lambda}{2a}\right)^2}$$

$$\times \sin\left[kh\sqrt{1-\left(\frac{m\lambda}{2a}\right)^2}\right] C_m^h, \qquad [23]$$

$$C_m^h \cos\left[kh\sqrt{1-\left(\frac{m\lambda}{2a}\right)^2}\right] \frac{a}{2d}(1+\delta_{m0})$$

$$= P_{m0}^* + \sum_{n=-\infty}^{\infty} P_{mn}^* B_n^h, \qquad [24]$$

$$P_{mn} = \frac{\lambda}{d}\frac{i}{2\pi}\frac{\gamma_n}{(m\lambda/2a)^2-\gamma_n^2}[\exp(ik\gamma_n a/2)$$

$$+ (-1)^{m+1}\exp(-ik\gamma_n a/2)]. \qquad [25]$$

One important difference between the E- and the H-polarizations should be noted. From Eqs. [14] and [22] it is clear that whereas the

lowest-order channel mode for the E-polarization is the $n = 1$ mode, for the H-polarization $n = 0$ is the lowest-order mode. Therefore, when $a < \lambda/2$, all the channel modes for the E-polarization, including the $n = 1$ mode, becomes evanascent (exponentially attenuated). However, for the H-polarization the $n = 0$ mode will never be evanascent no matter how small a becomes. As we will see in the next section, the existence of one nonattenuating channel mode for the H-polarization and none for the E-polarization accounts for qualitatively different diffracting characteristics when $a < \lambda/2$.

Once B_n^e or B_n^h is known, the intensity of the nth diffraction order is calculated as $|B_n^{e(h)}|^2$. In order to obtain the light flux of the nth order emanating from a unit area of the disc surface, we have to multiply the intensity by an area factor $\sqrt{1-\gamma_n^2} = \cos\theta_n$ as shown in Fig. 5. The normalized flux $I_n^{e(h)}$ is then given by

$$I_n^{e(h)} = |B_n^{e(h)}|^2 \frac{\cos\theta_n}{\cos\theta_i}. \qquad [26]$$

Eq. (26) is seen to reduce consistently to the scalar theory definition of the normalized flux, $I_n = |B_n|^2$, in the limit of $d \gg n\lambda$, so that $\cos\theta_n \simeq \cos\theta_i$ (see Eq. [8b]). In Appendix 1 it is shown that

$$\sum_n I_n^{e(h)} = 1, \qquad [27]$$

where the summation is over those n that give $|\theta_n| \leq \pi/2$. As distinct from the scalar theory, the noninclusion of the evanascent waves in Eq. [27] shows correctly that all the energy flux is carried by propagating modes.

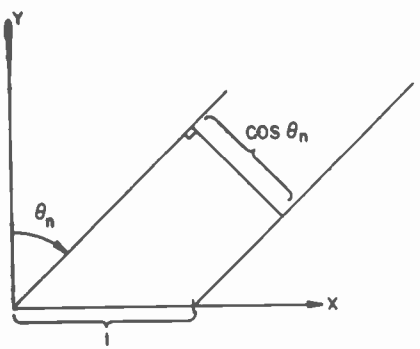


Fig. 5—For light emanating from a unit area of the disc and propagating at the direction θ_n, the total light flux is obtained by multiplying the light intensity with the area factor $\cos\theta_n$.

3.3 Numerical Calculation and Discussion

From the equations determining the amplitudes of the different diffraction orders and the channel modes, one set of variables can be eliminated to yield one infinite set of linear simultaneous equations for $C_n^{e(h)}$ or $B_n^{e(h)}$. Details of the equations for $C_n^{e(h)}$ are given in Appendix 2. Since it is expected that the magnitude of $C_n^{e(h)}$ will decrease rapidly for $n \to \infty$, in actual numerical calculation all those $C_n^{e(h)}$ for n beyond a certain number m are set equal to zero. The resulting $m \times m$ simultaneous equations are solved on computer using the Gauss elimination method. The error caused by truncation can be checked by varying the number m of channel modes included in the calculation and comparing the results. It is found in practice that accurate solution of the diffraction amplitudes can be obtained by including all the nonattenuating channel modes, i.e., $n < 2a/\lambda$, plus two or three evanascent modes.

In Fig. 6 the calculated results of $I_{0,1}^{e(h)}$ are plotted as a function of channel depth h/λ for $d = 1.5\lambda$, $a/d = 0.5$, and $\theta_i = 0°$. In this case $n = -1, 0, +1$ are the only propagating diffraction orders and θ_1 can be calculated to be ~42°. The numerical results for $I_1^{e(h)}$ confirms the general diffraction spectrometer data on the RCA VideoDiscs that $I_1^h > I_1^e$ for $h/\lambda \leq 0.25$. Two additional features of the result should be noted. First,

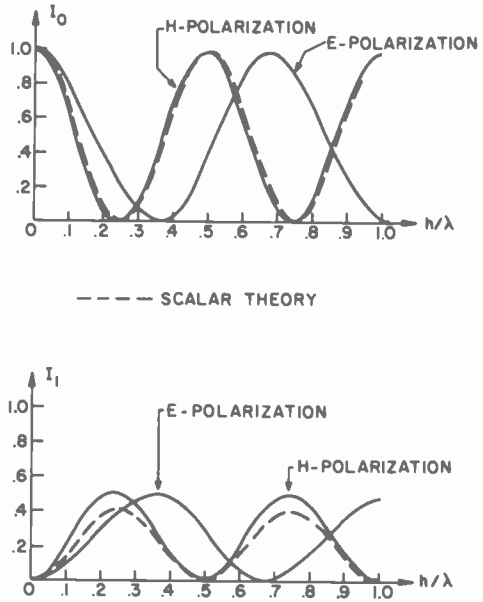


Fig. 6—I_0 and I_1 as a function of signal depth h for a plane wave normally incident on the square-wave grating with $d = 1.5\lambda$, $a/d = 0.5$.

the maximum value of I_1 given by the scalar theory is always smaller than that for the exact result. This is a direct consequence of energy nonconservation of the scalar theory as noted before. Second, the E-polarization result is qualitatively different from the others in that the nth maximum of I_1^e is seen to occur at larger value of h/λ than the similar peak of I_1^h. Physically, such behavior of the E-polarization diffraction can be ascribed to the boundary condition of vanishing E_z at the two sides of the signal channel. This condition effectively inhibits the propagation of the E-polarized light along the channel walls. As a result, penetration of the incident E-polarized light into the signal element is impeded by the "drag" of the channel edges. Compared to the H-polarization case, which does not have this "edge-dragging" effect (since H_z can be finite at metal surface due to the existence of surface currents), the average phase delay across the channel width is therefore smaller for the E-polarized reflected wave. Moreover, since the average phase delay ($4\pi h/\lambda$ for the scalar theory) can be taken as a measure of the effective optical depth of the signal, it follows that for the same signal element the E-polarized light would perceive a smaller effective depth than the H-polarized light. This explains why the first (or the nth) peak of I_1^e occurs at a larger value of h/λ than the same peak for I_1^h.

Obviously, since the effect of edge dragging decreases away from the channel walls, the importance of the effect should vary with the channel width a. It is expected that the smaller the width, the more conspicuous would be the difference between the E- and the H-polarizations and vice versa. This is indeed borne out in Fig. 7, where we show the case of $d = 1.5\lambda$, $a = 0.25d = 0.375\lambda$, and $\theta_i = 0°$. Since now a is less than $\lambda/2$, all the channel modes for the E-polarization are evanascent as noted in the previous section. This accounts for the asymptotic behavior of $I_{0,1}^e$, because beyond a certain depth, the channel bottom only has an exponentially small effect on the diffraction characteristics. In the H-polarization case, however, the existence of one nonattenuating mode insures the periodical behavior of $I_{0,1}^h$.

Another interesting result of the exact solution is shown in Fig. 8, where $d = 1.05\lambda$, $a/d = 0.75$, and $\theta_i = 0°$. From the given value of d, the angle θ_1 is calculated to be $\sim 72°$. It is seen that in the present case the maximum of I_1^e is considerably lower than the maximum of I_1^h. A simple explanation of this behavior can be obtained by noting that for the E-polarization wave, propagation along the disc surface ($\theta_n = \pm\pi/2$) is prohibited by the boundary condition of vanishing E_z at the metallic surfaces $(a/2) \leq |x - nd| \leq (d/2)$. Therefore, as d approaches λ so that $\theta_1 \to \pi/2$, I_1^e is expected to decrease toward zero. In fact, for $d \sim \lambda$ this effect can be utilized to determine the depth of the signal element. Fig. 9 gives the dependence of the ratio I_s^h/I_s^e on signal depth for various values of d. Here $I_s^{e(h)}$ stands for the total flux of the diffracted light

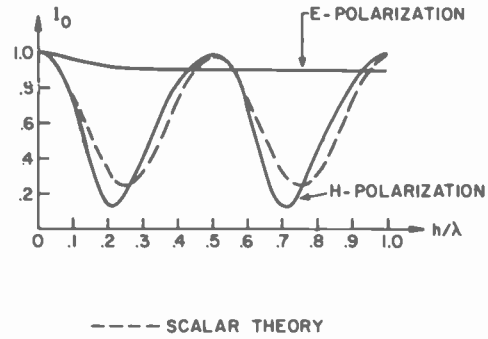


Fig. 7—I_0 and I_1 as a function of signal depth h for a plane wave normally incident on the square-wave grating with $d = 1.5\lambda$, $a/d = 0.25$.

emanating from the disc surface at angles between 10° and 80° as would be collected by the diffraction spectrometer. Values of the parameters used in the calculation are $a/d = 0.5$, $\lambda = 6328$Å for the HeNe laser, and $\theta_i = 0°$ to correspond to the operating condition of the diffraction spectrometer. It is seen that for value of d close to λ, the ratio I_s^h/I_s^e varies steeply with the signal depth. This general qualitative behavior is reproduced by the experimental data[8] obtained from the diffraction spectrometer as shown in Fig. 10. The fact that the experimental curve

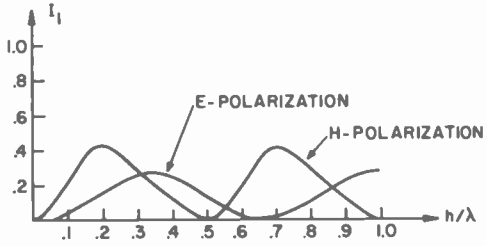


Fig. 8—I_1 as a function of signal depth h for a plane wave normally incident on the square-wave grating with $d = 1.05\lambda$, $a/d = 0.75$.

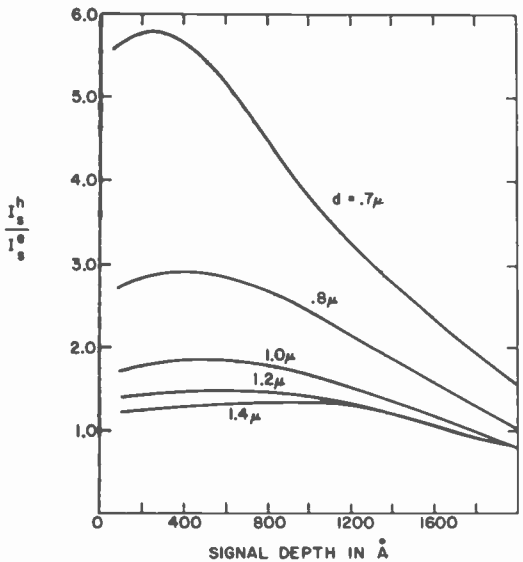


Fig. 9—Plot of the calculated ratio I_s^h/I_s^e versus signal depth for a coherent beam of wavelength $\lambda = 6328$Å normally incident on a square-wave, perfectly conducting grating with $a/d = 0.5$. Definition of $I_s^{e(h)}$ is given in the text.

for $d \simeq 0.7$ μm lies lower than the theoretical curve in Fig. 9 can be attributed to the deviation of the actual signal profile from the idealized model and to the finite disc conductivity. However, the curve does provide us with a means[8] for establishing the signal depth from a measurement of the ratio I_s^h/I_s^e. A more conventional mode of operation for the diffraction spectrometer is to use the unpolarized light for determining the signal depth. In this case the measured quantity is $(I_s/I_0)^{1/2}$, where $I_s = I_s^e + I_s^h$, $I_0 = I_0^e + I_0^h$, $\theta_i = 0°$, and $a/d = 0.5$. Fig. 11 gives the signal depth h as a function of the quantity $(I_s/I_0)^{1/2}$ for those values of d corresponding to the 5 MHz carrier frequency signal at different VideoDisc radii R indicated in the graph. The wavelength of the probe light is assumed to be 6328 Å. It is interesting to observe from Fig. 11 that the scalar theory agrees closely with the exact results up to $h = 800$ Å $\simeq \lambda/8$. This is rather remarkable in view of the fact that if polarized light is used, the deviation of the scalar theory prediction from the calculated results for either of the polarizations is fairly significant at $h \leq 800$ Å.

