OPTICS OF REFLECTIVE VIDEO DISC PLAYERS

Leonard J. Laub
Zenith Radio Corporation
1900 N. Austin Ave.
Chicago, Illinois 60639

Optics and television have come together in many ways, both in straightforward image-forming applications and, more recently, in "invisible" applications where the light handled by the systems doesn't come from natural sources and isn't received by human eyes. This second class includes not only process applications such as CRT phosphor screening and IC resistor trimming, but television signal recording applications as well. In this discussion, the word "signal" must be emphasized. There is a long history of the recording of television images, from "kinescopes" made by photographing a CRT faceplate up through the Image Transform system using a raster-scanning electron beam writing directly on film, where the intent has been the production of a projectible movie. Today we are concerned with systems which play back through a standard television receiver. It is possible to record two-dimensional images and then scan them back to produce a one-dimensional electrical comprehensible to the television set; this has been done in the EVR system of CBS with images written directly on film and in both the Holotape system of RCA and a recent disc system shown by Hitachi in which Fourier holograms are reconstructed to yield images. One basic problem of all such systems is the need to scan the recovered image; this requires the presence of flying spot scanners or vidicons in all players and is thus economically and technically unattractive. By concentrating on the television signal and not forming an image until the signal is displayed on the user's TV set, several groups around the world have realized great improvements in simplicity and performance of home video players.

In most of the current video disc systems, the one dimensional TV signal, in FM-encoded form, is placed along a spiral track wound tightly over the surface of a disc. The diversity of video disc players lies mostly in the method of pickup. Optics enters here in a completely "invisible" way, but with clear advantages over other approaches to signal pickup. Contact pickup systems require a carefully formed track containing carefully formed signal elements. All this care requires complex equipment, great skill and a lot of time. Electron beams or mechanical cutters must be employed to make the tight geometries. In contrast, recording machines for optical video disc systems are themselves optical, and in fact resemble scaled-up players in using a laser whose beam is focused onto the spinning disc. Real time operation is the rule in these machines. Contact systems have simple players by virtue of the engagement of the groove by the stylus; this simplicity is offset by the problems of stylus and disc wear and of limited flexibility of program display, and by the requirement that the information-bearing surface must be exposed and thus vulnerable to scratches and dirt. Optical systems read without physical contact or even close proximity; it is common practice for the information-bearing surface to lie more than a millimeter below the nearest optically important exposed surface. These advantages of optical systems come with one basic disadvantage; the absence of tangible engagement of the disc by the pickup necessitates the use of servo or other control devices for maintenance of tracking and focus. This paper will discuss the basic principles of signal pickup and of tracking and focus detection in reflective-mode optical video disc players.

Received July 1, 1976

Signal pickup deserves first attention. It is desirable to form a small reading spot to probe the disc. How small must this spot be? Given the practical constraint of a spiral track which contains exactly one TV frame per turn from inside to outside, we must deal with the worst case, which occurs at the inside (Figure 1). A practical carrier frequency for

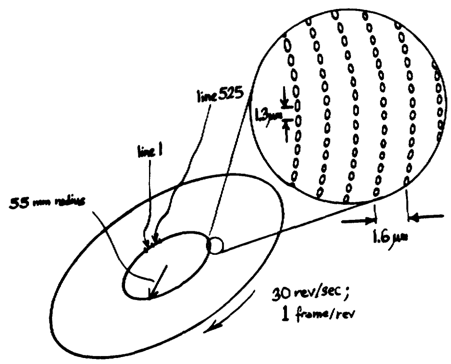


Figure 1

the FM-encoded video signal is 8 MHz; in one 33 msec frame, there are 267,000 cycles of the carrier. At the inside of the disc, all these cycles must fit around a track of 345 mm circumference, which gives the recorded carrier a wavelength of 1.3 μm. The practical requirement of thirty minute playing time requires a track-to-track spacing of 1.6 μm, so it is already clear that we need a more or less round spot whose diameter is no more than one micrometer (about forty microinches). Now we must examine how small spots of light are formed to find out the practical consequences of this requirement.

Much of practical optics falls under the heading "geometrical"; that is, light is assumed to travel in rays of zero transverse dimension, and focusing consists of causing all rays in a beam to intersect in a point (Figure 2). For many types of optical instruments, such as cameras and magnifiers, geometrical

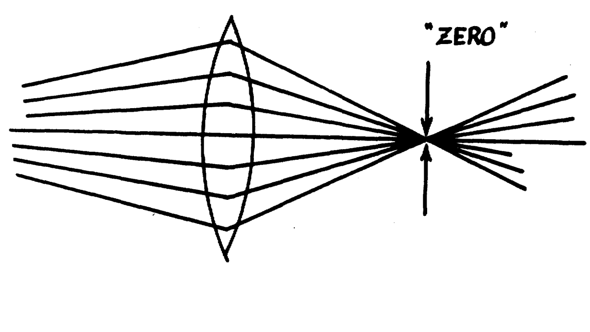


Figure 2

optics provides a satisfactory description, since the faults, or aberrations, of these instruments are large enough to be represented by imperfect intersection of rays traced through the instruments. In instruments with small aberrations, such as microscopes or good telescopes, ray tracing suggests that all the rays cross in so small a region as to approach zero size. Unfortunately, actual instruments do not show the enormous resolution predicted by geometrical optics. Even a superbly corrected microscope objective, whose aberrations are negligible, cannot produce a spot smaller than about one-half the wavelength of the light used (Figure 3). Furthermore, this spot is not a

