IMPROVED SYNTHESIS AND COMPUTATIONAL METHODS FOR COMPUTER-GENERATED HOLOGRAMS

by

James R. Fienup

IMPROVED SYNTHESIS AND COMPUTATIONAL METHODS FOR COMPUTER-GENERATED HOLOGRAMS

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF APPLIED PHYSICS

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

James Ray Fienup

May 1975

ABSTRACT

The principal goal of computer holography is the synthesis of a transparency that can modulate an optical wavefront according to a computed complex Fourier transform. A major problem has been that it is necessary to introduce spurious images into the reconstruction when using existing holographic materials, which can modulate only the intensity or only the phase of an incident wavefront. This dissertation describes a solution to this problem. By the use of multi-emulsion color film, a transparency can be synthesized that has controlled complex transmittance. Independent control over the amplitude and the phase is accomplished by absorption in one layer and thickness variations in another. Upon reconstruction, a single on-axis image is formed, without spurious images and with the maximum possible diffraction efficiency. This new kind of computer-generated hologram has other highly desirable properties, as well. The synthesis and calibration procedures necessary to gain complex control of the material are discussed in detail and experimental results are shown.

The properties and limitations of Kodachrome film as a holographic material are investigated. The modulation transfer function, phase response as a function of spatial frequency, dynamic range, and scattered

flux spectrum are studied. Various cross-talk effects between the different layers of the film are analyzed, and methods of compensating for cross-talk are presented.

Another major problem in computer holography is that the display devices and materials used are often limited in the number of gray levels and resolvable elements that can be achieved. Consequently, it is often not possible to represent any arbitrary complex Fourier coefficient, and quantization noise results. A new iterative technique is described that recomputes the hologram in such a way as to substantially reduce quantization noise in the reconstruction. Modified versions of this iterative method can be used to solve a number of other problems in optics as well, including spectrum shaping and the phase retrieval problem.

Full-color images can be obtained by combining the images from three color-separation holograms, one for each of the primary colors, each illuminated by its respective color of coherent light. False color images can be eliminated by some additional complexity in the optics of the reconstruction setup. However, a new method is described in which the holograms themselves reject the false color images by the use of color film. Both the wavelength-selective absorption of the different layers and phase effects in color film are used to eliminate false images.

Acknowledgements

I am indebted to Professor Joseph W. Goodman, my thesis advisor, for his support and encouragement of this work and for his introducing me to the rich field of coherent optics.

I am equally indebted to my fellow students at Stanford who generated a rich intellectual atmosphere and assisted me in both work and play. Foremost was David Chu, co-inventor of the ROACH, who provided many stimulating ideas. Also deserving mention are Robert Powers, Hisatoyo Kato, and Kalyan Dutta.

I am grateful for the encouragement and help of my wife, Patty, who typed the rough draft of this thesis. I appreciate the excellent job of typing this manuscript by Mrs. Judy Clark.

I appreciate the help of Professor Albert Macovski and thank him and Professor Calvin Quate for serving as readers for this thesis.

I appreciate the financial support of the National Science Foundation, of the Office of Naval Research, and of the Joint Services Electronics Program, and the generosity of Dr. William Streifer of Xerox Corporation for duplicating the page of color figures in Chapter 4.

TABLE OF CONTENTS

TABLES

TABLE OF FIGURES

CHAPTER 1

INTRODUCTION

Suppose that an object is illuminated by coherent light and that the light transmitted or reflected to the hologram plane is described by a complex-valued wavefront $F(u,v)$. The goal of holography [Gabor, 1948, 1949] is to produce a transparency that has at least one component of its complex transmittance proportional to $F(u,v)$. The classic way to produce such a hologram transparency is to expose a piece of photographic film to the interference of the wavefront $F(u,v)$ with an offset reference plane wave [Leith and Upatnieks, 1962, 1963]. Then, with the object absent, upon illuminating the processed hologram with a plane wave, the transmitted wave is modulated by the hologram to reconstruct a wavefront proportional to $F(u,v)$. Therefore, if we look through the hologram, it appears as though we were looking at the object itself.

In computer holography, the object need not exist physically; we need only a mathematical description of the object from which we can calculate $F(u,v)$ on a digital computer. For example, if a lens is inserted between the object and the hologram so that the object is in the front focal plane and the hologram is in the back focal plane of the lens, then there is a Fourier transform relationship between the wavefront $F(u,v)$ in the hologram plane and the object $f(x,y)$ [Goodman, 1968, Ch.5],

$$F(u,v) = \frac{1}{\lambda f} \int_{-\infty}^{\infty}\int f(x,y)\, e^{-i\frac{2\pi}{\lambda f}(ux+vy)}\, dxdy \qquad (1\text{-}1)$$

and the inverse relationship is given by

$$f(x,y) = \frac{1}{\lambda f} \int_{-\infty}^{\infty}\int F(u,v)\, e^{+i\frac{2\pi}{\lambda f}(ux+vy)}\, dudv \qquad (1\text{-}2)$$

where λ is wavelength of monochromatic light illuminating the object. We use

$$f(x,y) \supset F(u,v) \qquad (1\text{-}3)$$

as an abbreviated form of (1-1) and (1-2). After computing the wavefront $F(u,v)$ that will yield the desired image $f(x,y)$, the primary goal of computer holography is to synthesize a transparency that has at least one component of its complex transmittance proportional to $F(u,v)$. As with interferometrically generated holograms, illumination of a computer-generated hologram with a plane wave will cause the wavefront $F(u,v)$ to be "reconstructed". More precisely, the wavefront is constructed, and not re-constructed, since it never physically existed.

1.1 Applications of Computer-Generated Holograms

The ability of a computer hologram to reconstruct a wavefront allows it to form an image of any one, two, or three-dimensional object which can be described mathematically [Lesem, Hirsch, and Jordan, 1968]. The description can be a set of equations or a sequence or array of numbers obtained from such sources as a scanned photograph, an acoustic microscope, or a mathematical model. Color images can be obtained by making three color-separation holograms for red, green, and blue information and illuminating them with the respective colors of coherent light [Fienup, 1974a; Fienup and Goodman, 1974], as described in Chapter 4. The principal

drawback is that the large number of resolvable elements in a high quality image, typically 10^5 or 10^6 for a two-dimensional picture and many times that for a three-dimensional one, combined with the extra redundancy desirable in a hologram, requires time-consuming and expensive computation and display.

A promising use of computer-generated holograms is for archival storage of data from computer outputs. The data to be stored may exist in core, and in this case the digital Fourier transformation step would represent the only extra step in comparison with writing the data directly on film. The benefits of translation invariance, burst noise resistance and redundancy (see Appendix A) make this extra step worthwhile, particularly if the memory is to be machine-read later. The outstanding features of holographic memory are high storage density, quick accessibility due to parallel processing and low cost. A working memory based on computer holography has been developed at Radiation, Inc. [Kozma, Lee, and Peters, 1971; Kozma et al., 1971].

Another area in which computer holography has been put to practical use is in the testing of optical elements [Birch and Green, 1972]. Manufacturing errors in an optical element can be detected by observing the difference between a wavefront from the optical element and a reference wave in an interferometer. Computer-generated holograms can be used to generate the reference wavefront standard. This approach is particularly useful in the testing of aspherics, for which the complexity of the desired reference wave makes synthesis difficult by any other means.

In addition to their use in testing optical elements, computer-generated holograms can be the optical elements themselves. Examples of

possible uses include computer-generated Fresnel lenses, diffraction gratings, zone plates, and diffusers. Interferometrically generated holographic optical elements [Latta, 1972] have proven to be practical and do not have the limited space-bandwidth product of computer-generated holograms; but computer-generated optical elements would be far more versatile (in addition, on-axis computer-generated holograms, made, for example, by the technique to be described in this thesis, would avoid many of the aberrations encountered with off-axis designs). Recently computer-generated holograms were used to produce conical and helicoid-shaped reference waves for interferometry [Bryngdahl, 1973] and to accomplish the shearing operation for radial and azimuthal shearing interferometers [Bryngdahl and Lee, 1974].

With the ability to control amplitude as well as phase transmittance, a whole new class of optical elements is possible. One example is an "extrapolating pupil", which is an aperture mask that allows a lens to resolve beyond the diffraction limit [Frieden, 1969,1970]. The extrapolating pupil is extremely difficult to manufacture because it must be made very precisely to work at all. There is hope that a computer-generated hologram could be used as an extrapolating pupil, but it would have to be one without any spurious images, such as the new computer-generated hologram described in this thesis.

Still another application of computer-generated holograms is their use as optical spatial filters. Although many types of coherent optical spatial filters can be optically generated [Vander Lugt, 1964], computer generation is often the most convenient (and sometimes the only) way to make them. Spatial filters find use in image enhancement [Lohmann

and Paris, 1968], restoration of degraded images [Severcan, 1973], matched filtering [Brown and Lohmann, 1966], and code translation [Lohmann, Paris, and Werlich, 1967]. With computer-generated holograms, matched filtering with incoherent illumination is also possible [Lohmann and Werlich, 1971]. The versatility of computer-generated holograms allows other types of operations to be performed as well, such as geometric transformations of images [Bryngdahl, 1974].

The considerations that go into making computer-generated holograms for all the above-mentioned applications are very similar. Because the grant under which our work was performed was for the study of computer-generated holographic memories, the emphasis of this work is on image formation from computer-generated holograms. However, most of the contributions of this thesis can also be readily applied to the other applications as well.

Another emphasis found in this thesis is on Fourier transform holograms. The (widely known) reasons for preferring the Fourier transform geometry are mentioned in Appendix A. The contributions of this thesis can be readily applied to the Fresnel case as well.

1.2 Previous Approaches to Synthesizing Computer-Generated Holograms

Previous approaches to synthesizing computer-generated holograms varied greatly because of a wide choice of available display devices and hologram materials. After calculating $F(u,v)$ using (1-1), then manipulating it to put it in a form suitable to be recorded on a given material, that information must somehow be displayed and photographed to produce the hologram transparency. Display devices that have been used include CRT's, Cal Comp and gray-level mechanical plotters, laser scanning

systems, writing microdensitometers, and even scanning electron microscopes. Materials for computer-generated holograms usually fall into one of two classes: thin amplitude materials and thin phase materials, which modulate the intensity and the phase, respectively, of the transmitted wave. Thin amplitude materials include silver halide photographic films and plates, diazo, and magnetic materials [Mezrich 1970]. Thin phase materials include bleached silver halide films and plates [Lamberts, 1972], photoresists [Bartolini, 1974], dichromated gelatin, and thermoplastics [Collier et al., 1971, Ch.10]. Thick hologram materials using volume absorption and phase effects are generally less suitable than thin materials for computer holography, although they are extremely valuable if the hologram is to be generated interferometrically. Such materials can act like a stack of thin holograms [Alferness, 1975], show Bragg effects [Kogelnik, 1967,1969], and can have a net complex transmittance that varies in both amplitude and phase, even though the material may be purely absorbing or purely phase-shifting.

The main problem encountered in synthesizing a hologram transparency with one component of its complex transmittance proportional to $F(u,v)$ is that the thin amplitude and thin phase materials ordinarily used have real non-negative and pure phase transmittances, respectively. But $F(u,v) \equiv |F(u,v)|e^{i\psi(u,v)}$ is complex valued, where $|F(u,v)|$ is the amplitude, or modulus, and $\psi(u,v)$ is the phase. What is usually done is to add another complex function to $F(u,v)$ that makes the sum real and non-negative, or, alternatively, a pure phase function; but the Fourier transform of that added function must be spatially separated from $f(x,y)$, so that the desired image is not degraded. Fortunately, there

are many ways to accomplish this separation.

Let $\delta(x,y)$ be the Dirac delta function - the impulse function [Bracewell, Ch.5]. If A is a real number, then

$$\begin{aligned} A\delta(x,y) + f(x+s,y) + f^*(-x+s,-y) \supset H(u,v) &= \\ &= A + F(u,v)e^{i2\pi us/\lambda f} + F^*(u,v)e^{-i2\pi us/\lambda f} \\ &= A + 2|F(u,v)|\mathrm{Re}\left\{e^{i\psi(u,v)}e^{i2\pi us/\lambda f}\right\} \\ &= A + 2|F(u,v)|\cos(2\pi us/\lambda f + \psi(u,v)) \end{aligned} \tag{1-4}$$

is real and is non-negative if A is greater than or equal to the maximum value of $2|F(u,v)|$. Thus, this real and positive hologram function can be synthesized on ordinary "black and white" photographic film or any continuous-tone thin amplitude material, by exposing it in such a way that the processed transparency will have amplitude transmittance proportional to the right hand side of (1-4). However, upon reconstruction we will find an on-axis bright spot, $A\delta(x,y)$, and a conjugate image, $f^*(-x+s,y)$, along with the desired image. The desired image is shifted off-axis by a distance s in order to avoid overlap with the spurious images. For the convenience of using a thin amplitude material, we pay the price of spurious images in the reconstruction, an associated decrease in diffraction efficiency (the amount of light going into the desired image), and an increase in resolution required of both the hologram material and the display device that is needed to produce the high frequency sinusoidal carrier. Synthesizing a hologram transparency in the form of (1-4) is the most straightforward approach [Burch, 1967] using a thin amplitude material. This method has actually been used in a computer-hologram memory system [Kozma, Lee, and Peters, 1971; Kozma

et al., 1971]. A hard-clipped one-dimensional version of (1-4) was one of the first computer-generated holograms [Kozma and Kelley, 1965].

An abundance of other methods have been devised to encode $F(u,v)$ onto a thin hologram material. These methods can be divided (rather arbitrarily) into two classes, off-axis and on-axis ("in-line"), according to the location of their reconstructed images. The off-axis holograms produce conjugate images as in (1-4) and can be synthesized on either thin amplitude or thin phase materials. The off-axis holograms can be subdivided into two groups: those that encode the amplitude and phase by the carrier-frequency term $|F(u,v)|\cdot\cos(2\pi us/\lambda f + \psi(u,v))$ as in (1-4), and the detour-phase holograms, which incorporate a carrier frequency in a more subtle way. The on-axis holograms do not produce conjugate images (instead they produce other spurious images) and can be synthesized only on thin phase materials or thin amplitude-and-phase materials. A new on-axis hologram that produces no spurious images and can be synthesized only on a thin material with complex transmittance will be described in this thesis.

Substituting $|F(u,v)|^2$ for the term A on the right hand side of (1-4) eliminates the problem of picking the correct A, but replaces the on-axis spot with the autocorrelation of $f(x,y)$, which is wider and forces the desired image to be shifted farther from the optical axis to avoid overlap [Huang and Prasada, 1966]. Including both terms A and $|F(u,v)|^2$ in (1-4) results in a hologram similar to the interferometrically generated hologram [Lesem, Hirsch, and Jordan, 1967].

In the off-axis holograms described above, the carrier frequency results in a fringe structure with a mean fringe spacing of $\lambda f/s$, as

can be seen in the cosine term in (1-4). We also see from (1-4) that the phase, $\psi(u,v)$, of the desired wavefront modulates the positions of the fringes. Similar to the way that those holograms modulate the phase of the wavefront by the positions of the fringes, detour-phase holograms modulate the phase of the wavefront by the positions of apertures within the hologram.

Perhaps the most popular computer-generated hologram is the binary detour-phase hologram of Lohmann [Brown and Lohmann, 1966,1969; Lohmann and Paris, 1967]. Its primary advantage is ease of synthesis, since only binary control is needed, i.e., at any one point the hologram is either opaque or transparent. This hologram consists of an array of Fourier coefficient cells, each representing one sampled value of $F(u,v)$. Within each cell is a transparent aperture on an opaque background, with the area of the aperture proportional to the amplitude $|F(u,v)|$ and its lateral position within the cell (the detour-phase effect) related to the phase $\psi(u,v)$. It can be viewed as an approximate hard-clipped version of (1-4), but its binary nature introduces still more spurious images. A more exact hard-clipped version of (1-4) has also been developed [Lee, 1974]. Other binary detour-phase holograms have been proposed that use a number of apertures within each Fourier coefficient cell [Haskell and Culver, 1972; Haskell, 1973; Haskell and Tamura, 1974]. Still another detour-phase method uses continuous-tone materials [Lee, 1970; Burkhardt, 1970].

The off-axis holograms mentioned above can also be made on phase-only materials. For example, the pure-phase version of (1-4) is

$$H(u,v) = \exp\left\{i\alpha\left[A + 2|F(u,v)|\cos(2\pi us/\lambda f + \psi(u,v))\right]\right\} \qquad (1\text{-}5)$$

$$= e^{i\alpha A}\exp\left\{i2\alpha|F(u,v)|\cos(2\pi us/\lambda f + \psi(u,v))\right\} \ ,$$

where α is a phase modulation coefficient. Dropping the constant phase term $e^{i\alpha A}$ and expanding into a power series yields

$$H(u,v) \propto 1 + i2\alpha|F(u,v)|\cos(2\pi us/\lambda f + \psi(u,v)) - 2\alpha^2|F(u,v)|^2\cos^2(2\pi us/\lambda f + \psi(u,v)) + \ldots \quad . \qquad (1\text{-}6)$$

The third term $(\alpha^2\cdot)$ in (1-6) can be prevented from overlapping the desired image, which is preserved in the second term $(\alpha\cdot)$ of (1-6). However, the fourth $(\alpha^3\cdot)$ and higher-order terms cannot be prevented from overlapping the desired image, so it is necessary to have $\alpha^2|F(u,v)|^2 << 1$ to keep from degrading the desired image.

A method of accurately controlling the phase of a transparency in an arbitrary manner was used to make "kinoforms", which produce a single on-axis image [Lesem, Hirsch, and Jordan, 1969; Clair, 1972]. A kinoform is a pure phase transparency, and as such can have transmittance $e^{i\psi(u,v)}$, but cannot control $|F(u,v)|$. The result is an absence of spurious images of the type obtained from the hologram using (1-4); however, making the approximation, $|F(u,v)|$ = constant, is usually inaccurate and introduces noise into the reconstruction [Kermisch, 1970; Fillmore, 1972]. Surprisingly, a kinoform works well enough to give recognizable images, which indicates that the phase $\psi(u,v)$ carries considerably more information than does the amplitude $|F(u,v)|$. The kinoform does work especially well for shaping wavefronts for which $|F(u,v)|$ is nearly constant.

Another method using thin phase materials is the "parity sequence" hologram [Chu and Goodman, 1972; Chu, 1974]. A parity sequence image is added to the desired image such that, together, the two images have

a Fourier transform with constant modulus. The image has a dot-like structure, and rows of parity elements are interlaced with rows of the desired image. If used as a holographic memory, a readout-detector array would be placed so that only the desired image would fall on the detectors, and the parity elements would fall between detectors. A related hologram is the synthetic coefficient hologram [Chu, 1974] which does not require the parity elements to be interlaced with the desired data, but suppresses the parity elements near the optical axis and causes them to appear off-axis away from the desired image. Both of these thin-phase-material holograms control $|F(u,v)|$ by selectively diffracting light away from the desired image into the parity image. The same effect can be obtained by using a diffraction grating and a two step exposure sequence [Kirk and Jones, 1971].

A method that is in a category by itself requires the synthesis of two hologram transparencies, one producing the real component of $F(u,v)$ and the other the imaginary component [Ransom, 1972]. These two components are combined interferometrically during reconstruction with a $\frac{1}{4}$ wavelength relative phase shift (this is difficult to achieve experimentally). The reconstructed image is on-axis with a bright on-axis spot. Another two-transparency method uses polarized light and Vectograph film [Holladay and Galatin, 1966; Marathay, 1969].

Good reviews of computer holography can be found in two survey papers [Huang, 1971; Chu and Fienup, 1974] and in a chapter in the text Optical Holography [Collier et al., 1971].

This multitude of computer-generated holograms suggests the large number of trade-offs between various considerations such as materials,

display devices, diffraction efficiency, complexity, resolution, spurious images, noise, ease and cost of synthesis, ease of reconstruction, and so on. Each type of hologram mentioned differs in some respect from every other one, and each has a different set of spurious images associated with it. In all cases, the use of amplitude-only or phase-only materials to reconstruct an arbitrary complex-valued wavefront introduces spurious images into the reconstruction. These spurious images were tolerated because in most cases they do not overlap the desired image. Furthermore, there was not previously known a simple method to directly control the complex (both amplitude and phase) transmittance of a hologram material. Only through complex control is it possible to completely avoid spurious images. The closest attempt to arbitrary complex control was the very difficult and often impractical method of making two different transparencies, one an absorbing transparency to control the amplitude and one a phase-shifting transparency to control the phase, and sandwiching the two together [Tsujiuchi, 1963].

1.3 Overview of Thesis and Summary of Contributions

Chapter 2 of this thesis describes a new kind of computer-generated hologram, called the "Referenceless On-Axis Complex Hologram (ROACH) [Chu, Fienup, and Goodman, 1973]. The ROACH constitutes a method of controlling the arbitrary complex transmittance of a transparency; that is, the amplitude transmittance and the phase transmittance are controlled independently without resorting to a carrier frequency. Independent control is accomplished by using a multi-emulsion film, such as Kodak's Kodachrome color reversal film, the different layers of which can be addressed independently by different colors of light. As will be

explained, the complex transmittance of the ROACH can be made proportional to the desired complex wavefront $F(u,v)$, and so no spurious images need occur. The ROACH has a number of other extremely advantageous properties that are discussed in Chapter 2. The ROACH is the most direct and elegant of all holograms and represents a major advance in computer holography. Also discussed are the experimental steps and calibration required to succeed in obtaining a ROACH with the desired complex transmittance. Experimental results will be shown. The ROACH was invented jointly by the author and Dr. David Chu.

In Chapter 3 we discuss the properties and limitations of Kodachrome film that affect its use as a holographic material. There are cross-talk effects in both the spectral sensitivities of the different layers of the film, and in the dye spectral densities of the dye images in the processed transparency. The result is amplitude-phase cross talk that must be compensated for. The spatial frequency response of the film not only limits the information density achievable with the material, but also requires compensation if optimal results are desired. The dynamic range of the film limits the range of amplitudes that can be represented. The power spectrum of the noise due to scattering in the emulsion ultimately limits the signal-to-noise ratio of the reconstructed image.

Color images can be produced from three color-separation computer-generated holograms, one for each of the three primary colors. Three monochromatic images are added to form a full-color composite image. Two basic problems must be solved: scaling the images according to wavelength and avoiding false color images [Fienup and Goodman, 1974; Fienup, 1974a]. As discussed in Chapter 4, we solved the scaling problem

in three different ways: digitally, photographically during hologram synthesis, and optically by the reconstruction geometry. We also solved the problem of false images in three different ways: by optical means during reconstruction, by using the wavelength-selective absorption properties in color film, and by using phase effects in color film to produce what is called the "phase-null" effect. This is the first work in which the problems associated with producing color images from computer-generated holograms were clearly understood, stated, and solved in a systematic and straightforward manner.

Due to limitations in the display devices and materials used to synthesize computer holograms, often the complex values of the reconstructed wavefront $F(u,v)$ must be quantized to a certain set of values in the complex plane. In Chapter 5, we discuss an iterative procedure invented by the author that re-computes the hologram to significantly reduce noise in the reconstructed image due to quantization in the hologram [Fienup, 1974b]. Images from quantized holograms computed by this iterative method achieve error rates that are orders of magnitude better than the error rates from quantized holograms computed by previous methods. This type of iterative procedure is shown to be extremely versatile. Among other things, it can be used to find solutions to the phase retrieval problem.

In Chapter 6 we summarize our conclusions and make some final remarks. The remarks include ways in which computer holography could be improved still further and suggestions of recommended further research.

Appendix A contains a discussion of the Fourier transform geometry assumed in this thesis and the reasons for preferring that geometry. Appendix B contains a derivation of the discrete Fourier transform and the exact form of the image produced by a sampled hologram. Appendix C contains a discussion of a method used to produce sinusoidal patterns of intensity with variable modulation and spatial frequency, which were required for testing the spatial frequency response of Kodachrome. Appendix D contains a derivation of the phase-null effect.

CHAPTER 2

THE ROACH

In this chapter we describe a new kind of computer-generated hologram, called the ROACH, that is made with multi-emulsion film. The ROACH is the most elegant and direct of all holograms in that, if made properly, it reconstructs exactly the desired wavefront and produces no spurious images. The ROACH consists of two emulsion layers that independently control amplitude transmittance and phase transmittance. In this chapter we describe how ordinary color film with conventional processing can be used to synthesize a ROACH. Also described in detail are the calibration and synthesis procedures required to make a ROACH.

2.1 Basic Concept of the ROACH

The best possible hologram would have a complex transmittance directly proportional to the desired complex wavefront. To have a complex transmittance $F(u,v) = |F(u,v)|\, e^{i\psi(u,v)}$, the hologram must, at each point (u,v), absorb (or scatter) light according to $|F(u,v)|$ and vary in thickness or refractive index according to $\psi(u,v)$. The "referenceless on-axis complex hologram", or ROACH, is such a hologram. The ROACH consists of a two- (or multi-) layered material in which the different layers are independently addressable with radiation of different wavelengths. After exposure and processing, one layer responds with variations in amplitude

transmittance and another layer responds with variations in phase transmittance, according to the exposures by the radiations of the respective wavelengths (Figure 2.1). With the ability to independently control both amplitude and phase, we can, by a judicious pair of exposures with radiation of the two wavelengths, control the arbitrary complex transmittance of the material. Thus it is possible to synthesize a hologram transparency with transmittance exactly equal to the desired complex transmittance. When that hologram is illuminated during reconstruction, no more and no less than the desired complex wavefront is produced.

The advantages of the ROACH over other computer-generated holograms are significant. Since exactly the desired wavefront is produced, no conjugate image, on-axis spot, or other spurious images appear in the reconstruction. All the light transmitted by the ROACH would ideally go into the desired image, which is on-axis (in-line). Except for the sandwich hologram [Tsujiuchi, 1963], of all the computer-generated holograms, only the ROACH has no spurious images. In fact, the ROACH can be thought of as a sandwich hologram, only without the great difficulties encountered in aligning together two transparencies. Due to an absence of spurious images, the ROACH has a theoretical maximum efficiency of 100% (the diffraction efficiency of most computer holograms is only a fraction of this). Since the level of illumination is very often a problem in holography and in coherent optical spatial filtering, this high diffraction efficiency gives the ROACH an important advantage.

The space-bandwidth product (or number of picture or resolution elements) of most display devices is very limited - 1000×1000 is typical.

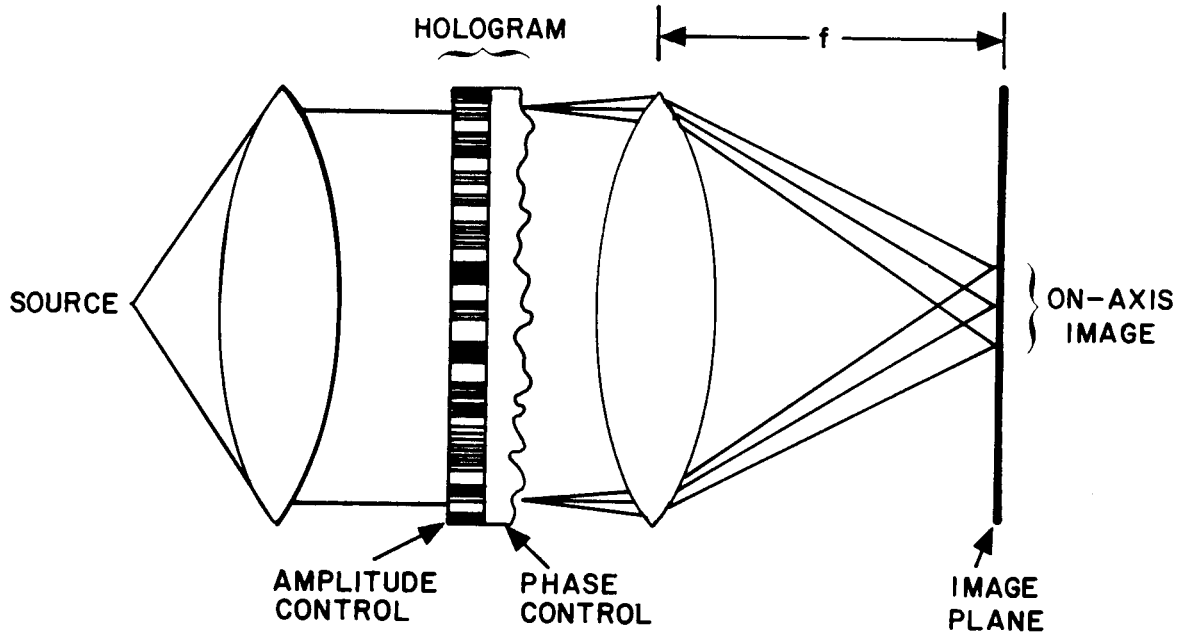


Fig. 2.1 Reconstruction of a ROACH. Amplitude and phase are controlled by different layers in the film.

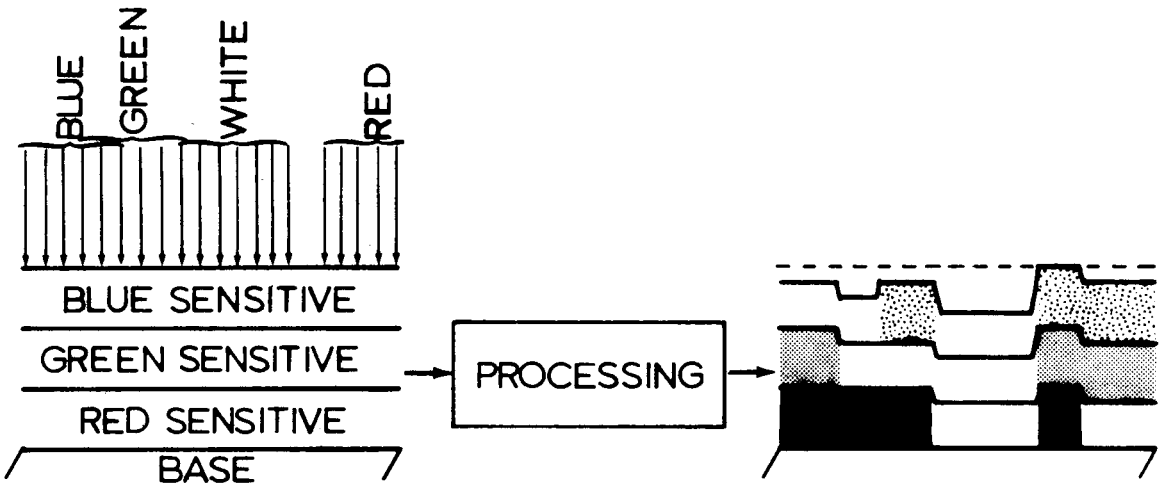


Fig. 2.2 Model of Kodachrome film. In the processed transparency, three dye images are formed, each of which absorbs the same color to which that layer had been sensitive before processing.

But as mentioned in Chapter 1, a redundant hologram for a high-quality image would require many times that display capability. This is particularly troublesome for binary detour-phase holograms [Brown and Lohmann, 1969] that require many display resolution elements to represent a single sampled value of $F(u,v)$. The ROACH, on the other hand, requires only one resolution element per sampled value of $F(u,v)$ on a gray-level display device. Thus, for a given display device with gray-level capability, the ROACH could produce the highest quality image possible from any computer-generated hologram, since it would include the largest number of sampled values of $F(u,v)$.

Since the ROACH produces $F(u,v)$ directly, no additional encoding calculations are required, although such calculations are needed with many of the other computer-generated holograms. The off-axis images produced by most computer-generated holograms require a high-frequency carrier on which the information is modulated in the hologram. Supporting a carrier requires a hologram material as well as a display device with far greater resolution than would be needed for the information in $F(u,v)$ alone. The ROACH, being in-line and without a carrier, keeps the bandwidth requirement to a minimum, so that hologram materials with lower resolution (and correspondingly higher-speed) can be used. In fact (as will be discussed in the next sections), we were able to use a hologram material of sufficiently high speed as to make it practical to expose it directly from a CRT.

2.2 Kodachrome Film

While the ROACH is the most direct and most elegant of all holograms, its synthesis was not accomplished until just two years ago [Chu, Fienup, and Goodman, 1973]. Since hologram materials are sensitive only to the

intensity of light, and not phase, it was not possible to make interferometrically-generated holograms in this direct way. Computer-generated holograms also were not made in this direct manner because there did not (and does not now) exist a hologram material designed to have independently addressable amplitude and phase layers. However, we discovered that ordinary color reversal film, particularly Kodachrome, did happen to behave in this manner and could be used to make a ROACH.

2.2.1 Amplitude and Phase Control with Kodachrome

A simplified model of Kodachrome film, sufficient for our purpose here, is that it consists of three emulsion layers, sensitive individually to blue, green, and red light, respectively (Figure 2.2) [Kodak, E-74]. After exposure to a pattern of multicolored light and processing, there are formed three corresponding dye images that are predominantly blue absorbing (yellow), green absorbing (magenta), and red absorbing (cyan), respectively. If the processed transparency is illuminated by monochromatic light, one layer will absorb, while the other layers will be predominantly transparent. For example, to the red 632.8 nm light of a helium-neon laser, the blue- and green-absorbing layers are transparent, and only the red-absorbing layer can attenuate the red light.

Our goal is the ability to control the complex transmittance of color film; therefore, we will discuss only those characteristics of color film that play a role in our gaining that ability. More complete descriptions of the manner in which color film works, its chemistry, and its physics, can be found in a number of references [Mees and James, 1966; Evans, Hansen, and Brewer, 1953; Neblette, 1962; Langford, 1974].

To control the amplitude transmittance to red light, the film is

exposed to a pattern of red light; and after processing, the transparency has greater amplitude transmittance where the exposure had been greater, and smaller amplitude transmittance where the exposure had been less (since color reversal processing is performed). By making trial exposures and making measurements on a color densitometer, one can determine what exposure is necessary to produce a desired amplitude transmittance. There is nothing unusual about trying to control the amplitude transmittance (optical density) of Kodachrome film, since its intended function is to be an absorbing material.