As noted before, the scalar theory should converge with the exact results in the limit of $d \gg \lambda$. However, the question remains: exactly how fast is this convergence? Fig. 12 gives a graphical answer to this question. For $a/d = 0.5$, $\theta_i = 0°$, and $h = 0.125\lambda$, it is shown that the agreement of the scalar theory value with I_1^e and I_1^h is better than 5% for $d > 5\lambda$. Besides the general trend of asymptotic convergence, we also note in Fig. 12 that

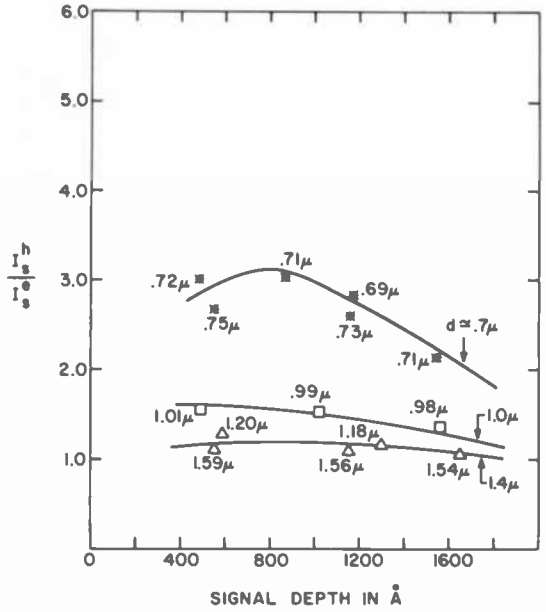


Fig. 10—Plot of the measured ratio I_s^h/I_s^e versus the signal depth for a coherent beam of wavelength $\lambda = 6328$Å normally incident on RCA VideoDisc signal elements with $a/d \simeq 0.5$. The signal depths are experimentally determined using the scanning electron microscope, and the measured values of d are marked beside each data point. The solid curves are drawn through the points to delineate the trend of the experimental data (after Ref. [8]).

the I_1^h curve exhibits well-defined dips known as the Wood anomaly.[9] The cause of these dips at $d = 3\lambda, 5\lambda, 7\lambda, \ldots$ can be traced to the appearance (disappearance) of 3rd, 5th, 7th, ... diffraction orders, respectively. Since the total diffracted power is conserved, the sudden increase or decrease of one diffraction order necessarily requires the readjustment of intensities for all the diffraction orders. In fact, the absence of dips at $d = 2\lambda, 4\lambda, 6\lambda, \ldots$ indicates that the even diffraction orders have zero intensity as can be verified by symmetry consideration and inspection of Eqs. [7] and [23]–[25]. In contrast, however, the I_1^e curve shows no dips at all because as d/λ is changed the highest diffraction orders always make their appearance or disappearance at $\pm 90°$, the directions at which the E-polarized light can not carry any energy as discussed in reference to Fig. 8. The same reason also explains why $I_1^e \rightarrow 0$ as $d \rightarrow \lambda$.

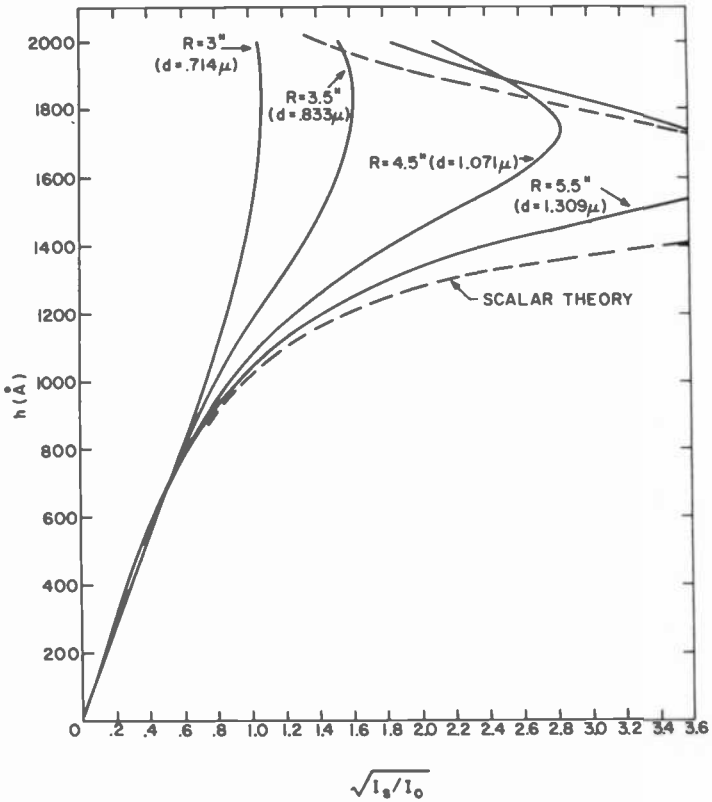


Fig. 11—Signal depth as a function of the measurable quantity $\sqrt{I_s/I_0}$ for an unpolarized plane wave of wavelength $\lambda = 6328$Å normally incident on a conducting square-wave grating with $a/d = 0.5$. The solid curves are for different values of d that represent the 5 MHz carrier signal at the VideoDisc radius R given in the figure. The dashed curve is the scalar theory result which represents the limiting behavior of the exact theory at $d \gg \lambda$.

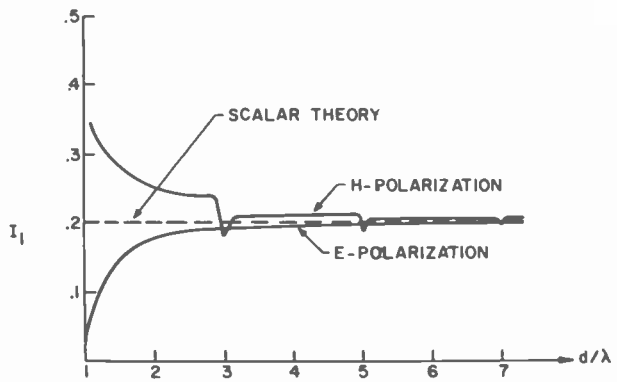


Fig. 12—A plot of I_1 versus periodicity d for a plane wave normally incident on a square-wave grating with depth $h = 0.125\lambda$ and $a/d = 0.5$.

4. Focused Spot Diffraction

In a previous paper in this issue of *RCA Review*,[2] it is seen that the consideration of focused spot diffraction by VideoDisc signal elements is basic to the principle of operation for the optical playback system. In order to relate the theoretical results of this section to the experimentally observed characteristics, it will be helpful to first give a schematic description of the optical readout scheme.

The optical playback system uses a focused light spot to resolve the edges of the signal elements so that, on a rotating disc, the varying spatial separation between the edges can be converted into frequency-modulated electrical signals. At present, optical resolution of the signal element can be accomplished by two different detection methods. In one method, a single axial detector collects all the light diffracted back from the VideoDisc surface through the same focusing lens. The principle of this detection scheme lies in the fact that whenever the focused spot moves over a signal element edge, some of the light is diffracted out of the lens aperture. As a result the detected light flux shows a drop, which can be interpreted as the desired signal. In the differential-phase-detection method used by the RCA optical playback system, a split detector takes as the signal the difference in collected light flux between the right and the left halves (defined along the x-direction) of the detector. This detection scheme essentially measures the asymmetry of the diffraction intensity when the incident spot moves over an edge of the signal element. As will be shown below, the two detection schemes have significantly different polarization dependences.

Before we embark on the formulation of the focused spot diffraction problems, it is necessary to first specify the focused spot profile. To simplify the calculation, we will assume that if $S(z,x)$ is the amplitude profile of the spot at the focal plane, then

$$S(z,x) = \beta(z)\alpha(x), \qquad [28]$$

where $\alpha(x)$ ($\beta(z)$) is usually a symmetric, peaked function of $x(z)$. Furthermore, we will adopt the basic approximation of Fourier optics that if $S(z,x)$ can be written in the form of Eq. [28], then the diffraction in the x-direction can be decoupled, or calculated independently, from the diffraction in the z-direction. These two approximations allow us to treat the focused spot diffraction calculation as a one-dimensional problem. That is, in the calculation of the diffraction characteristics in the x-direction, $\beta(z)$ can be treated as a multiplicative factor that can be cancelled out by normalizing the diffracted light flux to the total incident flux. Experimental evidences indicate that the decoupling of the x-direction and the z-direction diffraction is a good approximation, since large variation of the z-dimension of the focused spot produces no perceptible change in the detected signal for the optical readout system.

Due to the loss of periodicity in the problem (focused spot destroys the symmetry) and the addition of spot size s as a new parameter, an exact focused spot diffraction calculation is still complicated even in one dimension. The large amount of computing time required for its numerical solution makes it impractical to do extensive calculations. However, we note that the essential physical behavior of the general problem can be obtained by considering a mathematically simpler model of focused spot diffraction by a single signal element. It is easy to see why such a model is indeed a good physical approximation to the general case. From a previous paper,[2] it is known that optical resolution of the signal elements requires the focused-spot size to be less than d. On a rotating disc, the condition $s < d$ directly implies that, at any given instant of time, the incident light flux of the focused spot is mostly concentrated on one signal element. The interaction of the focused spot with a single signal element therefore clearly plays the dominant role in determining the resulting diffraction characteristics.

In this section, the single-element diffraction calculation will be examined by both the scalar approximation and the exact formulation. This is followed by the exact solution of the general problem of focused spot diffraction, which serves both as a check on the single-element diffraction results and as the starting point of investigation for higher-order diffraction phenomena involving the interaction of the focused spot with two or more signal elements.

4.1 Single-Element Case: Scalar Theory

The relative positions of the signal element and the focused spot are depicted in Fig. 13. The signal channel has width a and depth h. The focal plane of the lens is taken to be the disc surface, $y = 0$. In the coordinate system where $x = 0$ is defined as the center of the signal channel, the position of the center of the focused spot is denoted by x_0. The size of the focused spot is uniquely determined by the wavelength of light and the focusing angle θ_L defined in Fig. 13. Assuming that the incident laser beam illuminates the lens aperture with uniform intensity, we can calculate the amplitude profile of the focused spot in the x-direction as

$$\alpha(x) = \int_{-\eta}^{\eta} \exp[ik\sigma x]d\sigma = \frac{2\sin k\eta x}{kx}, \qquad [29]$$

where $\eta = \sin\theta_L$ is the numerical aperture of the lens. For $\eta = 0.85$, the intensity profile $|\alpha(x)|^2$ is plotted as a function of x/λ in Fig. 14. We will define the spot size s as

$$s = \frac{\lambda}{2\eta}. \qquad [30]$$

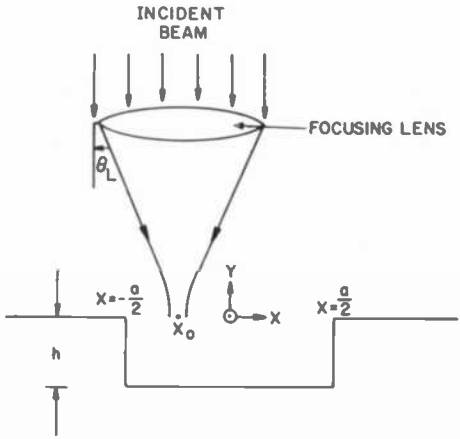


Fig. 13—The coordinate system and the definitions of the various parameters used in the focused-spot diffraction calculation.

As seen from Fig. 14, at $x = \pm s/2$ the intensity is roughly half the peak intensity.

In order to calculate the amplitude of the diffracted wave using the scalar theory, we will follow the approach described in Sec. 3.1. The effect of the signal element is treated as a phase factor multiplying the incident amplitude. With the focused spot centered at x_0, the reflected wave amplitude at $y = 0$ can be written as

$$R(x,x_0) = \alpha(x - x_0) \exp[ikG(x)], \qquad [31a]$$

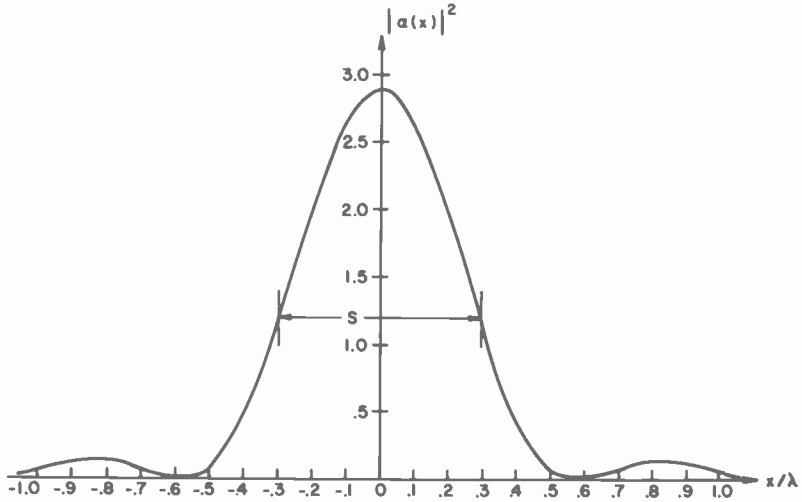


Fig. 14—Focused-spot intensity profile for numerical aperture $\eta = 0.85$.

$$G(x) = \begin{cases} 2h, & |x| \le a/2 \\ 0, & \text{otherwise.} \end{cases} \quad [31b]$$

From $R(x,x_0)$ the amplitude $B(\sin\theta, x_0)$ of the diffracted wave propagating at any particular direction θ can be obtained by simple superposition:

$$B(\sigma,x_0) = \lambda^{-1} \int_{-\infty}^{\infty} R(x,x_0) \exp[-ikx\sigma]dx, \quad [32]$$

where $\sigma = \sin\theta$. Substitution of Eqs. [29] and [31] into Eq. [32] yields

$$B(\sigma,x_0) = \exp[-ik\sigma x_0] \left\{ \text{rect}\left[\frac{\sigma}{2\eta}\right] + \frac{1}{\pi}(\exp[i2kh] - 1)(J_1 - iJ_2) \right\}, \quad [33a]$$

$$\text{rect}\left[\frac{\sigma}{2\eta}\right] = \begin{cases} 1; & |\sigma| \le \eta \\ 0; \text{ otherwise} \end{cases}, \quad [33b]$$

$$J_1 = \int_{-(a/2+x_0)k\eta}^{(a/2-x_0)k\eta} \frac{\sin y \cos(y\sigma/\eta)}{y} dy, \quad [33c]$$

$$J_2 = \int_{-(a/2+x_0)k\eta}^{(a/2-x_0)k\eta} \frac{\sin y \sin(y\sigma/\eta)}{y} dy. \quad [33d]$$

The intensity of the diffracted wave for $|\sigma| \le \eta$ is given by

$$|B(\sigma,x_0)|^2 = 1 + \left(\frac{2\sin(kh)}{\pi}\right)^2 (J_1^2 + J_2^2 - \pi J_1) + \frac{2}{\pi}\sin(2kh)J_2. \quad [34]$$

From $|B(\sigma,x_0)|^2$ the normalized responses for the single axial detector (S.A.D.) and the differential phase detector (D.P.D.) can be calculated as

$$\text{S.A.D. response} = \frac{1}{I_{inc}} \int_{-\eta}^{\eta} |B(\sigma,x_0)|^2 d\sigma, \quad [35]$$

$$\text{D.P.D. response} = \frac{1}{I_{inc}} \left[\int_0^{\eta} |B(\sigma,x_0)|^2 d\sigma - \int_{-\eta}^{0} |B(\sigma,x_0)|^2 d\sigma \right]. \quad [36]$$

Here $I_{inc} = 2\eta$ is the total incident flux.