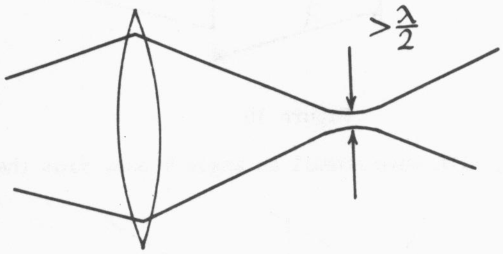


Figure 3

sharply bounded, uniformly bright circle; instead, it is concentrated in a small diameter but has a faint structure running out a great distance from the center. For the case of a perfect lens with a uniformly illuminated circular aperture, the spot is modeled by the Airy disc (Figure 4). Small changes in the shape or

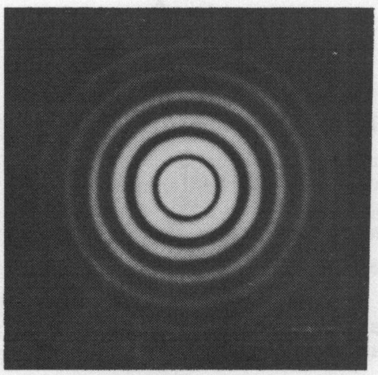


Figure 4

illumination of the entrance aperture produce large changes in the structure of the spot, as do small axial excursions from the plane of best focus (Figure 5).

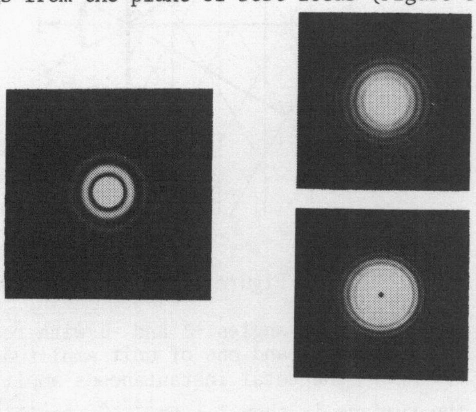


Figure 5

It is mathematically possible to apply the calculated or measured spot cross-section directly to the structure to be probed to predict what will be read, but this approach of convolution is quite laborious and masks some general principles. The convenient alternative results from the characterization of optical instruments as linear systems. Instead of temporal frequency, we use spatial frequency to lay out spectra and responses. If, as shown in Figure 6, many identical characters are spaced uniformly at intervals of

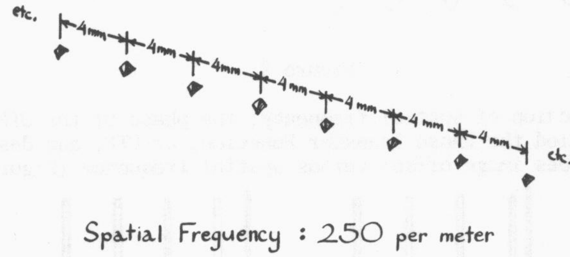


Figure 6

four millimeters along a line, we can describe this train by saying that characters occur at the rate of two hundred fifty characters per meter. This rate is also called the spatial frequency of the carrier train; the Elite type in the author's typewriter has a spatial frequency of twelve characters per inch. High quality photographic lenses can throw images containing structures with spatial frequencies up to eighty or one hundred cycles per millimeter; the very finest structures that can be seen through microscopes using visible light have spatial frequencies below five thousand cycles per millimeter. All of those figures represent spatial spectra, with two examples of single-frequency carriers followed by two examples of low-pass frequency responses. In general, the complex frequency response of an optical system which responds to light intensity or power is called its Optical Transfer Function, or OTF. The OTF of a system gives a complex coefficient describing magnitude and phase transmission by the system of a sine-wave object of arbitrary spatial frequency. The operation of an optical system is predicted by multiplying the spatial frequency spectrum of the object by the OTF of the system (Figure 7); the product is the

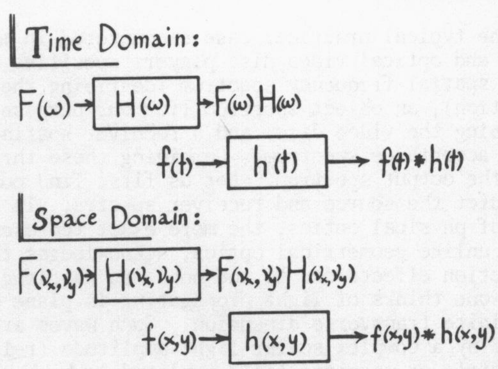


Figure 7

spatial frequency spectrum of the image. The magnitude of the OTF is called the Modulation Transfer Function, or MTF (Figure 8), and describes the transfer of contrast from object to image as a

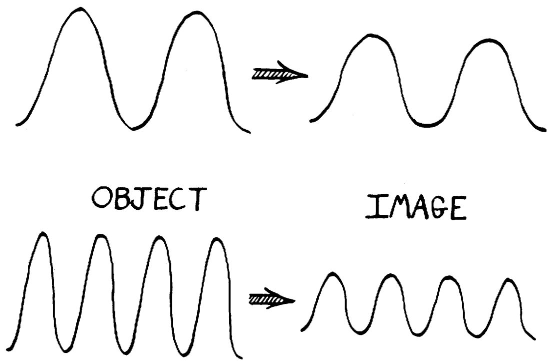


Figure 8

function of spatial frequency; the phase of the OTF is called the Phase Transfer Function, or PTF, and describes image offset versus spatial frequency (Figure 9).