The less obvious property of color film that can be used is its ability to control the phase structure of a transmitted wavefront. Approximately proportional to density differences in the three layers of Kodachrome are optical path differences due to variations in emulsion thickness and refractive index. If the film is exposed to a pattern of blue-green light and (in a second exposure) flooded with red light, then the processed transparency will be a pure-phase material to red light, with the blue- and green-absorbing layers providing the phase effects. The phase transmittance of Kodachrome to red light can be controlled in a way somewhat similar to the manner in which the amplitude transmittance is controlled. By making trial exposures and making measurements on a color densitometer (measuring the density to blue-green light) we can determine the relationship between exposure and density to blue-green, excursions of which are assumed to be proportional to the excursions of the phase delay. To find the proportionality constant, the film is exposed to simple patterns of blue-green light (and re-exposed to uniform red light) that result in phase gratings that have well-known diffraction patterns.

Then by measuring the amount of light that falls in the different orders of diffraction, the proportionality constant can be determined. We first used this method of making a phase transparency with Kodachrome film [Goodman and Gray, 1973] to make kinoforms [Lesem, Hirsch, and Jordan, 1969].

Combining the amplitude controlling and phase controlling layers gives us control over the complex transmittance of the material, and we have a ROACH. To synthesize a ROACH for illumination with red light, Kodachrome film is exposed to a pattern of red light in order to give it the proper amplitude transmittance and to a pattern of blue-green light to give it the proper phase transmittance. The red-absorbing layer will also give rise to an unwanted phase pattern, but, as will be explained later in more detail, that phase is compensated for in the phase controlling layers by adjusting the blue-green exposure.

2.2.2 Advantages of Kodachrome over Other Phase Materials

Ordinary color reversal film has many practical advantages over the more conventional thin phase materials that could be used to make on-axis holograms (besides the fact that it is a thin amplitude-and-phase material). Kodachrome, having a sensitivity of about one $(\text{erg/cm}^2)^{-1}$, is roughly ten to one hundred times as sensitive as thermoplastics, one thousand times as sensitive as 649-F (bleached), 10^5 times as sensitive as dichromated gelatin, and 10^5 times as sensitive as photoresist [Pennington, 1971]. Kodachrome has sufficient photographic speed for it to be conveniently exposed directly from a CRT (as will be explained in detail in the next section) in a reasonably short exposure time. The other commonly used phase materials mentioned above have such low sensitivity that they would require prohibitively long exposure

times (or require a different means of exposure, such as a laser scanning system).

Most thin phase materials, especially thermoplastics and bleached 649-F, have a surface-relief response that varies greatly with the spatial frequency of the pattern of light that exposes the material [Lamberts, 1972; Collier et al., 1971, Ch.10]. That is, the relationship between phase and exposure depends upon the spatial frequency of the exposure pattern. Consequently, to obtain the desired phase for a general pattern containing a wide range of spatial frequencies, a difficult compensation must be performed. As will be described in Section 3.5.2., the surface relief of Kodachrome also varies with spatial frequency. For our own ROACHs and phase-only holograms, the important range of spatial frequencies was 1 to 15 1/mm. For these spatial frequencies the response of Kodachrome does not vary as much as the response of bleached 649-F, and compensation is not a necessity.

Bleached 649-F has an advantage over other phase materials in that its phase pattern can be measured using a densitometer, since the phase after bleaching is linear with the density before bleaching [Lamberts, 1972]. Such an easy measurement of phase is extremely helpful in determining the exposure pattern (before processing) required to achieve the desired phase pattern (after processing). For Kodachrome the measurement of phase is even easier. Unlike bleached 649-F, for which the easy measurement of phase must be performed before the final processing steps (bleaching), the phase of Kodachrome can be measured on the completely processed transparency, as discussed in Section 2.3.5.

The use of Kodachrome is also advantageous because it is inexpensive,

readily available, and requires no darkroom facilities of the user, since it can be purchased almost anywhere and processed through any drug store or photo-counter, or by the manufacturer. The convenience of Kodachrome should make it extremely attractive to researchers who want to make computer-generated holograms, even if it is used just as a thin phase material.

We used Kodachrome II for most of our work. However, since the summer of 1974, Kodak has replaced it with a new film, Kodachrome 25. Both of these materials, as well as other color films, are suitable for making ROACHs; but Kodachrome II has the best combination of characteristics of the films we tried (these characteristics will be discussed in the following sections).

2.2.3 Different Modes of Use

A ROACH can be made for reconstruction with green or blue light, as well as with red light as discussed in Section 2.2.1. (Reconstruction with all the visible colors of light is possible, but we will consider here only the primary colors.) Furthermore, for reconstruction with red light, it is not necessary to use both the green-absorbing and the blue-absorbing layers to control the phase. There are many different modes in which a three-layered color film can be used to control amplitude and phase. For reconstruction with any one of the primary colors of light, the corresponding absorbing layer is used to control amplitude transmittance and either or both of the remaining two layers are used to control phase transmittance. The various modes and the colors of light needed in their synthesis are summarized in Table 2.1. When only one of the two transparent layers is used for phase control, then the other transparent layer can either be left unexposed, resulting in a uniformly

color of light, reconstruction	use of layers within the film: phase controlling	use of layers within the film: eliminated	filtered color of light: amplitude controlling exposure	filtered color of light: phase controlling exposure	filtered color of light: kinoform:phase controlling exposure
red	B_ℓ	G_ℓ	R or RG	B or BG	B or Wh
red	G_ℓ	B_ℓ	R or RB	G or GB	G or Wh
red	B_ℓ & G_ℓ	-	R	B&G equally	B&G equally
green	B_ℓ	R_ℓ	G or GR	B or BR	B or Wh
green	R_ℓ	B_ℓ	G or GB	R or RB	R or Wh
green	B_ℓ & R_ℓ	-	G	B&R equally	B&R equally
blue	G_ℓ	R_ℓ	B or BR	G or GR	G or Wh
blue	R_ℓ	G_ℓ	B or BG	R or RG	R or Wh
blue	G_ℓ & R_ℓ	-	B	G&R equally	G&R equally

Table 2.1. Different modes of use of Kodachrome film. R_ℓ,G_ℓ, and B_ℓ refer to the red-absorbing green-absorbing, and blue-absorbing layers, respectively. R, G, B, and Wh refer to synthesis exposures with red, green, blue, and "white" (i.e., no filter) light, respectively. For a given color of light used in reconstruction, the corresponding absorbing layer is always used for amplitude control.

thick dye image in that layer of the processed transparency, or be fully exposed (by flooding it with the corresponding color of narrowband light), resulting in an absence of dye in that layer of the processed transparency ("eliminating" the layer). In practice, if a layer is not used for either amplitude or phase control, then the results are better if that layer is "eliminated". In Table 2.1 it is assumed that unused layers are eliminated. When a phase-only transparency (kinoform) is desired, the layer corresponding to the color of the reconstructing light is eliminated.

For most of our experiments we made ROACHs for reconstruction with red light, using both the blue-absorbing and the green-absorbing layers together for phase control. In order to produce full-color images, as will be described in Chapter 4, it is necessary to synthesize ROACHs for reconstruction with blue light, green light, and red light. All of the modes were tested. Not all modes are equally useful. One consideration is the amount of phase variation obtainable from one layer (2π radians, or one wavelength, is required), and that depends on the specific type of film used. A comparison of the advantages and disadvantages of the various modes will be made in Section 2.3.5.

2.2.4 Dye Spectral Densities of Kodachrome

Figure 2.3 shows the spectral densities of the dyes in the three layers of Kodachrome II [Kodak Data]. The considerable amount of overlap of the three dye spectral densities is suitable for color photography since it reproduces colors that are interpreted by the eye to approximate the colors in the original scene. However, that overlap is highly undesirable for computer holography. There is little problem when the

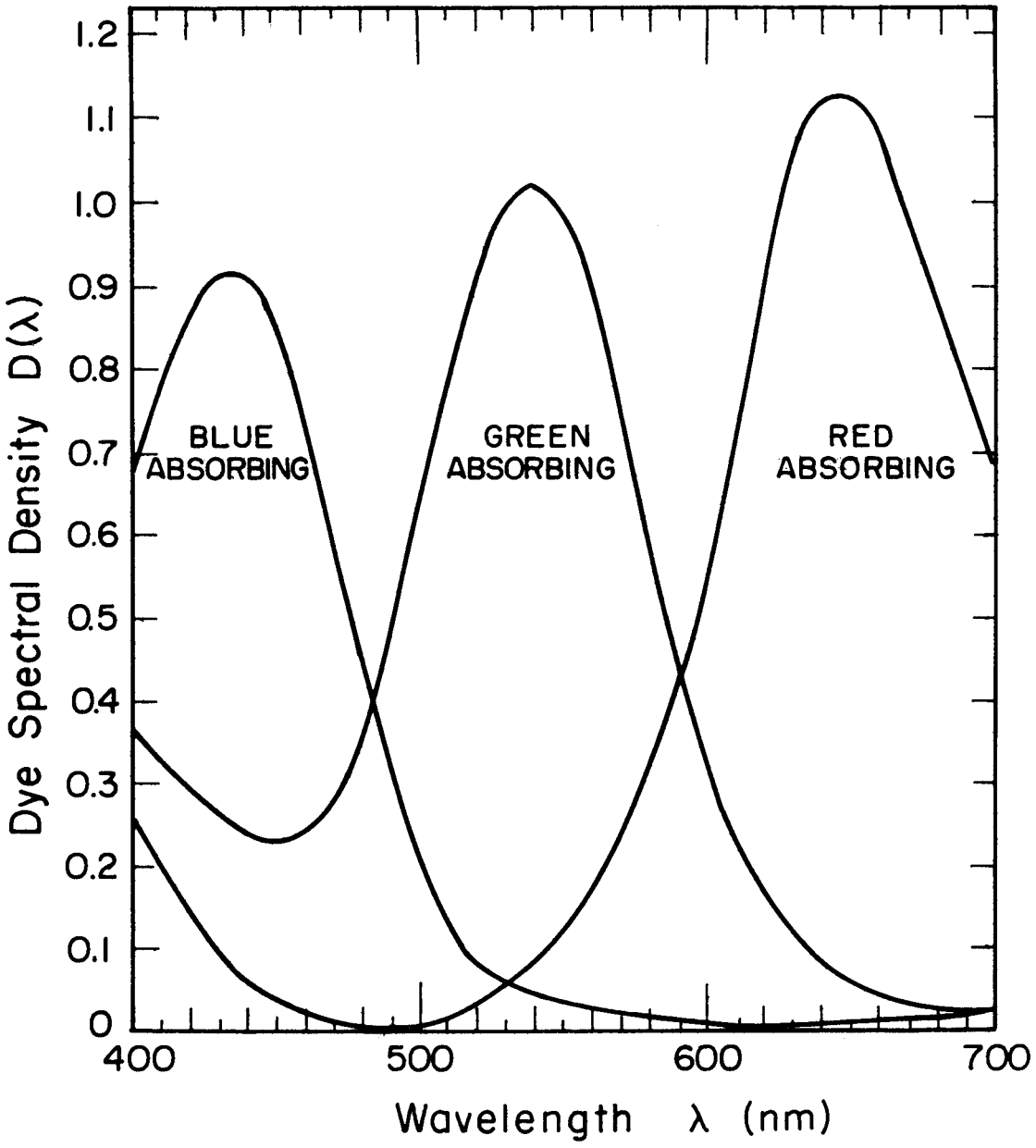


Fig. 2.3 Dye spectral densities of Kodachrome II, normalized for a visual density of 1.0.

processed transparency is illuminated with the red 632.8 nm light from a helium-neon laser, since the blue-absorbing and green-absorbing dyes absorb very little at that wavelength. However, when the transparency is illuminated by the blue 457.9 nm light from an argon-ion laser, the blue light is significantly absorbed by the green-absorbing and red-absorbing layers, causing an undesired change in the amplitude transmittance. If a ROACH were made for illumination with blue light, this cross-talk from the green-absorbing and red-absorbing layers could be compensated for by adjustments in the blue-absorbing layer. Since the green-absorbing and/or red-absorbing layers would be used for phase control, this is a phase-to-amplitude cross-talk. Fortunately, good results can be obtained even if we ignore this phase-to-amplitude dye-absorption cross-talk since the accuracy of the amplitude transmittance is much less important than the accuracy of the phase transmittance [Pearlman, 1974; Powers, 1975]. Similarly, the green 514.5 nm light from an argon-ion laser would be partially absorbed by the blue-absorbing and red-absorbing layers, but good results can be obtained if this cross-talk is ignored. For optimal results, however, compensation for this cross-talk must be performed, as will be discussed in Chapter 3. An effect of phase-to-amplitude cross-talk for which no compensation is possible is a reduction in the diffraction efficiency of the hologram due to absorption by the phase controlling layer.

We expect the phase contributed by each layer to be proportional [Lamberts, 1972] to the density of each individual layer, the analytical density. However, a measurement with one color on a color densitometer gives the sum of the contributions from all three layers, the integral

density, which differs somewhat from the analytical density [Evans, Hansen, and Brewer, 1953; Mees and James, 1966]. It is possible to determine the analytical densities from the integral densities [Evans, Hansen, and Brewer, 1953]; however, we used the integral densities for our experiments because they are easily measured and are sufficiently close to the analytical densities for our purposes. For optimal results, though, the analytical densities should be considered. Furthermore, since the densitometer available to us measures diffuse density, we used measurements of diffuse density for our H&D curves rather than specular density (which is more pertinent to Fourier transform imagery). The diffuse-density H&D curve is sufficiently similar to the specular-density H&D curve to warrant its use.

2.3 Synthesis and Calibration Procedures

In order to synthesize a hologram that will produce a desired image, it is necessary to first calculate the desired wavefront in the hologram plane. Ordinarily, the image is specified as an array of complex numbers, the moduli of which are the square roots of the desired intensity at a grid of sampled points in the image plane. A digital discrete Fourier transform is performed that yields sampled values $\{F_{p,q}\}$ (Fourier coefficients) of the Fourier transform of the sampled object. This will be explained in more detail in Appendix B. Then to synthesize a ROACH for reconstruction with red light, we expose Kodachrome film to two patterns of light, a pattern of red light to control the amplitude transmittance and a pattern of blue-green light to control the phase transmittance. We accomplish this by photographing a computer-controlled CRT display, the amplitude controlling exposure through a red filter and

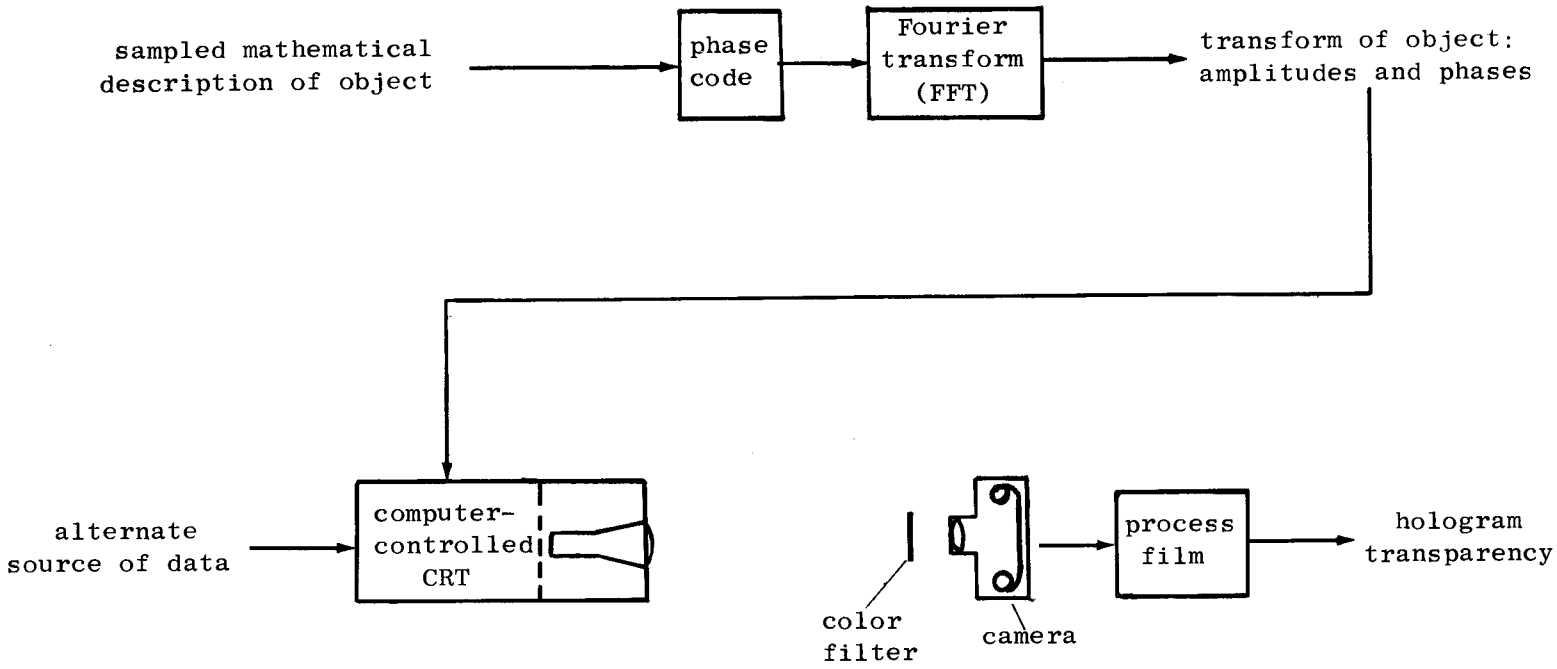
the phase controlling exposure through a blue-green filter. After standard processing, the ROACH can be reconstructed by illuminating it with coherent red light from a helium-neon laser. This synthesis procedure is summarized in Figure 2.4.

This work is the first successful attempt to control the complex transmittance of a material. The gaining of the ability to control complex transmittance requires considerable experimental effort; however, once the correct exposure parameters are found, then ROACHs can be made with relative ease. In this section we will describe in detail the CRT and photographic considerations, the choice of filters needed for independent amplitude and phase control, and the calibration procedure necessary to determine what exposures are needed to produce the desired complex transmittance. The considerations described in this section are of considerable practical importance in achieving the complex control needed to synthesize a ROACH.

2.3.1 CRT and Photographic Considerations

For our experiments, we use a Tektronix type 602 CRT display, controlled by a Xerox Sigma-5 computer through a digital-to-analog converter. Assembly language subroutines for controlling the CRT were written by Turner [Turner, 1972]. The CRT has 256×256 addressable points. On the CRT a pattern is scanned out once, the exposure controlled by the length of time that the CRT beam is turned on at each point, generating gray levels that are for all practical purposes continuous.

To minimize exposure times, the highest intensity that would not damage the phosphor is used. To get a repeatable intensity, after the CRT is fully warmed up, the brightness of a spot displayed on the CRT



2.4 Synthesis of a ROACH. A Fast Fourier Transform (FFT) routine is used to calculate the transform of the object. Alternatively, the amplitudes and phases may be obtained from an alternate source, for example, from an acoustic microscope or from the computed values of a desired spatial-filter function or optical element. The values of amplitude and phase are used to intensity modulate a CRT display which is photographed through color filters onto Kodachrome film.

is measured directly by putting a light meter up to the face of the CRT and adjusting the intensity knob until the light meter indicates the standardized intensity (20 μW in our experiments).

The ideal display would be a grid of adjoining squares with uniform illumination over each of the squares and no gaps between the squares. To approximate this, the X and Y gain control knobs on the CRT are adjusted so that adjoining lines just overlap. Unfortunately, a single dot on the CRT is round rather than square; and over its circular area the dot intensity is not uniform, but is peaked at the center and falls off toward the edges. Defocusing the electron beam can cause the spot to take a more rectangular shape, but the shape is difficult to control accurately. The method we use is to move the dots close together to make them overlap somewhat, then use four adjoining dots (in a 2×2 square) to represent one sampled location. That is, in effect we used the CRT as though it had only 128×128 resolution elements. Using 2×2 dots as one display element reduced the number of available display elements by a factor of four, but substantially increased the quality of the display.

Kodachrome II resolves spatial frequencies out to the range 96-135 lines/mm [Kodak, AF-1], and the 55 mm f/3.5 Micro-Nikkor lens that we used to photograph the CRT display resolves only 70 lines/mm (by our own measurements). To insure that the modulation transfer function (MTF) of our photographic process would not be a limiting factor, we chose a magnification that gave us a spacing of approximately 1/20 mm between Fourier coefficients in our holograms. Positioning the front of the lens 80 cm from the face of the CRT yields a magnification of 1/14.5 that

produces a hologram of width 5.3 mm (128×128 Fourier coefficients) from a CRT display of width 7.7 cm. The CRT is shielded from room light by a cardboard tube (Fig. 2.5).

2.3.2 Color Filters

It is necessary to photograph the CRT through color filters in order to independently address the different layers of the color film. The relative exposure of each layer (the photicity) is given by [Jones, 1931]

$$E_i = \int_0^\infty P(\lambda)T_f(\lambda)S_i(\lambda)d\lambda \qquad (2\text{-}1)$$

where $P(\lambda)$ is the spectral energy distribution from the CRT phosphor, $T_f(\lambda)$ is the spectral transmittance of the color filter, and $S_i(\lambda)$ is the spectral sensitivity of the i^{th} emulsion layer of the film.

The CRT display used for our experiments has a type P-31 phosphor, which has a spectral energy distribution shown in Figure 2.6a. Two filters typically used, Kodak Wratten filters No. 38A and No. 26, have spectral transmittances shown also in Figure 2.6a [Kodak, B-3]. Figure 2.6b shows the spectral energy distributions of the light from the CRT after it has passed through either of these two filters. Kodachrome II has spectral sensitivities shown in Figure 2.6c [Kodak Data].

As can be seen from Figure 2.6a, the P-31 phosphor emits little light energy with wavelengths greater than 620 nm; and as can be seen from Figure 2.6c, wavelengths longer than 620 nm are needed in order to expose the red-absorbing layer independently of the other two layers. The result is a need for very long exposure times when red light is required from the P-31 phosphor. A better choice of phosphor when working with color film would be a white phosphor, such as P-4 [Bell, 1970].

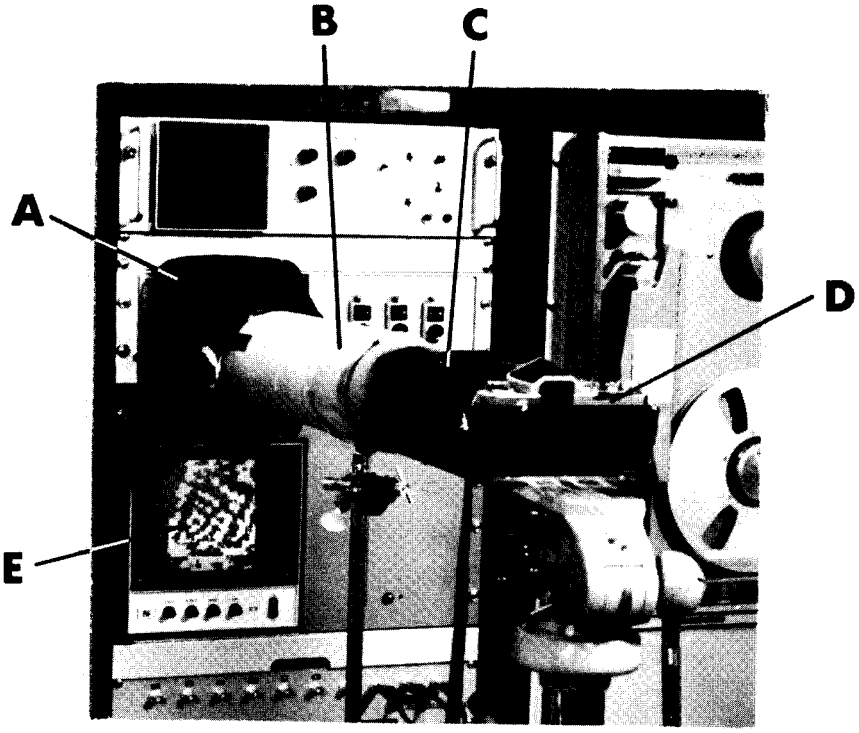


Fig. 2.5 Photographing the CRT. The CRT (A) is shielded from room light by a cardboard tube (B). Color filters are inserted in felt-lined slots (C) in the cardboard tube, between the CRT and the camera (D). A storage CRT (E) shows what is displayed on the primary CRT (A).

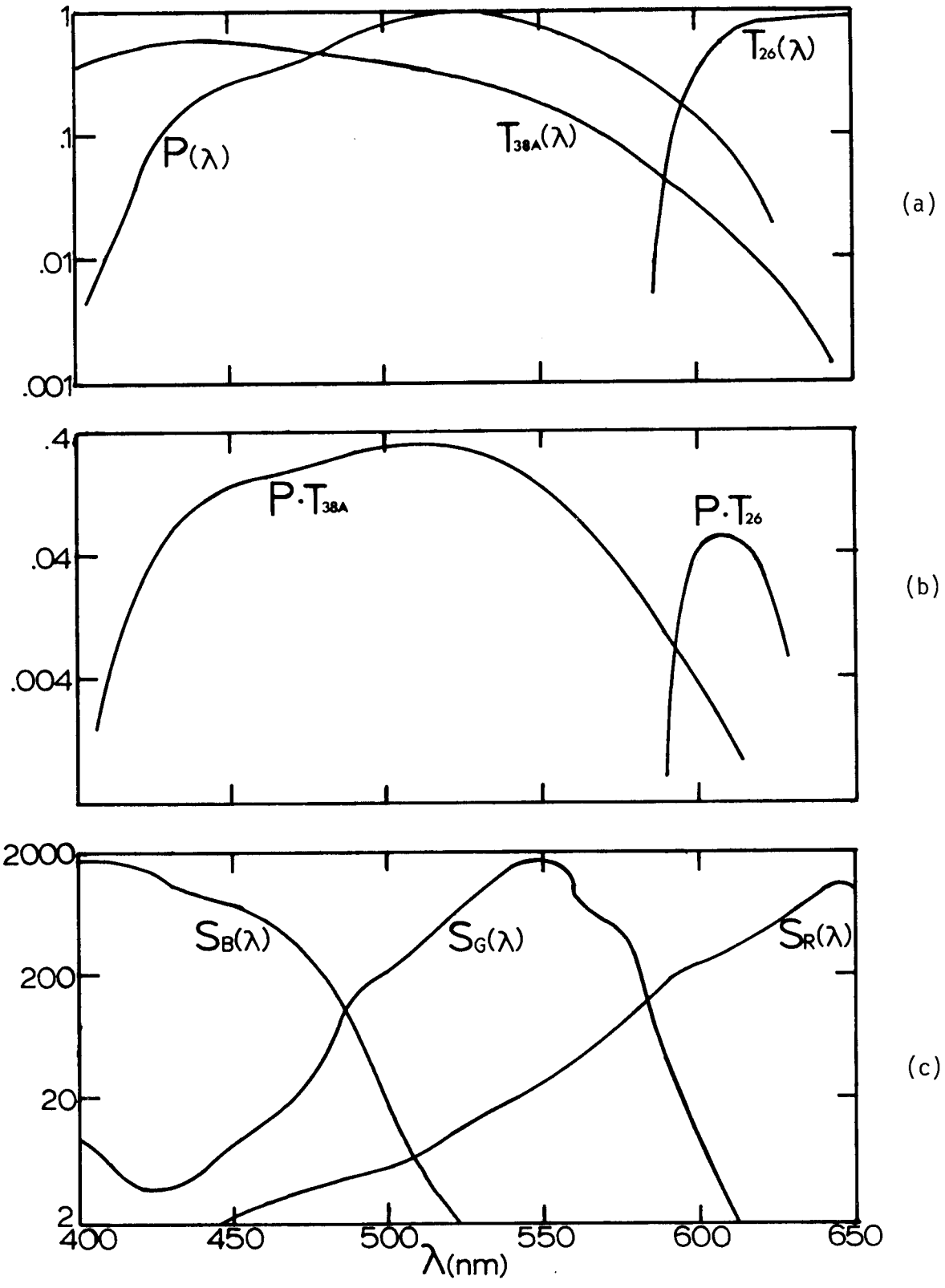


Fig. 2.6 Spectral distribution curves. (a) P-31 phosphor - $P(\lambda)$, and Wratten filters No. 38A and No. 26 - $T_f(\lambda)$; (b) $P(\lambda)T_f(\lambda)$; (c) Spectral sensitivities of blue-absorbing layer - $S_b(\lambda)$, green-absorbing layer - $S_g(\lambda)$, and red-absorbing layer - $S_r(\lambda)$, in $(ergs/cm^2)^{-1}$.

The spectral distribution curves of Figure 2.6 make apparent the interrelated problems of independently addressing the different layers of the film. Since there is considerable overlap of the spectral sensitivities of the different layers of the film, it is not possible to expose each layer completely independently of the others (the result being exposure cross-talk, which will be discussed further in Chapter 3.). Let the ratio of the relative exposure of the m^{th} layer to the relative exposure of the n^{th} layer be E_m/E_n . Then the requirement for layer m being addressable independent of layer n is only that $E_m/E_n >> 1$; in practice, E_m/E_n in the 50-100 range is sufficient. Competing with the desire to have E_m/E_n as large as possible is the desire to keep E_m , the relative exposure of the desired layer, as large as possible. This conflict arises from the fact that E_m/E_n can only be increased by choosing a narrow-band filter transmittance $T_f(\lambda)$; but, as can be seen from (2-1), a narrow-band $T_f(\lambda)$ will reduce E_m . Furthermore, narrow-band absorbing filters usually have a low peak transmittance, further reducing E_m . Higher peak transmittances can be obtained from dichroic interference filters, but at a much higher cost than Kodak Wratten filters. In order to find the Kodak Wratten filters with optimum E_m and E_m/E_n for a given mode, the integration (2-1) was performed on a digital computer for all Wratten filters of interest, using filter spectral transmittance data supplied by the manufacturer [Kodak, B-3]. The results are summarized in Table 2.2, a listing of the Kodak Wratten filters best suited for use with a P-31 phosphor and Kodachrome II film.

undesired color(s)	desired color		
	red	green	blue
Both other colors	26	61	98
red	-	61	98
green	26	-	98
blue	12	12	-

(a)

	desired colors (equally)		
third color	red and green	red and blue	green and blue
not allowed	22	none available	38A
allowed	22	CC50Y	CC50M

(b)

Table 2.2. Kodak Wratten filters for use with P31 phosphor and Kodachrome II. (a): Filters with maximum transmittance to the desired color and minimum transmittance to the undesired color(s). No filter is required for kinoforms using only one layer to control phase when both of the other layers are eliminated. (b): Filters for use when two layers are to be exposed equally.

2.3.3 Eliminating a Layer

Before attempting to control both amplitude and phase simultaneously, it is necessary to first concentrate on controlling just the phase. In order to make a phase-only transparency it is necessary to eliminate the absorbing dye from the processed transparency. This elimination can be accomplished by flooding the film (in addition to exposing it to the phase-controlling pattern) with the color of light to which the processed film is to be transparent. If the flooding exposure is sufficient, then the processed transparency will be sufficiently transparent to the reconstructing light; however, too great a flooding exposure will also expose the phase-controlling layers, reducing the ability to control the phase. The elimination of a layer by a flooding exposure is also needed when the film is used in a mode in which only one of the two transparent layers is used for phase control (Table 2.1), as discussed in Section 2.2.3.

In order to determine the optimum flooding exposures for eliminating undesired layers, a series of exposures are made through narrowband color filters, then the resulting densities to the three primary colors are measured. The exposures are made by photographing a uniform white surface through narrowband color filters. The white surface is illuminated by two No. 2 photofloods (with a color temperature of 3400^{o}K) in 12 inch-diameter reflectors, positioned on both sides of the camera, both at 45^{o} angles to insure even illumination. The distance from each of the photofloods to the white surface is adjusted so that the intensity of the light falling on the white surface is 520 foot-candles. Since minimization of exposure time is not a consideration

(as it is when exposing from a CRT), very narrowband filters can be used. The following Wratten filters were used for the flooding exposures: No. 92 (red), No. 93 (green), No. 98 (blue), and No. 21 (red and green = yellow-orange).

A Macbeth Quantalog TD-100A densitometer was converted for use as a color densitometer. Color filters are inserted into the illuminating beam to select the desired color for measurement. The filters used by us were taken from a Kodak 1-A visual densitometer: Wratten No. 92 (red), Wratten No. 61 (green), and Corning No. 5113 (blue). These filters are of relatively high transmittance, which facilitates densitometer readings. Measurements in better agreement with other color densitometers would result if Status A filters [Voglesong, 1973] were used.

Unfortunately, the tungsten light source of the densitometer emits very little blue light. Consequently, the density-to-blue readings are made at the highest sensitivity range of the densitometer and are unreliable beyond a density of 2.0 (using a 1 mm measuring aperture; with a 3 mm measuring aperture, useful measurements can be made up to a density of 3.0). An infrared-reflecting dichroic interference filter is also inserted in the illuminating beam in order to eliminate false readings at high densities due to infrared, which Kodachrome transmits freely [Powers and Miller, 1963].