It should be noted from Eq. [33] that J_1 is symmetric and J_2 is antisymmetric with respect to σ. Therefore, in Eq. [34], we can separate $|B(\sigma,x_0)|^2$ into a symmetric part, consisting of the first two terms, and an antisymmetric part, which is the last term. Such a separation allows

us to immediately deduce the h-dependence of the detector response. For example, since only the antisymmetric term contributes to the D.P.D. response, we see that the split detector signal would vanish at $h = \lambda/4$ according to the scalar theory.

Just as in the plane-wave-diffraction calculation, energy conservation poses a problem for the scalar theory in the present case. To check, we will let $\eta = 1$ so that all the scattered light is collected. By energy conservation it is obvious that the S.A.D. response should be 1. However, results of numerical integration show that when the spot is centered on the edge of the signal channel ($x_0 = \pm a/2$), S.A.D. response is only ~0.6–0.7 for $h = \lambda/8$, $a > \lambda$. Therefore, in this case, about 30–40% of the light flux is unaccounted by the scalar theory.

4.2 Single-Element Case: Exact Theory

The basic principle underlying the exact formulation of the focused diffraction problem is the same as that in the case of plane wave diffraction. For the E-polarization, E_z in the half space $y \geq 0$ is expressed as superposition of plane waves:

$$E_z^{>} = \int_{-\infty}^{\infty} A(\sigma,x_0) \exp[ik(\sigma x - \sqrt{1-\sigma^2}\, y)]d\sigma$$

$$+ \int_{-\infty}^{\infty} B^e(\gamma,x_0) \exp[ik(\gamma x + \sqrt{1-\gamma^2}\, y)]d\gamma, \qquad [37]$$

where

$$A(\sigma,x_0) = \begin{cases} \exp[-ik\sigma x_0]; & |\sigma| \leq \eta \\ 0; & \text{otherwise} \end{cases} \qquad [38]$$

is the Fourier amplitude of the focused spot centered at x_0, and $B^e(\sigma,x_0)$ is the amplitude of the diffracted wave to be determined. For $y \leq 0$, E_z can be expressed as a linear combination of channel modes,

$$E_z^{<} = \sum_{n=1}^{\infty} C_n^e f_n^e(x), \qquad [39]$$

where $f_n^e(x)$ is defined by Eq. [14b]. Requirements of continuity for E_z and $H_x = (1/ik)\partial E_z/\partial y$ at $y = 0$ yield

$$\int_{-\infty}^{\infty} [A(\sigma,x_0) + B^e(\sigma,x_0)] \exp[ik\sigma x]d\sigma$$

$$= \begin{cases} \sum_{n=1}^{\infty} C_n^e \sin\left[kh\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right] \sin\left[\frac{n\pi}{a}\left(x+\frac{a}{2}\right)\right]; |x| \le \frac{a}{2} \\ 0; \quad \text{otherwise} \end{cases},$$

[40]

$$\int_{-\infty}^{\infty} [A(\sigma,x_0) - B^e(\sigma,x_0)]\sqrt{1-\sigma^2}\exp[ik\sigma x]d\sigma$$
$$= i\sum_{m=1}^{\infty} C_m^e \sqrt{1-\left(\frac{m\lambda}{2a}\right)^2} \cos\left[kh\sqrt{1-\left(\frac{m\lambda}{2a}\right)^2}\right]$$
$$\times \sin\left[\frac{m\pi}{a}\left(x+\frac{a}{2}\right)\right]; |x| \le \frac{a}{2}. \qquad [41]$$

All functions in Eqs. [40] and [41] containing square roots have their usual meanings as defined in Sec. 3.2 when the arguments of square roots become negative. Also, the continuity of H_x as expressed by Eq. [41] is noted not to extend beyond $|x| \ge a/2$ due to the existence of surface currents at the metal surface. Multiplying Eq. [40] by $\exp[-ik\gamma x]$ and integrating with respect to x from $-\infty$ to $+\infty$, we get

$$B^e(\gamma,x_0) = -A(\gamma,x_0) + \sum_{n=1}^{\infty} \sin\left[kh\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right] Q_n(\gamma)C_n^e,$$

[42]

$$Q_n(\gamma) = \frac{1}{2\pi}\frac{(n\lambda/2a)}{(n\lambda/2a)^2-\gamma^2}[\exp(ik\gamma a/2)$$
$$+ (-1)^{n+1}\exp(-ik\gamma a/2)]. \qquad [43]$$

Similarly, multiplication of Eq. [41] by $\sin[(n\pi/a)(x + a/2)]$ and integrating with respect to x from $-a/2$ to $a/2$ yield

$$iC_n^e\sqrt{1-\left(\frac{n}{2a}\right)^2}\cos\left[kh\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right]\frac{a}{2\lambda}$$
$$= \int_{-\infty}^{\infty} [A(\sigma,x_0) - B^e(\sigma,x_0)]\sqrt{1-\sigma^2}\,Q_n^* d\sigma. \qquad [44]$$

For a given input $A(\sigma,x_0)$, Eqs. [42]–[44] uniquely determine all C_n^e's and $B^e(\sigma,x_0)$. The procedure for the H polarization is completely similar. Using the magnetic channel modes $f_n^h(x)$'s defined by Eq. [22] to expand H_z in the region $y \le 0$, we obtain the equations

$$B^h(\gamma,x_0) = A(\gamma,x_0) + \frac{i}{\sqrt{1-\gamma^2}} \sum_{n=0}^{\infty} \sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}$$

$$\times \sin\left[kh\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right] P_n(\gamma) C_n^h, \qquad [45]$$

$$C_n^h \cos\left[kh\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right] \frac{a}{2\lambda}(1+\delta_{n0})$$

$$= \int_{-\infty}^{\infty} [A(\gamma,x_0) + B^h(\gamma,x_0)] P_n^*(\gamma) d\gamma, \qquad [46]$$

$$P_n(\gamma) = \frac{i}{2\pi} \frac{\gamma}{(n\lambda/2a)^2 - \gamma^2} [\exp(ik\gamma a/2)$$

$$+ (-1)^{n+1} \exp(-ik\gamma a/2)]. \qquad [47]$$

The diffracted light flux propagating in a particular direction is given by $|B^{e(h)}(\gamma,x_0)|^2 \sqrt{1-\gamma^2}$, where $\sqrt{1-\gamma^2}$ is the area factor defined by Fig. 5. By using procedures similar to those in Appendix 1, it can easily be shown that

$$\int_{-1}^{1} |B^{e(h)}(\sigma,x_0)|^2 \sqrt{1-\sigma^2}\, d\sigma = \int_{-1}^{1} |A(\sigma,x_0)|^2 \sqrt{1-\sigma^2}\, d\sigma \qquad [48]$$

as required by conservation of energy. From the knowledge of $|B^{e(h)}(\sigma,x_0)|^2$ the normalized detector response can be calculated:

$$\text{S.A.D. response} = \frac{1}{I_{inc}} \int_{-\eta}^{\eta} |B^{e(h)}(\sigma,x_0)|^2 \sqrt{1-\sigma^2}\, d\sigma, \qquad [49]$$

$$\text{D.P.D. response} = \frac{1}{I_{inc}} \left[\int_0^{\eta} |B^{e(h)}(\sigma,x_0)|^2 \sqrt{1-\sigma^2}\, d\sigma \right.$$

$$\left. - \int_{-\eta}^{0} |B^{e(h)}(\sigma,x_0)|^2 \sqrt{1-\sigma^2}\, d\sigma \right], \qquad [50]$$

$$I_{inc} = \int_{-1}^{1} |A(\sigma,x_0)|^2 \sqrt{1-\sigma^2}\, d\sigma$$

$$= \eta\sqrt{1-\eta^2} + \arctan\frac{\eta}{\sqrt{1-\eta^2}}, \qquad [51]$$

where $A(\sigma,x_0)$ is defined by Eq. [38].

4.3 Numerical Calculation and Discussion

The calculation of diffraction amplitudes in the scalar theory involves the numerical evaluation of integrals J_1 and J_2 defined by Eq. [33]. This is done on computer using Gaussian integration routines. For the exact theory, the determination of $C_n^{e(h)}$ and $B^{e(h)}(\sigma,x_0)$ requires further re-

the equations. Appendix 2 gives the results of this procedure in terms of an infinite set of equations for $C_n^{e(h)}$. To calculate $C_n^{e(h)}$, the infinite set of equations is truncated to m equations and m $C_n^{e(h)}$'s. The resulting finite number of equations are solved on computer using the Gauss elimination method. Once $C_n^{e(h)}$'s are known, the diffraction amplitude $B^{e(h)}(\sigma,x_0)$ can be directly calculated from Eqs. [42] and [45]. It is found in practice that the number of channel modes required for the accurate solution of diffraction amplitudes varies directly with the channel width a. In the following, the empirical rule $m = (2a/\lambda) + 2$ will be followed in deciding the number of channel modes m to be used in the calculation.

Fig. 15 shows the diffraction light flux as a function of angle for the case of $a = 2\lambda$, $h = 0.125\lambda$, $\eta = 0.85$, and $x_0 = -a/2$. The quantities plotted are $|B(\sin\theta,x_0)|^2 \cos\theta$ for the scalar theory and $|B^{e(h)}(\sin\theta,x_0)|^2 \times \cos^2\theta$ for the exact solution, normalized by the respective incident flux I_{inc}. The additional $\cos\theta$ factors arise because we are plotting the fluxes as a function of θ and $d(\sin\theta) = \cos\theta d\theta$. Two features of Fig. 15 should be noted. First, the total diffracted light flux for the scalar theory is obviously less than the exact results. This reflects the problem of energy conservation in the scalar theory as mentioned earlier. Second, at $\pm 90°$, the E-polarization gives no diffracted light whereas the H-polarization has a finite amount of diffracted light. This phenomenon is similar to the one observed in the plane wave diffraction (see Figs. 9 and 10) and can be attributed to the inability of the E-polarized light to propagate along the surface because of the condition of vanishing E_z at the metallic boundary.

In the plane wave diffraction we have seen that the deviation of the results for the scalar theory from those for the exact solution becomes significant when the grating periodicity d or the channel width a becomes comparable to λ. For the focused spot diffraction there is an additional length s, the spot size, and it is natural to ask whether the accuracy of the scalar theory also varies with s. To examine this question, let us consider the different cases of focused spot diffraction where the channel width a and the spot size s ($<a$) stay at a fixed ratio. Careful examination of Eqs. [33]–[36] shows that under this condition (a/s = const. or, since $s = \lambda/2\eta$, $a\eta$ = const.) the scalar theory predicts a detector response that is an invariant function of x_0/a. In Figs. 16 and 17 the scalar theory prediction of the normalized D.P.D. response is compared with the exact solution for different spot sizes. We have chosen $h = 0.125\lambda$ and $a > 2\lambda$ so that the deviation of the scalar theory can be mainly attributed to spot-size effect. It is seen that for $\eta = 0.85$, the numerical aperture used for the optical readout of RCA VideoDiscs, the difference between the scalar theory and the exact solution is quite significant for either of the polarizations. Therefore, it can be concluded that apart from the sig-

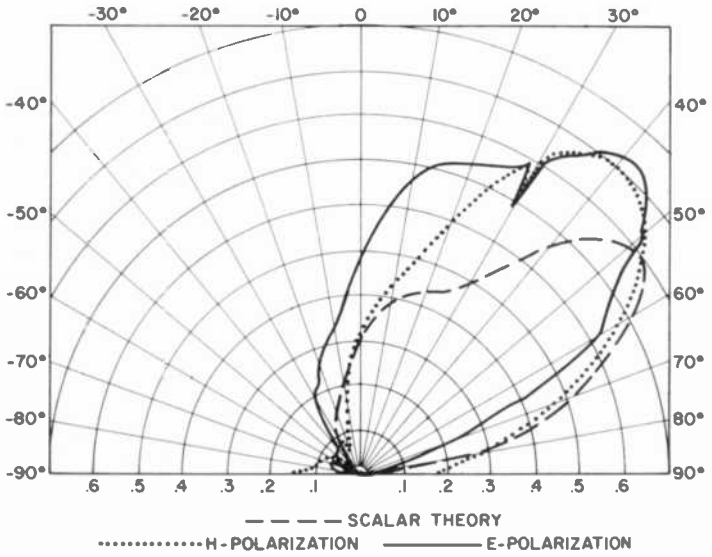


Fig. 15—Angular diffraction spectrum when a focused spot with $\eta = 0.85$ is centered at the left edge ($x_0 = -a/2$) of the square-wave signal channel with $a = 2\lambda$ and $h = 0.125\lambda$.