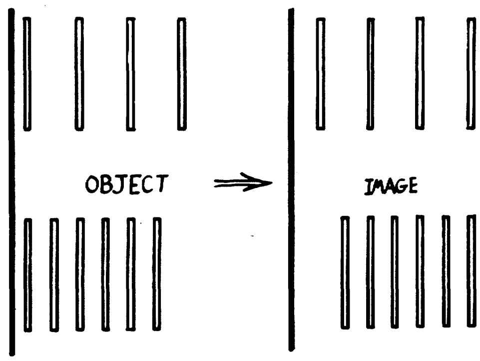


Figure 9

All of this is based on the principle that any reflecting or transmitting object, as well as any image, can be represented by or decomposed into a spectrum of sine and cosine waves or of complex exponentials. This is just the same Fourier analysis used to describe the time functions used in electronics, but the formalism is changed so that space, instead of time, is the input and output variable; the space of objects and images has two dimensions compared to the one of electrical signals. The basic work leading to a spatial Fourier analysis of optical systems has come in the last thirty years, both from the work of physicists, led by Duffieux, and of electrical engineers, led by Schade (see references).

The typical practical case encountered in microscopes and optical video disc players specifies a source spatial frequency spectrum (desribing the illumination), an object spectrum (for our purposes, describing the video disc) and a receiver spatial frequency acceptance spectrum. Combining these three gives the output spectrum. Let us first find out how to predict the source and receiver spectra. In the world of physical optics, the more exact representation which, unlike geometrical optics, acknowledges the diffraction effects due to the non-zero wavelength of light, one thinks of light propagating in plane waves of infinite transverse dimension. Such waves are described by a complex scalar light amplitude (related to electric or magnetic field amplitude) which varies longitudinally as $e^{i(-kr+\omega t)}$ and transversely not at all (Figure 10). Thus, at any instant, a surface receiving a normally incident plane wave has uniform complex amplifude everywhere over its extent. If a plane wave falls obliquely onto a surface, as shown in

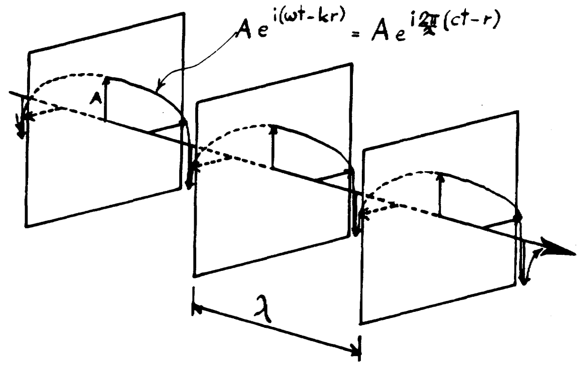


Figure 10

Figure 11, with wave normal an angle θ away from the

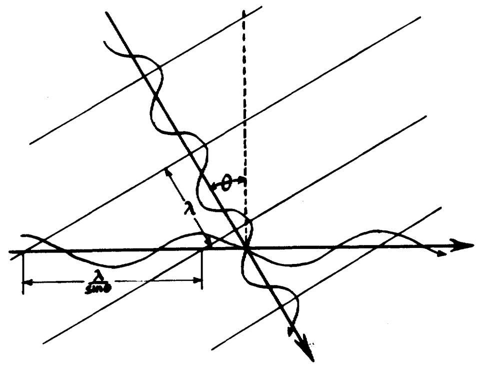


Figure 11

surface normal, the instantaneous complex light ampliude on the surface varies as $e^{ikx \sin\theta}$, where $k = \frac{2\pi}{\lambda}$, with λ the wavelength of the light used, and x is distance along the surface following the projection of the wave propagation vector. To simplify matters, we may define the quantity $\nu = \frac{\sin\theta}{\lambda}$ as the spatial frequency of the projection of the plane wave onto the surface. To exercise this idea, let us send three plane waves, as shown in Figure 12, onto a surface;

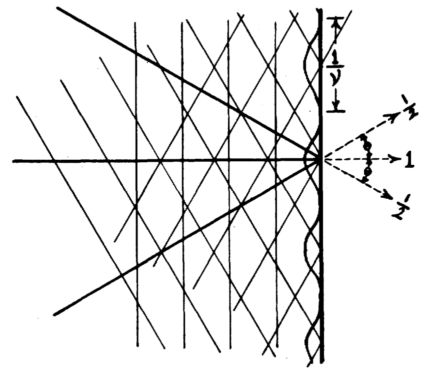


Figure 12

two of amplitude ½ at angles +θ and -θ with respect to the surface normal and one of unit amplitude normal to the surface. The total instantaneous amplitude across the surface is then $\underbrace{1 + \frac{1}{2}e^{i2\pi\nu x} + \frac{1}{2}e^{-i2\pi\nu x}}_{= 1+\cos 2\pi\nu x}$ with ν as defined above. Alternatively (Figure 13), if we construct a screen with amplitude transmittance