Figure 2.7 shows the three-color integral densities resulting from the series of flooding exposures (density vs. log of exposure, also known as the H&D curves [Hurter and Driffield, 1890] or the characteristic curves). As mentioned in Section 2.2.4, these measurements differ from the analytical densities, due to the overlap of the dye spectral

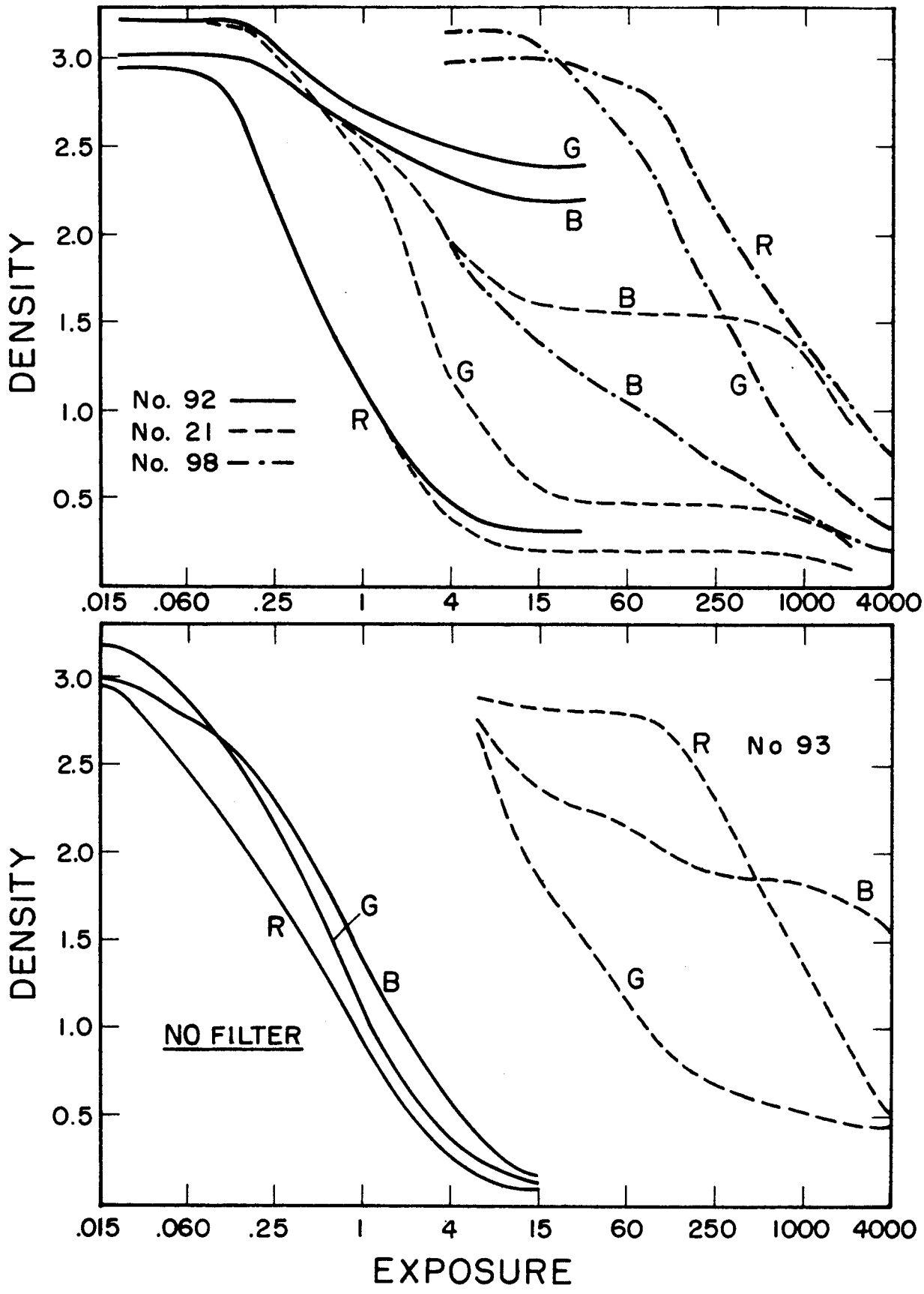


Fig. 2.7 H&D curves for flooding exposures. One unit of exposure is equivalent to 1 msec at f/4 with an illumination of 520 footcandles.

densities in the three layers. In no case does the density go to zero, due to the density of the film base.

Figure 2.7 makes apparent the absorption cross-talk effects mentioned in Section 2.2.4. For the Wratten filter No. 21 case, the densities to green and red level off at 0.45 and 0.20, respectively, at high exposure levels (even though the green-absorbing and red-absorbing dyes are eliminated) due to a small amount of green and red absorption by the remaining blue-absorbing dye layer. For the Wratten filter No. 92 case, the drop in density to blue and to green is due to the decreasing blue and green absorption by the red-absorbing layer as it is increasingly eliminated. For the Wratten filter No. 98 case, the decrease in density to blue results from a decrease in absorption in the blue-absorbing layer, up to an exposure level of about 250 units (see Fig. 2.7), at which point the blue-absorbing layer is eliminated; above an exposure level of 250 units, the decrease in density to blue results from a decrease in blue absorption by the green-absorbing and red-absorbing layers. Taking into consideration the absorption cross-talk effects mentioned above, the exposures at which the various layers are eliminated are given in Table 2.3.

As mentioned in Section 2.3.2, another problem is exposure cross-talk which results from the overlap of the spectral sensitivities of the three layers (Fig. 2.6d). As can be seen from the exposures through Wratten filter No. 92, the red-absorbing layer can be exposed completely independently of the blue-absorbing and green-absorbing layers (i.e., no exposure cross-talk). Through Wratten filter No. 93, the green-absorbing layer can be exposed completely independently of the blue-

Filter	Layers eliminated	Exposure
92	red	8
93	green	250
98	blue	125
21	red and green	60

Table 2.3. Flooding exposures for eliminating layers. Units of exposure are the same as in Figure 2.7.

absorbing layer and almost completely independently of the red-absorbing layer. But for Wratten filter No. 98, if the blue-absorbing layer is exposed to the point of elimination, then the green-absorbing and red-absorbing layers are significantly exposed. However, the exposure of the red-absorbing and green-absorbing layers resulting from the elimination of the blue-absorbing layer only adds a bias to the red and green exposures and limits the maximum attainable densities to red and to green. The result is to limit the dynamic range available in the red-absorbing and green-absorbing layers, but not to prohibit their use.

2.3.4 Density vs. Log of Exposure Curves from the CRT

In order to determine the film characteristic (H&D) curves for the particular CRT, filters, and photographic set-up, grey scales are displayed on the CRT and photographed through the filters. Flooding exposures are also made when necessary. The densities of the resulting transparencies are measured on the densitometer and the H&D curves are plotted. Figure 2.8 shows H&D curves for a few filters typically used for amplitude control and for phase control. In Figure 2.8, as well

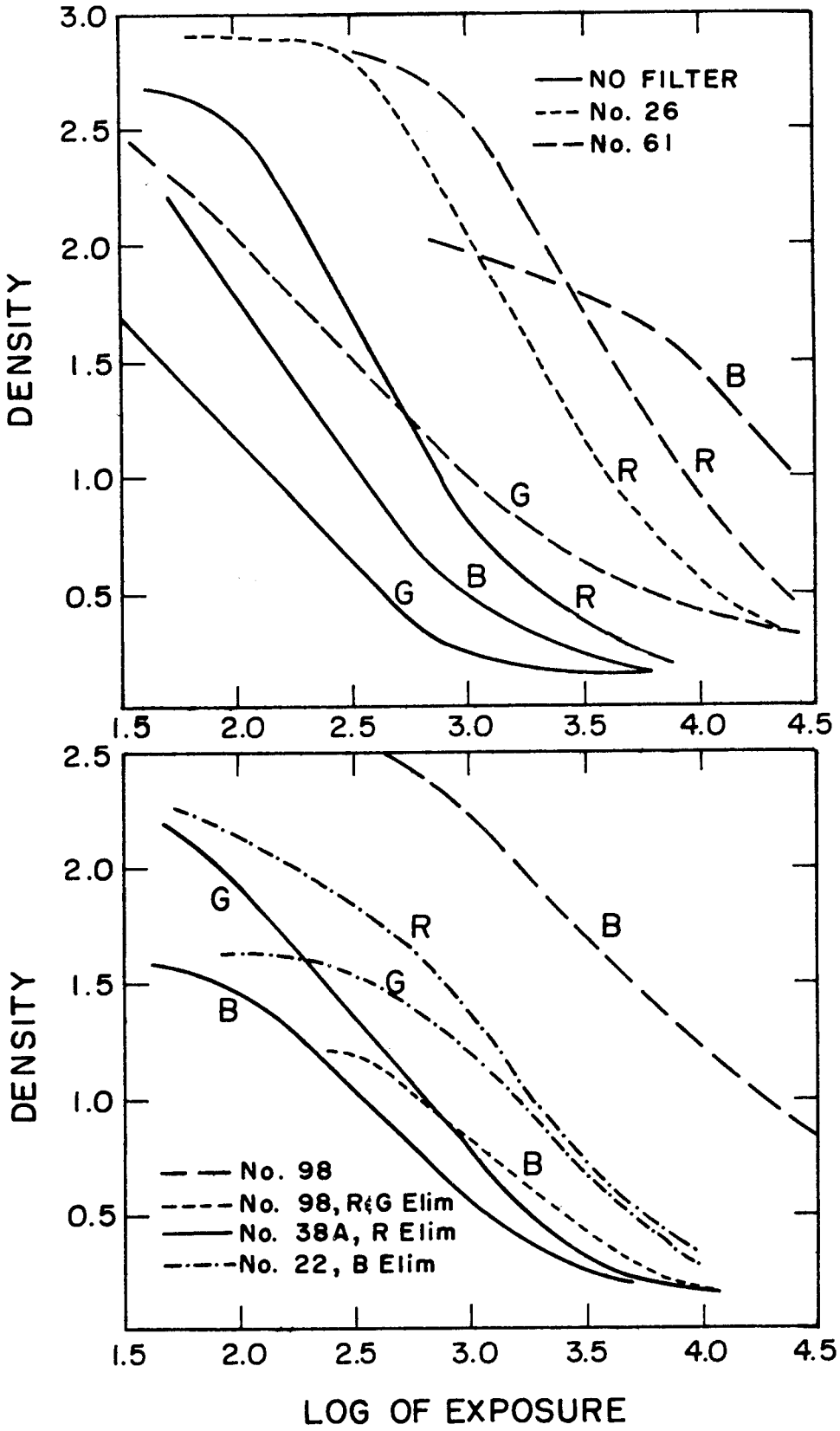


Fig.2.8 H&D curves for CRT exposures. One unit of exposure is equivalent to one computer cycle at f/8 with a CRT spot brightness of 20μW.

as in some other figures in this thesis showing measured characteristics of color films, we show only the most representative or most often used measurements made by the author. The large number of color filters and different modes of use of the film actually tested would require too much space and only serve to confuse the reader.

2.3.5 Establishing the Phase Controlling Exposure

Depending on its mode of use (Table 2.1), Kodachrome film is exposed through one of the Wratten filters listed in Table 2.2 in order to control the phase of the processed transparency. We will refer to the phase controlling filter as the "phase filter", to the corresponding layer(s) within the film that control(s) the phase transmittance as the "phase layer", and to the effective exposure received by that layer as the "phase exposure", E_p. As discussed in Section 2.3.3, to produce as pure-phase transparency it is necessary to flood the film (in a second exposure) to the color of light with which the processed transparency is to be illuminated, which we will refer to as the "flooding color".

It is assumed that phase excursions are proportional to density excursions [Lamberts, 1972], an assumption that is shown in Chapter 3 to be accurate. The more heavily exposed parts of Kodachrome II are thinner (as well as less dense), yielding by our convention a positive phase shift. Therefore, we can use the H&D curves of Figure 2.8 to determine the phase as a function of exposure, to within a proportionality constant. It simplifies the mathematics if the phase exposure is made over the linear part of the H&D curve. From Figure 2.8 it can be seen that the H&D curves do generally have a linear region of considerable extent.

If, on the linear region of the H&D curve, an exposure range E_{pmin} to E_{pmax} results in 2π radians of phase shift, then the phase shift resulting from an exposure of E_p is

$$\psi = \frac{2\pi}{\Delta \log E_{pm}} \cdot (\log E_p - \log E_{pmin}) \qquad (2\text{-}2)$$

where $\Delta \log E_{pm} = \log E_{pmax} - \log E_{pmin} = \log(E_{pmax}/E_{pmin})$ and the phase shift at exposure E_{pmin} is arbitrarily defined to be zero. Since the phase shift is linear with the log of exposure, all exposure ranges E_{pmin} to E_{pmax} on the linear region of the H&D curve that result in 2π radians of phase shift have the same value of $\Delta \log E_{pm} = \log(E_{pmax}/E_{pmin})$. Hence, the determination of the phase matching condition reduces to finding the value of $\Delta \log E_{pm}$. Solving (2-2) for $\log E_p$, the exposure that results in the desired phase shift ψ is given by

$$\log E_p = \log E_{pmin} + (\psi/2\pi)\Delta \log E_{pm} \qquad (2\text{-}3a)$$

or

$$E_p = (E_{pmin}) \cdot 10^{(\psi/2\pi)\Delta \log E_{pm}} . \qquad (2\text{-}3b)$$

The phase-matching condition can be determined by displaying and photographing an appropriate pattern through the phase filter (subsequently, the film must be re-exposed to the flooding color). Equation (2-3) is used to determine the exposure pattern that gives the desired pattern of phase shift. Since $\Delta \log E_{pm}$ is not known, exposures are made with test values of E_{pmin} and E_{pmax}; and $\Delta \log E \equiv \log E_{pmax} - \log E_{pmin}$ is used in place of $\Delta \log E_{pm}$ in (2-3). The resulting phase pattern can be a sawtooth pattern, a sinusoidal pattern, or any phase-only hologram with a known diffraction pattern. By measuring the intensities in the different orders of diffraction in the reconstruction

of the phase transparencies, the phase-matching condition (the value of $\Delta \log E_{pm}$) can be established.

Let $t(x)$ be the complex amplitude transmittance of the transparency, which is assumed here to be constant in the y direction. If $t(x)$ is periodic, with period L, then it can be decomposed into the Fourier series [Matthews and Walker, 1970] (physically, the transmitted wave is decomposed into plane waves):

$$t(x) = \sum_{n=-\infty}^{\infty} \alpha_n \, e^{i2\pi nx/L} \qquad (2\text{-}4)$$

where

$$\alpha_n = \frac{1}{L} \int_{-L/2}^{L/2} t(x) \, e^{-i2\pi nx/L} \, dx \quad . \qquad (2\text{-}5)$$

The intensity of the light from the illuminated transparency falling into the n^{th} order of diffraction (in the Fourier plane) is proportional to $|\alpha_n|^2$. As discussed in Appendix B, each spot of light in the Fourier plane is a point convolved with a sinc function of width depending upon the width of the transparency, giving each spot a finite diameter. For the sawtooth phase-grating pattern shown in Figure 2.9a, the transmittance is $t(x) = e^{i\psi(x)}$, where $\psi(x) = (2\pi/L)M\cdot(x+L/2)$ over one period (that is, $x = -L/2$ to $L/2$), and M is the phase modulation coefficient for a particular grating. The phase-matching condition is met when $M = 1$, since $\psi(x)$ would then vary by 2π over one period. Using (2-5), for the sawtooth pattern we have

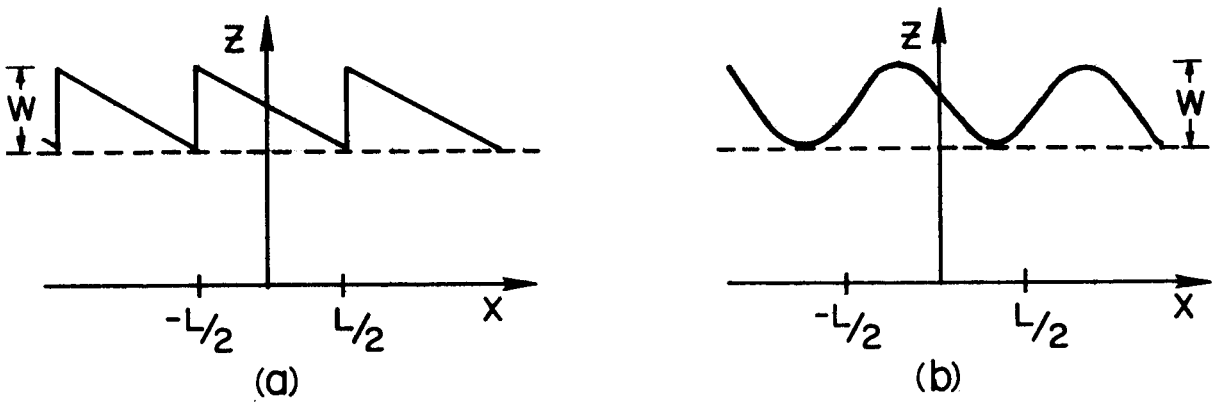


Fig. 2.9 (a) Sawtooth phase grating pattern;
(b) Sinusoidal phase grating pattern.

$M = (\text{refractive index} - 1)\, W/\lambda$

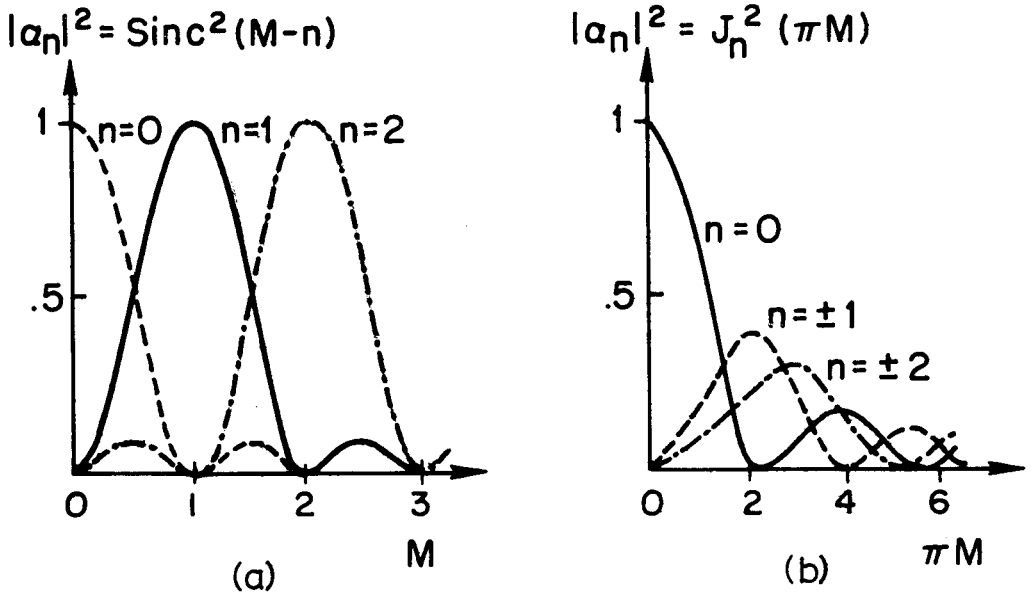


Fig. 2.10 Reconstructed intensity of n^{th} order.
(a) For sawtooth phase grating, $|\alpha_n|^2 = \text{sinc}^2(M-n)$;
(b) For sinusoidal phase grating, $|\alpha_n|^2 = J_n^2(\pi M)$.

$$\alpha_n = \frac{1}{L}\int_{-L/2}^{L/2} \exp\left\{i\,\frac{2\pi M}{L}\,(x+L/2)\right\} e^{-i2\pi nx/L}\,dx$$

$$= \frac{2}{L} e^{i\pi M}\cdot\int_{0}^{L/2} \cos[2\pi(M-n)x/L]\,dx \qquad (2\text{-}6)$$

$$= e^{i\pi M}\,\frac{\sin\pi(M-n)}{\pi(M-n)} = e^{i\pi M}\,\mathrm{sinc}(M-n) \quad .$$

Therefore, the intensity falling into the n^{th} order of diffraction will be proportional to

$$|\alpha_n|^2 = \mathrm{sinc}^2(M-n) \qquad (2\text{-}7)$$

for the sawtooth pattern. Similarly, for the sinusoidal phase-grating pattern of Figure 2.9b, where the phase is $\psi(x) = (2\pi M/2)\cdot\sin(2\pi x/L)$,

$$\exp\left\{i\left(\frac{2\pi M}{2}\right)\sin(2\pi x/L)\right\} = \sum_{n=-\infty}^{\infty} J_n(\pi M)\,e^{i2\pi nx/L} \qquad (2\text{-}8)$$

[Goodman, 1968] where J_n is the n^{th} order Bessel function of the first kind. Then, for a sinusoidal pattern, the intensity falling into the n^{th} order of diffraction is proportional to

$$|\alpha_n|^2 = J_n^2(\pi M) \quad . \qquad (2\text{-}9)$$

Figure 2.10 shows $\mathrm{sinc}^2(M-n)$ and $J_n^2(\pi M)$ as functions of M for various orders of diffraction, n. To determine the actual M of a given transparency, the measured intensities in different orders of diffraction could be compared to the curves in Figure 2.10. However, to use Figure 2.10 directly, the proportionality constant between $|\alpha_n|^2$ and the measured intensity of the n^{th} order, I_n would have to be known (since there is some residual absorption by the phase grating). Instead, we consider the ratios I_0/I_1, and I_2/I_1 (other orders

could be considered, too). Let $R_{mn}(M)$ be the calculated value of $|\alpha_m|^2/|\alpha_n|^2$ for a given value of M. The ratios I_0/I_1 and I_2/I_1 are then compared to R_{01} and R_{21}, respectively, to determine M. Plots of R_{01} and R_{21}, as a function of M (which are used to determine M), are shown in Figure 2.11 for the sawtooth pattern and in Figure 2.12 for the sinusoidal pattern.

When phase-matching is achieved, then a perfect sawtooth phase pattern would be just a prism modulo 2π radians of phase, and 100% of the light will go into the first order ($R_{01} = R_{21} = 0$), which would be very easy to interpret. A sinusoidal phase pattern yields intensities in the zero, first, and second orders in the ratios of $J_0^2(\pi) : J_1^2(\pi) : J_2^2(\pi)$, or $0.0928 : 0.0810 : 0.235$, when phase-matching is achieved, a condition which is less easy to determine experimentally.

For a sawtooth pattern, $\psi(x)$ contains many different spatial frequencies (although the fundamental frequency dominates); and, as discussed in Chapter 3, the phase response (which is proportional to M) varies somewhat with spatial frequency. Consequently, phase-matching cannot be exactly achieved for a wide band of spatial frequencies simultaneously. Furthermore, various non-linearities and noise also prevent the ratios I_0/I_1 and I_2/I_1 from going to zero even when the best phase-matching is achieved. The non-linearities and noise tend to boost the intensities I_0 and I_2 compared to I_1 when M is near 1. Consequently, the ratio I_0/I_1 tends to predict too low a value of M when $M \lesssim 1$. Similarly, the ratio I_2/I_1 tends to predict too high a value of M when $M \gtrsim 1$. When the value of R_{01} (or R_{21}) is below

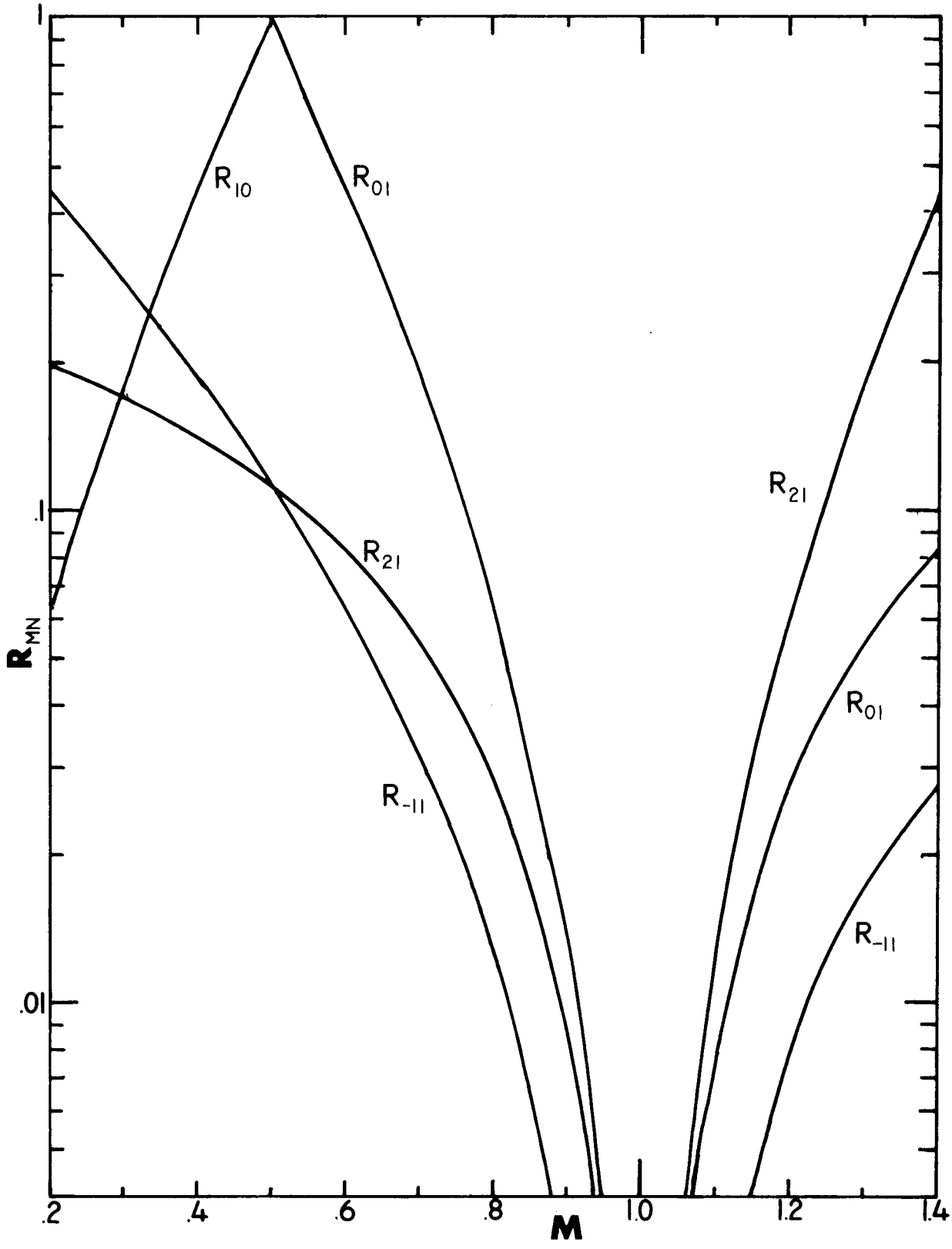


Fig. 2.11 Sawtooth phase grating: ratios of intensities of orders of diffraction. $R_{mn} = |\alpha_m|^2/|\alpha_n|^2 = \text{sinc}^2(M-m)/\text{sinc}^2(M-n) = (M-n)^2/(M-m)^2$.

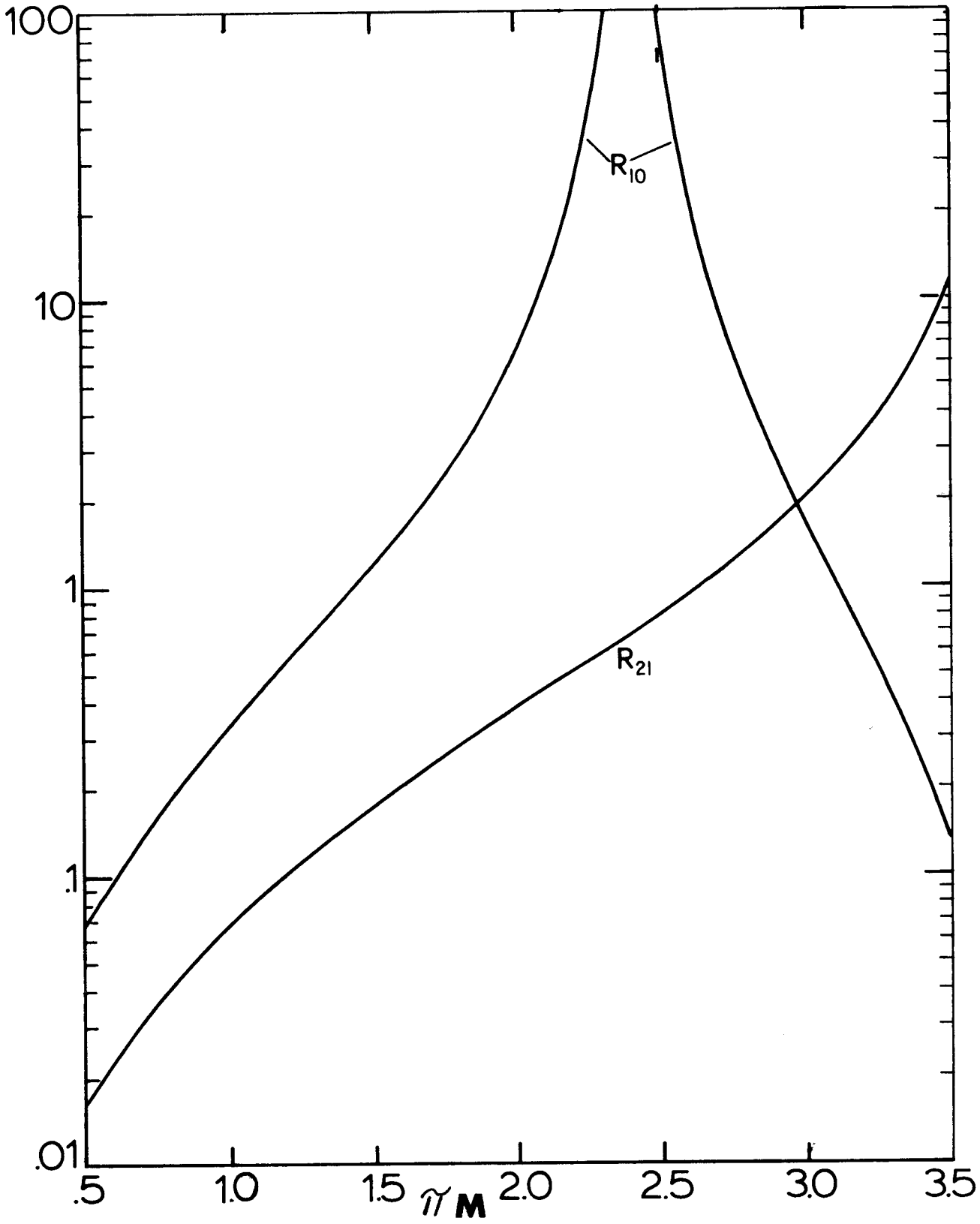


Fig. 2.12 Sinusoidal phase grating: ratios of intensities of orders of diffraction.
$R_{mn} = |\alpha_m|^2/|\alpha_n|^2 = J_m^2(\pi M)/J_n^2(\pi M)$.

about 0.05 (depending upon the amount of noise present), then I_0/I_1 (or I_2/I_1) no longer gives an accurate estimate of M, except to indicate that M is between 0.82 and 1.18 (see Figure 2.11).

Using only the linear part of the H&D curve to control the phase, the phase-matching condition is achieved by finding the value of $\Delta\log E \equiv \log(E_{pmax}/E_{pmin})$ that makes $M = 1$ (that value being $\Delta\log E_{pm}$). Theoretically, only one trial exposure would be needed to find M for a given $\Delta\log E$; then $\Delta\log E_{pm}$ would be equal to $(\Delta\log E)/M$. However, since non-linearities and noise limit the accuracy of our estimate of M, it is necessary to make a series of trial exposures in order to pick the optimum $\Delta\log E_{pm}$.

Table 2.4 shows typical experimental data used for determining the phase-matching condition. In this case, the phase filter was Wratten No. 22 (using both the green-absorbing and red-absorbing layers for phase control), and the film was re-exposed to blue light through a Wratten filter No. 98 to eliminate the blue-absorbing layer. The transparencies were reconstructed by 457.9 nm blue light from an argon-ion laser. The indicated values of M were obtained by comparing the values of I_0/I_1 and I_2/I_1 with the values of R_{01} and R_{21}, respectively, using Figures 2.10 and 2.11. Since I_2/I_1 gives a very inaccurate estimate of the value of M when I_2 is very small (for $\Delta\log E \leq .5$), those values should be ignored. From the data in Table 2.4, a best guess for the value of $\Delta\log E_{pm}$ would be $1.10 \pm .05$. To get acceptable images, phase-matching should be accurate to within about 10%, as discussed in Chapter 3.

Tables 2.5 and 2.6 summarize the experimentally determined phase-

		measured quantities		indicated M		indicated $\Delta\log E_{pm} = (\Delta\log E)/M$	
	$\Delta\log E$	I_0/I_1	I_2/I_1	by I_0/I_1	by I_2/I_1	by I_0/I_1	by I_2/I_1
	.50	1.25	.22	.48	-	1.04	-
	.80	.175	.052	.70	.70	1.14	1.14
	.90	.082	.027	.78	~1	1.15	~.90
sawtooth	.95	.066	.036	~1	~1	$\gtrsim$.95	~.95
patterns	1.00	.052	.023	~1	~1	1.00	~1.00
	1.05	.041	.021	~1	~1	1.05	~1.05
	1.10	.033	.033	~1	~1	1.10	~1.10
	1.20	.021	.080	~1	1.23	$\lesssim$1.20	.98
	.20	10.40	.082	.18	.35	1.10	.57
	.40	2.22	.12	.35	.41	1.14	.97
sinusoidal	.60	.54	.25	.53	.56	1.14	1.08
patterns	.80	.089	.50	.67	.70	1.20	1.15
	1.0	.29	1.20	.91	.88	1.10	1.15
	1.2	1.32	2.23	1.01	.96	1.19	1.24

Table 2.4. Typical experimental data for phase-matching

reconstructing wavelength, color	phase controlling layer(s)	$\Delta \log E_{pm}$	$\Delta \log E$ available	phase-matching possible?
632.8 nm red	blue	1.90±.15	1.4	no
" "	green	2.20±.20	1.6	no
" "	blue and green	0.85±.10	1.8	yes
514.5 nm green	blue	1.40±.20	1.4	yes
" "	red	~ 2.8	1.4	no
" "	blue and red	0.80±.10	1.4	yes
457.9 nm blue	green	~ 1.2	1.4	yes
" "	red	1.70±.20	1.6	marginal
" "	green and red	1.10±.05	1.4	yes

Table 2.5. Experimental phase-matching conditions for all modes of Kodachrome II

reconstructing wavelength, color	phase controlling layer(s)	$\Delta \log E_{pm}$	$\Delta \log E$ available	phase-matching possible?
632.8 nm red	blue	2.45±.10	1.3	no
" "	green	3.70±.40	1.4	no
" "	blue and green	1.20±.10	1.6	yes
514.5 nm green	blue	1.75±.05	1.4	no
" "	red	2.20±.20	1.4	no
" "	blue and red	-	-	yes
457.9 nm blue	green	4.30±.30	1.4	no
" "	red	1.85±.15	1.4	no
" "	green and red	1.50±.10	2.0	yes

Table 2.6. Experimental phase-matching conditions for all modes of Kodachrome 25

matching conditions for all modes of Kodachrome II and Kodachrome 25, respectively. The values of $\Delta\log E_{pm}$ were determined by comparing with Figure 2.11 the measured values of I_0/I_1 and I_2/I_1 that resulted from sawtooth phase patterns. "$\Delta\log E$ available" is the log of the dynamic range in exposure (over the approximately linear part of the H&D curve) of the phase controlling layer. If $\Delta\log E_{pm}$ is somewhat greater than $\Delta\log E$ available, then (2-3) can no longer be used to determine the exposure required to produce the desired phase: the exact shape of the H&D curve must be taken into account. If for a given mode $\Delta\log E_{pm}$ is much greater than $\Delta\log E$ available, then it is not possible to achieve phase-matching (2π radians of phase) for that mode; consequently, that mode could not be used to make either a ROACH or a kinoform (except for the special case mentioned in Section 3.6). As noted in the last column of Tables 2.5 and 2.6, only a few of the modes are capable of achieving phase-matching. The use of both transparent layers for phase control makes phase-matching possible in all cases. Hence it is possible to make ROACHs for all three of the primary colors (see Chapter 4).