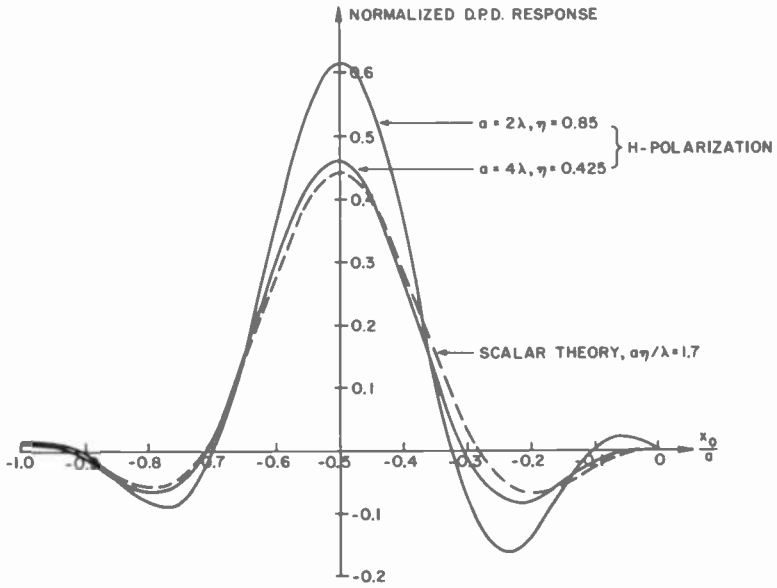


Fig. 16—Normalized differential phase detector response when an *H*-polarized focused spot traverses the left edge ($x = -a/2$) of the square-wave signal channel with $h = 0.125\lambda$. The values of a and η for each curve are indicated in the figure. The dashed curve gives the invariant scalar-theory result.

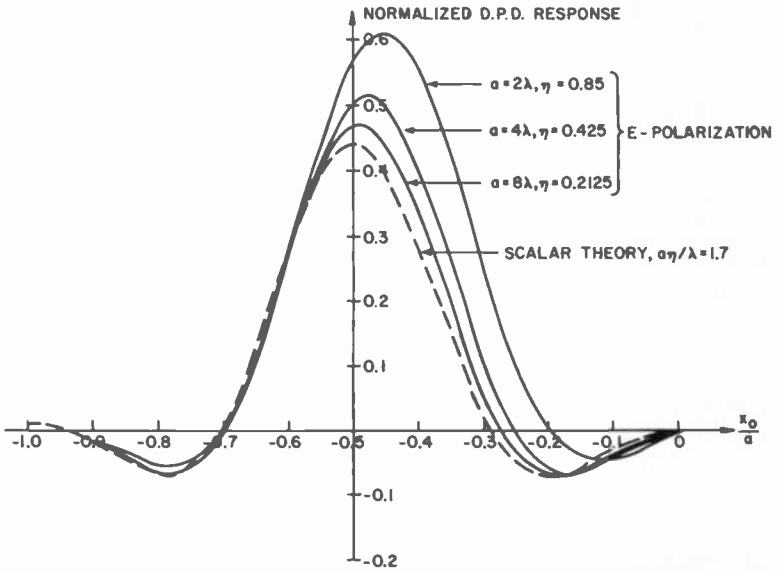


Fig. 17—Normalized differential phase detector response when an E-polarized focused spot traverses the left edge ($x = -a/2$) of the square-wave signal channel with $h = 0.125\lambda$. The values of a and η for each curve are indicated in the figure. The dashed curve gives the invariant scalar theory result.

duction of Eqs. [42]–[47] by eliminating either $C_n^{e(h)}$ or $B^{e(h)}(\sigma,x_0)$ from nal-element dimension, the size of the focused spot also plays an important role in determining the validity of the scalar theory.

Figs. 18, 19, and 20 show the normalized D.P.D. response as a function of spot position for various channel widths and depths. The numerical aperture value $\eta = 0.85$ is used in all the graphs. To avoid crowding, the scalar theory results, which behave similar to the H-polarization case, are not plotted. We note that for $h = 0.125\lambda$, the two edges of the signal channel are well resolved by both polarizations for $a = 4\lambda$, λ, and 0.5λ. However, at $h = 0.25\lambda$ the two polarizations give drastically different D.P.D. responses. It is recalled from the discussion in Sec. 4.1 that the scalar theory predicts zero D.P.D. response when $h = 0.25\lambda$ regardless of the value for a. In Fig. 18, it can be seen that the H-polarization signal fluctuates around zero for $a = 4\lambda$. At $a = \lambda$ and 0.5λ, the D.P.D. response for the H-polarization reverses in sign when h is increased from 0.125λ to 0.25λ. Further calculation has shown that the H-polarization response nearly vanishes when $h = 0.21\lambda$–0.22λ. Therefore, the overall behavior of the H-polarization case conforms with the scalar theory prediction. The same thing can not be said about the E-polarization, however, which has the maximum signal at $h \simeq 0.25\lambda$ instead of the predicted zero response. In fact, calculation has indicated that the sign reversal for the E-polarization signal will not occur until h reaches about 0.5λ. Therefore,

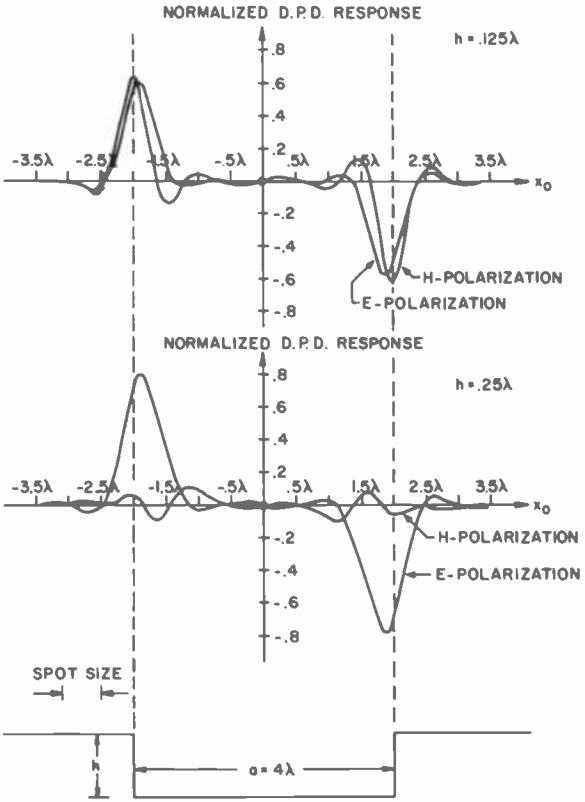


Fig. 18—Normalized differential phase detector response when a focused spot ($\eta = 0.85$) traverses a single square-wave signal channel with $a = 4\lambda$. The two cases shown are for different values of the signal depth h indicated in the figure.

in terms of depth insensitivity of the readout signal, the E-polarization is expected to be superior to the H-polarization.

Why does the E-polarization behave so differently from the H-polarization? In order to answer this question, let us examine the amplitude and the phase delay of the reflected wave at $y = 0$ when the incident focused spot is centered at $x = -a/2$. According to the scalar theory approximation, the reflected wave is different from the incident wave only by a constant phase factor across the signal channel. However, Fig. 21 shows that for $h = 0.125\lambda$, $a = 2\lambda$, the phase delay for the E-polarized reflected wave is significantly less than that of the H-polarization or the scalar approximation near the channel edge, $x = -a/2$. This is understandable, since the E-polarized wave is prohibited from propagating along the channel wall due to the boundary condition of vanishing E_z at the metallic surface. The resulting edge-dragging effect, which has

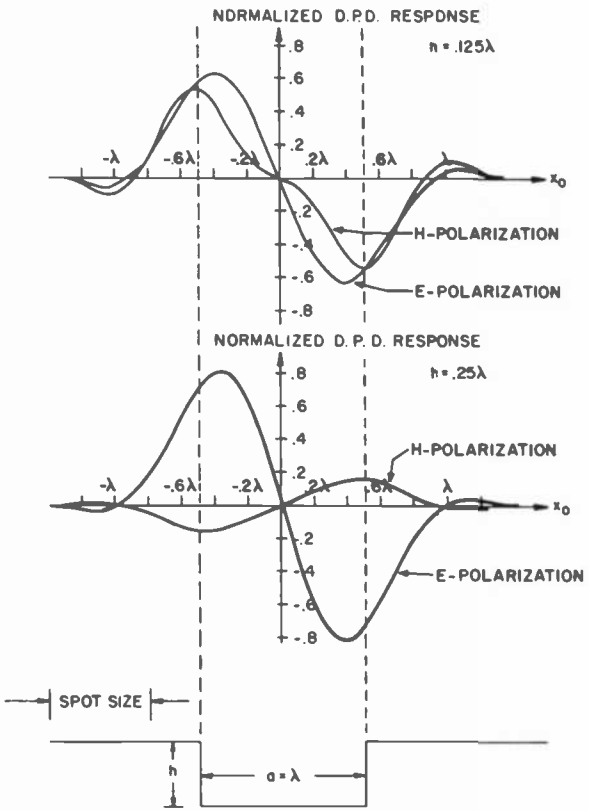


Fig. 19—Normalized differential phase detector response when a focused spot ($\eta = 0.85$) traverses a single square-wave signal channel with $a = \lambda$. The two cases shown are for different values of the signal depth h indicated in the figure.

been discussed earlier in connection with the plane wave diffraction problem, causes the average phase delay of the E-polarization case to fall far short of the scalar theory value near the channel edge. In Fig. 22 we can see that for $h = 0.25\lambda$, the average phase delay in the region of highest reflected light intensity (extends about 0.6λ from the channel edge) is about $\lambda/4$ for the E-polarization instead of $\lambda/2$ as given by the scalar approximation. It is therefore not surprising that the D.P.D. response for the E-polarization is maximum at $h = 0.25\lambda$. Besides phase delays, amplitudes of the reflected waves are also plotted in Figs. 21 and 22. The scalar theory assumption, that the reflected amplitude is identical to the incident spot profile, is seen to be qualitatively descriptive of the exact results for $h = 0.125\lambda$. However, as the signal depth is increased to 0.25λ the reflected amplitudes for both the E- and the H-polarizations lose any resemblance to the incident spot, suggesting that reflections from the channel wall play an increasingly important role.

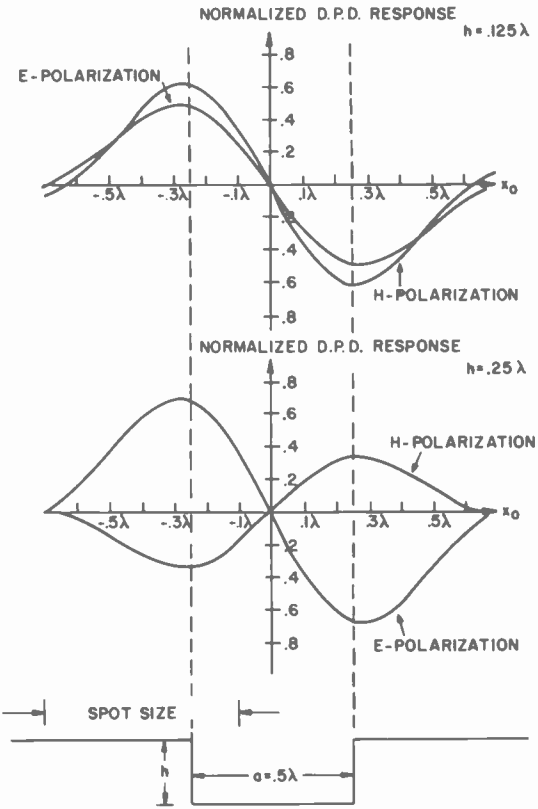


Fig. 20—Normalized differential phase detector response when a focused spot ($\eta = 0.85$) traverses a single square-wave signal channel with $a = 0.5\lambda$. The two cases shown are for different values of the signal depth h indicated in the figure.

To be sure that the effects described above are actually associated with the channel edge and not due to something else, we show in Fig. 23 the amplitude and the phase delay of the reflected wave at $y = 0$ when the spot is focused at the center of the signal channel with dimensions $a = 2\lambda$ and $h = 0.125\lambda$. In this case the scalar theory is indeed seen to be an excellent approximation. It is therefore clear from Figs. 21–23 that there is a "boundary region," extending about 0.5λ from the channel wall, in which the scalar approximation breaks down. In this picture the fraction of total light flux concentrated in the "boundary region" is a measure of invalidity for the scalar theory.

The normalized S.A.D. responses for $\eta = 0.85$ and various values of a and h are plotted in Figs. 24–26. For $a = 4\lambda$, the two edges of the signal channel are well resolved by both the E- and the H-polarizations. However, the E-polarization signal is much smaller than that of the

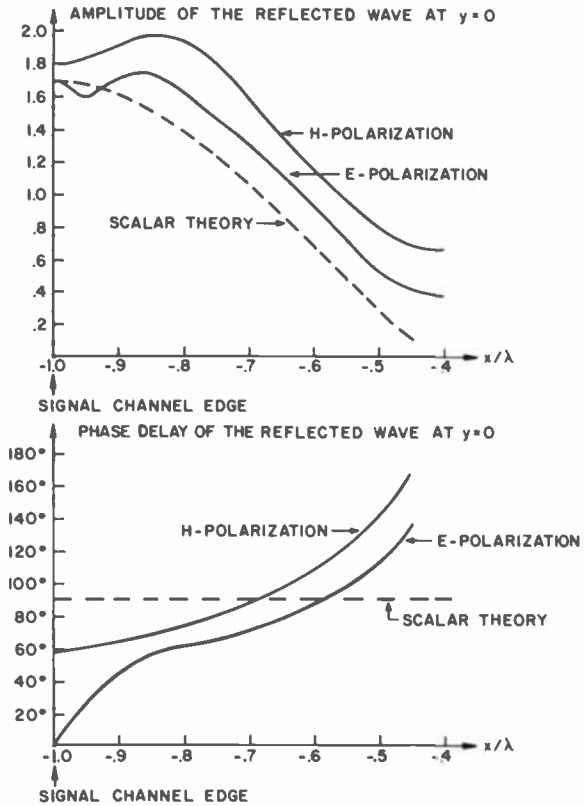


Fig. 21—The amplitude and the phase delay of the reflected wave at $y = 0$ when the incident focused spot ($\eta = 0.85$) is centered at the left edge ($x_0 = -a/2$) of the square-wave signal channel with $a = 2\lambda$, $h = 0.125\lambda$.

H-polarization. Since the S.A.D. signal strength reflects the amount of light scattered out of the lens aperture, the cause of the smaller response for the E-polarization can be easily understood by referring to Fig. 15, where it is shown that due to the boundary condition, the E-polarized light has great difficulty in being diffracted into angles close to $\pm 90°$. When the channel width is decreased to λ, Fig. 25 shows that whereas the H-polarized light can still resolve the two edges, the E-polarization has almost lost the resolution. From these results it can be concluded that in the limit of large numerical aperture (and therefore small spot size), the H-polarization is superior to the E-polarization in terms of signal strength and resolution whenever S.A.D. is used.