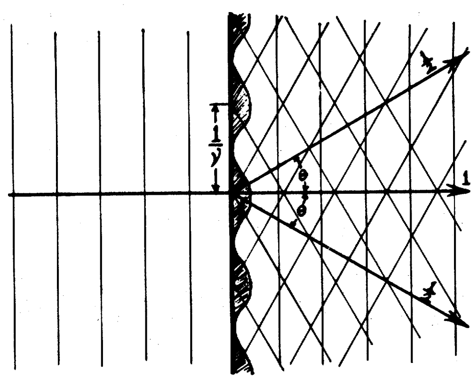


Figure 13

$\frac{1}{2}(1 + \cos 2\pi\nu x)$ and illuminate it with a normally incident plane wave so that the instantaneous amplitude immediately before the screen is 2 everywhere, the light amplitude immediately after the screen is $1 + \cos 2\pi\nu x$ which can be decomposed into $1 + \frac{1}{2}e^{i2\pi\nu x} + \frac{1}{2}e^{-i2\pi\nu x}$ which in turn may be seen to represent three plane waves, one of unit amplitude propagating normally and two of amplitude $\frac{1}{2}$ propagating at angles $+\theta$ and $-\theta$. This mathematical result describes the phenomenon of diffraction, which may be observed easily by placing a grating in a laser beam. The direct connection of spatial frequency and angle of inclination of the propagation vector of a plane wave continues to the specification of an illuminating or receiving optical device by its numerical aperture, defined by $NA = \sin\theta_{max}$, where θ_{max} is the inclination of the most oblique plane wave sent or received by the device (Figure 14). If the device is free of

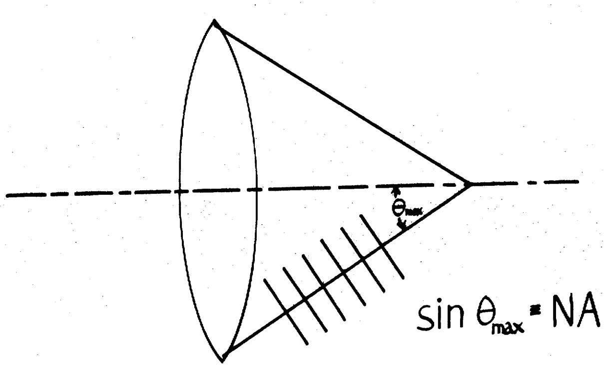


Figure 14

aberration and its aperture uniformly illuminated, its spectrum is uniform from zero spatial frequency to $\nu_{max} = \frac{NA}{\lambda}$. In the context of a microscope or a video disc player, the basic practical question now becomes: what is the response of the system of a given illuminator and a given receiver, said receiver understood to be feeding film or an eye or a photodetector, all devices which respond to light power or intensity, represented by the squared modulus of the complex amplitude.

In the particular case of an illuminator with numerical aperture NA_i and a similar receiver with NA_r, the answer comes from recognizing that only plane waves propagating within the acceptance angle of the receiver will get into it and be used. If, as shown in Figure 15, NA_i were zero (normally incident plane wave illumination), all object spatial frequencies up

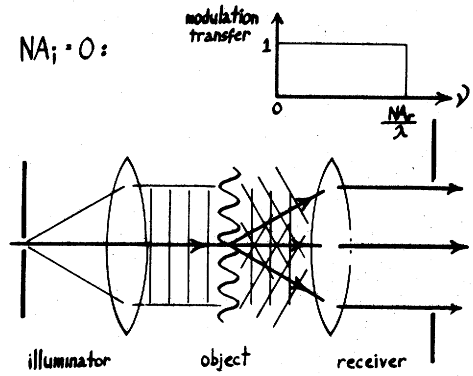


Figure 15

to NA_r/λ would be received, and all higher object spatial frequencies would be rejected. The complementary case of $NA_r = 0$ (scanning focused spot and small on-axis photodetector, as shown in Figure 16) accepts

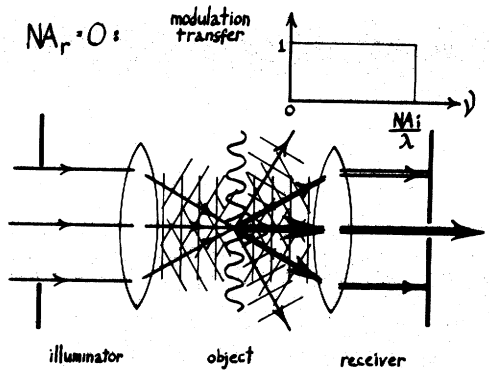


Figure 16

spatial frequencies up to NA_i/λ and no higher. Away from either of these extremes, we find that the object is illuminated with a broad, uniform spectrum of plane wave whose projections have spatial frequencies up to NA_i/λ and that the receiver accepts plane waves whose projections have spatial frequencies up to NA_r/λ (Figure 17). Each of the incident plane waves acts