A comparison of Table 2.5 with Table 2.6 shows that Kodachrome II is much better suited for computer holography than is Kodachrome 25, since the former exhibits greater phase variations than the latter. While a ROACH for green light can be made with Kodachrome II using only the blue-absorbing layer for phase control, it can be made with Kodachrome 25 only if both the blue-absorbing and the red-absorbing layers are used for phase control. No single Wratten filter exists that can cause the blue-absorbing and the red-absorbing layers to be equally

exposed without also significantly exposing the green-absorbing layer. Consequently, to make a ROACH for green light on Kodachrome 25 it is necessary to make two exposures just for phase control, one through a blue (Wratten No. 98) filter and one through a red (Wratten No. 26) filter. For this mode of Kodachrome 25 the double-exposure $\Delta\log E_{pm}$ was not determined; however, since use of the blue-absorbing layer alone (see Table 2.6) comes close to phase-matching, the combination of the blue-absorbing and the red-absorbing layers should make phase-matching possible. The value of $\Delta\log E_{pm}$ given for the corresponding mode of Kodachrome II in Table 2.5 was obtained from single exposures through Wratten filter No. CC50Y which also transmits green light; thus that value of $\Delta\log E_{pm}$ is meaningful only for kinoforms, for which the green-absorbing layer is eliminated.

Of the modes listed in Tables 2.5 and 2.6, only those that were actually used to make ROACHs were accurately tested. Only rough estimates of $\Delta\log E_{pm}$ were made for other modes, particularly those for which phase-matching is not possible.

2.3.6 Adding Amplitude Control

In order to produce the desired calculated wavefront, $F(u) = |F(u)|e^{i\psi(u)}$, a ROACH should have a complex transmittance, $H(u)$, that is proportional to $F(u)$. Since the maximum amplitude transmittance of any passive hologram material is 1, $F(u)$ is normalized to have a maximum value of 1 by $t_a(u) = |F(u)|/\text{Max}\{|F(u)|\}$. If t_{max} is the actual maximum amplitude transmittance of the hologram material, then the desired hologram transmittance is $H(u) = t_{max}\, t_a(u)\, e^{i\psi(u)}$. The optical density of the amplitude controlling layer, D_a, corres-

ponding to t_{max} t_a is given by

$$D_a - D_{amin} = -\log(t_a^2) = -2\log(t_a) \qquad (2\text{-}10)$$

where D_{amin} is the density corresponding to $t_{max} = 10^{-D_{amin}/2}$ and to the maximum exposure, E_{amax}. The amplitude controlling exposure, E_a, needed to achieve the desired density, D_a (and the desired t_a), is determined from the H&D curve (Fig.2.8), and(2-10).

If, for simplicity, only the linear part of the H&D curve is used, then

$$\log E_a = \log E_{amax} + B \cdot \log t_a \qquad (2\text{-}11)$$

or

$$E_a = E_{amax} \cdot t_a^B$$

where $B = 2(\log E_{amax} - \log E_{amin})/(D_{amax} - D_{amin})$ and D_{amax} is the density corresponding to the minimum exposure, E_{amin}. The hologram is limited to a dynamic range of $10^{D_{amax}-D_{amin}}$ in intensity or $10^{(D_{amax}-D_{amin})/2}$ in amplitude.

The addition of amplitude information also results in undesired phase effects, since the amplitude controlling layer contributes to the phase delay. As discussed in Chapter 3, the phase delay contributed is linear with density. To compensate for this undesired phase delay, the phase controlling exposure, E_p, is adjusted. We replace (2-3a) with

$$\log E_p = \log E_{pmin} + (\psi/2\pi)\Delta\log E_{pm} - \beta\log(E_a/E_{amin}) \qquad (2\text{-}12)$$

where the factor β is determined experimentally. If $\log E_p$ computed by (2-12) is less than $\log E_{pmin}$, then $\Delta\log E_{pm}$ is added to $\log E_p$ in order to preserve the phase modulo 2π radians in the phase controlling layer.

As discussed in Chapter 3, for optimal results the effect of the phase controlling layer on the amplitude transmittance may also be compensated for; however, good results can be obtained without that compensation.

2.4 Experimental Results: Images

Due to inaccuracies in the measured densities (integral vs. analytical) and chemical and physical interactions between the different layers (interimage effects) during processing [Evans, Hansen, and Brewer, 1953], the measured values of E_{amax}, E_{amin}, D_{amax}, and D_{amin} used in (2-11) are considered to be only approximate. To determine the best set of exposure parameters, the measured values of E_{amax}, $\log(E_{amax}/E_{amin})$, and $(D_{amax}-D_{amin})$ are used as a starting point in a trial-and-error search. The search for the best set of exposure parameters (including β) is accomplished by varying one exposure parameter at a time to converge on the best image. Once the optimal exposure parameters are found, the calibration procedure is completed; and the same exposure parameters can be used to make ROACHs reliably and with good repeatability.

Figure 2.13 shows an image from a ROACH illuminated by red light. The object was specified by a 32×32 array of 1's and 0's and was phase-coded using the "hyperbolic" phase-code [Chu, 1974]. The basic 32×32 element ROACH, shown in Figure 2.25, was replicated 4×4 times, resulting in the dot structure of the image (see Appendix B). ROACHs for illumination by blue light, by green light, and by red light, and the full-color image produced by them, will be shown in Chapter 4.

Figure 2.14 shows a similar image from a kinoform. The phase controlling layer of the kinoform was exposed in the same manner as the

Fig. 2.13 Reconstructed image from a ROACH.

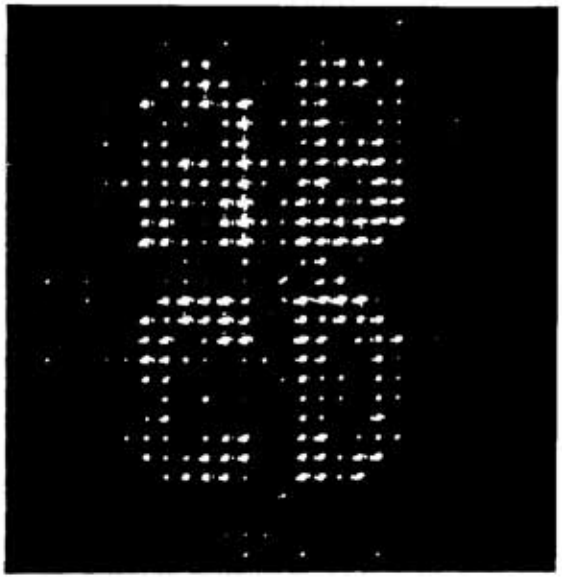


Fig. 2.14 Reconstructed image from a kinoform.

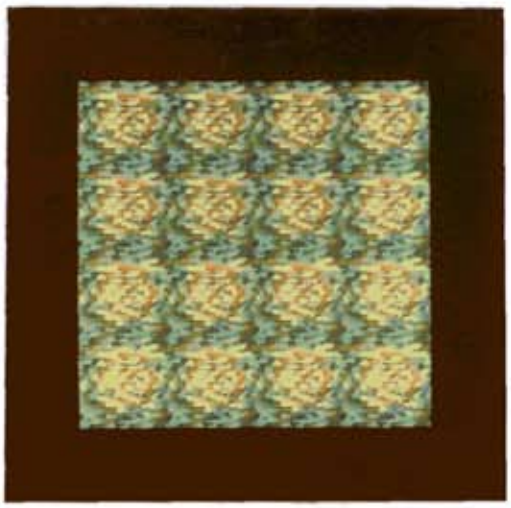


Fig. 2.15 A 32×32 element ROACH repeated in a 4×4 array (magnified 8×).

phase controlling layer in the ROACH, except that no compensation was required for the phase due to the amplitude controlling layer. The amplitude controlling layer in the kinoform was eliminated, as discussed in Section 2.3.3. The variations in the intensities of the dots in the kinoform image are primarily a result of the loss of the hologram amplitude information. The improvement of the image from the ROACH (Fig. 2.13) over the image from the kinoform (Fig. 2.14) is a result of the control of amplitude as well as phase.

CHAPTER 3

PROPERTIES AND LIMITATIONS OF KODACHROME AS A HOLOGRAPHIC MATERIAL

In Chapter 2, references were made to various considerations that effected the use of Kodachrome film as a holographic material, such as resolution, cross-talk and noise. In this chapter we will investigate these and other considerations in more detail, determine to what extent they limit the image quality obtainable from a ROACH, and describe methods of compensating for some of the deficiencies.

3.1 The Relationship Between Phase and Density

In Chapter 2, it was assumed that phase variations in a layer are proportional to the density variations of that layer (Eq.2-2) [Lamberts, 1972; Smith, 1968]. For example, to blue light, the phase variations due to the red-absorbing layer are assumed proportional to the density variations in that layer, as measured with red light (ideally, the red-absorbing layer is transparent to blue light, so there would be no variations in density to blue). This assumption was tested in a manner similar to the method of determining the phase-matching condition described in Section 2.3.5. The assumption of proportionality between phase and density is equivalent to assuming that M is proportional to $\Delta \log E$ over the linear portion of the H&D curve. That M is, indeed, proportional to $\Delta \log E$ can be seen from Figure 3.1a, a plot of

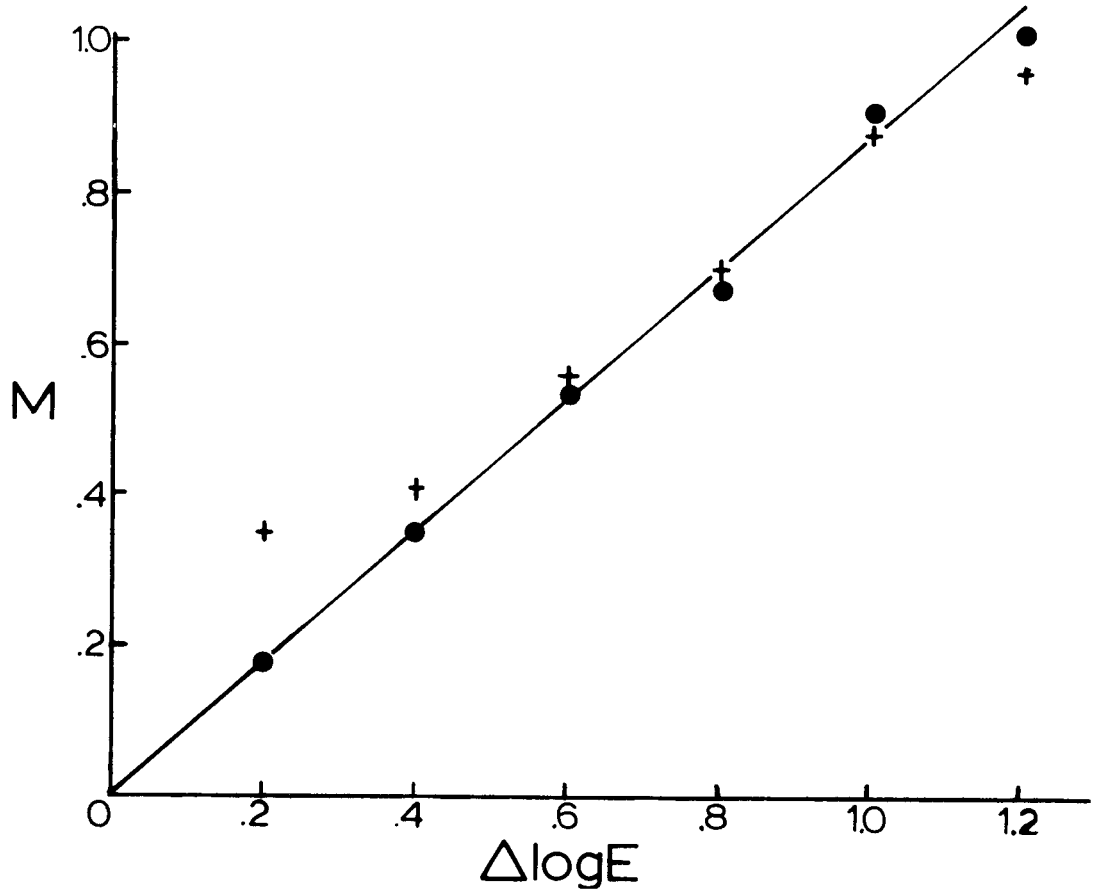


Fig. 3.1a Phase modulation coefficient M vs. ΔlogE for the green-absorbing and red-absorbing layers of Kodachrome II, as indicated by the values of I_0/I_1 ("·") and by the values of I_2/I_1 ("+").

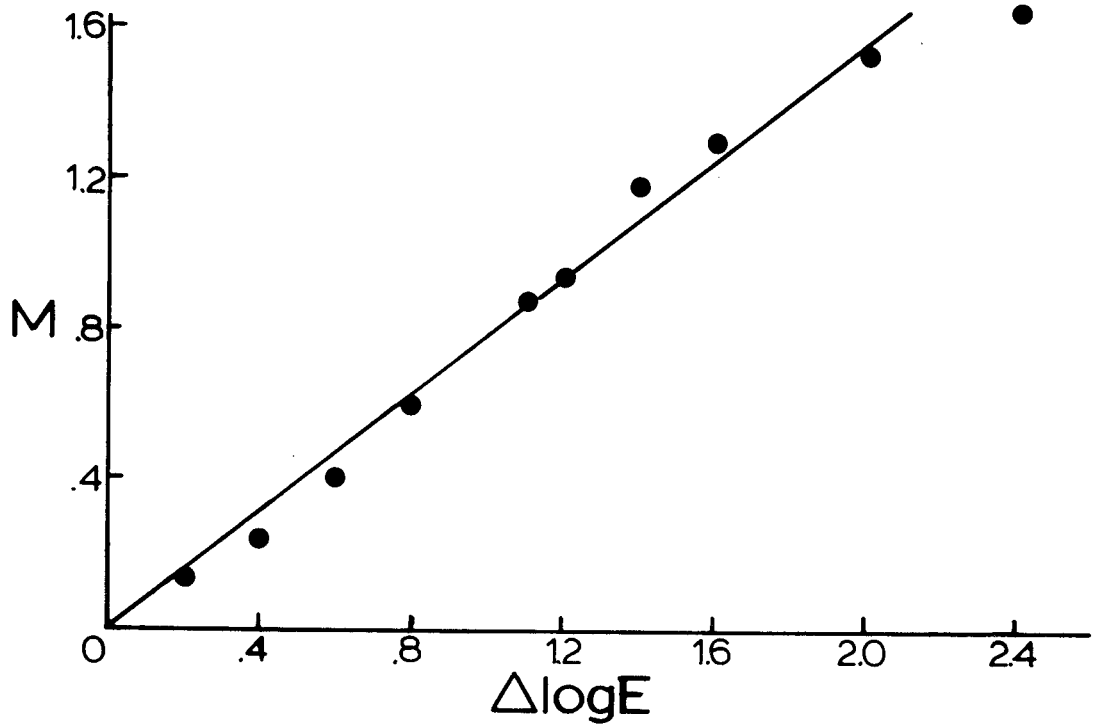


Fig. 3.1b M vs ΔlogE for the blue-absorbing and green-absorbing layers of Kodachrome 25.

M vs. Δlog E for the sinusoidal patterns listed in Table 2.4 (The I_2/I_1 data point at Δlog E = 2 is meaningless because of the large error due to noise in measuring a very weak I_2.) This particular test was performed on Kodachrome II using both the red-absorbing and the green-absorbing layers for phase control. A similar test was performed on Kodachrome 25, using both the blue-absorbing and the green-absorbing layers for phase control for illumination with red light. The results are graphed in Figure 3.1b, from which it is seen that the proportionality rule holds also for those two layers of Kodachrome 25.

Another test was made to see if the proportionality rule held in the shoulder and toe regions of the H&D curve as well as on the linear part. Sawtooth patterns were exposed using a constant $\Delta\log E = \log(E_{pmax}/E_{pmin})$, but E_{pmax} was varied over the entire range of the H&D curve. Two different constant values of Δlog E were used, 0.3 and 0.6. Exposures were made on Kodachrome 25 using both the blue-absorbing and the green-absorbing layers for phase control, and the resulting sawtooth phase gratings were read out with red light. The density excursions, ΔD, due to the exposure ranges Δlog E were determined from the H&D curve. Figure 3.2 shows the values of the phase modulation, M, plotted against the density excursions, ΔD. Since all of the data points fall near a straight line passing through the origin, the rule that phase excursions are proportional to density excursions is shown to hold true for the entire extent of the H&D curve.

The mechanisms causing variations in phase are only partially understood. Both variations in refractive index (the index image) and variations in surface relief height (the relief image) contribute to

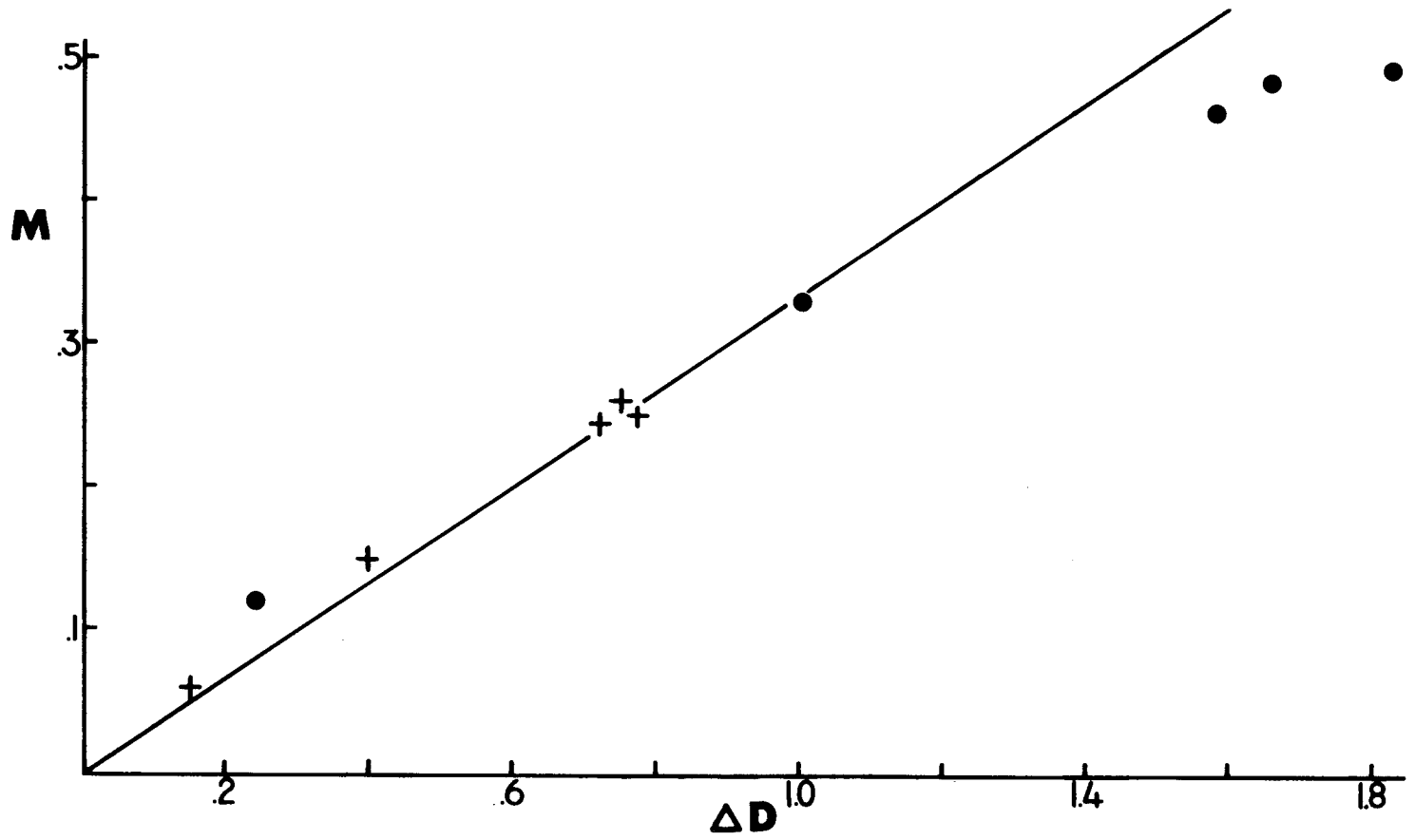


Fig. 3.2 Phase modulation coefficient M vs ΔD over the entire H&D curve. ΔlogE was fixed to 0.3 ("+") or 0.6 ("·"). ΔD is the sum of the density excursions to blue light and to green light.

the phase, and for commonly used photographic emulsions, the relief image dominates [Lamberts, 1972]. The relief image is the result of a tanning action that occurs during processing [Smith, 1968].

We determined the phase modulation coefficient, M, for phase gratings inserted in a Xylene liquid gate. Assuming Xylene has a refractive index that is nearly equal to the refractive index of the film emulsion, then the effect of the relief image is cancelled, leaving only the contribution due to the index image, allowing the measurement of the phase due to the index image alone. In this way it was found that the relief image dominates for all modes of Kodachrome II, contributing more than twice the phase variations of the index image.

3.2 The Effect of Phase Mismatch

Suppose that the desired phase transmittance is $e^{i\psi(u,v)}$, but that, due to inaccurate phase-matching, the actual phase transmittance is $e^{iM\psi(u,v)}$, where $M \neq 1$. The purpose of this section is to discuss the degree of image deterioration when M differs from 1.

The simplest case is an image that consists of a single off-axis point, such as that produced by the sawtooth phase pattern of Figure 2.9a. The results of a phase mismatch are given by (2-7):

$$|\alpha_n|^2 = \text{sinc}^2(M-n) \tag{3-1}$$

which is plotted in Figure 2.10a. For M near 1, $\text{sinc}^2(M-1) = \sin^2\pi(M-1)/\pi^2(M-1)^2 \simeq 1 - (\pi^2/3)(M-1)^2$. Therefore, for a slight phase mismatch, the percentage of the light that goes into noise is approximately $(\pi^2/3)(M-1)^2$. That noise is 12.5% if M is off by 20%, 3.2% if M is off by 10%, and 0.82% if M is off by 5%. Consequently, very good results can be obtained if phase-matching is within 10%.

A more general derivation is given by Kermisch [1970] for the case of a kinoform. He concludes that the intensity of the desired image is proportional to $\text{sinc}^2(M-1)$ and that the intensity of the n^{th} spurious image due to phase mismatch is proportional to $\text{sinc}^2(M-n)$ (his factor $T_m(c)$ in both cases). These results are in agreement with the sawtooth phase-grating case. The zero order spurious image is simply an on-axis spot and is proportional to $\text{sinc}^2 M$. The minus first order image (referred to by Kermisch as a virtual image) is the conjugate of the desired image. Higher order spurious images are multiple convolutions of the desired image with itself.

The results from our sawtooth phase grating example and from Kermisch's analysis were born out by our experiments. Figure 3.3 shows images from kinoforms, each with a different phase modulation, M. The images shown in Figures 3.3a, b, and c are of the letters STAN above the optical axis and FORD below the optical axis, with a space between the letters STAN and the letters FORD. In the image shown in Figure 3.3a, which resulted from $M \simeq .2$, the on-axis spot (the zero order image) contains most of the energy and the minus first order is clearly visible (an upside-down FORD between the desired STAN and FORD), as predicted by the $\text{sinc}^2(M-n)$ factor (see Fig.2.10a). In the image shown in Figure 3.3b, which resulted from $M \simeq .5$, both the on-axis spot and the minus first order are substantially reduced. In the image shown in Figure 3.3c, which resulted from $M \simeq .9$, both the on-axis spot and the minus first order are greatly reduced. The image shown in Figure 3.3d is of the letters STAN and FORD, but without any space between them (as in Figures 3.3a, b, and c). In this case M

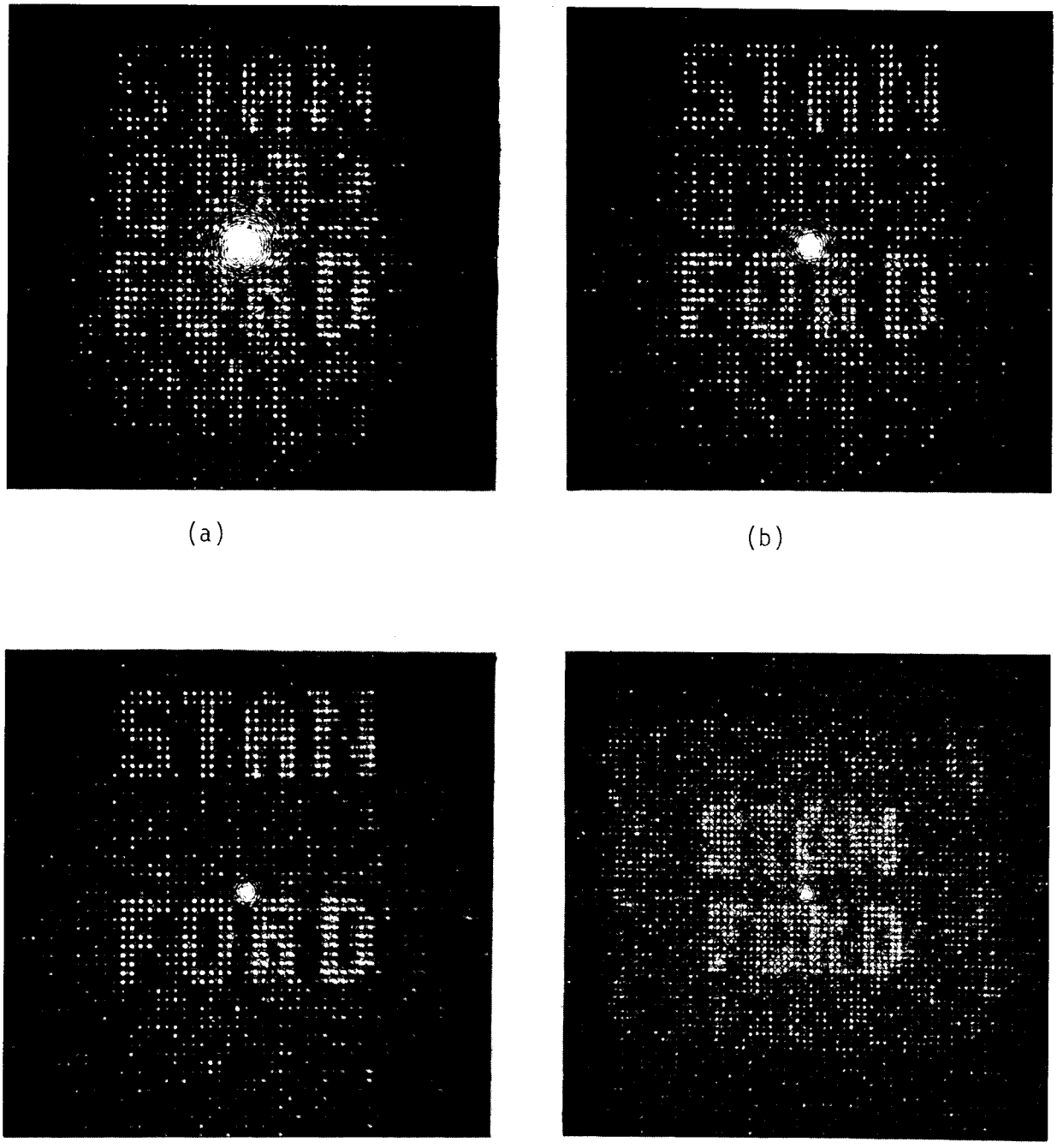


Fig. 3.3 Images from kinoforms with phase mismatch.
(a) M < .5 (severely undermodulated); (b) M ~ .5 (undermodulated);
(c) M ~ 1 (near phase matching); (d) M > 1 (overmodulated).

is greater than 1.4, and the second order image is clearly visible. The second order image is the convolution of the first order (desired) image with itself. Consequently, the second order image is double the width of the first order image and has the appearance of a rectangle of noise.

3.3 Amplitude-Phase Cross-Talk

Amplitude-phase cross-talk occurs when amplitude information results in variations in phase (amplitude-to-phase cross-talk) or when phase information results in variations in amplitude (phase-to-amplitude cross-talk). The two major sources of amplitude-phase cross-talk are the inability to address the amplitude controlling and the phase controlling layers completely independently (exposure cross-talk) and the non-ideal nature of the amplitude controlling and the phase controlling processed layers (dye layer cross-talk).

3.3.1 Amplitude-to-Phase Cross-talk

Let the desired hologram complex transmittance be given by $F = |F|e^{i\psi}$ and the desired image be given by f $(F \supset f)$, where the spatial coordinates are understood. When amplitude-to-phase cross-talk is present, then the hologram will have a complex transmittance of the form

$$H = |F|e^{i[\psi+\beta C(|F|)]} = F\, e^{i\beta C(|F|)} \qquad (3\text{-}2)$$

where $C(|F|)$ determines the nature of the cross-talk and β the strength of the cross-talk. Transforming, and using the convolution theorem [Bracewell, 1965],

$$H = F\, e^{i\beta C(|F|)} \supset h = f * \mathfrak{F}\left\{e^{i\beta C(|F|)}\right\} \qquad (3\text{-}3)$$

where h is the reconstructed image, $\mathfrak{F}\{\cdot\}$ denotes the Fourier trans-

form operation, and * denotes the convolution operation. For weak cross-talk, $\beta C(|F|) \ll 1$, and the exponential can be expanded,

$$e^{i\beta C(|F|)} \simeq 1 + i\beta C(|F|) - \tfrac{1}{2}\beta^2 C^2(|F|) + \ldots \qquad (3\text{-}4)$$

where the higher order terms may be dropped. Inserting (3-4) into (3-3), we obtain

$$h = f + i\beta f * \mathfrak{F}\{C(|F|)\} - \tfrac{1}{2}\beta^2 f * \mathfrak{F}\{C(|F|)\} * \mathfrak{F}\{C(|F|)\} + \ldots . \qquad (3\text{-}5)$$

For simplicity, suppose that $C(|F|) = |F|$. Then it is necessary to know the shape of $\mathfrak{F}\{|F|\}$. Let $e^{i\psi} \supset k$. Then the convolution theorem yields

$$F = |F|e^{i\psi} \supset \mathfrak{F}\{|F|\} * \mathfrak{F}\{e^{i\psi}\} = \mathfrak{F}\{|F|\} * k = f . \qquad (3\text{-}6)$$

But k is just the image obtained from a kinoform. Kinoforms of diffuse (random phase-coded) objects are known to produce easily recognizable images in which the noise is only 21% of the total energy [Kermisch, 1970]. That is, k does not differ drastically from f. Therefore, since by (3-6), $\mathfrak{F}\{|F|\} * k = f$, then $\mathfrak{F}\{|F|\}$ must be a very sharply peaked function (somewhat like a delta function with broad, irregular sidelobes). Consequently, not only is the first term in (3-5) exactly the desired image, but also the second term in (3-5) should look very much like the desired image. Image quality would be drastically reduced only if at least a few terms in (3-5) were significant, which would occur if the excursions of $\beta C(|F|)$, that is, excursions in phase due to amplitude-to-phase cross-talk, were on the order of 1 radian. In practice, these phase excursions are usually greater than 1

radian; therefore, compensation for amplitude-to-phase dye-layer cross-talk is a necessity (and was performed, as mentioned in Section 2.3.6).

3.3.2 Phase-to-Amplitude Cross-talk, Sawtooth Grating Example

When phase-to-amplitude cross-talk is present, then the complex transmittance of the hologram is of the form

$$H = |F|e^{i\psi-\beta\psi} = F\,e^{-\beta\psi} \tag{3-7}$$

where we have used the phase-to-density proportionality rule and the fact that amplitude transmittance equals $10^{-\text{density}/2}$. Then we have

$$H = F\,e^{-\beta\psi} \supset f*\mathfrak{F}\{e^{-\beta\psi}\} \simeq f-\beta f*\mathfrak{F}\{\psi\} + \tfrac{1}{2}\beta^2 f*\mathfrak{F}\{\psi\}*\mathfrak{F}\{\psi\} + \ldots \tag{3-8}$$

for small β.