The S.A.D. response for the case where the channel width $a = 0.5\lambda$ is smaller than the spot size ($s \simeq 0.62$) is given in Fig. 26. The loss of edge resolution in the present case should be compared to the D.P.D. response

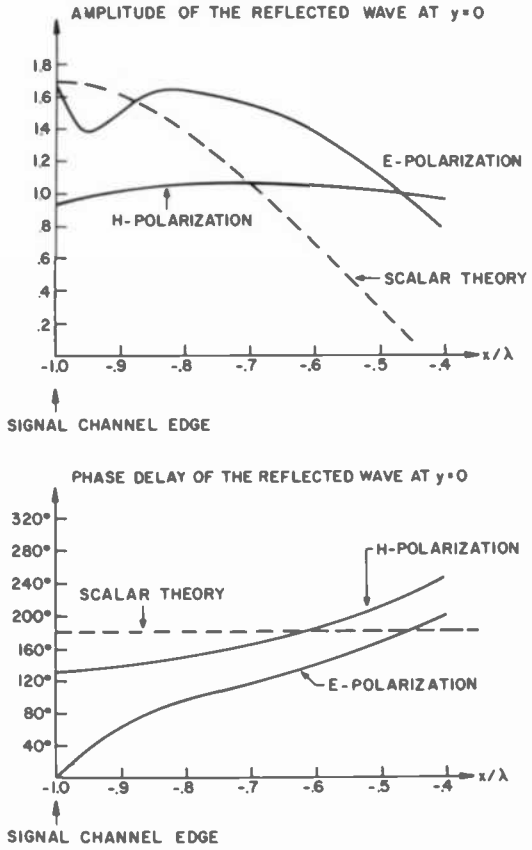


Fig. 22—The amplitude and the phase delay of the reflected wave at $y = 0$ when the incident focused spot ($\eta = 0.85$) is centered at the left edge ($x_0 = -a/2$) of the square-wave signal channel with $a = 2\lambda$, $h = 0.25\lambda$.

of Fig. 20 showing the resolution of the edges for the same channel width. The inability of the S.A.D. to resolve the signals when $a \leq s$ is in fact the cause for the development of the D.P.D. for optical readout of the RCA VideoDisc. It is interesting to note, however, that there is a certain similarity between the diffraction geometries considered in Fig. 26 and those encountered in the optical readout of Philips video discs.[10] Fig. 27 gives a schematic picture of the Philips video disc signal elements. In contrast to the RCA VideoDisc, the signal elements of the Philips video disc are narrow in the radial direction and long in the direction of motion. In the Philips optical readout scheme with S.A.D., the focused spot size is required to be larger than the narrow dimension of the signal element, and the decrease in the detected light flux as the spot traverses the long dimension of the signal element is taken to be the readout signal.

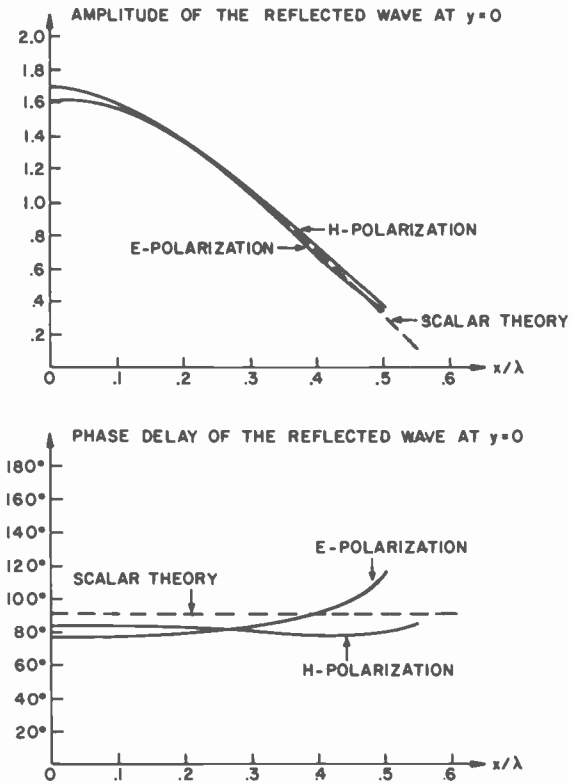


Fig. 23—The amplitude and the phase delay of the reflected wave at $y = 0$ when the incident spot ($\eta = 0.85$) is focused at the center ($x_0 = 0$) of the square-wave signal channel with $a = 2\lambda$, $h = 0.125\lambda$.

It is easy to see that when the spot is focused at the center of the signal element as shown in Fig. 27, the diffraction geometry is completely similar (apart from an interchange of x and z directions) to the case considered in Fig. 26 with $x_0 = 0$. As a result, the S.A.D. response at $x_0 = 0$ in Fig. 26 could be interpreted as indicating the readout signal magnitude in the Philips scheme. If the E(H)-polarization is still defined as the configuration in which the E(H) vector of the light lies in the radial plane, the results of Fig. 26 clearly indicate the superiority of E-polarization (in terms of signal strength) in the Philips optical readout format. It should be remarked, however, that the size of the numerical aperture plays an important role in this large polarization effect. Calculation has shown that the two polarizations give almost identical signal strength when both the spot size and the channel width are increased by a factor of two.

When the channel width is equal to or less than 0.5λ, qualitatively

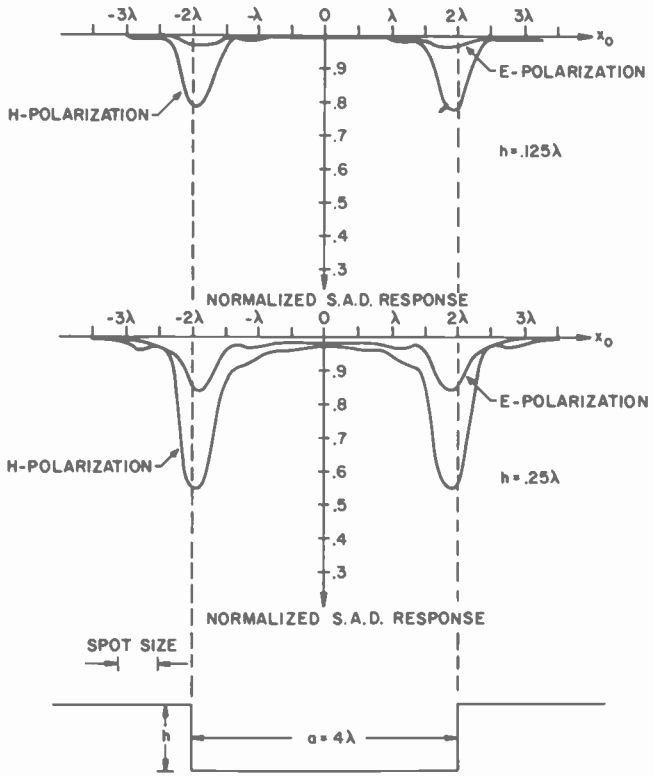


Fig. 24—Normalized single axial detector response when the focused spot ($\eta = 0.85$) traverses a single square-wave signal channel with $a = 4\lambda$. The two cases shown are for different values of the signal depth h indicated in the figure.

different dependences of the D.P.D. response on channel depth are calculated for the two polarizations as shown in Fig. 28. With $x_0 = -a/2$ and $\eta = 0.85$, we observe an asymptotic behavior in the E-polarization signal which reflects the evanascent character of all the E-polarization channel modes when $a \leq \lambda/2$. A completely analogous effect is encountered in the plane-wave diffraction calculation (see Fig. 7). For the H-polarization, the existence of one nonattenuating zeroth order magnetic channel mode, $f_0^h = \cos[k(y + h)]$, insures its oscillating dependence on h. However, the difference between the H-polarization and the scalar theory result in this case appears puzzling at first sight because the dominance of f_0^h, which has the free space wave number k in the y direction, should result in phase delays for the reflected wave closely resembling that predicted by the scalar theory. To get a better understanding of this problem, we show in Fig. 29 the amplitude and the phase delay of the reflected H-polarized wave when $x_0 = -a/2$ and $\eta = 0.85$,

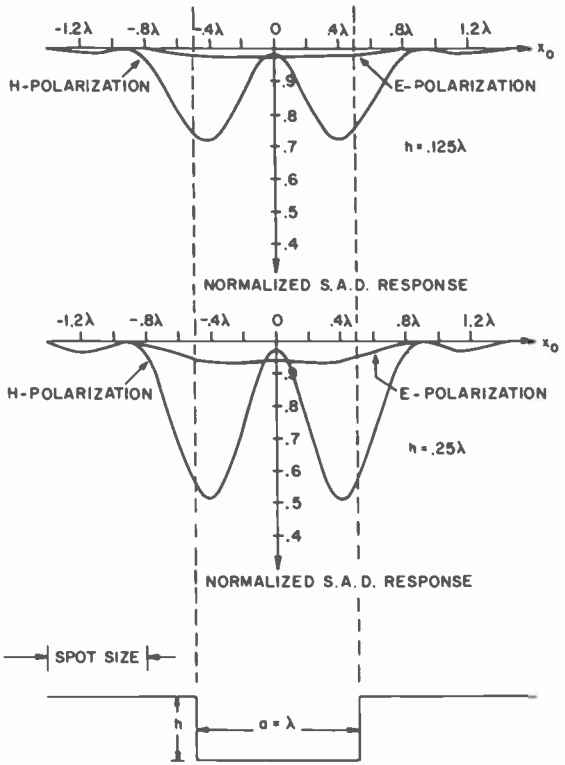


Fig. 25—Normalized single axial detector response when the focused spot ($\eta = 0.85$) traverses a single square-wave signal channel with $a = \lambda$. The two cases shown are for different values of the signal depth h indicated in the figure.

$h = 0.25\lambda$, $a = 0.5\lambda$. It is interesting to observe that the phase delay of the H-polarized reflected wave for $x \geq -a/2$ is very close to the scalar theory prediction as we have expected. However, in the region $x \leq -a/2$, i.e., on the metallic surface $y = 0$, the H-polarized reflected wave amplitude differs significantly from the incident amplitude (and therefore the scalar theory prediction). This difference is due to the existence of surface currents, which makes it possible to rearrange the incident light energy at the surface. The reflection from the metal surface in this case is therefore not a local effect, as in the case of E-polarization, but is actually influenced by the disturbance introduced by the nearby channel edge.

The dependence of the D.P.D. response on the channel width a for $x_0 = -a/2$, $h = 0.125\lambda$, and $\eta = 0.85$ is plotted in Fig. 30. The position of the initial maximum in the response curves is directly correlated with the size of the focused spot. However, the smaller bumps observable in

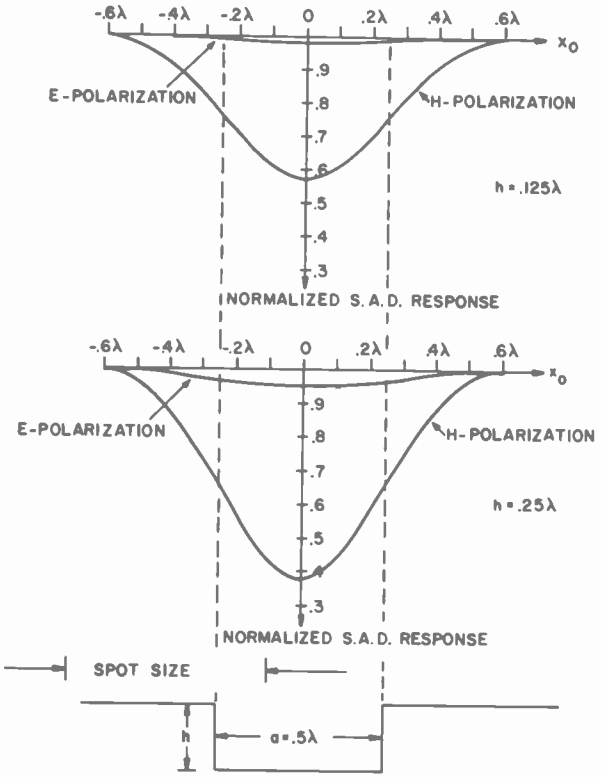


Fig. 26—Normalized single axial detector response when the focused spot ($\eta = 0.85$) traverses a single square-wave signal channel with $a = 0.5\lambda$. The two cases shown are for different values of the signal depth h indicated in the figure.

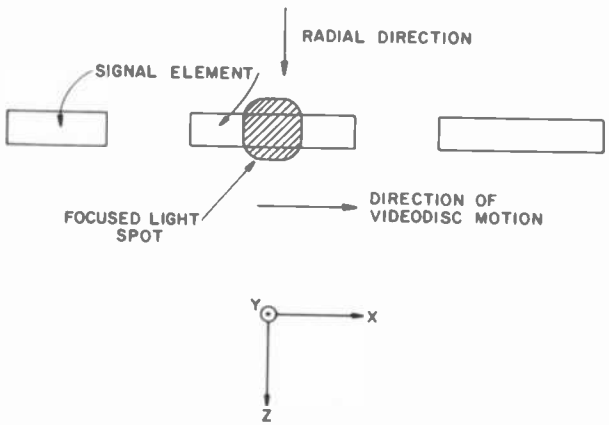


Fig. 27—Schematic representation of the Philips video disc signal elements.

the H-polarization case are caused by the varying magnitude of the surface wave ($\theta = \pm 90$) as the channel width increases.