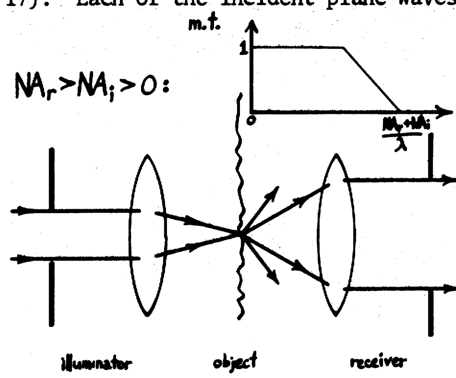


Figure 17

as a carrier for the object; the result is that object spatial frequencies up to $\frac{NA_i + NA_r}{\lambda}$ send plane waves into the receiving aperture, as spatial modulation sidebands of the plane waves in the illuminating spectrum. Low object spatial frequencies send more plane waves into the receiving aperture than do high frequencies, so the system response is not uniform, but falls with increasing frequency. Regardless of intermediate frequency behavior, however, the system response cuts off absolutely at an object spatial frequency of $\frac{NA_i + NA_r}{\lambda}$; there is no response at all to higher frequencies. In the very particular case of $NA_i = NA_r = NA$, the system response falls gradually throughout the intermediate range of spatial frequencies and cuts off at an object spatial frequency $2\ NA/\lambda$ (Figure 18). This case describes most metallurgical microscopes and, of current importance, most reflective video disc players, where the illuminator

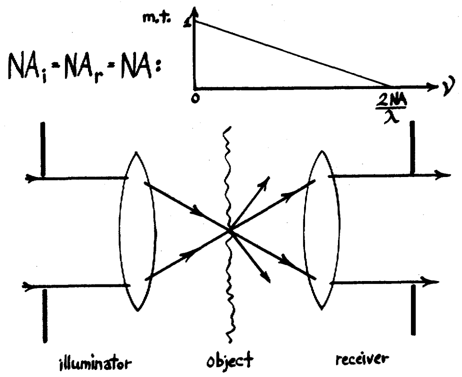


Figure 18

and the receiver are the same objective lens. In general, sweeping the object spatial frequency to measure the system response appears physically to move an image or projection of the illuminator's aperture over the receiver's aperture (Figure 19); the mathematical expression of this is that the OTF of the

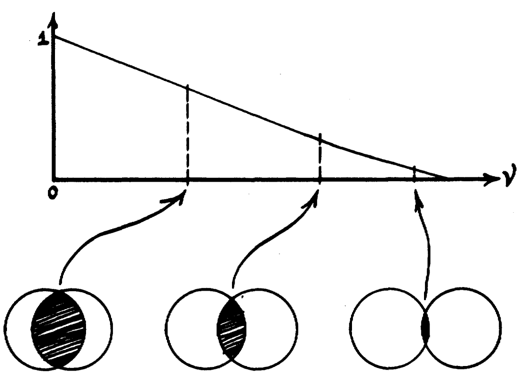


Figure 19

illuminator-receiver system is the cross-correlation of the two apertures, or the autocorrelation in case the two apertures are the same. The round, uniformly lit, practically unaberrated equal apertures of the reflective video disc player case give a tent- or cone-shaped OTF, peaked at zero spatial frequency and bounded by a zero-response circle at $\nu = 2\ NA/\lambda$ (Figure 20).