To give a more concrete example, consider the sawtooth phase grating pattern of Figure 2.9a. In the presence of phase-to-amplitude cross-talk and a mismatch of phase, the complex transmittance would be of the form (over one period, $L/2 \leq x \leq L/2$)

$$t(x) = e^{-(x+L/2)D_s/2}\, e^{i(2\pi M/L)(x+L/2)} \tag{3-9}$$

where $D_s L$ is the density excursion due to phase-to-amplitude cross-talk. Inserting (3-9) into (2-5) and evaluating the integral, we obtain

$$\alpha_n = \frac{-e^{-D_s L/2+i\pi M}}{(D_s L/2)+i2\pi(M-n)}\left[e^{-D_s L/4-i\pi(M-n)} - e^{+D_s L/4+i\pi(M-n)}\right]. \tag{3-10}$$

Therefore

$$|\alpha_n|^2 = \frac{e^{-D_s L/2}}{(D_s L/2)^2 + \left[2\pi(M-n)\right]^2}\left[e^{-D_s L/2} + e^{+D_s L/2} - e^{-i2\pi(M-n)} - e^{+i2\pi(M-n)}\right]$$

$$= \frac{e^{-D_s L/2}}{(D_s L/2)^2 + \left[2\pi(M-n)\right]^2}\left[2\cosh(D_s L/2) - 2\cos 2\pi(M-n)\right]$$

$$= \frac{2\, e^{-D_s L/2}}{(D_s L/2)^2 + \left[2\pi(M-n)\right]^2}\left[(\cosh(D_s L/2) - 1) + 2\sin^2\pi(M-n)\right] . \qquad (3\text{-}11)$$

Note that for $D_s L = 0$ (no phase-to-amplitude cross-talk), (3-11) reduces to (2-7). For perfect phase-matching $(M = 1)$, (3-11) reduces to

$$|\alpha_n|^2 = \frac{2\, e^{-D_s L/2}\left[\cosh(D_s L/2) - 1\right]}{(D_s L/2)^2 + \left[2\pi(n-1)\right]^2} . \qquad (3\text{-}12)$$

The most immediately noticeable effect of increasing $D_s L$ is the decrease in intensity of the reconstruction, due to increased absorption by the hologram. This decrease in intensity, however, does not degrade the quality of the image. The total intensity transmitted by the hologram is given by

$$\frac{1}{L}\int_{-L/2}^{L/2} |t(x)|^2 dx = \frac{1}{L}\int_{-L/2}^{L/2} e^{-(x+L/2)D_s} dx = \frac{e^{-D_s L/2}\sinh(D_s L/2)}{(D_s L/2)} . \qquad (3\text{-}13)$$

The normalized intensity (the fraction of the total intensity) of the n^{th} order, $|\alpha_n'|^2$, is obtained by dividing (3-12) by (3-13):

$$|\alpha_n'|^2 = \frac{2\left[\cosh(D_s L/2) - 1\right]}{\sinh(D_s L/2)} \cdot \frac{(D_s L/2)}{(D_s L/2)^2 + 2\pi(n-1)^2} . \qquad (3\text{-}14)$$

Figure 3.4 shows a plot of $|\alpha_n'|^2$ vs. D_sL for $n = 1, 2$ and 3. The values of $|\alpha_n'|^2$ for $n = 0, -1, -2$, etc., are the same as $|\alpha_n'|^2$ for $n = 2, 3, 4$, etc., respectively. Also shown in Figure 3.4 is $1 - |\alpha_1'|^2$, the fraction of light going into noise. For cross-talk that is not too strong ($D_sL < 1$), (3-14) reduces to

$$|\alpha_n'|^2 \simeq \frac{(D_sL/2)^2 - (D_sL/2)^4/12}{(D_sL/2)^2 + 2\pi(n-1)^2} \qquad (3\text{-}15)$$

where use was made of the expansions $\cosh x \simeq 1 + x^2/2 + x^4/24$ and $\sinh x \simeq x + x^3/6$. Then the fraction of light going into the desired image is given by

$$|\alpha_1'|^2 \simeq 1 - (D_sL/2)^2/12 \quad . \qquad (3\text{-}16)$$

Therefore, the fraction of light going into noise is $(D_sL)^2/48$. For example, if phase-to-amplitude cross-talk results in density excursions of 1.0 (equivalent to an amplitude transmittance factor of $\sqrt{10} \simeq 3.16$), then only 2% of the light goes into noise.

We conclude from this analysis that except for extreme cases (density excursion $D_sL > 1$), phase-to-amplitude cross-talk can be ignored (except, perhaps, for certain applications for which the amplitude transmittance must be very accurate). While this analysis treated only the special case of the sawtooth phase grating, we believe that it is representative of the general case. The sawtooth phase grating example did prove to be representative of the general case for the phase mismatch analysis of Section 3.2.

The density excursions due to the phase controlling layer can be predicted from the dye spectral density curves of Figure 2.3. For example, when the green-absorbing and red-absorbing layers both have

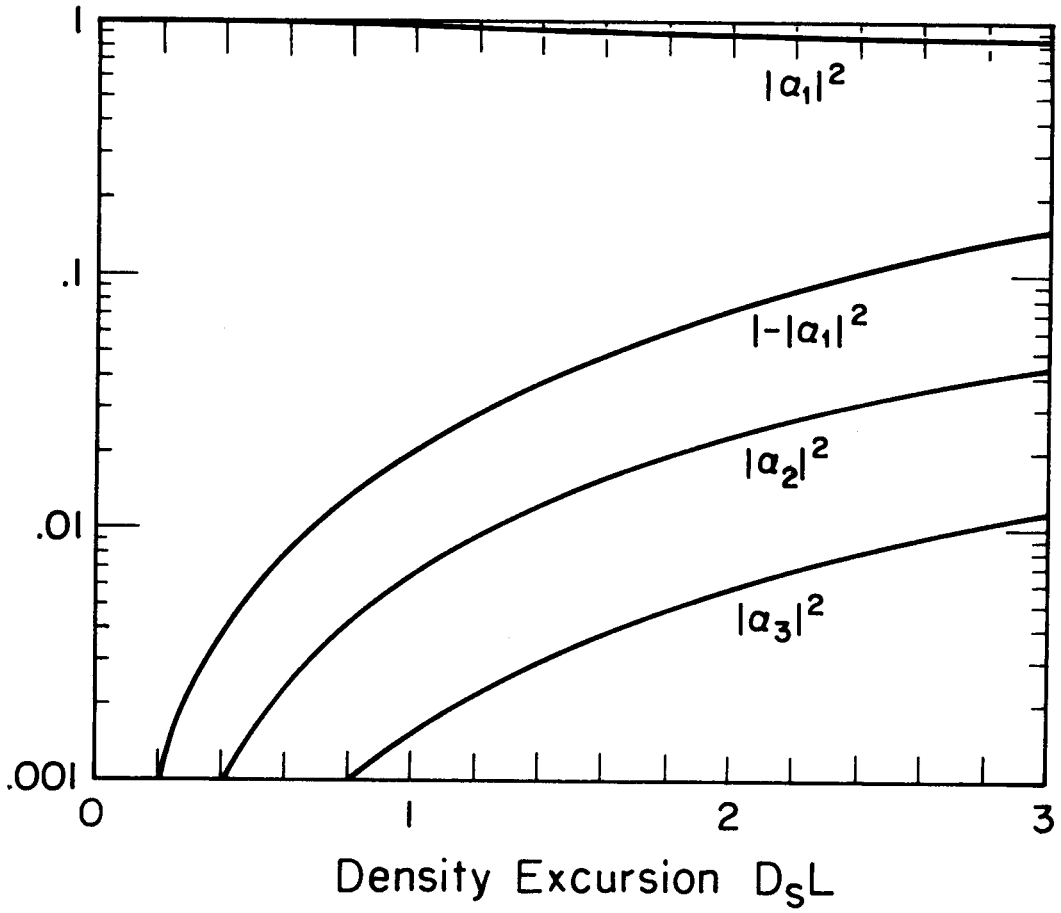


Fig. 3.4 Sawtooth phase grating suffering phase-to-amplitude cross-talk: relative intensity of the n^{th} order of diffraction vs density excursion. $1 - |a_1|^2$ is the total fraction of noise.

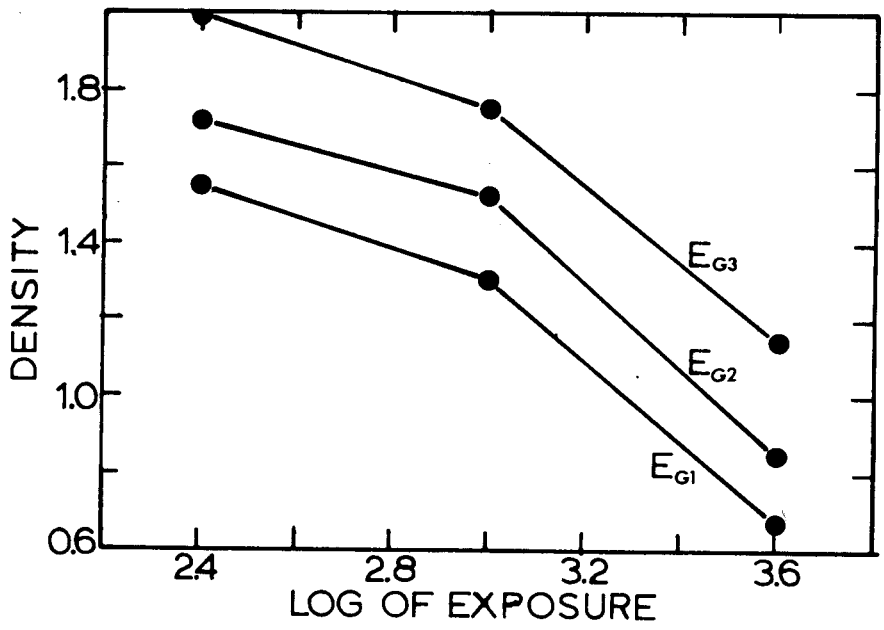


Fig. 3.5 Density to blue light vs log of exposure to blue light, for different constant exposures to green light.

a density of 1.0 to their respective colors of light, then they also contribute 0.24 and 0.03, respectively to the density to blue 457.9 nm light.

The density excursions due to the phase controlling layer can also be determined from the measured H&D curves, such as those in Figure 3.5, which shows portions of H&D curves representing density to blue light vs. log of exposure through a blue Wratten filter No. 98. For this experiment, Kodachrome II was also (in addition to the blue exposure) exposed to green light through a Wratten filter No. 61 and was flooded with red light. In Figure 3.5, each curve of the family of H&D curves resulted from a different exposure to green light. The difference in density to green light between the samples represented by the top and bottom curves is 1.66, and the average difference in density to blue light between the top and bottom curves is .456. Therefore, the change in density to blue due to a 1.0 change in density to green (in the green-absorbing layer) is .456/1.66 = 0.27. This value is greater than the predicted value of 0.24 (which we consider to be more accurate) because a band of wavelengths (rather than monochromatic light) was used to measure the densities for the H&D curves.

The amount of phase-to-amplitude cross-talk can be determined as follows. The determination of the density excursions due to the phase controlling layer (as described above) yields the density to the reconstruction wavelength as a function of the density of the phase-controlling layer to its corresponding color of light. The relationship between the density of the phase controlling layer (to its corresponding color of light) and the phase controlling exposure is given by the H&D

curve (as in Figure 2.8). The relationship between the phase controlling exposure and the resulting phase is determined as in Section 2.3.5. By combining the relationships above, the relationship between the density due to cross-talk and the phase (or the phase controlling exposure, if desired) is determined.

3.4 Compensation for Amplitude-Phase Cross-Talk

Of the various types of cross-talks mentioned in Section 3.3, the most severe is the amplitude-to-phase cross-talk due to phase excursions in the absorbing layer (a dye laser cross-talk). In our experiments, that was the only cross-talk for which we compensated, as discussed in Section 2.3.6. For optimum results, the phase-to-amplitude cross-talk can also be compensated for.

For simplicity, suppose that both the amplitude controlling and the phase controlling layers are made on the linear parts of their respective H&D curves. We will use the same notation here as in Section 2.3.5 and 2.3.6. Using the phase-to-density proportionality rule, the total phase contributed by both layers is

$$\psi = 2\pi \log(E_p/E_{pmin})/\Delta\log E_{pm} + \beta_a' \log(E_a/E_{amin}) \qquad (3\text{-}17)$$

where $\Delta\log E_{pm} = \log(E_{pmax}/E_{pmin})$ if E_{pmin} and E_{pmax} are the minimum and maximum phase controlling exposures that accomplish phase-matching, β_a' is an amplitude-to-phase cross-talk coefficient, and the first and second terms are the contributions to phase due to the phase layer and the amplitude layer, respectively. Solving (3-17) for $\log E_p$ results in (2-12):

$$\log E_p = \log E_{pmin} + (\psi/2\pi)\Delta\log E_{pm} - \beta\log(E_a/E_{amin}) \quad (2\text{-}12)$$

where $\beta = (\beta_a'/2\pi)\Delta\log E_{pm}$. Similarly, the density contributed by both layers is given by

$$D_a - D_{min} = -2\log t_a = -\frac{2}{B}\log(E_a/E_{amax}) - \beta_p\log(E_p/E_{pmax}) \quad (3\text{-}18)$$

where the first and second terms are the contributions due to the amplitude layer and to the phase layer, respectively, $B = 2(\log E_{amax} - \log E_{amin})/(D_{amax} - D_{amin})$, and β_p is a phase-to-amplitude cross-talk coefficient. β_p can be determined by the procedure described in Section 3.3. Solving (3-18) for $\log E_a$, we obtain

$$\log E_a = \log E_{amax} + \frac{B}{2}\left[2\log t_a - \beta_p\log(E_p/E_{pmax})\right]. \quad (3\text{-}19)$$

Inserting (3-19) into (2-12) and solving for $\log E_p$, we find that

$$\log E_p = \text{const.} + \frac{(\psi/2\pi)\Delta\log E_{pm} - \beta B\log t_a}{(1 - \beta\beta_p B/2)} \quad (3\text{-}20)$$

where const. $= \log E_{pmin} - \left[(\beta\beta_p B/2)\Delta\log E_{pm} + \beta\log(E_{amax}/E_{amin})\right]/(1-\beta\beta_p B/2)$, and the relationship $\log E_{pmax} = \log E_{pmin} + \Delta\log E_{pm}$ was used.

The amplitude-to-phase cross-talk term can cause $\log E_p$ computed from (3-20) to fall outside the linear region of the H&D curve. $\log E_p$ can be brought back to the linear region of the H&D curve by substituting $(\psi \pm 2\pi)$ for ψ in (3-20). After computing it by (3-20), $\log E_p$ is inserted into (3-19) to compute $\log E_a$. In this manner the phase controlling and amplitude controlling exposures are computed to compensate for amplitude-phase dye-layer cross-talk. If the nonlinear regions

of the H&D curves are used, then (3-19) and (3-20) cannot be used in the nonlinear regions, but the exact shape of the H&D curve must be taken into account.

When synthesizing computer-generated holograms, small changes in the absolute level of exposure from one day to the next are difficult to avoid. These small changes in the absolute (effective) level of exposure arise from changes in the level of brightness of the display device, in the aperture (f - stop) of the lens used to photograph the display, in the film speed from one emulsion to the next, and in photographic processing. The result is the multiplication of all the exposure parameters E_p , $E_{pmin/max}$, E_a, and $E_{amin/max}$ by a constant that is somewhat greater than (if overexposure occurs) or less than (if underexposure occurs) unity. An advantage of using the linear regions of the H&D curves of the amplitude and phase layers is that the relative phase delay (3-17) and the relative amplitude transmittance (3-18) are unaffected by a small change in the absolute level of exposure. (A constant term is added to the phase delay and the amplitude transmittance is scaled to a new D_{min} .)

It is possible to expose the different layers of Kodachrome II with sufficient independence to make exposure cross-talk negligible, so compensation for exposure cross-talk is not necessary. However, if other materials are used, then such compensation might be required. A generalized compensation procedure for both dye-layer cross-talk and exposure cross-talk is possible. A pair of exposures (E_a, E_p) will result in a complex transmittance $t_a e^{i\psi}$, which can be specified by the pair (t_a,ψ) . When only the linear parts of the H&D curves are used, then

the mapping $(E_a, E_p) \rightarrow (t_a, \psi)$ is described by (3-17) and (3-18); and the inverse mapping $(t_a, \psi) \rightarrow (E_a, E_p)$, which is necessary to determine the required exposures, is given by (3-19) and (3-20). In the general case, the mapping of $(E_a, E_p) \rightarrow (t_a, \psi)$ can be determined by making a number of exposures (E_a, E_p), then measuring t_a and ψ with a densitometer or by interferometric techniques [Lamberts, Eisen, and Kaspar, 1973] and listing the results in tabular form. Then for a desired (t_a, ψ), the corresponding (E_a, E_p) is determined by interpolation from the values in the table. Such a procedure would automatically compensate for every type of cross-talk.

3.5 Resolution and Spatial Frequency Response

3.5.1 Resolution and Packing Density

The resolution (resolving power) of a material, usually expressed in lines per millimeter (l/mm), indicates the closest practical spacing of resolvable elements on the material. The resolution of a holographic material indicates the packing density of Fourier coefficients that can be achieved (that is, the number per unit area). The maximum packing density is particularly important in the computer memory application, for which it is desired to put the maximum amount of information in the smallest possible space.

Since the image produced by a computer-generated hologram is the spectrum of the hologram, the resolution required of a material for a computer-generated hologram is proportional to the width of the image. Since a ROACH produces only the desired image, the resolution required by a ROACH is simply proportional to the width of the desired image. Holograms on thin amplitude materials must produce a conjugate image

as in (1-4), and, in order to avoid overlap, the total (desired and conjugate) image has at least double the width of the desired image alone, requiring at least double the resolution of the hologram material. If, in addition, the image contains the autocorrelation term mentioned in Section 1.2, then the resolution required is four times that required by a ROACH. Of all the types of computer-generated holograms, the ROACH requires the least resolution of a material for a given desired image. Thus, a ROACH can be made on a material (ordinary color film) that is of low resolution compared to the usual holographic materials.

As will be discussed in Appendix B, the number of Fourier coefficients in a computer-generated hologram equals the number of sampled points (picture elements) in the image. For the ROACH, only one resolution element of the hologram material is required to encode one Fourier coefficient. Therefore, a ROACH made on Kodachrome II, which has a resolution of approximately 100 l/mm [Kodak, AF-1], can reconstruct an image of no more than approximately 1000×1000 picture elements per square centimeter of the film. Since pushing a film to the limit of its resolution usually results in a decrease in quality of the image, it is inadvisable to require more than 50 l/mm resolution from Kodachrome II. Materials suitable for the ROACH with far greater resolution than Kodachrome films are being developed [MacAnally, 1974], but are not readily available at the time of this writing.

3.5.2 Spatial Frequency Response

Containing more information about a material than just its resolution is the spatial frequency response, or, its modulation transfer function (MTF) [Goodman, 1968, Ch.7]. The resolution of the film is

given by the spatial frequency at which the MTF goes below a certain (somewhat arbitrary) cutoff value (5 or 10%). The MTF of Kodachrome 25 is shown in Figure 3.7 [Kodak Data].

Analogous to the MTF is the phase response as a function of spatial frequency. Phase excursions caused by a sinusoidal pattern of exposure depend upon the spatial frequency as well as on the modulation of the exposure pattern. Consequently, if phase-matching is achieved for one spatial frequency, then there will be phase mismatch for other spatial frequencies. Therefore, for optimal results compensation for the variations of the phase response with spatial frequency is necessary. The compensation procedure is described by Smith [1969] and requires knowledge of the particular shape of the phase response vs. spatial frequency curve.

A rough measurement of the phase response as a function of spatial frequency was made. Phase gratings were produced by photographing the CRT, as described in Chapter 2. However, due to limitations of the resolution of the CRT and the photographic magnifications possible, only gratings of low spatial frequency (less than 10 l/mm) could be made in this way. In order to cover a much wider range of spatial frequencies, phase gratings were produced interferometrically. Sinusoidal intensity patterns were produced by the interference of two coherent plane waves of 501.7 nm light from an argon-ion laser. The sinusoidal intensity patterns were recorded on Kodachrome 25, in which the blue-absorbing and red-absorbing layers were equally exposed by the 501.7 nm light. Also, the film was flooded with red light to make it a pure phase grating after processing. Unfortunately, interferometrically generated phase

gratings are not exactly sinusoidal since there is a nonlinear relationship between phase and exposure over most of the dynamic range of the film. Phase is instead linear with the log of exposure over the linear region of the H&D curve (which could be taken into account for the patterns produced by the CRT). Consequently, in determining the phase modulation from measurements of the different orders of diffraction of the phase gratings produced, it is necessary to take into account the nonlinear nature of the recording.

Let Ae^{ik_1x} and Be^{ik_2x} represent the two interfering plane waves, where A^2 and B^2 are the intensities and k_1 and k_2 are the x-components of the wave vectors of the respective plane waves. Then the intensity of the interference pattern of the two plane wave is

$$I(x) = \left| Ae^{ik_1x} + Be^{ik_2x} \right|^2 = A^2 + B^2 + 2AB\cos kx$$

$$= I_0(1 + m\cos kx) \qquad (3\text{-}21)$$

where $I_0 = A^2 + B^2$, $k = k_1 - k_2$ is the grating vector, and $m = 2AB/I_0$. If the film is exposed for a time τ, then the exposure pattern on the film is

$$E(x) = \tau I(x) = E_0(1 + m\cos kx) \qquad (3\text{-}22)$$

where $E_0 = \tau I_0$ is the bias exposure and $m = 2AB/(A^2 + B^2)$ is the modulation of the exposure. If the exposure is made over the linear region of the H&D curve, then, using the phase-to-density proportionality rule, the resulting phase pattern is

$$\psi(x) = 2\pi \log E(x)/\Delta \log E_{pm} + \text{const.}$$

$$= \frac{2\pi}{\Delta \log E_{pm}} \log(1 + m\cos kx) + 2\pi \log E_0/\Delta \log E_{pm} + \text{const.} \quad (3\text{-}23)$$

where $\Delta \log E_{pm}$ is the excursion of the log of exposure needed to achieve a phase excursion of 2π radians. Using the identity $\log x = (\ln x)/\ln 10$, the complex transmittance of the phase grating can be expressed as

$$H(x) = e^{i\psi(x)} = e^{i\alpha \ln(1 + m\cos kx)}$$
$$= (1 + m\cos kx)^{i\alpha} \quad (3\text{-}24)$$

where the constant phase terms are dropped and $\alpha = 2\pi/(\Delta \log E_{pm} \cdot \ln 10)$ is a phase modulation coefficient. Our goal in this section is to determine how α varies with the spatial frequency of $\cos kx$. Using the power series expansion $(1 + x)^q = 1 + qx + q(q-1)x^2/2! + q(q-1)\cdot(q-2)x^3/3! + \ldots$ [Gradshteyn and Ryzhik, 1965, p.21], (3-24) becomes

$$\begin{aligned} H(x) = {} & 1 + i\alpha m\cos kx + i\alpha(i\alpha-1)m^2\cos kx/2! \\ & + i\alpha(i\alpha-1)(i\alpha-2)m^3\cos^3 kx/3! \\ & + i\alpha(i\alpha-1)(i\alpha-2)(i\alpha-3)m^4\cos^3 kx/4! \\ & + i\alpha(i\alpha-1)(i\alpha-2)(i\alpha-3)(i\alpha-4)m^5\cos^5 kx/5! \\ & + \ldots\ldots \end{aligned} \quad (3\text{-}25)$$

From Table 2.6 it can be seen that the value of $\Delta \log E_{pm}$ (measured at 2.5 l/mm) is approximately 1.2 when both the blue-absorbing and the green-absorbing layers are used for phase control in Kodachrome 25. Therefore, for this experiment, the value of α is approximately $2\pi/(1.2\ln 10) = 2.27$. Using 2.27 for the value of α, the modulus of the coefficient of the term $m^n\cos^n kx$ in (3-25) is approximately 2.8

for $n = 1$ and decreases slowly with increasing n. Consequently, the expansion (3-25) is not useful for values of $\mathfrak{m}$ near 1.0, but is good for $\mathfrak{m} \lesssim 0.5$.

Using the identities $\cos^2 x = (\cos 2x + 1)/2$, $\cos^3 x = (\cos 3x + 3\cos x)/4$, $\cos^4 x = (\cos 4x + 4\cos 2x + 3)/8$, and $\cos^5 x = (\cos 5x + 5\cos 3x + 10\cos x)/16$ [Gradshteyn and Ryzhik, 1965, p.26], we find that

$$\begin{aligned} H(x) \simeq & \left[1 + (-\alpha^2 - i\alpha)\mathfrak{m}^2/4 + (\alpha^4 - 11\alpha^2 + i6\alpha^3 - i6\alpha)\mathfrak{m}^4/64\right] \\ & + 2\cos kx\left[i\alpha\mathfrak{m}/2 + (3\alpha^2 - i\alpha^3 + i2\alpha)\mathfrak{m}^3/16 \right. \\ & \qquad \left. + (-10\alpha^4 + 50\alpha^2 + i\alpha^5 - i35\alpha^3 + i24\alpha)\mathfrak{m}^5/384\right] \\ & + 2\cos 2kx\left[(-\alpha^2 - i\alpha)\mathfrak{m}^2/8 + (\alpha^4 - 11\alpha^2 + i6\alpha^3 - i6\alpha)\mathfrak{m}^4/96\right] \\ & + 2\cos 3kx\left[(3\alpha^2 - i\alpha^3 + i2\alpha)\mathfrak{m}^3/48 \right. \\ & \qquad \left. + (-10\alpha^4 + 50\alpha^2 + i\alpha^5 - i35\alpha^3 + i24\alpha)\mathfrak{m}^5/768\right] \\ & + \ldots \end{aligned} \tag{3-26}$$

where terms of order $\mathfrak{m}^6$ and higher are dropped. Since $2\cos nkx = e^{inkx} + e^{-inkx}$ represents two plane waves going into the n^{th} and $-n^{th}$ orders, respectively, the complex amplitude of the n^{th} order of diffraction is given by the coefficient (in brackets) of $2\cos nkx$ in (3-26). The intensity falling into the n^{th} order of diffraction is given by the modulus squared of the coefficient of $2\cos nkx$ in (3-26). Thus, the intensity falling into the zero order (on-axis) is

$$\text{zero order} \simeq 1 - \alpha^2\mathfrak{m}^2/2 + (3\alpha^4 - 9\alpha^2)\mathfrak{m}^4/32 \tag{3-27}$$

where again terms of order $\mathfrak{m}^6$ and higher are dropped. Similarly, the intensity falling into the first order is

$$\text{first order} \simeq \alpha^2\mathfrak{m}^2/4 + (-\alpha^4+2\alpha^2)\mathfrak{m}^4/16 + (5\alpha^6-85\alpha^4+60\alpha^2)\mathfrak{m}^6/768 \qquad (3\text{-}28)$$

and the intensity falling into the second order is

$$\text{second order} \simeq (\alpha^4+\alpha^2)\mathfrak{m}^4/64 + (-\alpha^6+5\alpha^4+6\alpha^2)\mathfrak{m}^6/384 \qquad (3\text{-}29)$$

where terms of order $\mathfrak{m}^8$ and higher are dropped.

Using the expansion $(1 + x)^{-1} = 1 - x + x^2 - x^3 + \ldots$ [Gradshteyn and Ryzhik, 1965, p.21], we find that the ratio, R_{10}, of the intensity falling into the first order to the intensity falling into the zero order is

$$R_{10} \simeq \alpha^2\mathfrak{m}^2/4 + (\alpha^4+2\alpha^2)\mathfrak{m}^4/16 + (11\alpha^6+17\alpha^4+60\alpha^2)\mathfrak{m}^6/768 \qquad (3\text{-}30)$$

where terms of order $\mathfrak{m}^8$ and higher are dropped. Figure 3.6 shows plots of R_{10} given by (3-30) vs. α for various values of $\mathfrak{m}$.

A purely sinusoidal grating with transmittance $e^{i\alpha\mathfrak{m}\cos kx}$ has intensity proportional to $J_n^2(\alpha\mathfrak{m})$ falling into the n^{th} order (as discussed in Section 2.3.5). Then the ratio of the intensity of the first order to that of the zero order would be

$$J_1^2(\alpha\mathfrak{m})/J_0^2(\alpha\mathfrak{m}) = \alpha^2\mathfrak{m}^2/4 + \alpha^4\mathfrak{m}^4/16 + 11\alpha^6\mathfrak{m}^6/768 + \ldots \qquad (3\text{-}31)$$

which is obtained by using the expansions $J_1(x) = x/2 - x^3/(2^3 2!) + x^5/(2^5 2!3!) - \ldots$ and $J_0(x) = 1 - x^2/2^2 + x^4/(2^4(2!)^2) - x^6/(2^6(3!)^2) + \ldots$. A comparison of (3-30) with (3-31) shows the effect of the non-linear nature of the recording process; and, as expected, (3-30) agrees with (3-31) for $\mathfrak{m} << 1$.

Phase gratings of the form (3-24) were produced interferometrically. The spatial frequency, $\nu = k/2\pi$, and the modulation, $\mathfrak{m}$, of the

exposure pattern were varied in the manner described in Appendix C. The phase gratings were reconstructed with 632.8 nm light from a helium-neon laser, and the intensities falling into the zero (on-axis) and first orders were measured. R_{10}, the ratio of the measured intensities of the first order to the zero order, was compared with Figure 3.6 to determine α for various values of $\mathfrak{M}$ and ν. For a fixed ν, the values of α determined for different values of $\mathfrak{M}$ were in good agreement, indicating the validity of (3-30).

Figure 3.7 shows the resulting phase response vs. spatial frequency curve for the combined blue-absorbing and green-absorbing layers of Kodachrome 25. Interestingly, the phase response curve closely follows the MTF curve. This characteristic contrasts with 649-F plate, which has a constant MTF for this range of spatial frequencies, but has a very sharply peaked phase response (due to surface relief) at 10 l/mm [Smith, 1968]. Both Figure 3.7 and similar results from CRT-generated patterns indicate that Kodachrome shows significant phase response even at very low spatial frequencies ($\lesssim$ 1 l/mm). The MTF of Kodachrome II is considerably flatter in the 1 to 15 l/mm range [Kodak Data], making it even more suitable for computer holography.

3.6 Dynamic Range and ROACHs on Materials of Limited Dynamic Range

The dynamic range of a material is defined as the ratio of the maximum amplitude (or intensity) transmittance to the minimum amplitude (or intensity) transmittance that can be achieved. Desired transmittances that are below (or above) the achievable dynamic range are set equal ("clipped") to the minimum (or maximum) achievable value, and noise in

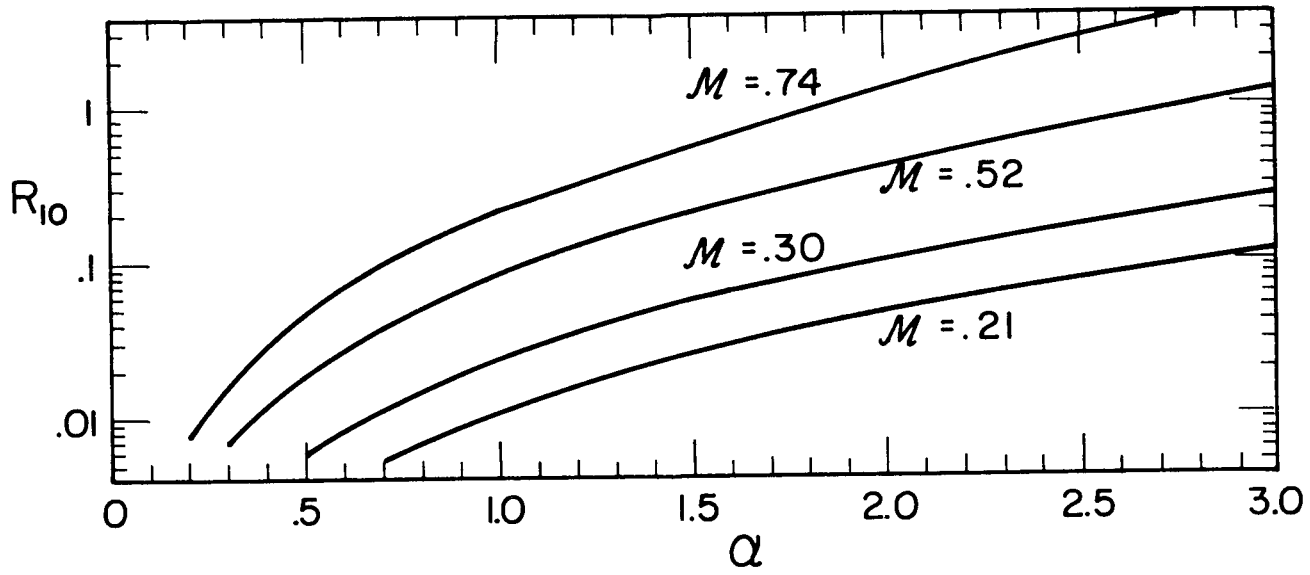


Fig. 3.6 Non-linear sinusoidal grating: ratio of first order to zero order intensity vs phase modulation coefficient, α.