4.4 The General Case: Exact Theory

In this section we consider the general problem of focused spot diffraction by a square-wave metallic grating. From Eqs. [29], [37], and [38], it is seen that a focused spot is composed of a linear combination of plane waves. Since the Maxwell's equations are linear and the same boundary conditions hold for all plane-wave states, the general problem of focused-spot diffraction can be solved by considering the diffraction of each plane wave individually and then superposing the diffracted amplitudes. That is, the diffracted amplitude at angle θ from the normal when an E(H)-polarized focused spot is incident on the grating is given by

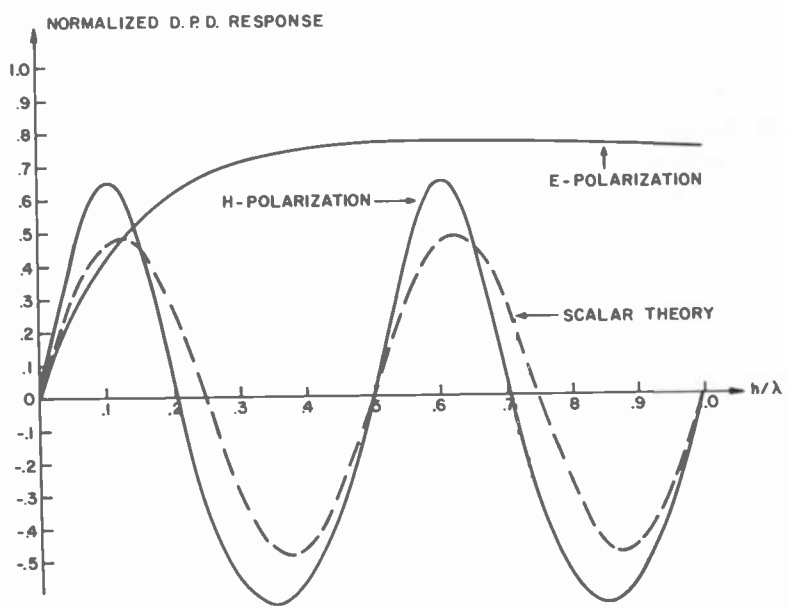


Fig. 28—Normalized differential phase detector response as a function of signal depth h when a focused spot ($\eta = 0.85$) is centered at the left edge ($x_0 = -a/2$) of the square-wave signal channel with $a = 0.5\lambda$.

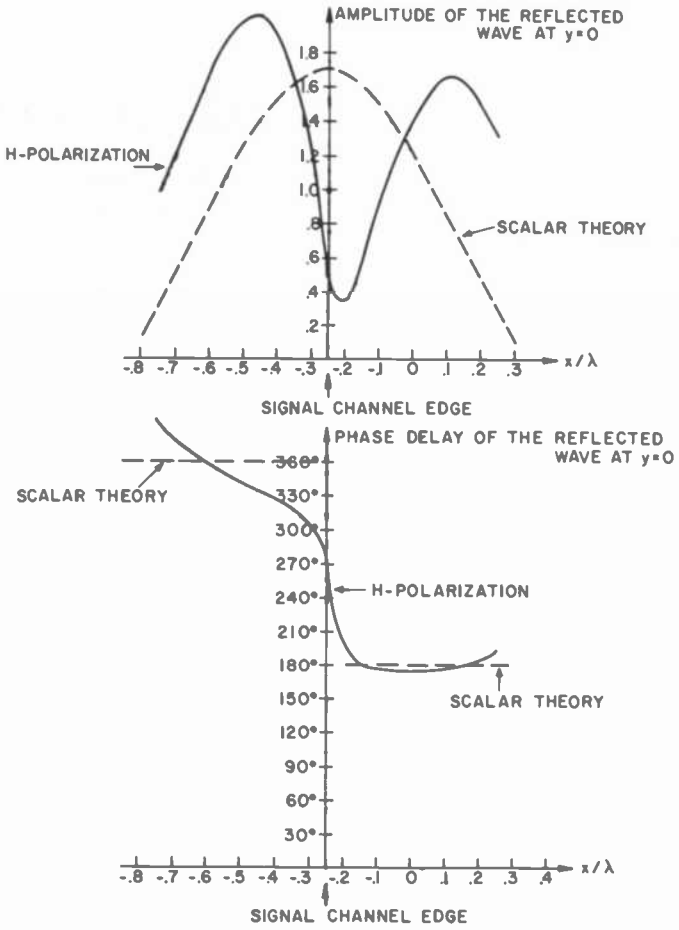


Fig. 29—The amplitude and the phase delay of the reflected wave at $y = 0$ when a focused spot ($\eta = 0.85$) is centered at the left edge ($x_0 = -a/2$) of the square-wave signal channel with $a = 0.5\lambda$, $h = 0.25\lambda$.

$$B_F^{e(h)}(\sin\theta, x_0) = \sum_{n=-\infty}^{\infty} A\left(\sin\theta - \frac{n\lambda}{d}, x_0\right) B_n^{e(h)}\left(\sin\theta - \frac{n\lambda}{d}\right), \quad [52]$$

where the subscript F denotes focused spot diffraction, function A is defined by Eq. [38], and $B_n^{e(h)}(\sin\theta_0)$ is the (complex) amplitude of the nth order diffracted wave when an E(H)-polarized plane wave is incident on the grating at an angle $\theta_i (= -\theta_0)$ from the normal. Since the plane-wave diffraction problem has already been solved in Sec. 3, substitution of Eq. [19] (for E-polarization) or Eq. [23] (for H-polarization) into Eq. [52] completes the solution. From $B_F^{e(h)}$ the detector responses can be

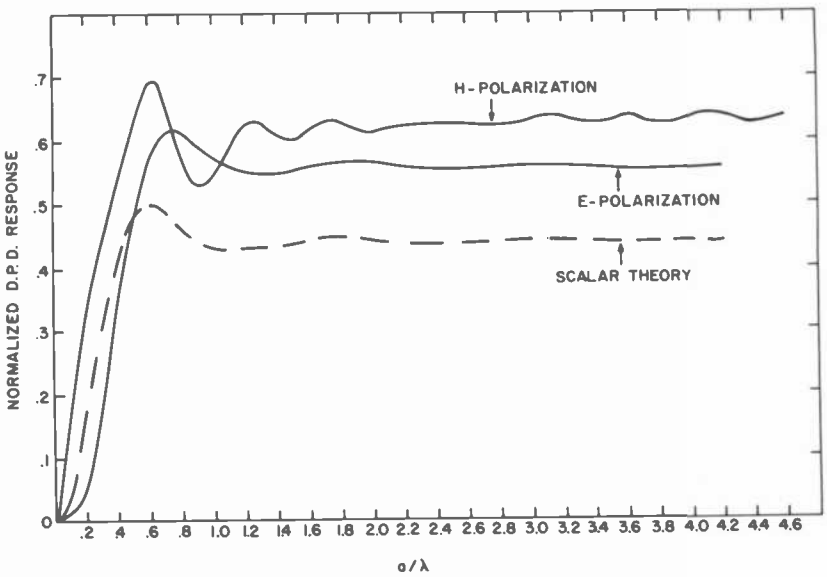


Fig. 30—Normalized differential phase detector response as a function of the channel width a when a focused spot ($\eta = 0.85$) is centered at the left edge ($x_0 = -a/2$) of the square-wave signal channel with $h = 0.125\lambda$.

calculated from Eqs. [50]–[51]. Although simple in formation, the numerical calculation involved in evaluating $B_F^{e(h)}$ and the resulting detector responses is much more tedious than the single-element case, because at every value of θ the $B_n^{e(h)}$'s have to be obtained by solving a set of simultaneous equations. In Fig. 31 the calculated D.P.D. response at $x_0 = -a/2$ is plotted as a function of the dimensionless frequency parameter $\nu = \lambda/\eta d$ for $\eta = 0.85, h = 0.125\lambda$ and $a/d = 0.5$. In the same graph we also show the single-signal-element results where the signal-element width is taken to be $0.5d$. It can be seen that the single-element case is a good approximation to the general problem in the low frequency regime. At values of ν higher than 0.8 ($d \simeq 1.5\lambda$), the general case begins to show features, such as the many bumps in the H-polarization case, that can be attributed to interaction of the focused spot with two or more signal elements. However, apart from these additional oscillations in the frequency response, the single-element results still mimic the qualitative behaviors of the general case. Fig. 32 shows the D.P.D. response at $x_0 = -a/2$ as a function of signal depth h for $\eta = 0.85, \nu = 1$, and $a/d = 0.5$. The single-element results are also plotted in the same graph for comparison. Besides a small difference in magnitude, it is clear that the single-element calculation gives an excellent description of the depth variation of the D.P.D. response for the general focused spot diffraction problem.

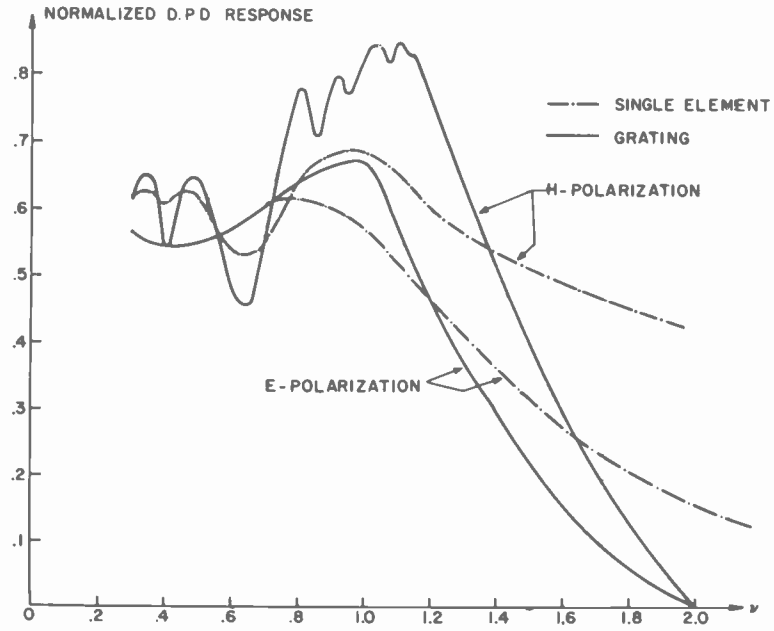


Fig. 31—Normalized differential phase detector response as a function of the dimensionless frequency parameter ν when a focused spot ($\eta = 0.85$) is incident on a square-wave grating with $a/d = 0.5$, $h = 0.125\lambda$. The focused spot is centered at $x = -a/2$ in the coordinate system defined by Fig. 3. For comparison, the single element results with $a = 0.5d$, $x_0 = -a/2$ are also plotted.

Concluding Remarks

In this paper optical diffraction from RCA VideoDisc signals is examined theoretically by exact model calculations. The results obtained not only confirm the experimentally found inadequacies of the scalar diffraction theory but also provide general criteria for the optimal design of optical readout and quality control systems. Furthermore, by determining the various causes for the breakdown of the scalar diffusion theory in the limit of short signal wavelength, a physical picture of optical diffraction is obtained that directly relates the various polarization-dependent properties of the diffraction amplitudes to the boundary conditions at the metal surface.

Acknowledgments

The author gratefully acknowledges many valuable discussions with A. Firester, I. Gorog, and W. R. Roach in the course of this work.

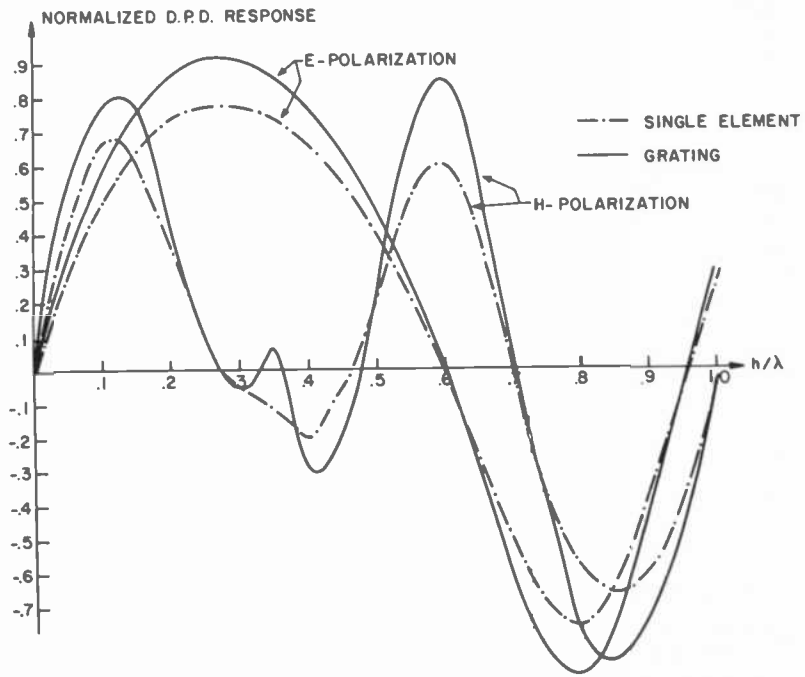


Fig. 32—Normalized differential phase detector response as a function of signal depth h when a focused spot ($\eta = 0.85$) is incident on a square-wave grating with $a/d = 0.5$, $\nu = 1$. The focused spot is centered at $x = -a/2$ in the coordinate system defined by Fig. 3. For comparison, the single-element results with $x_0 = -a/2$, $a = 0.588\lambda$ (deduced from $\nu = 1$, $a/d = 0.5$) are also plotted.