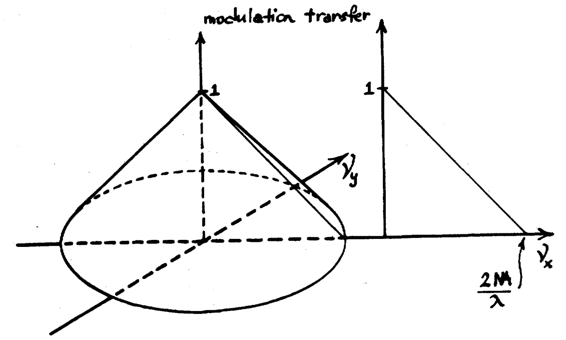


Figure 20

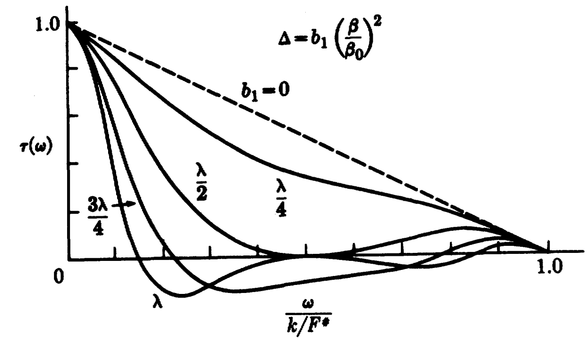


Figure 21

The effect of defocusing is described by applying a quadratic (more accurately, spherical) phase factor to either or both apertures. Small defocus produces phase rotations in the OTF; larger defocus produces nulls of response at intermediate spatial frequencies (Figure 21). Larger apertures show these effects with shallower phase curvatures; i.e., with smaller absolute amounts of devocus. For the case of round, uniform, equal apertures, defocus of $\pm \frac{\lambda}{2(NA)^2}$ takes the system halfway to the first null of response; this range is frequently referred to as the depth of focus of the system. If we denote the depth of focus by $d = \frac{\lambda}{(NA)^2}$ and the cutoff spatial frequency by $\nu_c = \frac{2\ NA}{\lambda}$, then we can derive the practical relationships $d = \frac{4}{\lambda(\nu_c)^2}$, which shows how the depth of focus drops rapidly with increasing cutoff frequency, and $\nu_c = \frac{2}{\sqrt{\lambda d}}$, which shows how the maximum useful cutoff frequency falls as focus error or variation increases (Figure 22).

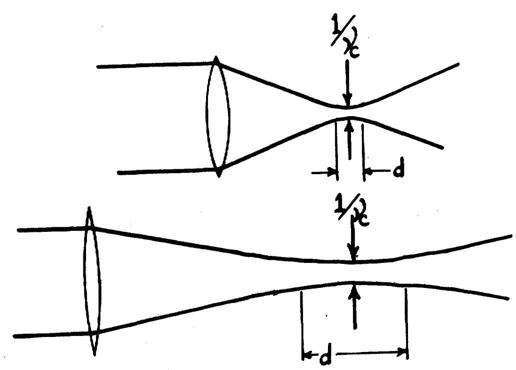


Figure 22

All of the foregoing has omitted any description of the object, for the very good reason that, in the terms used above, any optical system is linear and has a response (OTF) which is independent of the object. However, now that we know how to predict the OTF, it is a good time to characterize a typical object to be used in the system of interest. Most video discs are phase objects; that is they have pits or bumps which locally shift the phase of reflected light without changing the intensity of the light. Since these objects do not locally absorb light, it is not immediately clear that an image perceivable to an intensity detector can be formed of such an object. It turns out to be difficult to avoid forming such an image. Any corrugation of a reflecting surface will cause the reflected light to fan out over a larger angle than did the incident light (Figure 23); the simple result of this is that light will be lost, and the image of the corrugations will be dark. For reflective video discs, the optimum corrugation is deep enough to produce a local phase shift of an odd multiple of π radians, and as wide as one-half the period of the cutoff spatial frequency. In the actual working system, these optimum structures are pits about 0.12 µm deep and 0.5 µm wide. All this is intended to make these phase objects look dark against a light background, so that the player sees the record as if it were a density or photographic object. This is handy, as it permits the use of actual density objects without modification of the player.

Reflective video discs show their greatest utility when they are read through the body of the disc (Figure 24a); this mode of use puts the information-bearing surface out of harm's way, a distance

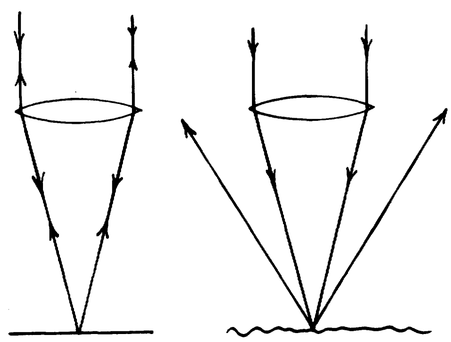
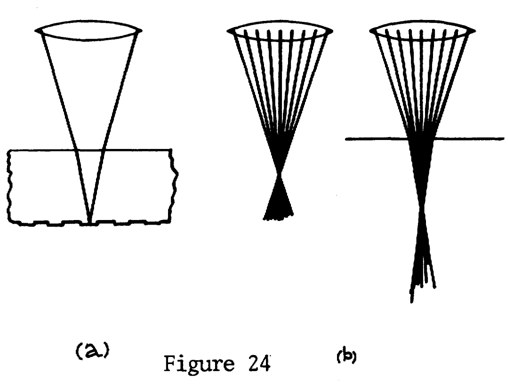


Figure 23

many times the depth of focus from the nearest optically effective exposed surface. This has the benefit of holding dirt and scratches out of focus to reduce or remove their effect on signal playback; it has the drawback of requiring compensation elsewhere in the optical system. Focusing a beam of light through a slab of refracting material makes the point of focus move axially (Figure 24b) and introduces spherical aberration. Above a certain tolerance limit, this aberration degrades the focused spot enough so that



(a) Figure 24 (b)

the signal is no longer useful. The objective lens can be corrected to compensate completely for any disc thickness, but the tolerance limit (±0.1 mm) at the numerical aperture typically used in video disc players is less than the range of disc thicknesses proposed in the working standards. A lens designed for use with the "rigid" standard disc (1.1 ± 0.1 mm) is measurably off when used with the "flexible" standard disc (0.2 ± 0.1 mm); this can be overcome by inserting extra optical path length when playing the flexible disc, for instance by supporting the flexible disc on a 0.9 mm spacer of smooth, clear plastic.