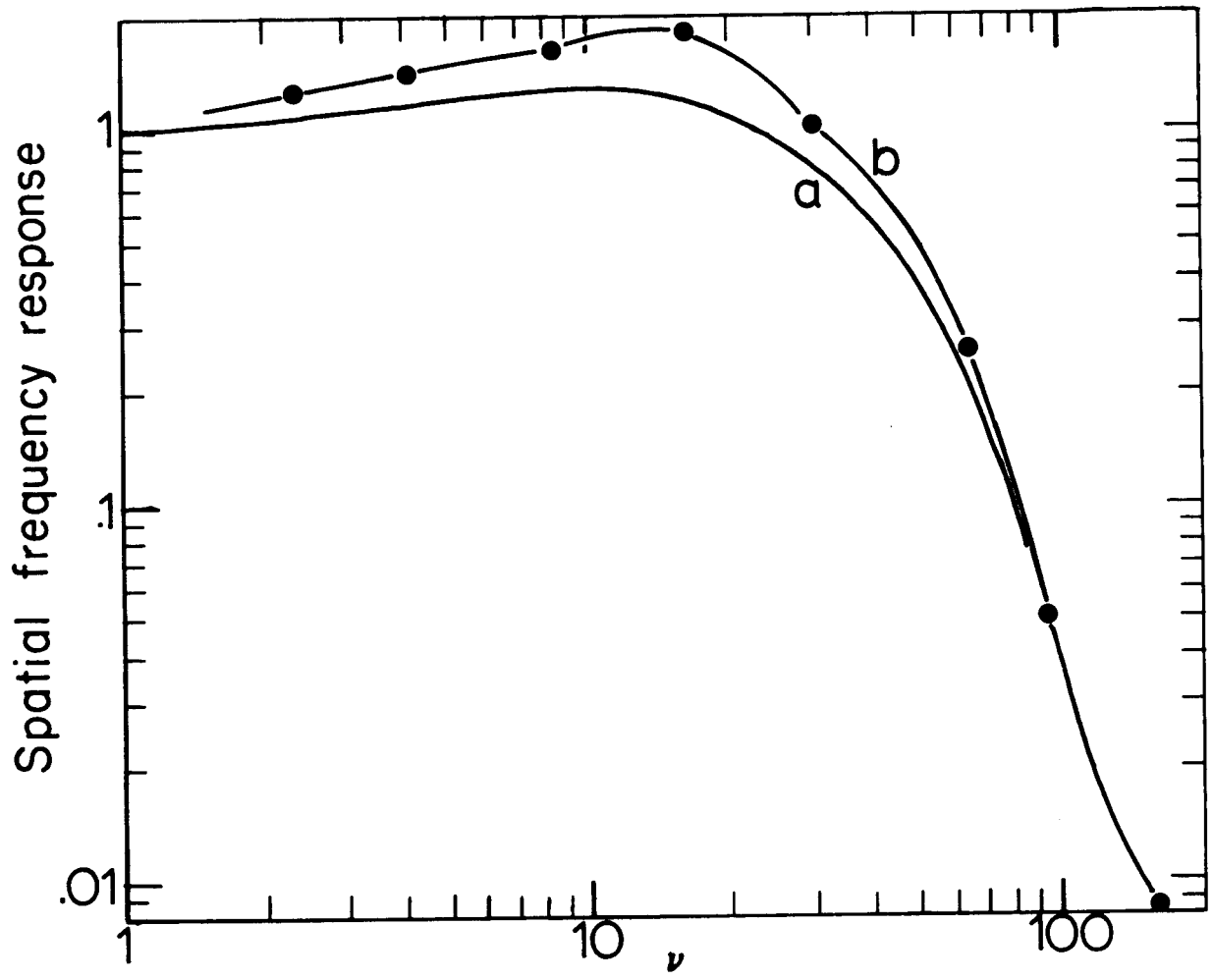


Fig. 3.7 Spatial frequency response of Kodachrome 25: MTF (a) [Kodak Data], experimentally determined phase modulation coefficient, α, vs spatial frequency (b).

the reconstructed image results from this approximation. For our own experiments, the range of desired transmittances was always scaled so that the maximum desired transmittance was achievable, and only the lower values beyond the dynamic range of the film were clipped. The effect of clipping has been investigated by Powers [1975, Ch. 4] and will not be discussed here. Since the maximum density achievable in Kodachrome film is approximately 3, the maximum possible dynamic range is 1000 : 1 in intensity or 30 : 1 in amplitude. For blue light, however, the blue-absorbing layer alone contributes a density of only 2 (the other two layers contribute considerable density to blue to bring the total density of 3 for ordinary photographic purposes); so in practice the dynamic range for blue light is only 100 : 1 in intensity or 10 : 1 in amplitude.

Analagous to the dynamic range in amplitude mentioned above, is the "phase dynamic range", the number of radians of phase excursion obtainable (at a given wavelength). A phase dynamic range of 2π radians is required for phase-matching. The phase dynamic ranges of the different modes of Kodachrome II and Kodachrome 25 are approximately given by $2\pi \cdot (\Delta \log E \text{ available})/\log E_{pm}$, which can be determined from Tables 2.5 and 2.6, respectively.

Ordinarily, if there is a considerable phase mismatch in a ROACH, then, just as in the case of the kinoform discussed in Section 3.2, spurious images are produced that degrade the desired image. Therefore, it was initially thought that a material would have to be capable of producing phase excursions of at least 2π radians in order to be useful as a ROACH material. However, as we will show in this section, it

is possible to shape the complex spectrum (the Fourier transform) of the object in such a way as to reduce the required dynamic range both in amplitude and in phase, and, therefore, allow the use of materials with limited dynamic range.

Let the object and its Fourier transform be given by $f(x,y) \supset F(u,v)$. The dynamic range of $|F(u,v)|$ can be reduced by phase-coding $f(x,y)$, as will be discussed in Chapter 5. But here we consider a more general problem: given $F(u,v)$, how can we encode it on a holographic material in such a way as to minimize the dynamic range required of the material?

First, consider the manner in which dynamic range can be reduced in off-axis holograms. As described in Section 1.2, the addition of an on-axis spot and a conjugate term to the object yields a hologram function $H(u,v)$ that is real and nonnegative and can be synthesized on a thin amplitude material:

$$H(u,v) = A + 2|F(u,v)|\cos(2\pi us/\lambda f + \psi(u,v)) \quad (1\text{-}4)$$

where $A \geq 2 \cdot \text{Max}\{|F(u,v)|\}$. Since the maximum amplitude transmittance is unity, then it is required that the hologram function be normalized so that $A + 2\text{Max}\{|F(u,v)|\} \leq 1$. By normalizing A and $F(u,v)$ so that $\text{Max}\{|F(u,v)|\}/A$ is small, the dynamic range required of the hologram material can be made arbitrarily small. The analagous procedure (and one that is commonly used) for interferometrically-generated holograms is to use a large reference beam to object beam ratio to reduce dynamic range. Two undesired effects of reducing the dynamic range in this manner are a reduction in diffraction efficiency and an increase in the intensity of the undiffracted wave (the on-axis spot for the

Fourier transform geometry). The above holds true for all off-axis computer-generated holograms, including thin phase as well as thin amplitude holograms.

The method of reduction of dynamic range for the ROACH is less obvious. As described in Section 2.3.6, by multiplying $F(u,v)$ by the constant $1/\text{Max}\{|F(u,v)|\}$, a function is obtained, the modulus of which is less than or equal to unity everywhere. That is, if $H(u,v) = F(u,v)/\text{Max}\{|F(u,v)|\}$, then all values of $H(u,v)$ lie within a unit circle centered at the origin of the complex plane (see Fig. 3.8a). The diffraction efficiency of an ideally synthesized ROACH is then

$$\eta_0 = \overline{|F(u,v)|^2} \Big/ \text{Max}\{|F(u,v)|^2\} \qquad (3\text{-}32)$$

where the overbar indicates the mean value. η_0 is less than one because the maximum amplitude transmittance of a ROACH is one, and the amplitude transmittance is controlled by the absorption of light. Only when $|F(u,v)|$ is constant can a ROACH have 100% diffraction efficiency (then it is indistinguishable from a kinoform). At this point, we redefine $f(x,y)$ so that $f(x,y) \supset F(u,v)/\text{Max}\{|F(u,v)|\}$. If we set

$$H(u,v) = BF(u,v)/\text{Max}\{|F(u,v)|\} \qquad (3\text{-}33)$$

which would correspond to the object $Bf(x,y)$, where $0 \leq B \leq 1$, then all values of $H(u,v)$ lie within a circle of radius B, as shown in Figure 3.8b, and the diffraction efficiency becomes $B^2\eta_0$. This hologram function can be synthesized on a ROACH material of maximum amplitude transmittance B, where B can be arbitrarily small, but at the expense of reduced diffraction efficiency.

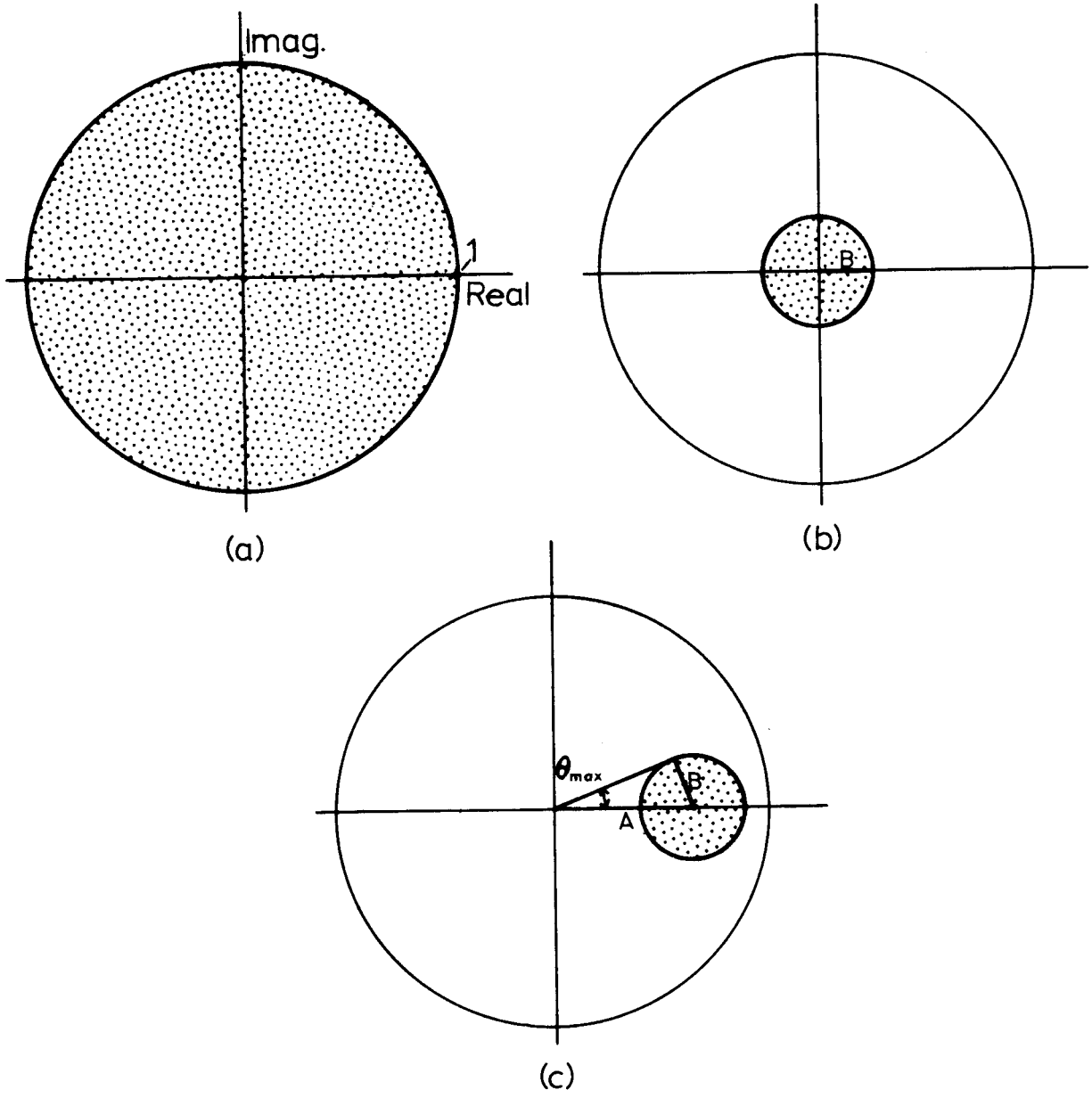


Fig. 3.8 Dynamic range required of a ROACH. Shaded areas indicate complex values that must be achieved. (a) Values ordinarily required; (b) Values reduced by a factor of B; (c) Values reduced by a factor of B and shifted by A.

If an on-axis bright spot, $A\delta(x,y)$, is added to the object, so that the new object is $A\delta(x,y) + Bf(x,y)$, then the corresponding hologram function is

$$H(u,v) = A + BF(u,v)/\text{Max}\{|F(u,v)|\} \qquad (3\text{-}34)$$

where $0 \leq A \leq 1 - B$. Then all values of $H(u,v)$ lie within a circle of radius B centered at point A in the complex plane. As shown in Figure 3.8c, the amplitudes of $H(u,v)$ lie between $A - B$ and $A + B$ and the phases of $H(u,v)$ lie between $-\theta_{max}$ and $+\theta_{max}$. Consequently, this hologram function can be synthesized on a ROACH material that has a dynamic range in amplitude of only $(A+B)/(A-B)$ and a phase dynamic range of only $2\theta_{max}$. From the geometry of Figure 3.8c we see that $\sin \theta_{max} = B/A$. Apparently, a ROACH can be synthesized on a material with arbitrarily small dynamic range, both in amplitude and in phase, by choosing a small enough B. The only effect on the image is a decrease in diffraction efficiency and an on-axis spot. No other spurious images are produced.

For a given value of B, the diffraction efficiency is $B^2\eta_0$, independent of the value of A. Referring to Figure 3.8c, the effect of decreasing the value of A would be to move the circle (within which lie the values of $H(u,v)$) toward the origin, reducing the intensity of the on-axis spot, A^2 (by increasing the absorption by the ROACH). A reduction in A is desirable since a very intense on-axis spot will obscure nearby image points. Another effect of reducing A is an increase in $2\theta_{max}$, the required phase dynamic range. If A is less than B, then the entire 2π radians of phase dynamic range is required. Thus there is a trade-off between the intensity of the on-axis spot and

the dynamic range, for a given diffraction efficiency, $B^2\eta_0$. The minimum dynamic range is required when $A = 1 - B$; then the dynamic range in amplitude is $1/(1 - 2B)$ and the phase dynamic range is given by $\sin \Theta_{max} = B/(1 - B)$. Conversely, the value of B necessary to achieve a desired dynamic range $2\Theta_{max}$ is $B = \sin \Theta_{max}/(1 + \sin \Theta_{max})$. Figure 2.10 shows a plot of the dynamic range required of a ROACH as a function of B , for $A = 1 - B$.

The method described above for reducing the dynamic range required of a ROACH can also be used to reduce the phase dynamic range required of a parity sequence hologram [Chu and Goodman, 1972]. For the parity sequence hologram, the hologram complex transmittance for the $(m,r)^{th}$ Fourier coefficient is given by [Chu, 1974, Eq.2-13]

$$H_{m,r} = \exp\left\{i\left[\arg\left(F'_{m,r}\right) \pm \cos^{-1}\left(F'_{m,r}\right)\right]\right\} \qquad (3\text{-}35)$$

where $F'_{m,r} \equiv F_{m,r}/\text{Max}\{|F_{m,r}|\}$. Ordinarily, since the values of $F'_{m,r}$ lie within a unit circle centered at the origin in the complex plane, the phase dynamic range of $H_{m,r}$ is 2π radians. However, if $F'_{m,r}$ in (3-35) is replaced by H in (3-34), the values of which lie within a circle of radius B centered at point A in the complex plane, then the phase dynamic range of $H_{m,r}$ can be significantly reduced. As in the case of the ROACH, the diffraction efficiency is $B^2\eta_0$ and the intensity of the on-axis spot is A^2 .

The reduced value of the phase dynamic range of $H_{m,r}$ can be determined as follows. Referring to Figure 3.9a, $H_{m,r}$ is found by adding to $F_{m,r}$ a phasor $P_{m,r}$ that is perpendicular to $F'_{m,r}$ and is of such a length that $H_{m,r} = F'_{m,r} + P_{m,r}$ falls on the unit circle. Referring to Figure 3.9b, the value within the circle of radius B

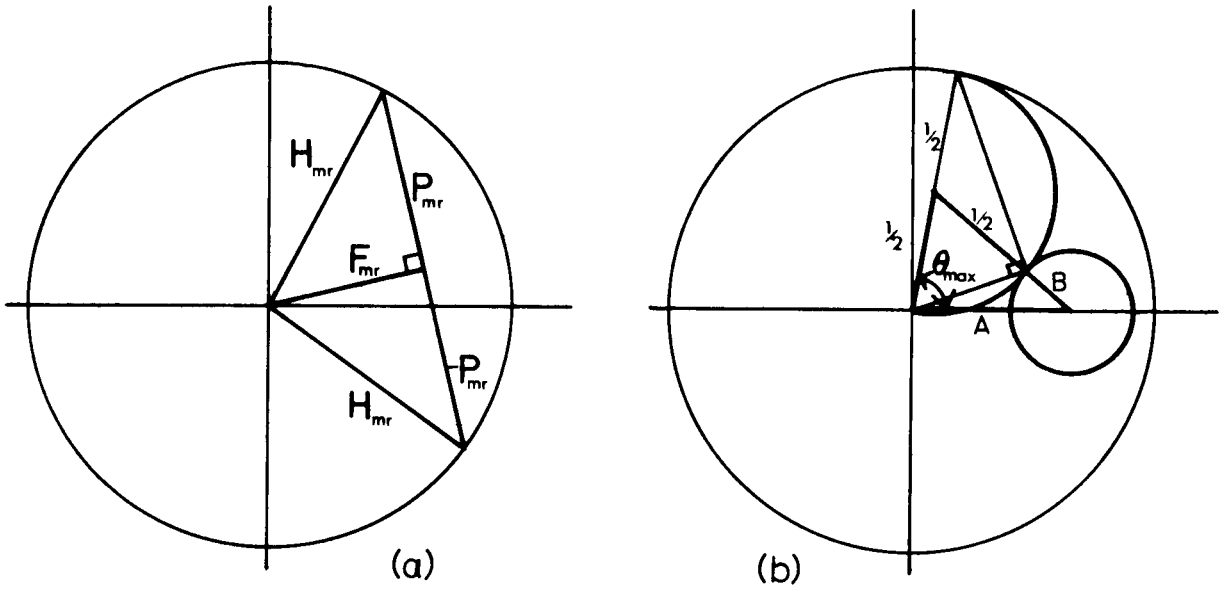


Fig. 3.9 (a) Geometry for determining the complex coefficient H_{mr} for the parity sequence hologram; (b) Geometry for determining $2\Theta_{max}$, the reduced phase dynamic range for the parity sequence hologram.

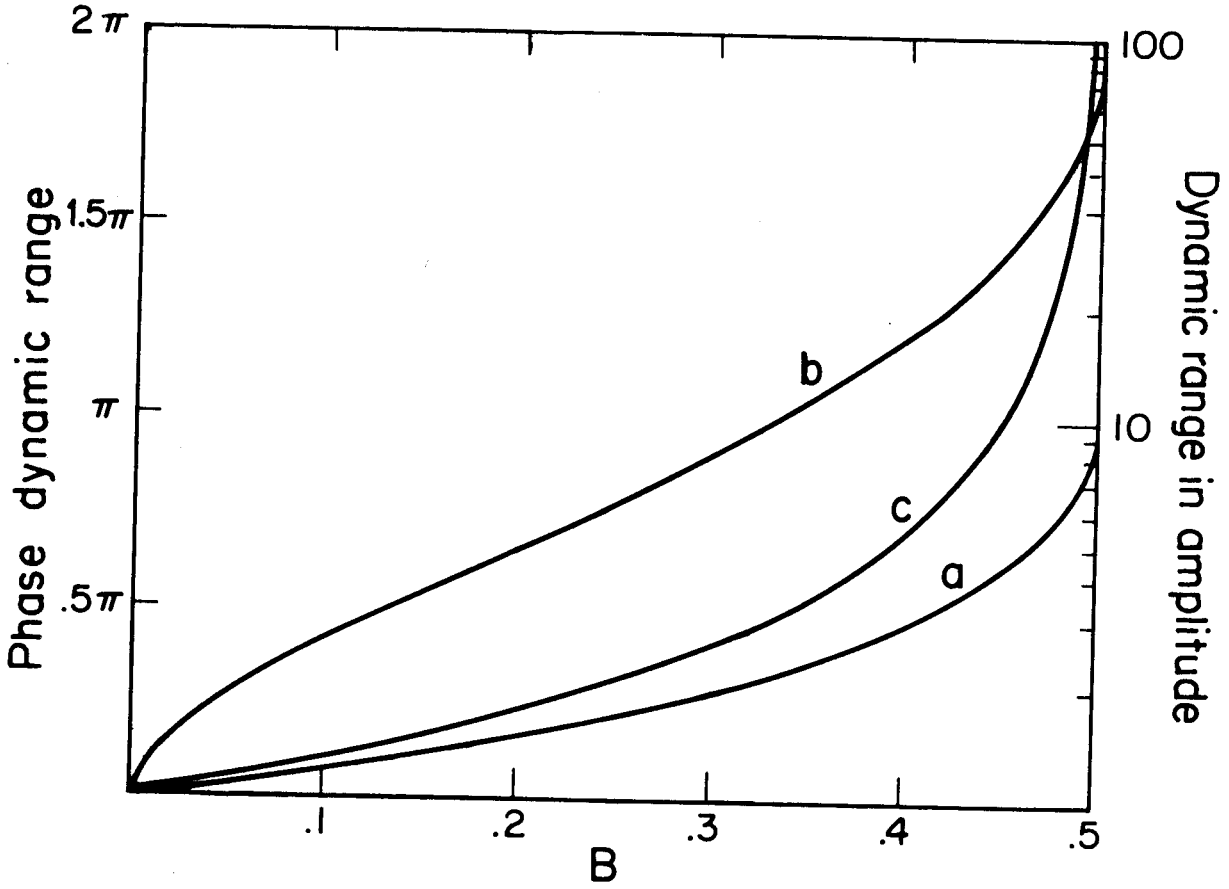


Fig.3.10 Reduced phase dynamic range, $2\Theta_{max}$, for the ROACH (a) and for the parity sequence hologram (b); and reduced dynamic range in amplitude for the ROACH (c), all for A = 1-B.

that contributes to the largest phase angle of $H_{m,r}$, Θ_{max}, is the value at the intersection of the circle of radius B with the half-circle of radius ½ that is tangent to the circle of radius B. (All values on the circumference of that half-circle yield the same value of $H_{m,r}$). From the geometry of Figure 3.9b, using the law of cosines, the maximum phase angle, Θ_{max}, is given by

$$\cos \Theta_{max} = \frac{(\frac{1}{2})^2 + A^2 - (\frac{1}{2}+B)^2}{2(\frac{1}{2})A} = \frac{A^2 - B^2 - B}{A} \quad , \quad A > B \tag{3-36}$$

where $A + B \leq 1$. The value of Θ_{max} is decreased by increasing A and decreasing B. (Increasing A also increases the intensity of the on-axis spot, and by the same amount decreases the intensity of the parity sequence image.) Since the motivation for these manipulations is the reduction of the phase dynamic range, $2\Theta_{max}$, the maximum allowable value of $A = 1 - B$ is chosen. Then (3-36) reduces to

$$\cos \Theta_{max} = \frac{1 - 3B}{1 - B} \tag{3-37}$$

or

$$\sin^2(\Theta_{max}/2) = \frac{B}{1-B} \quad .$$

The phase dynamic range, $2\Theta_{max}$, is shown in Figure 3.10 as a function of B. Note that a phase dynamic range of π radians (that is, half the usual value) is obtained when $B = 1/3$. By inverting (3-37), the value of B necessary to achieve a desired phase dynamic range of $2\Theta_{max}$ is

$$B = \frac{1 - \cos \Theta_{max}}{3 - \cos \Theta_{max}} = \frac{\sin^2(\Theta_{max}/2)}{1+\sin^2(\Theta_{max}/2)} \quad . \tag{3-38}$$

An additional disadvantage of reducing the dynamic range required of the hologram material in the manner described in this section is an increase in quantization noise. Quantization noise, which will be discussed in more detail in Chapter 5, results from the fact that only certain discrete values of $H(u,v)$ can be encoded on the hologram. For example, since we control the exposure by the number of computer cycles for which the electron beam of the CRT is turned on at a given point, our exposure times are limited to integer multiples of one computer cycle. So if we were to use an exposure time range of,say, 10 to 210 computer cycles, then there would be 201 quantized values of exposure time. This number of quantized values is so large that it is for all practical purposes continuous. However, if the ROACH were made with a small dynamic range by making B small in (3-37), then there could be a range of only a few quantized values of exposure time, and quantization noise could be significant.

Two methods could be used to reduce quantization noise due to reduced dynamic range. One would be to use the iterative procedure that will be described in Chapter 5. Another method would be to optimize the display gray level capability with respect to the limited dynamic range of exposures. The manner in which this latter method would be accomplished depends upon the specific display device, but we will use our particular display device as an example. Suppose that the range of exposure times is, say, from 20 to 25 computer cycles, only six quantized values. If we were to stop down the lens of the camera used to photograph the CRT by 2 f-stops, then four times the exposure time would be required. Then the range of exposure times would be from 80 to 100

computer cycles, or 21 quantized values. In this manner quantization noise would be greatly reduced.

3.7 Scattered Flux Spectrum

The most fundamental limit to the amount of information that can be stored in a ROACH is set by the scattering of light by the film. It is possible to compensate for the other potential sources of error mentioned in this chapter, but not for scattered light. Contributing to scattering are microscopic variations in amplitude transmittance ("film grain"), variations in refractive index, and surface roughness. Figure 3.11 shows a scanning electron micrograph of the surface of Kodachrome II, magnified 1000×. Surface roughness detail in the one to two micron range is evident and contributes substantially to scattering. The scattered flux appearing in the Fourier plane is simply the squared modulus of the Fourier transform of the complex transmittance of the transparency. The scattered flux spectrum, also known as the power spectrum or the noise spectral density, has been measured for most holographic materials by a number of authors [Biedermann, 1970; Brandt, 1970; Thomas, 1972]. Theoretical studies relating film-grain noise to the ultimate signal-to-noise ratio and error rate in an image from a hologram were reported by Goodman [1967], Kozma [1968], and Lee [1972].

Figure 3.12 shows the experimental setup used to measure the scattered flux spectrum. In order to avoid edge-diffraction effects, the apertures before the film plane (used to reduce stray light) were made wider than the laser beam. Since surface relief effects are used to control the phase delay for the ROACH, the scattered flux spectrum was measured without a liquid gate. The spatial frequency in the Fourier plane is

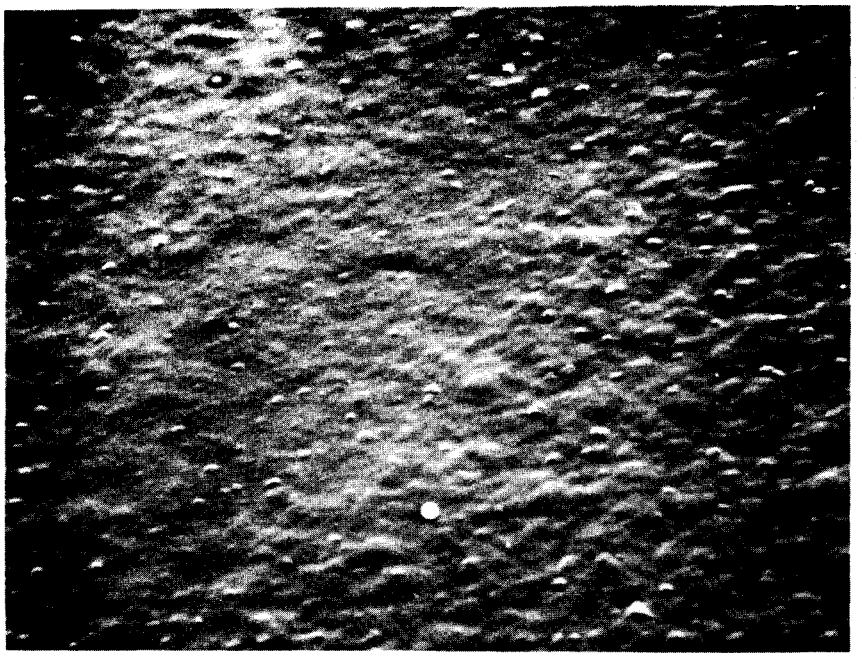


Fig. 3.11 Scanning electron micrograph of the surface of Kodachrome II, magnified one thousand times.

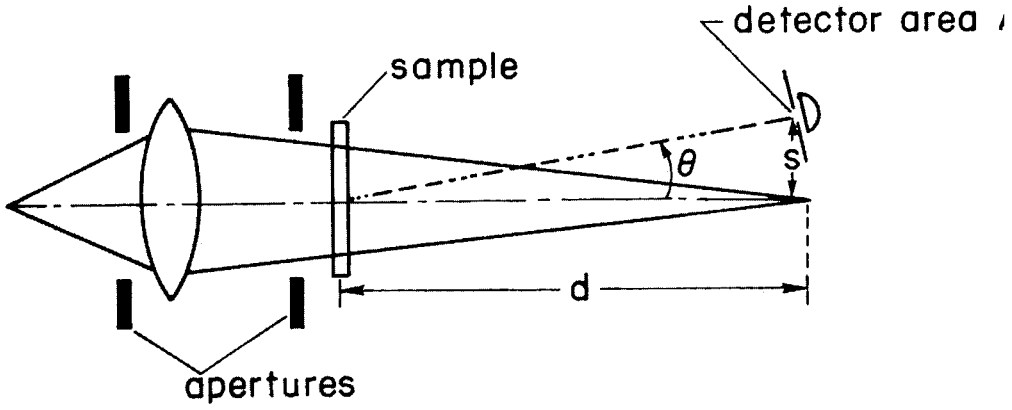


Fig. 3.12 Schematic of experimental apparatus used to measure the scattered flux spectrum (not to scale).

$$\nu = \frac{\sin\theta}{\lambda} = \frac{s}{d\lambda} \quad . \qquad (3\text{-}39)$$

Values of d used were 1020mm and 2040mm; and the area, A, of the aperture at the detector was either 3.33mm^2 or 31.7mm^2. Measurements were made on samples of the uniformly exposed Kodachrome II transparencies that had been used to determine the H&D curves of Figure 2.7. The area of illumination was a circle of diameter 18mm.

Figure 3.13 shows the resulting scattered flux spectrum for a number of different samples. Table 3.1 lists the filter through which the film had been exposed before processing and the measured densities of each of the processed transparencies used for Figure 3.13. The value of Φ, the scattered flux spectrum, is the measured value of flux divided by the unscattered (on-axis) component, Φ_0, and normalized to a detector aperture of 1 $(1/\text{mm})^2$ (that is, multiplied by $d^2\lambda^2/A$). The units of Φ are $(1/\text{mm})^{-2}$. Some authors [e.g., Biedermann, 1970] divide the measured value of flux by the incident flux, Φ_{inc}, rather than by Φ_0. The curves in Figure 3.13 can be converted to this second type of normalization by dividing each $\Phi(\nu)$ by the corresponding value of Φ_{inc}/Φ_0 (which also is listed for each sample in Table 3.1). Our choice of normalization is based on the fact that it is more pertinent to compare the scattered flux with the unscattered component (which, if the transparency were a hologram, would be proportional to the intensity of the image) than with the incident flux, most of which would simply be absorbed if the optical density of the transparency were greater than 0.3.

The scattered flux spectra of all the samples tested have the same shape, decreasing steadily with increasing ν. This shape is similar

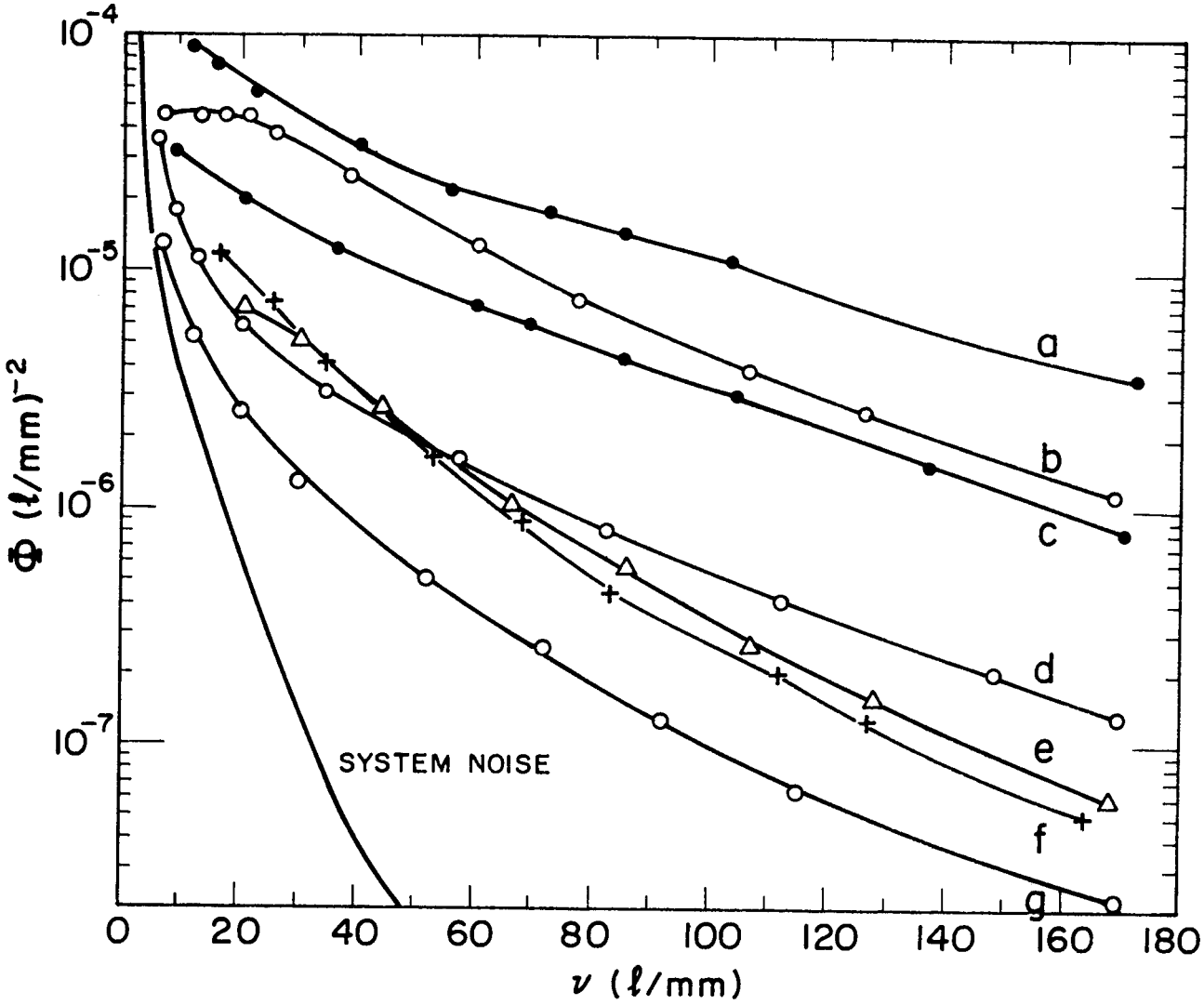


Fig. 3.13 Scattered flux spectra of various samples of Kodachrome II (see Table 3.1)

Sample (see Fig. 3.13)	Wratten Filter No.	Density D_R	D_G	D_B	Φ_{inc}/Φ_0	Φ at 50 l/mm
a	21	0.18	0.48	1.57	4.2	$2.7\ 10^{-5}$
b	none	1.54	1.88	1.96	168.0	$1.8\ 10^{-5}$
c	21	.12	.26	.95	2.6	$9\ 10^{-6}$
d	none	.41	.57	.82	4.3	$1.9\ 10^{-6}$
e	36	.22	.94	.43	2.5	$2.0\ 10^{-6}$
f	98	1.04	.46	.27	17.4	$1.9\ 10^{-6}$
g	none	.08	.14	.17	1.5	$5.5\ 10^{-7}$

Table 3.1 Kodachrome II samples measured for Fig. 3.13.

to the shape of the scattered flux spectra of high-resolution films and plates, such as Kodak 649-F, rather than low-resolution films such as Pan-X [Brandt, 1970]. A comparison of samples g, d, and b in Figure 3.13 shows that scattering increases with increasing dye density. A comparison of samples c, e, and f shows that the scattering due to the blue-absorbing layer (the outermost layer in the film) is considerably greater than the scattering due to either the green-absorbing or the red-absorbing layer; and the scattering due to the green-absorbing and red-absorbing layers is about equal. Much of the density to blue light of sample b is due to absorption of blue light by the green-absorbing and red-absorbing layers. Therefore, there is more blue-absorbing dye present in sample a than in sample b, accounting for the greater scattering by sample a than by sample b.