Appendix 1

In this appendix we prove that for the E-polarization,

$$\cos\theta_i = \mathrm{Re}\left[\sum_{n=-\infty}^{\infty} |B_n^e|^2 \sqrt{1-\gamma_n^2}\right]. \qquad [53]$$

From Eqs. [17] and [18], which express the continuity conditions,

$$E_z^>|_{y=0} = E_z^<|_{y=0} \qquad [54]$$

and

$$H_x^>|_{y=0} = H_x^<|_{y=0}, \qquad [55]$$

we can form a new equation

$$E_z^>(H_x^>)^*|_{y=0} = E_z^<(H_x^<)^*|_{y=0}, \qquad [56]$$

which is essentially a statement of energy conservation. Integration of Eq. [56] over $|x| \leq d/2$ yields

$$-\cos\theta_i + \sum_{n=-\infty}^{\infty} |B_n^e|^2 (\sqrt{1-\gamma_n^2})^* + \cos\theta_i[(B_0^e)^* - B_0^e]$$
$$= -i\frac{a}{4d}\sum_{n=1}^{\infty} |C_n^e|^2 \left(\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right)^*$$
$$\times \sin\left[2kh\sqrt{1-\left(\frac{n\lambda}{2a}\right)^2}\right], \quad [57]$$

where

$$(\sqrt{1-u_n^2})^* = \begin{cases} \sqrt{1-u_n^2};\ |u_n| \le 1 \\ -i\sqrt{u_n^2-1};\ |u_n| > 1 \end{cases}, \quad [58]$$

and u_n is any arbitrary function of n. The important point to note is that the product $(\sqrt{1-(n\lambda/2a)^2})^* \sin[2kh\sqrt{1-(n\lambda/2a)^2}]$ is always real. Therefore the right hand side of Eq. [57] is completely imaginary. Equating the real parts of right- and left-hand sides gives

$$-\cos_i + \mathrm{Re}\left[\sum_{n=-\infty}^{\infty} |B_n^e|^2 (\sqrt{1-\gamma_n^2})^*\right] = 0,$$

which is identical to Eq. [53]. The proof of energy conservation for the H-polarization is completely analogous and therefore will not be repeated.

Appendix 2

In this appendix we give the equations for the determination of $C_n^{e(h)}$ in the case of plane wave or focused spot diffraction.

The equations can be written in the general form

$$\sum_n [U_{mn}^{e(h)} + \delta_{mn} T_m^{e(h)}] C_n^{e(h)} = K_m^{e(h)}. \quad [59]$$

For plane wave diffraction,

$$U_{mn}^e = \sin[kh\sqrt{1-(n\lambda/2a)^2}] \sum_{p=-\infty}^{\infty} \sqrt{1-\gamma_p^2}\, Q_{mp}^* Q_{np}, \quad [60]$$

$$T_m^e = i\frac{a}{2d}\sqrt{1-(m\lambda/2a)^2}\cos[kh\sqrt{1-(m\lambda/2a)^2}], \quad [61]$$

$$K_m^e = 2Q_{m0}^* \cos\theta_0, \quad [62]$$

$$U_{mn}^h = -i\sqrt{1-(n\lambda/2a)^2}\sin[kh\sqrt{1-(n\lambda/2a)^2}]$$
$$\times \sum_{p=-\infty}^{\infty} \frac{P_{mp}^* P_{np}}{\sqrt{1-\gamma_p^2}}, \quad [63]$$

$$T_m^h = \frac{a}{2d}(1+\delta_{m0})\cos[kh\sqrt{1-(m\lambda/2a)^2}], \quad [64]$$

$$K_m^h = 2P_{m0}^*, \quad [65]$$

where γ_p, Q_{np}, P_{np}, and all the functions with square roots have their usual meanings as defined in the text. The summation over n on the left hand side of Eq. [59] extends from 1 to ∞ for the E-polarization and from 0 to ∞ for the H-polarization. We note that all the terms in Eq. [59] have simple physical meanings. From the definitions of Q_{np} and P_{np} it is clear that they are simply inner products, or overlap integrals, between the pth diffraction-order plane-wave state and the nth electric and magnetic channel modes, respectively. We can therefore regard them as representing the coupling strengths between these channel modes and the free-space plane-wave states. With this interpretation of Q_{np} and P_{np}, the meanings of $K_m^{e(h)}$ and $U_{mn}^{e(h)}$ become immediately clear. $K_m^{e(h)}$ is the driving term for the mth channel mode. $U_{mn}^{e(h)}$ gives the interaction between the mth and the nth channel modes. It is interesting to observe from the definition of $U_{mn}^{e(h)}$ that the nature of this interaction is second-order, i.e., indirect interaction via their mutual couplings to the plane wave states. Finally, $T_m^{e(h)}$ represents the admittance of the mth channel mode, defined as the ratio between the magnitudes of the tangential (parallel to the $y = 0$ plane) E and H fields of the mth channel mode.

From these interpretations of the various terms we can immediately generalize them to the case of focused spot diffraction by a single signal element:

$$U_{mn}^{e} = \sin[kh\sqrt{1 - (n\lambda/2a)^2}] \int_{-\infty}^{\infty} d\sigma\sqrt{1 - \sigma^2}\, Q_m^*(\sigma)Q_n(\sigma), \tag{66}$$

$$T_m^{e} = i\frac{a}{2\lambda}\sqrt{1 - (m\lambda/2a)^2}\cos[kh\sqrt{1 - (m\lambda/2a)^2}], \tag{67}$$

$$K_m^{e} = 2\int_{-\infty}^{\infty} A(\sigma,x_0)\sqrt{1 - \sigma^2}\, Q_m^*(\sigma)d\sigma, \tag{68}$$

$$U_{mn}^{h} = -i\sqrt{1 - (n\lambda/2a)^2}\sin[kh\sqrt{1 - (n\lambda/2a)^2}] \times \int_{-\infty}^{\infty} d\sigma \frac{P_m^*(\sigma)P_n(\sigma)}{\sqrt{1 - \sigma^2}}, \tag{69}$$

$$T_m^{h} = \frac{a}{2\lambda}(1 + \delta_{m0})\cos[kh\sqrt{1 - (m\lambda/2a)^2}], \tag{70}$$

$$K_m^{h} = 2\int_{-\infty}^{\infty} A(\sigma,x_0)P_n^*(\sigma)d\sigma. \tag{71}$$

References:

[1] W. R. Roach, C. B. Carroll, A. H. Firester, I. Gorog, and R. W. Wagner, "Diffraction Spectrometry for RCA VideoDisc Quality Control," *RCA Review*, **39**, p. 472, Sept. 1978 (this issue).

[2] A. H. Firester, C. B. Carroll, I. Gorog, M. F. Heller, J. P. Russell, and W. C. Steward, "Optical Readout of the RCA VideoDisc," *RCA Review*, **39**, p. 392, Sept. 1978 (this issue).

[3] I. Gorog, "Optical Techniques Developed for the RCA VideoDisc," *RCA Review*, **39**, p. 162, March 1978.

[4] J. D. Jackson, *Classical Electrodynamics*, p. 297, John Wiley & Sons, N.Y. (1962).

[5] J. W. Goodman, *Introduction to Fourier Optics*, p. 48, McGraw-Hill Book Co., N.Y. (1968).

[6] A. Hessel, J. Schmoys, and D. Y. Tseng, "Bragg-Angle Blazing of Diffraction Gratings," *J. Opt. Soc. of Am.*, **65**, p. 380 (1975).

[7] L. N. Deriugin, *Radiotekhnika*, **15**, p. 9 (1960).

[8] W. R. Roach and P. Sheng, private communication.

[9] R. W. Wood, "Films of Minute Metallic Particles," *Phil. Mag.*, **3**, p. 396 (1902); Rayleigh, "Dynamical Theory of Gratings," *Proc. R. Soc.*, **A 79**, p. 399 (1907); J. E. Stewart and W. S. Gallaway, "Diffraction Anomalies in Grating Spectrophotometers," *Appl. Opt.*, **1**, p. 421 (1962).

[10] K. Compaan and P. Kramer, "The Philips VLP System," *Philips Tech. Rev.*, **33**, p. 178 (1973); W. van den Bussche, A. H. Hoogendijk, and J. H. Wessels, "Signal Processing on the Philips VLP System," *Philips Tech. Rev.*, 33, p. 181 (1963); P. J. M. Janssen and P. E. Day, "Control Mechanism on the Philips VLP Record Player," *Philips Tech. Rev.*, 33, p. 190 (1973).

Patents Issued to RCA Inventors Second Quarter 1978

April

A. A. Ahmed Self-Starting Amplifier Circuit (4,085,359)
F. Aschwanden and T. E. Bart Secam Subcarrier Generator (4,084,177)
J. Avins Automatic Control of Free Wheeling (4,084,672)
A. L. Baker VideoDisc Player Mechanism Control System (4,086,617)
V. S. Ban and S. L. Gilbert Method for Chemical Vapor Deposition (4,082,865)
A. Bloom, D. L. Ross, L. K. Hung, and R. A. Bartolini Organic Volume Phase Holographic Recording Media Using Sucrose Benzoate (4,084,970)
A. G. F. Dingwall Method of Making a Substrate Contact for an Integrated Circuit (4,081,896)
F. C. Easter Video Digital Display Device with Analog Input (4,086,579)
J. Evans, Jr. Plural-Gun Cathode Ray Tube Having Parallel Plates Adjacent Grid Apertures (4,086,513)
I. Gorog and B. F. Williams Optical Filter (4,084,188)
J. N. Hewitt and V. Cristiano Method of Forming a Metal Pattern on an Insulating Substrate (4,083,710)
W. Hinn Video Amplifier Including an A-C Coupled Voltage Follower Output Stage (4,082,996)
K. Knop Zero-Order Diffractive Subtractive Filter Projector (4,082,438)
K. Knop Fabrication of Rectangular Relief Profiles in Photoresist (4,082,453)
M. A. Leedom VideoDisc Package (4,084,691)
M. L. McNeely and H. Rees Method for Producing Injection Molded and Centrally Apertured Disc Records (4,085,178)
A. Miller Optical Coupler (4,082,425)
L. Muhlfelder Roll/Yaw Body Steering for Momentum Biased Spacecraft (4,084,772)
J. C. Peer Circuit for Correcting Setup Error in a Color Television Receiver (4,084,115)
C. F. Pulse VideoDisc Package Having a Center Post (4,084,690)
L. N. Reed and J. R. Khalifeh Overcurrent Protection Circuit (4,084,070)
G. E. Schmidt, Jr. and L. Muhlfelder Magnetic Control of Spacecraft Roll Disturbance Torques (4,084,773)
T. D. Smith Switch Mechanism for a Calculator Type Keyboard (4,084,071)
J. Tults Digital Frequency Deviation Detector Useful in a Television Tuning System (4,084,127)
Z. Turski and A. Rosen Microwave Frequency Discriminator Comprising a One Port Active Device (4,085,377)
L. C. Upadhyayula Planar Transferred Electron Logic Device with Improved Biasing Means (4,086,501)
L. W. Varettoni Voltage Limiter Circuit (4,085,432)
J. C. Williams and W. B. Sisco Method and Detection of Phase and Frequency Modulation (4,085,367)
H. A. Wittlinger Voltage Monitoring Circuit (4,084,156)
O. M. Woodward Circularly-Polarized Antenna System Using Tilted Dipoles (4,083,051)

May

F. Aschwanden Master Oscillator Synchronizing System (4,092,672)
P. K. Baltzer Parallel Access Memory System (4,092,728)
R. L. Barbin Convergence Apparatus for In-Line Beams (4,091,347)
C. A. Berard Guard Circuit for High Impedance Signal Circuits (4,091,439)
S. Berkman, K. M. Kim, and H. E. Temple Si_3N_4 Coated Crucible and Die Means for Growing Single Crystalline Silicon Sheets (4,090,851)
R. H. Brader Scrambler and Unscrambler for Serial Data (4,087,626)
J. Dresner and B. Goldstein Magnesium Oxide Dynode and Method of Preparation (4,088,510)
J. Dresner and K. W. Hang Article with Electrically-Resistive Glaze for Use in High-Electric Fields and Method of Making Same (4,091,144)
M. Ettenberg Multi-Layer Reflector for Electroluminescent Device (4,092,659)
N. H. Farhat Electronically Steered Antenna System Using a Reflective Surface Formed of Piezoelectric Transducers (4,090,204)
F. C. Farmer, Jr. Electrical Continuity Test Apparatus Having a Forward Biased Diode Across the Test Terminals (4,088,947)
A. M. Goodman and C. E. Weitzel Method for Adjusting the Leakage Current of Silicon-on-Sapphire Insulated Gate Field Effect Transistors (4,091,527)

P. E. Haferl Pincushion Correction Circuit (4,088,931)
S. A. Harper Photographic Method for Printing Particle Pattern with Improved Adherence Utilizing Vanadates (4,089,687)
L. F. Heckman, Jr. Coaxial Cavity Microwave Oscillator with Manually Adjustable Capacitive Feedback Element (4,091,337)
E. P. Herrmann Charge Coupled Device with Diode Reset for Floating Gate Output (4,090,095)
M. V. Hoover Complementary Symmetry FET Mixer Circuits (4,090,139)
A. C. Ipri Silicon Implanted and Bombarded with Phosphorus Ions (4,092,209)
L. A. Kaplan Circuit for Reducing Ripple Voltage (4,092,609)
K. R. Keller Forming Patterned Polycrystalline Silicon (4,090,915)
M. L. Levene Creating a Closed Image From Segments (4,088,867)
W. N. Lewis Combination Glass/Low Temperature Deposited $Si_wN_xH_yO_z$ Passivating Overcoat with Improved Crack and Corrosion Resistance for Semiconductor Devices (4,091,406)
H. F. Lockwood and H. Kressel Stripe Contact Providing a Uniform Current Density (4,092,561)
S. Miko Ferroresonant Transformer Structure (4,088,942)
J. Ollendorf and F. J. Cestone Method of Performing Contactless Photolithography (4,088,406)
B. M. Pradal and P. L. Buess Frequency Linearization and Sensitivity Equalization of a Frequency Modulated Crystal Oscillator (4,088,968)
C. E. Profera Beam Forming Network (4,091,387)
P. M. Russo Priority Vectored Interrupt Using Direct Memory Access (4,090,238)
O. H. Schade, Jr. Amplifier Circuits (4,092,612)
F. N. Sechi Connection of a Plurality of Devices to a Circular Waveguide (4,091,334)
W. W. Siekanowicz, C. H. Anderson, and T. L. Credelle Flat Display Device with Beam Guide (4,088,920)
H. Sorkin Process for Filling Dynamic Scattering Liquid Crystal Cells (4,091,847)
M. Toda and S. Osaka Method of Producing Optical Image on Chromium or Aluminum Film with High-Energy Light Beam (4,087,281)
M. Toda and S. Osaka Surface Acoustic Wave Absorber (4,090,153)
C. C. Wang, T. C. Lausman, and R. F. Bates Method of Regenerating a Lead Monoxide Target Layer of a Camera Tube (4,090,758)
C. F. Wheatley, Jr. Voltage Reference Circuits (4,088,941)
C. F. Wheatley, Jr. Semiconductor Device Having Symmetrical Current Distribution (4,091,409)
R. P. Williams and M. A. Polinsky Combination Glass/Low Temperature Deposited $Si_wN_xH_yO_z$ Passivating Overcoat with Improved Crack and Corrosion Resistance for a Semiconductor Device (4,091,407)
J. P. Wittke Method of Aligning Optical Fibers with Light Emitting Optical Devices (4,090,777)