At this point, we can finally draw the curtain on an actual reflective-mode optical video disc player (Figure 25). The heart of the player resembles a metallurgical microscope in that light is

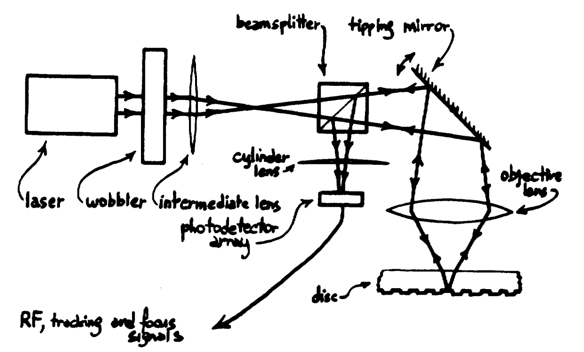


Figure 25

brought to the object through the same objective lens used to examine the object. Typically, a simple lens is used to diverge the collimated (parallel) beam from the laser; downstream an objective lens, which can resemble a microscope objective in its optical design, brings the beam to focus on the disc. The numerical aperture of this lens is usually 0.4; along with the 0.633 µm wavelength of the laser light, this leads to a spatial frequency cutoff of 1,264 cycles per millimeter, or about 13.1 MHz at the innermost (55 mm radius) track of a disc spinning at 1800 rpm. Included in the path between the first and second lenses are a controllable tipping mirror and a beamsplitter. The beamsplitter performs the basic function of separating light returning from the disc and light heading toward the disc. The continuation of the return path past the beamsplitter proceeds through a cylinder lens and a photodetector array. Before the beamsplitter in the incident path may be found a low-amplitude beam deflector, known as a "wobbler". The laser, the two lenses, the beamsplitter and one or two sections of the photodetector array are needed to pick up the signal; all the rest of the components listed perform functions needed because of practical imperfections in the disc.

One such practical imperfection is departure from perfect flatness. For a typical player, the acceptable depth of focus is less than 4 µm. Typical pressed discs may be wavy by ½ mm (500 µm) or so, and thus begins the need for a focus servo. To operate a focus servo, we need a bipolar focus error signal. Ideally, this signal should be derived optically from the same beam which carries the signal back from the disc. A method to do this, invented at Thomson-CSF, dictates the arrangement of the cylinder lens and the photodetector array (Figure 26). The first lens in the optical system produces an intermediate spot which is imaged at a reduction onto the disc. When the system is in focus, the returning light will attempt to retrace the path of the incident light and re-form the intermediate spot.

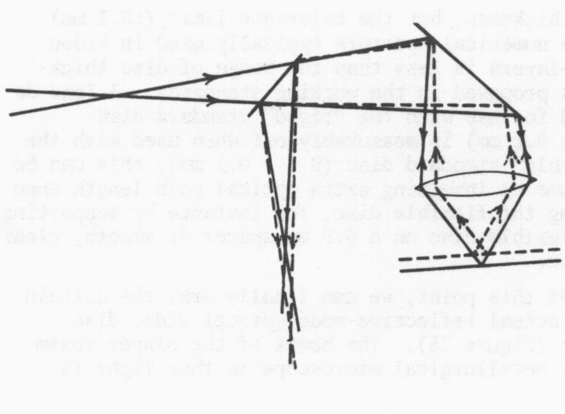


Figure 26

If the system is out of focus, this image of the intermediate spot will appear somewhere upstream or downstream of its "in focus" position. This displacement is a sensitive function of disc displacement: 1 μm motion of the disc displaces the image of the intermediate spot about 1.3 mm. With the introduction of the cylinder lens after the beamsplitter (Figure 27),

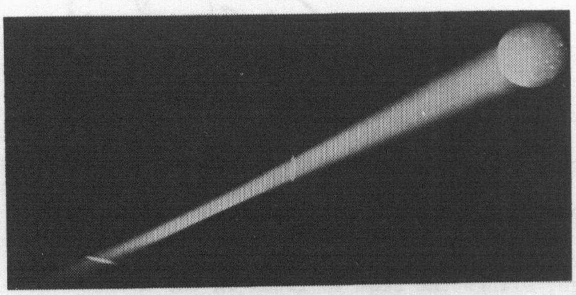


Figure 27

this image of the intermediate spot becomes astigmatic (not so the spot which reads the disc; there are no cylinder lenses in the incident path); this means that the small round image spot is replaced by two elongated line images, perpendicular to each other, one ahead of and the other behind the original round image. As the disc moves up and down through focus, the photodetector array intercepts differently shaped beam cross-sections, from linear horizontally through diamond-shaped to linear vertically. The use of an appropriately matrixed quadrant array permits shape detection so as to yield the desired bipolar error signal (Figure 28). This signal, suitably equalized and amplified, goes to the servo motor which moves the objective lens.