The scattering of Kodachrome II is approximately ten times that of 649-F, is comparable to that of Pan-X below 40 l/mm, and is less than that of Pan-X above 40 l/mm [Brandt, 1970].

3.8 Binary-Amplitude ROACH

The ROACH described so far in this thesis modulates the amplitude of a transmitted wavefront by variations in optical density. An unfortunate side effect is the amplitude-to-phase cross-talk due to phase variations, caused by variations in thickness of the amplitude controlling layer. One can compensate for this cross-talk, as discussed in Section 3.4. However, it is possible to eliminate this amplitude-to-phase cross-talk (without compensation) by modulating the amplitude of the transmitted wave in a binary manner. The amplitude is modulated in the same manner as the binary detour-phase hologram of Lohmann [Brown and Lohmann, 1966].

That is, the amplitude is controlled by the area of a transparent aperture on an opaque background. No amplitude-to-phase cross-talk results, because the transparent areas are all exposed identically. An additional advantage is that the binary amplitude exposure is less critical and not subject to film nonlinearity problems. Furthermore, since the apertures have low optical density (ideally zero), the scattered flux spectrum is at a minimum, as discussed in Section 3.7. The phase transmittance of the binary amplitude ROACH is controlled in the same manner as discussed in Chapter 2.

In the ordinary ROACH, a Fourier coefficient is encoded as a square cell of width d and transmittance $t_a e^{i\psi}$, uniform over the area of the cell. The contribution to the image due to one cell centered at (u_0,v_0) is

$$\text{contribution } (x,y) = \frac{1}{\lambda f}\int_{u_0-d/2}^{u_0+d/2}\int_{v_0-d/2}^{v_0+d/2} t_a e^{i\psi}\, e^{-i(2\pi/\lambda f)(ux+vy)}\,du\,dv$$

$$= \frac{t_a e^{i\psi}}{\lambda f}\int_{u_0-d/2}^{u_0+d/2} e^{-i(2\pi/\lambda f)ux}\,du \cdot \int_{v_0-d/2}^{v_0+d/2} e^{-i(2\pi/\lambda f)vy}\,dy$$

$$= \frac{t_a e^{i\psi}}{\lambda f}\cdot d^2 \operatorname{sinc}\left(\frac{dx}{\lambda f}\right)\operatorname{sinc}\left(\frac{dy}{\lambda f}\right) e^{-i(2\pi/\lambda f)(u_0x+v_0y)} \qquad (3\text{-}40)$$

where the amplitude is controlled by the factor t_a and the cell width d is fixed (equal to the spacing between cells). For the binary-amplitude ROACH, on the other hand, t_a is fixed equal to 1 and the area of the aperture is used to control the amplitude. If the apertures were always square, then the area (and therefore the amplitude) would be controlled by the d^2 factor in (3-40). In both cases, an unwanted

attenuation near the edges of the image is caused by the sinc factors in (3-40). In the case of the ordinary ROACH, the sinc factors depend only on the location in the image (since d is fixed); hence this attenuation factor can be compensated for by multiplying the desired image by $\text{sinc}^{-1}(dx/\lambda f)\text{sinc}^{-1}(dy/\lambda f)$ before Fourier transforming. Then during reconstruction the sinc and sinc^{-1} factors cancel, leaving the desired image unattenuated. In the binary-amplitude ROACH, on the other hand, the value of d in the sinc factors is different for every Fourier coefficient. Therefore, instead of a simple attenuation there is noise added to the image near its edges and exact compensation is not possible. (Near the center of the image the sinc factors are near unity and have negligible effect). For apertures with shapes other than square, this source of noise may be even more severe. If the aperture is not symmetric about (u_0,v_0), then phase errors are introduced in addition to amplitude errors.

The intensity transmittance of one cell in a binary-amplitude ROACH (as well as its contribution to amplitude) is proportional to the area of the cell. Let the maximum amount of flux transmitted by any cell be 1.0 (that is, an aperture transparent over the entire area of the cell). Then, for a binary-amplitude ROACH, a cell with, say, 0.25 times the maximum amplitude transmittance will transmit a flux of 0.25. A cell of amplitude transmittance 0.25 in an ordinary ROACH will, however, transmit a flux of only $(.25)^2 = .0625$. Therefore, a binary-amplitude ROACH will transmit considerably more light than an ordinary ROACH, even though the images produced by them are equally bright. That additional flux transmitted by the binary-amplitude ROACH goes into higher orders of diffraction,

away from the desired image. The higher-order images due to sampling (see Chapter 6) are increased in intensity due to diffraction from the sharp edges of the apertures of the binary-amplitude ROACH, where the transmittance jumps between zero and one.

Another disadvantage of a binary-amplitude ROACH, is that, for amplitude control, it requires a number of resolution elements per Fourier coefficient on a display device (which must still have grey-level control for the phase exposure). The number of amplitude levels that may be achieved is equal to the number of resolution elements per Fourier coefficient plus one. Since all the resolution elements are used to determine the amplitude (unlike a Lohmann hologram, for which there is a trade-off between amplitude and phase quantization for a fixed number of resolution elements [Gabel & Liu, 1970]), only a few resolution elements per Fourier coefficient are required for good amplitude control. An ordinary ROACH, however, requires only a single resolution element per Fourier coefficient.

We successfully synthesized a binary-amplitude ROACH. We feel that for most purposes the advantages of the binary-amplitude ROACH (as compared to the ordinary ROACH) are outweighed by its disadvantages.

CHAPTER 4

COLOR IMAGES FROM COMPUTER-GENERATED HOLOGRAMS

4.1 General

In order to produce a full-color image, it is necessary to synthesize three color-separation holograms, one for each of the three primary colors: red, green, and blue. Each of the three color-separation holograms is illuminated by the corresponding color of light to produce three monochromatic images which are superimposed to form a full-color composite image. If the images are reconstructed using the 632.8 nm red line from a helium-neon laser and the 514.5 nm green and 457.9 nm blue lines from an argon-ion laser, then all the colors on a C.I.E. chart bounded by the triangle defined by these three wavelengths can be produced (Fig. 4.8).

Any type of computer-generated hologram can be used to produce color images, and the considerations that are important in ordinary monochromatic computer holography are equally important for color imagery. In addition to the ordinary considerations of hologram synthesis, two fundamental problems present themselves. First, if a hologram for one color also has two other colors passing through it, then there appear, in addition to the desired image, false images. Second, since the angle of diffraction is proportional to wavelength, the three desired images will have

Fig. 4.1 Overlaying corresponding Lohmann-type cells of different colors. Although the equation holds for intensity transmittance, it does not hold for the complex transmittance, due to a phase error in the region of overlap, where the film is thinner.

Fig. 4.2 A portion of a multiplexed binary detour-phase hologram for a color image. Since the different colors do not overlap, no undesirable phase variations result.

Fig. 4.3 Image from multiplexed binary detour-phase hologram. The desired and conjugate images are in the upper right and lower left corners, respectively. Also seen are zero order and other spurious images.

Fig. 4.4 Three ROACHs for red, green, and blue light, scaled according to wavelength.

Fig. 4.5 (a) Image from ROACHs.
(b) Image from parity-sequence holograms. The difference in the spacing of the dots is a result of scaling the digital representation of the object according to the wavelength.

Fig. 4.6 (a) A ROACH for blue light using the phase-null effect, illuminated by blue light. (b) the same ROACH illuminated by red light.

Fig. 4.7 Reconstruction from a phase-null ROACH. The desired central blue image is undisturbed while the undesired red image is diffracted off to the sides.

Fig. 4.1

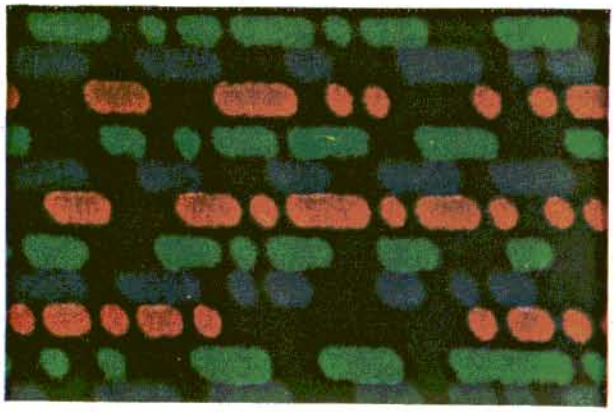


Fig. 4.2

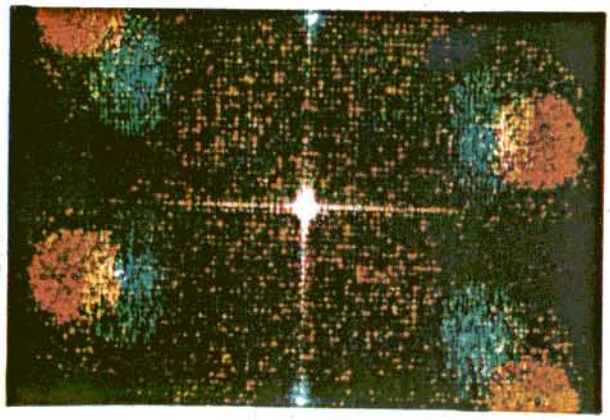


Fig. 4.3

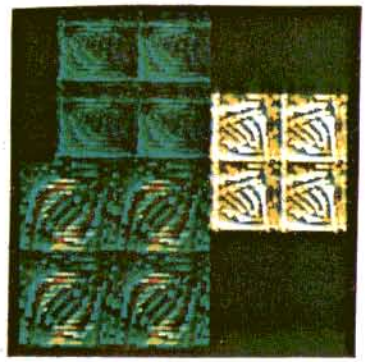
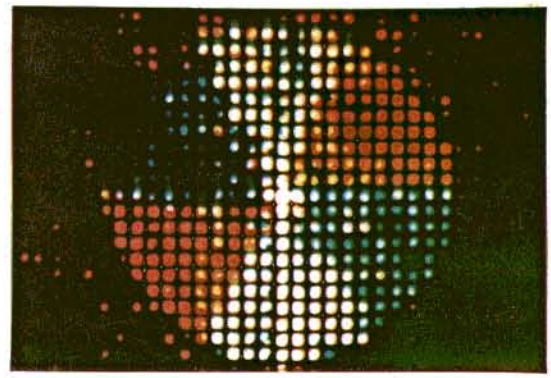


Fig. 4.4



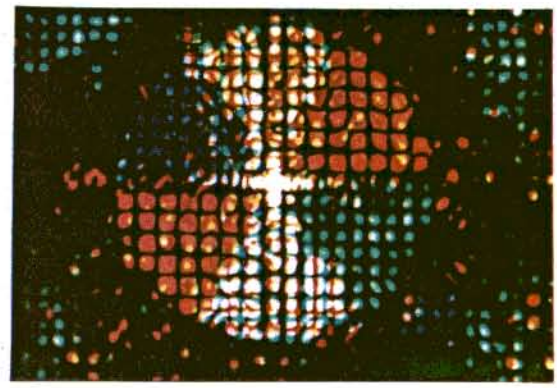


Fig. 4.5a,b

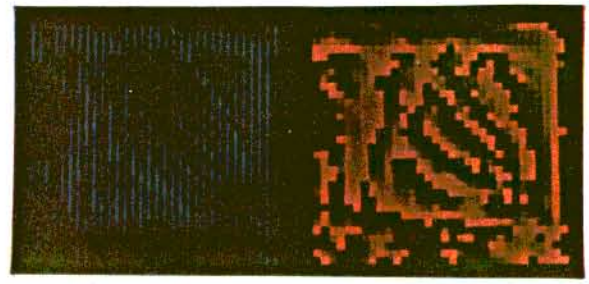


Fig. 4.6a,b

Fig. 4.7

three different magnifications, as illustrated in Figure 4.9, requiring a scaling procedure if they are to superimpose in register.

Scaling according to wavelength can be accomplished in a number of ways. Since image size is inversely proportional to the final hologram size, exact scaling can be done by adjusting the sizes of the hologram transparencies during their synthesis by adjusting the scale of the display device or the magnification when photographing the display. Alternatively, scaling can be done on the three sets of color-separation digitized data before Fourier transforming, effectively scaling the sampling intervals for the different colors. This latter scaling method can distort and misalign the images by up to one-half of a sampling interval, as will be demonstrated in Section 4.3; also, the image field sizes remain unchanged, even though the images within them are scaled, so part of the potential image space may go to waste (see Fig. 4.9).

Yet another method of scaling is accomplished by the readout geometry of Figure 4.10. If a Fourier transform hologram is placed in a converging beam, then the desired image is obtained in the focal plane of the lens, with a magnification that is proportional to d, the distance from the hologram to the image plane [Goodman, 1968]. Thus, the desired scaling can be accomplished by choosing the three hologram-to-image distances according to the wavelengths. Furthermore, since each hologram is illuminated only by the color of light appropriate to its image, no false images occur. Unfortunately, this setup requires the accurate positioning of three separate holograms, each in a different location. Other complicated reconstruction geometries based on spatial frequency multiplexing have also been devised [Dallas, Ichioka, and Lohmann, 1972].

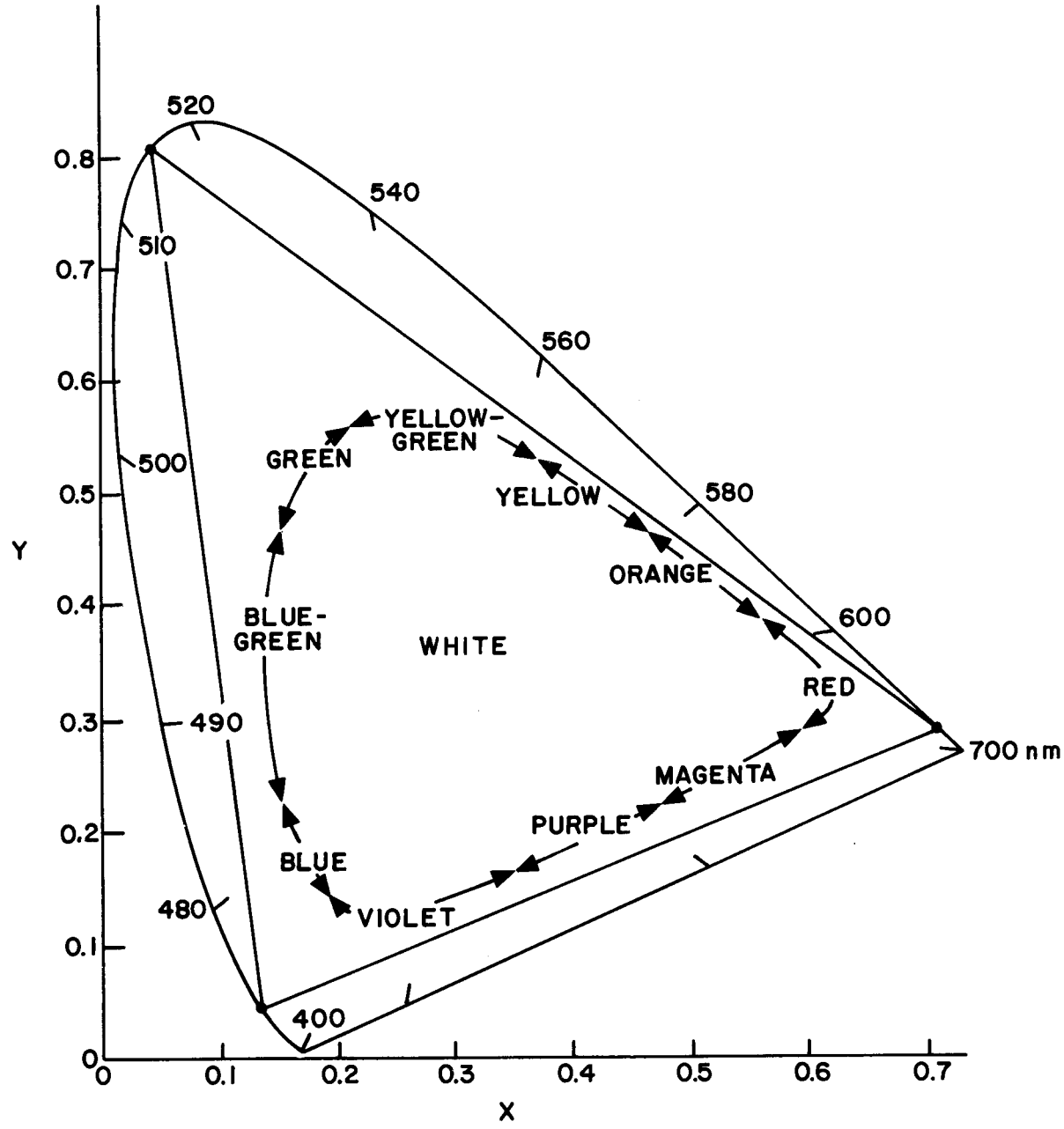


Fig. 4.8 The CIE chromaticity diagram: Any color within the triangle can be obtained by mixing red (632.8nm), green (514.5nm), and blue (457.9nm) light.

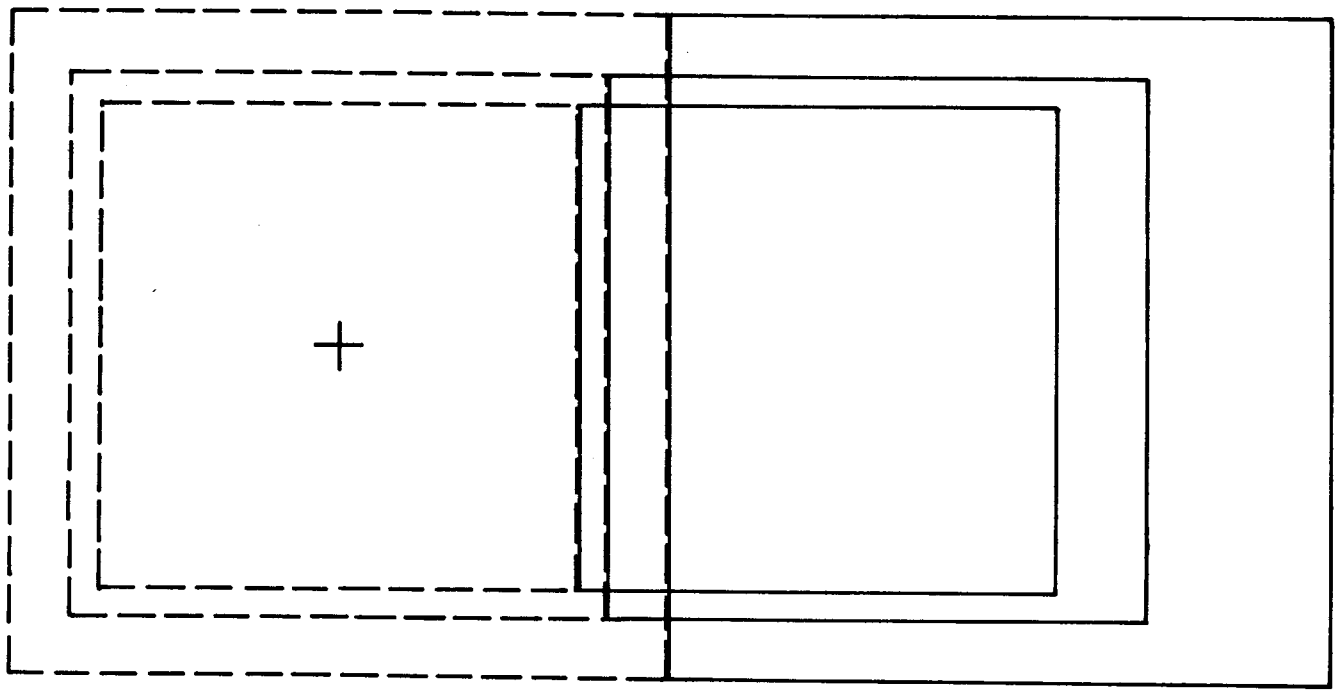


Fig. 4.9 Image areas for on-axis (broken lines) and first-order (solid lines) reconstructions. "+" denotes the optical axis. Largest squares indicate the image area in red (632.8nm), smaller squares in green (514.5nm), and smallest squares in blue (457.9nm). Note the difference in position as well as size in the first-order reconstruction.

Figure 4.11 shows a much simpler setup which is particularly desirable if the same laser is used to produce more than one color, since all three planes waves have a common axis. All three holograms can be made side-by-side as a unit on a single transparency. If the holograms are properly scaled, then the three desired images superimpose correctly, due to the shift invariance of this Fourier transforming geometry (see Appendix A). Three color filters are inserted or some other optical means may be used so that each hologram is illuminated only by light of the appropriate color; otherwise, six false images would appear in addition to the three desired images.

The setup of Figure 4.11 would be made even more desirable if the color filters could be eliminated without introducing false images, that is, if the holograms themselves could be made to reject the false images. If that were possible, then the holograms could be spatially multiplexed (interleaving rows of Fourier coefficient cells of the different colors, for example) instead of being positioned side by side. Spatial multiplexing is desirable when the image is viewed directly by the eye (i.e., the lens in Figure 4.11 is the lens of the eye) and is necessary for Fresnel holograms, which do not possess the shift-invariance property (Appendix A). In this chapter it will be shown that the extra degrees of freedom in multi-emulsion color film do allow the holograms themselves to reject false images.

4.2 Binary Detour-Phase Holograms

The obvious property of color film that can be used to eliminate false images is the wavelength-selective absorption of its different layers. One method of eliminating false images would be to make three

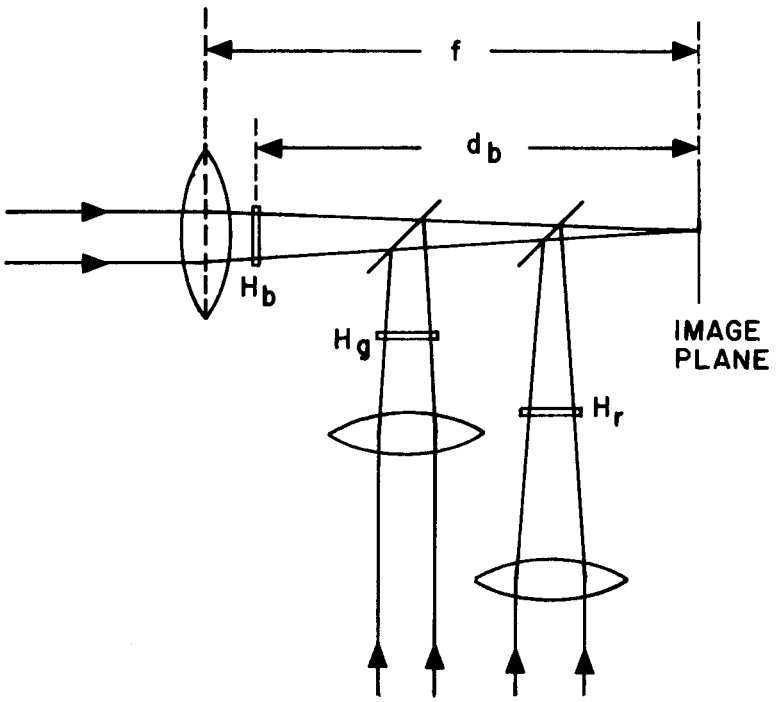


Fig. 4.10 Reconstruction geometry allowing scaling. H_b, H_g , and H_r are holograms illuminated by blue, green, and red light, respectively.

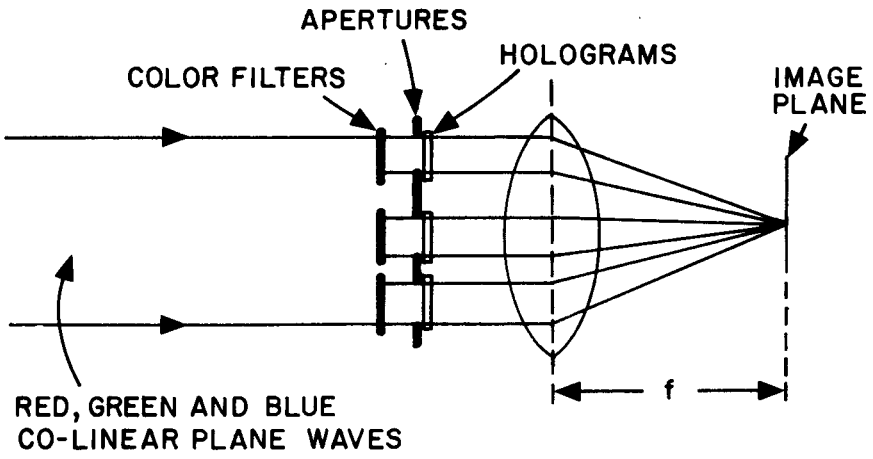


Fig. 4.11 Reconstruction geometry using three co-linear beams.

binary detour-phase holograms on color film, each exposed through a color filter appropriate to the color of its image. Then false images are avoided because only light of the desired color is transmitted by each hologram: the undesired colors are absorbed.

A disadvantage of binary detour-phase holograms is that they require many resolution elements of the display device to represent one Fourier coefficient; and if three color-separation binary detour-phase holograms were spatially multiplexed, then three times the display capability would be required. Attempting to reduce this extra display requirement, one might simply overlay cells of different colors, as shown in Figure 4.1*. However, at least for Kodachrome, the film would be thinner where two apertures overlap (the yellow area in Fig. 4.1), resulting in undesired phase variations. Thus, if cells are to be overlayed, then it is necessary to design the hologram cell in such a manner that apertures for different colors do not overlap.

As demonstrated by Haskell [1973], the opening of a combination of small binary apertures, or "subcells", within a cell can be used to represent a Fourier coefficient. Powers [1975] in particular studied the use of only one row of subcells within a cell to represent a Fourier coefficient. Referring to Figure 4.12, if one uses only a combination of five subcells along one line in a cell, then at the off-axis angle to the hologram corresponding to the first order of diffraction, light coming from each of these subcells is 1/5 wavelength out of phase with its neighbor. Any one of these apertures (subcells) contributes a unit phasor

*Figures 4.1 to 4.7 are located on the color plate on page 109.

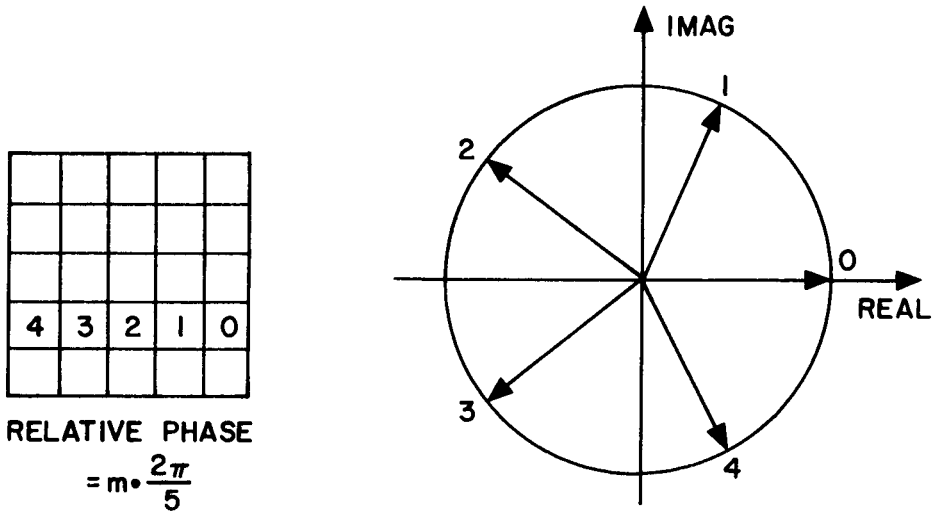


Fig. 4.12 Each open subcell in a binary detour-phase hologram contributes a unit phasor to the complex Fourier coefficient. If there are M subcells per coefficient, then each phasor is one distinct M^{th} root of unity.

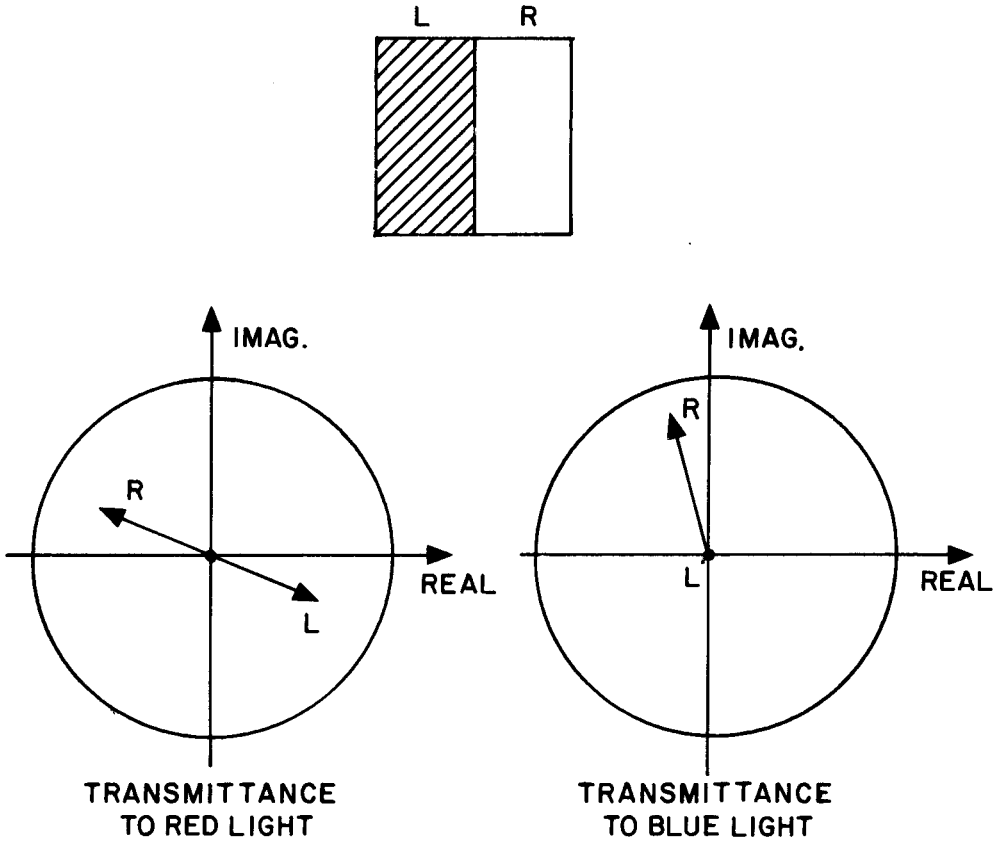


Fig. 4.13 Phase-null method.

(Fig. 4.12). By opening a combination of the five apertures, the corresponding unit phasors are added, allowing 31 points in the complex plane to be addressed. The attractive feature of this type of hologram is that there is enough unused space in the cell to multiplex in two more colors, without overlap and without requiring three times the display capability. For multiplexed holograms, scaling is performed on the digitized object. Figure 4.2 shows a (greatly magnified) portion of such a multiplexed binary detour-phase hologram, which has a cell of five sub-cells, 40×50 coefficients, and was replicated four times (in a 2×2 array) to give the image shown in Figure 4.3

Three problems plague the detour-phase holograms: their low diffraction efficiency (theoretically 10.3% maximum, but usually less than 1% in practice), the large number of resolvable elements of the display device needed to encode a single Fourier coefficient, and the abundance of spurious images (which are seen in Figure 4.3). All three of these problems are solved by the use of on-axis holograms.

4.3 On-Axis Holograms and the Phase-Null Effect

As discussed in Section 2.2.3, by using different combinations of emulsions (modes) of Kodachrome, ROACHs can be made for illumination with any of the three primary colors. Shown in Figure 4.4 are three ROACHs, one each for illumination with red light (lower left), green light (upper left) and blue light (right). Scaling for the three wavelengths was accomplished by adjusting the x and y gain controls on the CRT, giving three different sizes of holograms. Each ROACH is replicated 4 times (in a 2×2 array), giving the image (Figure 4.5a) a dot structure. (The original color picture, from which the color Xerox of Figure 4.5

was made, is yellow in the upper center section, blue in the upper left section and green in the lower right section.)