June

A. A. Ahmed Voltage Supply Regulated in Proportion to Sum of Positive- and Negative-Temperature-Coefficient Offset Voltages (4,095,164)
G. A. Alphonse Variable Acoustic Wave Energy Transfer-Characteristic Control Device (4,096,756)
F. Aschwanden Phase Locked Loop Tuning System with a Preset Channel Memory (4,097,810)
L. R. Avery Sync Separator Circuit (4,097,896)
G. W. Bain, A. G. Morris, and S. V. Forgue Infrared Sensitive Photoconductive Pickup Tube (4,097,775)
W. L. Behrend Amplitude and Delay Equalization of Surface Acoustic Wave Filters in Amplitude Modulation System (4,096,454)
R. R. Brooks and J. E. Wojslawowicz Direction Reversing Direct Current Motors and Their Control (4,095,155)
R. H. Buckley Ultraviolet Radiation Detector (4,096,387)
G. Denes Random Access-Erasable Read Only Memory Cell (4,095,281)
M. S. Fisher SCR Trigger Circuit (4,097,770)
L. A. Goodman Liquid Crystal Matrix Display Device with Transistors (4,094,582)
W. H. Groeneweg, L. A. Harwood, and A. V. Tuma Controlled Oscillator with Increased Immunity to Parasitic Capacitance (4,095,255)
P. E. Haferl Switched Vertical Deflection Circuit (4,096,415)
J. M. Hammer Optical Waveguide Coupler Employing Deformed Shape Fiber-Optic Core Coupling Portion (4,097,118)
W. J. Hannan Format for Color Diffractive Subtractive Filters (4,094,584)
F. Z. Hawrylo and H. Kressel Electroluminescent Semiconductor Device with Passivation Layer (4,095,011)
F. Z. Hawrylo and H. Kressel Metallized Device (4,097,636)
L. V. Hedlund Chrominance Signal Transcoding Apparatus (4,093,959)
M. L. Henley Vertical Deflection Circuit with Retrace Switch Protection (4,096,416)
W. Hinn Video Amplifier (4,096,517)

R. J. Hollingsworth Sense Circuit for an MNOS Array Using a Pair of CMOS Inverters Cross-Coupled via CMOS Gates Which are Responsive to the Input Sense Signals (4,096,401)
C. W. Horsting and W. B. Hall Liquid Crystal Device and Method for Preparing Same (4,095,876)
W. Kern and C. E. Tracy Combination Glass/Low Temperature Deposited $Si_wN_xH_yO_z$ Passivating Overcoat with Improved Crack and Corrosion Resistance for a Semiconductor Device (4,097,889)
C. C. Neil, R. A. Bartolini, and J. M. Hammer Optical Coupler Having Improved Efficiency (4,097,117)
S. E. Ozga Hardware Control for Repeating Program Loops in Electronic Computers (4,097,920)
K. D. Peters Disc Caddy and Disc Player System Therefor (4,093,152)
E. S. Poliniak Method of Transferring a Surface Relief Pattern from a Poly(1-Methyl-1-Cyclopropene Sulfone) Layer to a Non-Metallic Inorganic Layer (4,097,618)
M. A. Polinsky Monolithic Light Detector (4,096,512)
B. M. Pradal High Signal-to-Noise Ratio Negative Resistance Crystal Oscillator (4,096,451)
D. H. Pritchard Electronic Signal Processing Apparatus (4,096,516)
M. D. Ross Color Video Signal Processing Circuits (4,096,513)
J. E. Ruedy and G. A. Morton Image Tube Cathode (4,097,761)
K. M. Schlesier, J. M. Shaw, and C. W. Benyon, Jr. Method of Making a Sapphire Gate Transistor (4,097,314)
F. W. Spong Multilayer Optical Record (4,097,895)
A. V. Tuma, W. H. Groeneweg, and L. A. Harwood Average Beam Current Limiter (4,096,518)
A. V. Tuma and W. H. Groeneweg Standard/Nonstandard Internal Vertical Sync Producing Apparatus (4,096,528)
P. N. Yocum Rare Earth Activated Lanthanum and Lutetium Oxy-Chalcogenide Phosphors (RE29662)
J. P. Yu Video Record Player Switching System (4,097,899)

AUTHORS

Charles B. Carroll graduated cum laude from the University of Florida in 1960, with a B.S. degree in Mechanical Engineering. He joined the Astro-Electronics Division of RCA and designed sensor and camera components for spacecraft application. Also, environmental tests were designed and conducted using these devices. In 1966, Mr. Carroll received the MSME degree from Drexel University. In the same year, he transferred to RCA's David Sarnoff Research Center, where he worked, initially, on the development of facsimile printers for the Homefax project. He was instrumental in the mechanical design of a holographic recording and playback system. In 1973, Mr. Carroll transferred to the Palm Beach Division and worked on products using mini and microcomputers for controlling a motel. He rejoined the Research Center in 1975, where he worked on equipment for optical scanning and read out of the RCA VideoDisc. He received an RCA Laboratories Achievement Award for this work.

Arthur H. Firester received a B.A. degree, cum laude, with Honors in Physics from Brandeis University in 1962, and M.A. and Ph.D. degrees from Princeton University in 1964 and 1967, respectively. His doctoral dissertation research was on the modulation of light by optically pumped alkali metal vapors. Before joining RCA Laboratories, he was an Instructor at Princeton University for the academic year 1966–67. Dr. Firester's research at RCA Laboratories has been in the areas of nonlinear optical phenomena and coherent light optical problems, and their applications, lasers, holography, and VideoDisc optical systems. He was awarded a 1969 RCA Achievement Award for his work on parametric conversion of infrared to visible images, a 1975 RCA Achievement Award for his contribution to a team effort resulting in the development of a VideoDisc Defect Detector, and a 1977 RCA Achievement Award for his contribution to a team effort on optical recording and reading techniques compatible with the VideoDisc format. In 1973 he was a visiting scientist at the RCA Zurich Laboratories and in 1974–75 was a visiting Lecturer in the graduate physics department of Princeton University.

Dr. Firester is a member of the Society of Sigma Xi, the American Physical Society, and Optical Society of America and the IEEE.

J. James Gibson received the B.Sc. degree (Civilingenjor) in electrical engineering from the Royal Institute of Technology, Stockholm, Sweden and the M.Sc. degree (Teknologie Licentiat) from Chalmers Institute of Technology, Gothenburg, Sweden. From 1946 to 1952 he was employed by the Swedish Department of Defense engaged in research on antennas, microwave circuits and radar systems. As a fellow of the American Scandinavian Foundation he worked for a year and a half at RCA Laboratories, Princeton, New Jersey, where he contributed to the development of TV transmitting antennas and to transistor circuits for radio re-

ceivers. Upon his return to Sweden in 1954 he initiated and led a transistor circuits research group at the Royal Institute of Technology, where he also instructed in solid state circuits.

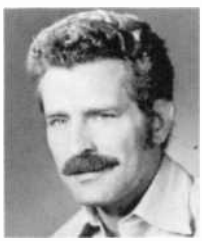
In 1956 Mr. Gibson joined the technical staff of RCA Laboratories where he has been involved in a variety of projects related to broadcast systems, solid state circuits, and consumer electronics including TV bandwidth compression, high speed computer circuits, integrated memories with field effect transistors, facsimile broadcasting, cable television systems, TV recieving antennas, holographic recording of television, and FM broadcast systems for stereophonic and quadraphonic sound. His current research activities are primarily related to the RCA VideoDisc system. For his contributions he has received three RCA Laboratories achievement awards and became a Fellow of the Technical Staff of 1969. From 1971 to 1973 he was also a lecturer in communication theory at LaSalle College, Philadelphia, and he is currently visiting professor in the graduate electrical engineering department at Rutgers University.

Mr. Gibson is a senior member of the IEEE, a member of the AAAS, the Audio Engineering Society, and the Society of Motion Picture and Television Engineers.

Istvan Gorog received the B.Sc. (1961), M.Sc. (1962), and Ph.D. (1964) degrees in Electrical Engineering from the University of California at Berkeley. In 1964, Dr. Gorog joined the technical staff of the RCA Laboratories, Princeton, N. J., where his main areas of interest have been quantum electronics and electro-optical systems. His research activities have included lasers and laser systems, holography, pre-recorded-video recording and playback techniques, displays, and investigation of the psychophysical aapects of electronic imaging. His current activities include product and process development related to the RCA VideoDisc, manufacturing instrumentation, and electrochromics. Dr. Gorog is Head of the Optical Electronics Research Group at RCA Laboratories.

During 1968 he was on leave of absence from RCA as a National Science Foundation Post-Doctoral Fellow at the Laboratori Nazionali di Frascati and the European Space Research Institute in Frascati, Italy, where he worked on the problems of production of high-temperature plasmas by laser irradiation of solid targets and scattering of laser radiation by collective plasma fluctuations.

Dr. Gorog is a member of the American Physical Society and of Eta Kappa Nu.



Macy E. Heller came to RCA Laboratories in 1958 after graduating RCA Institutes advanced electronics technology course. Mr. Heller contributed to the development of masers and lasers and the use of lasers in spectroscopy. Mr. Heller made the first RCA hologram and hologram master, and is the inventor of thermal holography. His optical work was associated with the RCA SelectaVision project. In 1971 he worked on the design of a system to detect faults in the mesh used in vidicon tubes. In 1976 Mr. Heller received an RCA Achievement Award for contributions to a team effort in the developments of optical recording and reading techniques for the RCA VideoDisc.

W. Ronald Roach received a B.S. degree in Engineering Physics from Cornell University in 1961 and a Ph.D. degree in Physics from the University of Illinois in 1966. He joined the RCA Laboratories as a Member of the Technical Staff in 1966, and since then his research interests have included non-linear optics, 2-photon spectroscopy, mode locking of lasers, holographic storage, and photoconductively controlled deformable mirror light valves. Recently he has been involved with the development of optical instrumentation for use in VideoDisc manufacturing.

Dr. Roach has received two RCA Laboratories Outstanding Achievement Awards and is a member of the American Physical Society and Sigma Xi.

John P. Russell received the B.Sc. in physics from Queens University, Belfast, N. Ireland in 1958 and the D. Phil. from the University of Oxford, England, in 1962. After a year of post-doctoral research at Oxford, he joined RRE Malvern in 1963, where he worked on light scattering in solids. In 1967 he joined RCA Laboratories, Zurich. There he continued his work on light scattering, making the first observation of raman scattering in solids with laser sources. Later he participated in the development of the Holocard system and color encoded F-I-H for micropublishing. In 1974, he transferred to RCA Laboratories, Princeton, N. J., where he has worked on the development of optical readers and recorders for the RCA VideoDisc system.

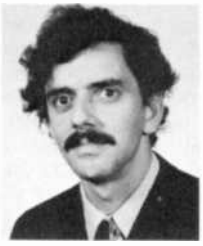
Dr. Russell received RCA Laboratories Achievement Awards in 1969 and in 1977. He is a member of the Institute of Physics (GB), the European Physical Society, and the American Optical Society.

Ping Sheng received his BSc degree in Physics (with honors) from the California Institute of Technology in 1967. He received his PhD in Physics from Princeton University in 1971. After spending two years at the Institute for Advanced Study, Princeton, N. J., as a visiting member of the School of Natural Sciences, Dr. Sheng joined the RCA David Sarnoff Research Center in 1973 as a Member of the Technical Staff. Dr. Sheng has done research on a variety of subjects. As a graduate student, his research effort was mainly directed toward polaron theory and ferromagnetic transmission resonance. As a result of his work on the transmission resonance, a new technique was developed for measuring the magnetic critical exponent β. While at the Institute for Advanced Study, Dr. Sheng did research on the theory of liquid crystals and developed a theory for electron transport in granular metals which can account for all the observed temperature and electric field dependences of the granular-metal dc conductivity. After joining RCA, Dr. Sheng continued his interest in liquid crystals and granular metals, while at the same time broadening his research effort into hologram noise analysis, Fourier optics calculations, contact problems of viscoelastic bodies, and rigorous diffraction calculations. Currently, he is engaged in research aimed at understanding electrical properties of inhomogeneous materials.

Dr. Sheng is a member of the American Physical Society.

W. C. Stewart earned the BSEE, MS, and PhD degrees at Duke University. In 1964, he joined a group at RCA Laboratories that was exploring the application of superconductivity to digital computers. His research encompassed various characterization studies of persistent-current devices and systems, and the Josephson tunnel junction. More recently, his activities have included research in optical holography, high resolution optical systems, and studies of television systems and devices. He was visiting scientist at RCA Laboratories in Switzerland in 1974.

Dr. Stewart has received four RCA Laboratories awards for outstanding achievement. He is a member of the Institute of Electrical and Electronics Engineers, the Optical Society of America, Sigma Xi, Phi Beta Kappa, Tau Beta Pi, and Eta Kappa Nu.



Robert W. Wagner received the B.S. degree in Electronic Physics from LaSalle College in 1976 and M.S. degree in Electrical Engineering from Drexel University in 1978. After several years at the research and development laboratory of Weyerhaeuser Corporation, he joined RCA Laboratories in 1976, where he worked on the development of instrumentation using laser techniques for optical inspection equipment involved in the RCA VideoDisc project. In 1978 he joined the engineering staff of Metrologic Corporation, where is is working on high voltage switching problems involved with laser power supply design, laser power modulation circuitry, and product development.