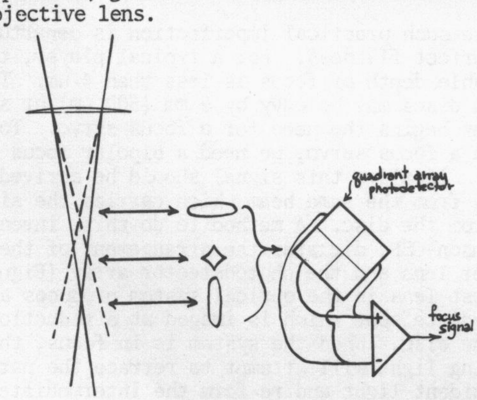


Figure 28

Another practical imperfection of typical discs is eccentricity and distortion of the spiral track. The tracks are spaced apart by 1.6 μm, and the light spot must be no more than two or three tenths of a micrometer off track center to read the signal properly, but typical discs may show ±125 μm eccentricity and distortion. Again, we need a servo, which in turn needs bipolar error detection. One attractive solution to this problem gives rise to the "wobbler". With its assistance, the light spot is "wobbled" radially sinusoidally with an amplitude of one or two tenths of a micrometer at a frequency of a few tens of kilohertz (Figure 29). This wobble causes amplitude

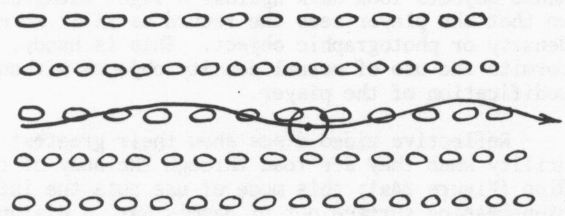


Figure 29

modulation of the "DC" and "RF" signals from the photodetector. Either of these AM signals can be detected synchronously with the signal driving the wobbler. When the wobble pattern is centered on the track (no tracking error; Figure 30a), the AM is

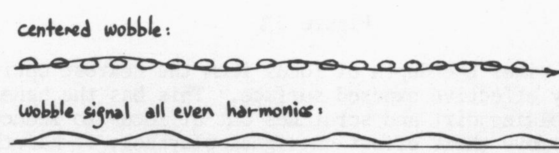


Figure 30a

purely even harmonics of the wobble frequency; offset of the wobble pattern (Figure 30b) gives a fundamental component whose sign tells the direction of the

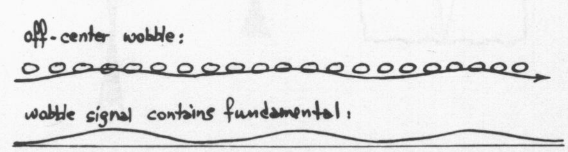


Figure 30b

tracking error and whose magnitude tells the amount of the tracking error. The synchronous detector represents these cases by a positive or negative DC output which controls the radial axis of the tipping mirror so as to keep the spot on the track. The wobble amplitude is intended to be small enough to have no degrading effect on signal pickup or on rejection of crosstalk from neighboring tracks.

One more practical imperfection of discs is the non-uniformity of the time base of the recovered signal. This is due to a combination of disc distortions and spin motor errors, and is an equal plague to all video playback methods. Optical playback systems having enough field of view under the objective lens to allow for radial tracking have the same field of view in the tangential direction. This permits the use of a tipping mirror, either similar or identical to that used for radial tracking, to run the spot back and forth along the track to regulate the instantaneous track velocity relative to the spot. Excellent time base correction is achieved this way.

The partnership of electronics and optics, from the theory of modulation transfer to the actuality of tracking and focus servo loops, has produced a system in which a tiny focused spot of light with no effective mass becomes routinely useful as a massless, inertialess, non-wearing stylus. The technology is advanced, but concentrated effort has made it quite practical.

References:

1. Duffieux, P.M.: L'Intégrale de Fourier et ses Applications à L'Optique, Faculté des Sciences, Besançon, 1946.

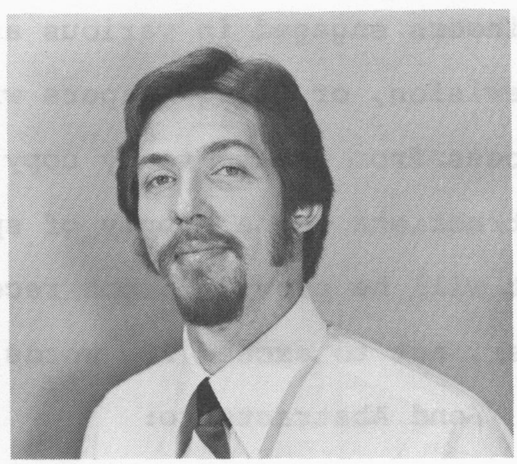
2. Schade, O.H.: Electro-Optical Characteristics of Television Systems, RCA Review, IX, 5 (part I), 245 (part II), 490 (part III), 653 (part IV) (1948).

3. O'Neill, E.L.: Introduction to Statistical Optics, Addison-Wesley Publishing Co., Inc. Reading, Mass., 1963.

4. Goodman, J.R.: Introduction to Fourier Optics, McGraw-Hill, New York, 1968.

BIOGRAPHY

LEONARD J. LAUB serves as Section Manager, Electro-Optics in the Research Division of Zenith Radio Corporation in Chicago. He has been with Zenith since 1965, when he started as a Physics Co-Op student. Since then he has worked in optical system design, signal and image processing and recording, holography, acousto-optics, ultrasonics, magnetic recording and acoustics. He holds the B.Sc. degree in Physics from Illinois Institute of Technology, is a member of the Optical Society of America and holds four patents.



L. J. Laub