Figure 4.5b shows an image from three parity-sequence holograms [Chu and Goodman, 1972] made on Kodachrome used in phase-only (kinoform) modes. The considerable noise in the corners of the image arises from the parity sequence elements added to the desired image to level the spectrum. In this case scaling was done on the digitized object, in order to demonstrate the difference in sampling rates for the three different colors.

Both the ROACHs and the parity-sequence holograms that produced the images shown in Figure 4.5 required color filters in the read-out geometry (Fig. 4.11) in order to avoid false images. As will now be explained, it is possible to remove this requirement and have the on-axis holograms themselves reject the false images. For a given color-separation hologram, one incident color produces the desired image, while the two other incident colors produce false images. Of these latter two colors, one can be completely absorbed by one layer of the three-layer film, still leaving two layers to control both amplitude and phase of the desired color. For example, for a ROACH for illumination with blue light, the blue-absorbing layer could be used for amplitude control and the red-absorbing layer could be used for phase control, leaving the green-absorbing layer unexposed during synthesis so that the processed transparency absorbs green light (unlike the usual manner in which the green-absorbing layer would be eliminated--see Section 2.2.3). Similarly, the green false color image from a parity sequence hologram for blue light could be eliminated by using only the red-absorbing layer for phase control.

The second color leading to a false image can be diffracted away from the desired image by using the "phase-null" method [Chu, 1973]. Consider a single Fourier coefficient cell in a hologram designed for a blue image. During ordinary synthesis, the cell is exposed uniformly over its area to blue and red light to give the processed transparency the proper complex transmittance to blue light. For the phase-null method, the cell is divided into two halves, as illustrated in Figure 4.13. The right-hand half is exposed as usual to blue and red light to give the processed transparency the proper complex transmittance to blue light. The left-hand half is exposed the same to red but much less to blue light than is the right half. This difference in exposure to blue light causes, in the processed transparency, a phase difference to red light between the two halves of the cell. If exposed properly, this phase difference can be made to be π radians, that is, $\frac{1}{2}$ wavelength, of red light. Then for red light, the net contribution on-axis due to this cell is zero, since the two halves cancel one another, being equal in amplitude but opposite in phase (Fig. 4.13). The red light is diffracted off to the sides, away from the desired image. To blue light, however, the left-hand half is opaque compared to the right-hand half, so only the right-hand half makes a significant contribution to the phasor; and since the right-hand half was exposed to give the proper complex transmittance, then the blue image reconstructs as it should. Therefore, the result of the phase-null effect is to attenuate a false image without disturbing the desired image. So for a given hologram, of the two colors that could form false images, one can be absorbed by the film (green in the example above) and the other (red) can be diffracted off to the sides without

disturbing the desired (blue) image. If this is done for all three holograms, then no disturbing false images would appear, and the holograms could be spatially multiplexed, if desired. A more detailed mathematical treatment of the phase-null effect is given in Appendix D.

Figure 4.6a shows such a phase-null ROACH for a blue image illuminated with blue light. Notice the alternating vertical bands of transmitting and opaque blue half-cells. Figure 4.6b shows the same phase-null ROACH illuminated with red light. Both halves of each square cell transmit red light equally, but due to the difference in exposure to blue, the two halves are out of phase with one another. Figure 4.7 shows the reconstruction of such a phase-null ROACH illuminated with both red and blue light. The blue image reconstructs on-axis as expected, while the false red image is diffracted to both sides, away from the desired blue image.

Unfortunately, the phase-null method is difficult to employ. Since only one layer is used for phase control, that one layer alone must be capable of phase-matching (producing a phase excursion of 2π radians). As seen from Table 2.5, phase-matching is possible with a single layer of Kodachrome II only if it is illuminated by blue or green light. Phase-matching for red light cannot be accomplished strictly by the blue-absorbing layer alone. However, if exposures are made through a blue Wratten filter No. 98, then exposures above the value needed to eliminate the blue-absorbing layer still cause an increased phase delay since, in that region of exposure, the green-absorbing layer is also being exposed (see Section 2.3.3). Thus, phase-matching for red light can be achieved by using the blue-absorbing layer over the entire range of its

H&D curve, together with the green-absorbing layer in the shoulder region of the green H&D curve. In that region of the green H&D curve, green light is still heavily absorbed, so the green false image would still be greatly attenuated. Consequently, the phase-null method can be successfully employed with Kodachrome II. Kodachrome 25, on the other hand, is not suitable for the phase-null method since phase-matching cannot be achieved for any color with a single layer (Table 2.6). If color images are produced using only two of the primary colors (capable of producing all the colors along one line of the triangle in Figure 4.8), then only one false image per hologram must be avoided, and two layers can again be used for phase control.

Implementation of the phase-null effect requires very accurate control over the photographic process. The phase-null effect relies on two phasors being exactly equal in amplitude and opposite in phase in order to cancel one another. A small error in one of the phasors will destroy that cancellation. Making exact cancellation even more difficult is dye-density cross-talk. To the undesired color, in addition to a difference in phase of π radians between the two half-cells, there is also an amplitude difference, which requires compensation in the left half of the cell if exact cancellation is to occur. Since the blue-absorbing dye absorbs very little red light, no compensation was required for the example shown in Figures 4.6 and 4.7, but in general compensation is required.

4.4 Conclusion

The additional problems involved in the production of color images from computer-generated holograms can be solved in a number of ways. We

have described three methods for taking care of the scaling problem: by a digital step, during hologram synthesis, and during reconstruction. We have also shown three ways to eliminate false color images: by the optics of the reconstruction set-up, by wavelength-dependent absorption of color film, and by the phase-null method.

When full-color imagery is required, the advantages of on-axis holograms over binary detour-phase holograms (particularly their higher diffraction efficiency and economical use of the display device) are magnified (as is the principal advantage of binary holograms--ease of synthesis). Table 4.1 compares the binary detour-phase hologram, the parity sequence hologram, and the ROACH.

Hologram type[a]	Material required	Plotter points per coefficient	Maximum diffraction efficiency (%)	Inherent noise
Binary Detour-phase	Black/white[b]	Many[c]	10.3	Quantization[c] and detour-phase approx.
Parity Sequence	Phase only[b]	2	100	Parity noise
ROACH	Multi-emulsion	1	100	None

[a]Holograms using the phase-null method require multi-emulsion film, require double the number of plotter points per Fourier coefficient, and have a quarter of the diffraction efficiency listed.

[b]Multi-emulsion film is required for color multiplexing.

[c]Quantization noise is decreased by increasing the number of plotter points per Fourier coefficient.

Table 4.1 Comparison of binary detour-phase hologram, parity sequence hologram, and ROACH.

CHAPTER 5

ITERATIVE PROCEDURE FOR REDUCING QUANTIZATION NOISE

Due to the limitations of the display devices and materials used to synthesize computer holograms, it is often not possible to exactly represent any arbitrary complex Fourier coefficient. Instead, only a certain set of quantized values can be represented, and all other values are approximated by the quantized values. For example, as illustrated in Figure 5.1a, Lohmann's binary detour-phase hologram [Brown and Lohmann, 1969] can represent only a discrete (quantized) set of complex values, depending upon the number of resolution elements of the display device used to form one cell to represent a Fourier coefficient. The kinoform (Fig. 5.1b) [Lesem, Hirsch, and Jordan, 1969] allows nearly continuous phase control, but quantizes all the amplitudes to a single level. If a gray-level display device used to synthesize a kinoform has a finite number of gray-levels, then the phase is quantized as well. Similarly, a ROACH or any other continuous-tone (not binary) computer-generated hologram is quantized if a display device with a finite number of gray-levels is used for synthesis.

Let the sampled object be $\{a\}$ and its (discrete) Fourier transform be $\{A\} : \{a\} \supset \{A\}$. A desired $(p,q)^{th}$ complex coefficient A_{pq}

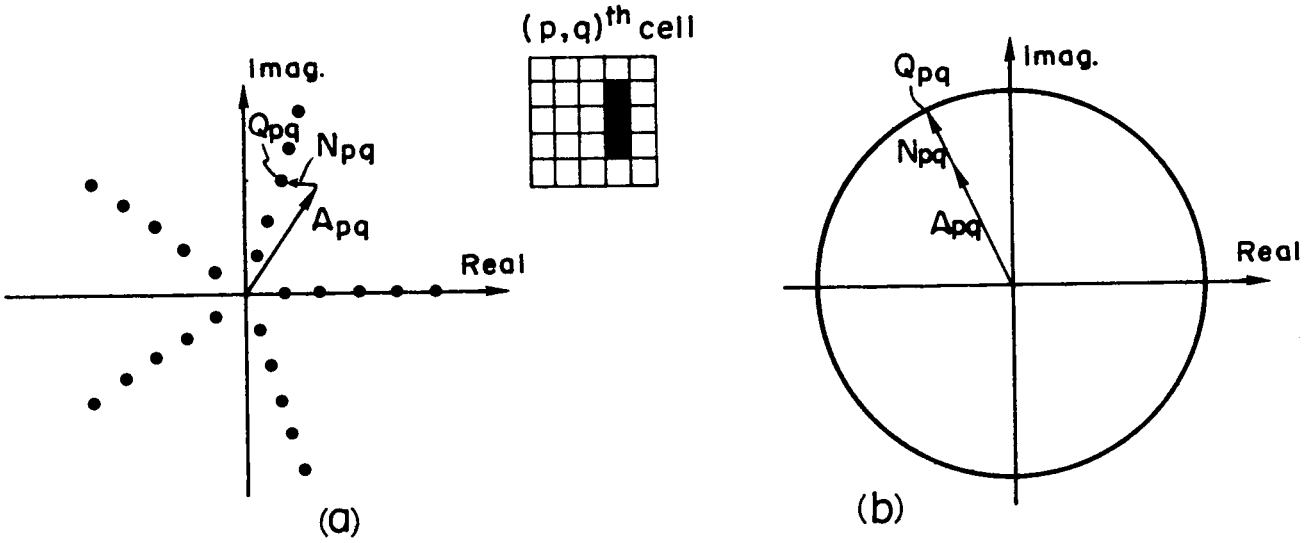


Fig.5.1 (a) Lohmann binary detour-phase hologram. Amplitude and phase are determined by the area and position, respectively, of an aperture within the cell. For 5×5 subcells per Fourier coefficient, only 26 points in the complex plane can be addressed. (b) Kinoform. Phase is determined by the thickness of the film, and the amplitude is quantized to a single level.

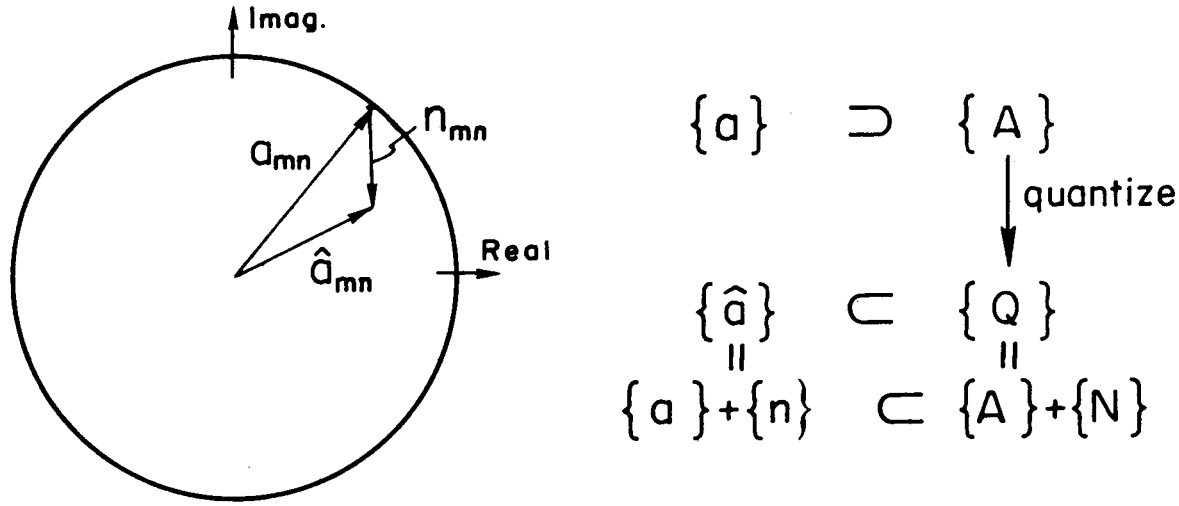


Fig.5.2 Quantization noise in the image produced by a computer-generated hologram.

(see Fig. 5.1) is only approximated by a quantized value, Q_{pq}, which equals $A_{pq} + N_{pq}$. N_{pq} is the Fourier-domain quantization noise for the $(p,q)^{th}$ Fourier coefficient. In the image plane (Fig.5.2), the result is quantization noise, $\{n\}$, the transform of $\{N\}$, added to the desired image. The actual image is $\{\hat{a}\} = \{a\} + \{n\}$, where $\{Q\} \subset \{\hat{a}\}$. Studies of quantization noise have been performed by a number of authors [Powers, 1975; Powers and Goodman, 1975; Gabel and Liu, 1970; Goodman and Silvestri, 1970].

5.1 Previous Methods, Phase Coding

One method of reducing (or eliminating, in certain cases) quantization noise is by the parity-sequence method [Chu and Goodman, 1972; Chu, 1974], but at the expense of additional complexity in the hologram and the interlacing of parity noise with the desired image.

Another approach, using the fact that only the intensity of the image is observed, is to assign various phase angles to the different points in the object (phase code the object) to help level the spectrum, which reduces the quantization noise in kinoforms. Better than random phase coding or other deterministic phase codes (see, for example, [Akahori, 1973]) is an iterative phase-coding method developed at IBM [Hirsch et al., 1971] and later discussed by Gallagher and Liu [1973]. That method, which is outlined by the block diagram of Figure 5.3, starts by random phase coding the object, then Fourier transforming, then setting all the Fourier amplitudes to a constant, and then inverse transforming, thereby calculating the image that would have resulted from a kinoform. Then the phases of that image are used as a new phase code for the object (and the cycle is repeated for a few iterations). This method substantially improves the image from a kinoform by finding

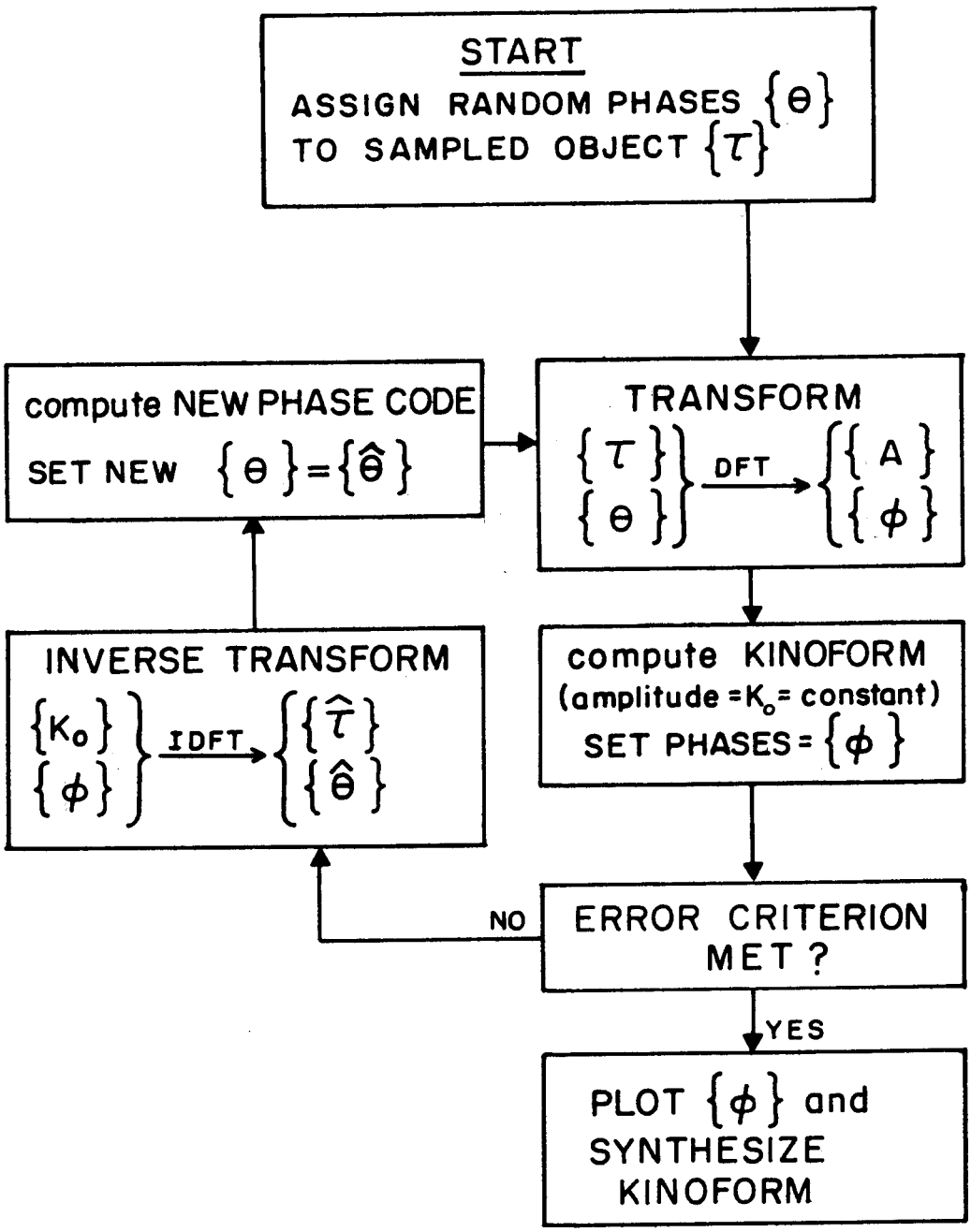


Fig.5.3 Previous iterative phase-coding method.

a very good phase code for the object.

We discovered a deficiency in the iterative phase-coding method that is particularly important to the computer memory application. After a number of iterations, the integrated squared error (noise) decreases, apparently asymptotically approaching a certain minimum value [Gallagher and Liu, 1973]. We followed the progress (at each iteration) of each individual point in the image and noticed that, although the error at most points generally decreased as the number of iterations increased, the error at some points failed to decrease, stabilizing at a high level. That is, the overall quality of the image improves, except at just a few points at which the error remains high. For the computer memory application, the error at those few points prevents the error rate (the fraction of incorrectly detected bits) from becoming arbitrarily small, even after an arbitrarily large number of iterations.

In an attempt to correct that deficiency in the iterative phase-coding method, the problem was attacked with a different approach, and a new method was developed that goes beyond mere phase coding.

5.2 Input-Output Approach: Manipulating Amplitude and Phase

Referring to Figure 5.4, we view the problem in this way: what we have is an input, $\{a\}$ (the object), and an output, $\{\hat{a}\}$ (the image), of a nonlinear computer-generated hologram system. The purpose of all our manipulations is to have the amplitudes of the output approximate as closely as possible some desired amplitude pattern $\{\tau\}$ (and we are not concerned with the phases of the output). The phase-coding methods all choose the amplitudes of the input equal to $\{\tau\}$ and manipulate the phases of the input in order to optimize the output. However, we claim

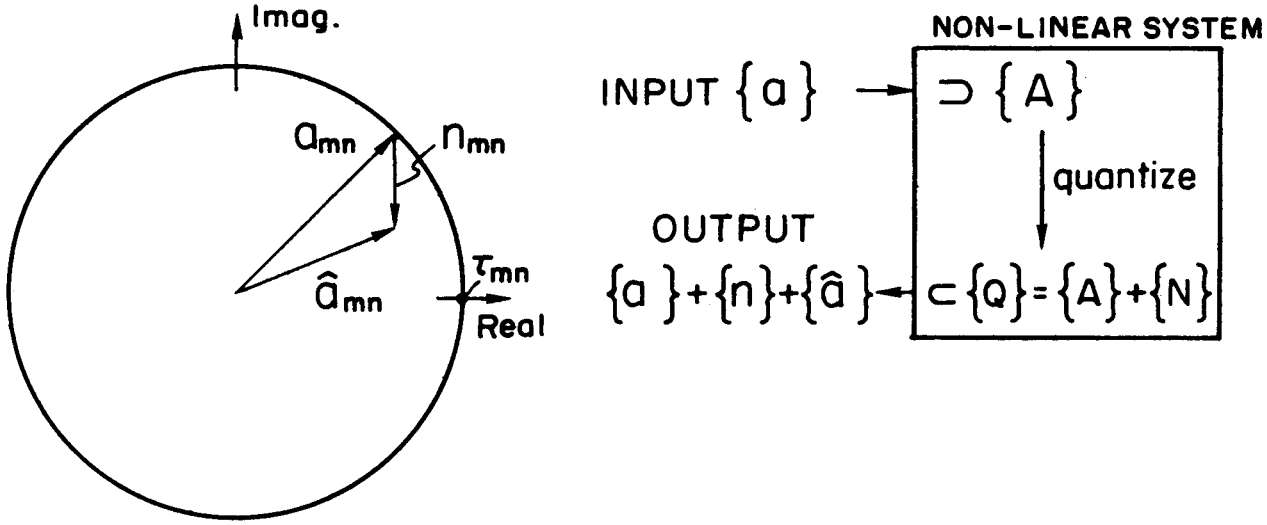


Fig.5.4 Input-output viewpoint. The desired output has $\{|\hat{a}_{mn}|\} = \{\tau_{mn}\}$.

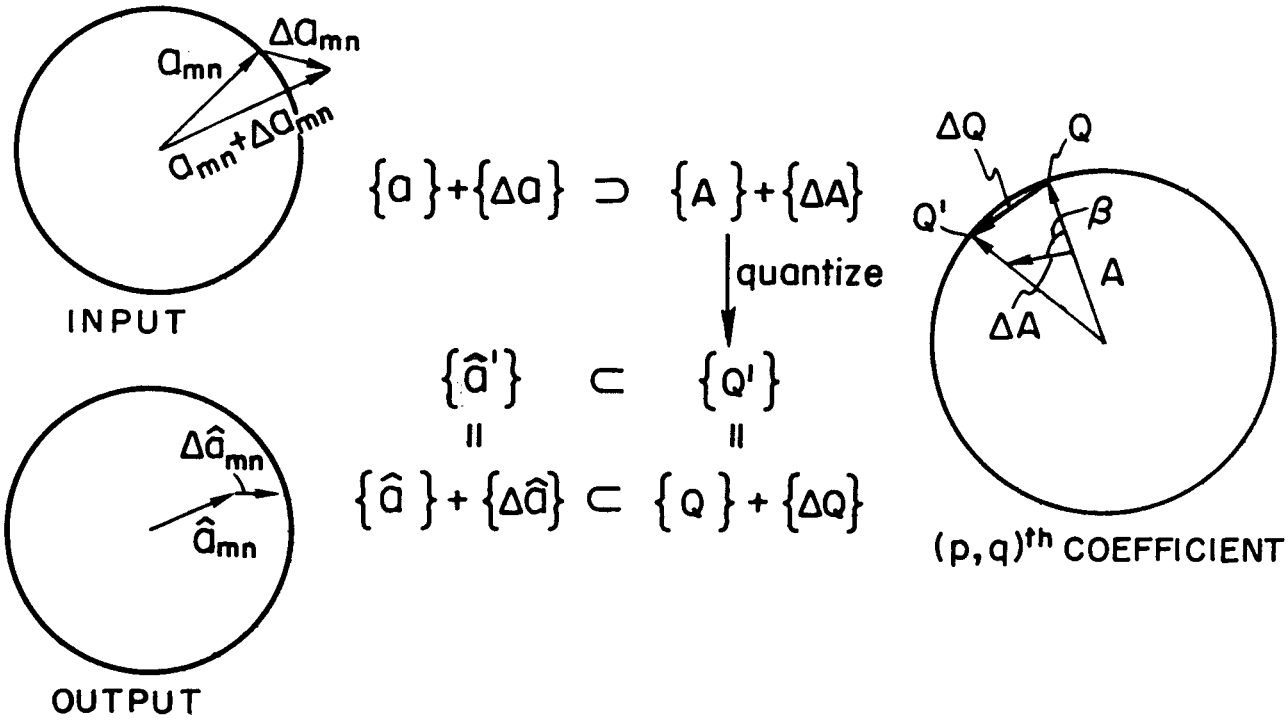


Fig.5.5 A change $\{\Delta a\}$, in the input results in a change, $\{\Delta Q\}$, in the kinoform and a change, $\{\Delta\hat{a}\}$, in the output.

that the input that gives the best output may have amplitudes different from $\{\tau\}$; that is, we should manipulate not only the phases, but also the amplitudes of the input in order to optimize the output.

5.2.1 Changes in Kinoform Image

In order to see in what manner the amplitudes and phases should be manipulated, we consider what happens in the kinoform system: if the input is changed in some way, what is the effect on the output? Referring to Figure 5.5, we start with an input $\{a\}$ that results in an output $\{\hat{a}\}$. The input is changed by adding $\{\Delta a\}$ to it. The corresponding change in the Fourier domain is that $\{\Delta A\}$, the transform of $\{\Delta a\}$, is added to $\{A\}$, and the sum is used to determine a new quantized value $\{Q'\}$. The change in the kinoform is $\{\Delta Q\} = \{Q'\} - \{Q\}$. The change in the output due to the change in the input is $\{\Delta\hat{a}\}$, the transform of $\{\Delta Q\}$. In this section we will investigate the expected value of $\{\Delta\hat{a}\}$.

Let K_0 be the level to which all the Fourier amplitudes, $\{|A|\}$, are set equal. Also, let $A_{pq} = |A_{pq}|e^{i\xi_{pq}}$. From Figure 5.5, we see that $Q_{pq} = K_0\, e^{i\xi_{pq}} = K_0 \cdot A_{pq}/|A_{pq}|$, and the changed $(p,q)^{th}$ coefficient in the changed kinoform is

$$Q'_{pq} = Q_{pq} + \Delta Q_{pq} = K_0(A_{pq} + \Delta A_{pq})/|A_{pq} + \Delta A_{pq}| \ . \qquad (5\text{-}1)$$

At this point we assume that $|\Delta A_{pq}| << |A_{pq}|$ for all (p,q) . This assumption will be true for nearly all (p,q) if the changes in the input are small, that is, if $\sum_{mn} |\Delta a_{mn}|^2 << \sum_{mn} |a_{mn}|^2$. Then, by Parseval's theorem, $\sum_{pq} |\Delta A_{pq}|^2 << \sum_{pq} |A_{pq}|^2$. Let the angle between ΔA_{pq} and A_{pq} be β_{pq} . Then $\Delta A_{pq}/|\Delta A_{pq}| = e^{i(\xi_{pq}+\beta_{pq})}$. Also, as illustrated in

Figure 5.6a, ΔA_{pq} can be broken up into two components. We define the following orthogonal components:

$$\Delta A^r_{pq} = |\Delta A_{pq}| \cos \beta_{pq} \, e^{i\xi_{pq}} \tag{5-2}$$

is the component of ΔA_{pq} parallel to A_{pq}, and

$$\Delta A^t_{pq} = |\Delta A_{pq}| \sin \beta_{pq} \, e^{i(\xi_{pq}+\pi/2)} \tag{5-3}$$

is the component of ΔA_{pq} orthogonal to A_{pq}, where

$$\Delta A_{pq} = \Delta A^r_{pq} + \Delta A^t_{pq} = |\Delta A_{pq}| \, e^{i(\xi_{pq}+\beta_{pq})} . \tag{5-4}$$

From the geometry of Figure 5.5, using the law of cosines and the identity $\cos(\pi - \beta_{pq}) = -\cos \beta_{pq}$, we find that

$$\begin{aligned} |A_{pq} + \Delta A_{pq}|^{-1} &= (|A_{pq}|^2 + |\Delta A_{pq}|^2 + 2|A_{pq}| \cdot |\Delta A_{pq}| \cos \beta_{pq})^{-\frac{1}{2}} \\ &\simeq (1 - |\Delta A_{pq}| \cos \beta_{pq}/|A_{pq}|)/|A_{pq}| \\ &= (1 - |\Delta A^r_{pq}|/|A_{pq}|)/|A_{pq}| \end{aligned} \tag{5-5}$$

for $|\Delta A_{pq}| << |A_{pq}|$. Inserting (5-5) into (5-1), we obtain

$$\begin{aligned} Q'_{pq} &\simeq K_0 (A_{pq} + \Delta A_{pq}) \cdot (1 - |\Delta A^r_{pq}|/|A_{pq}|)/|A_{pq}| \\ &= K_0 \frac{A_{pq}}{|A_{pq}|} + \frac{K_0}{|A_{pq}|} \Bigg[\underbrace{\Delta A_{pq} - |\Delta A^r_{pq}| \frac{A_{pq}}{|A_{pq}|}}_{\Delta A^t_{pq}} - \Delta A_{pq} \cdot \frac{|\Delta A^r_{pq}|}{|A_{pq}|} \Bigg] \\ &\simeq Q_{pq} + \Delta A^t_{pq} \left(\frac{K_0}{|A_{pq}|} \right) \end{aligned} \tag{5-6}$$

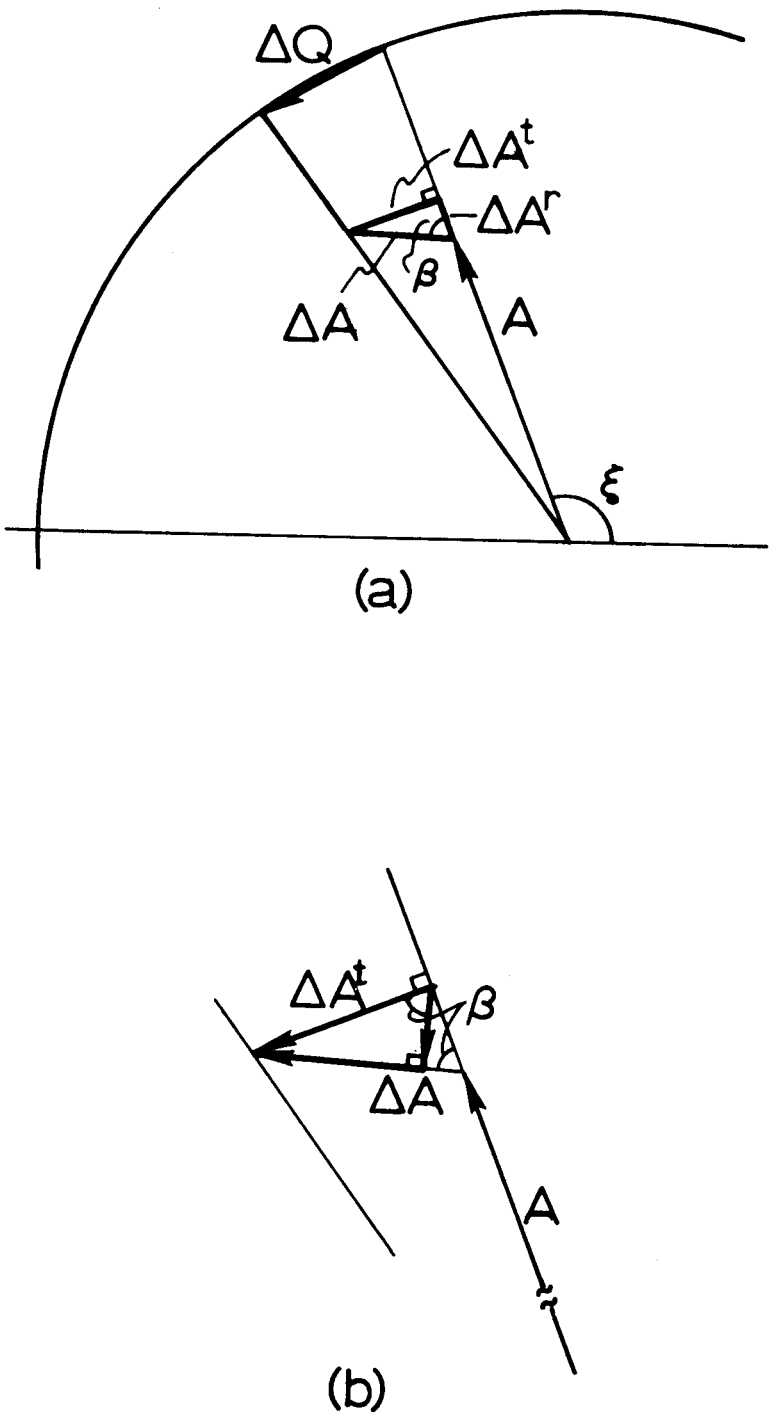


Fig.5.6 Geometry for determining expected change.

where the term of order $|\Delta A|^2$ was dropped. Therefore, the change in the kinoform due to the change in the input is (for small changes)

$$\Delta Q_{pq} \simeq \Delta A^t_{pq} \left(\frac{K_0}{|A_{pq}|} \right) \quad . \tag{5-7}$$

This result (5-7) can also be seen from the geometry of Figure 5.6a (similar triangles). Thus, we see that only one of two orthogonal components of ΔA_{pq} contributes to ΔQ_{pq}, the change in the kinoform.

We will investigate the average or expected change, $E\{\Delta \hat{a}\}$, treating $\{\Delta A\}$ as given and treating the Fourier coefficients A_{pq} and the phase angles β_{pq} as random variables. For example, if $\{a\}$ is random phase coded, then the amplitudes $|A_{pq}|$ are identically distributed Rayleigh random variables, and the phases ξ_{pq} are uniformly distributed on $[0,2\pi)$ [Powers and Goodman, 1975]. For this analysis we will assume that the A_{pq} are independent of each other, that the amplitudes $|A_{pq}|$ are identically distributed, and that the angles ξ_{pq} are uniformly distributed on $[0,2\pi)$. We will also assume that the angles β_{pq} are uniformly distributed on $[0,2\pi)$ and are independent of each other and of the A_{pq}. ΔQ_{pq} and ΔA^t_{pq} will also be treated as random variables since they depend on β_{pq}. A justification of these assumptions will not be given here; and, in fact, they are true only for special cases (see, for example, [Pearlman, 1974]). However, the results based on these assumptions are borne out by computer trials.

$\{\Delta \hat{a}\}$ is related to $\{\Delta Q\}$ through the (inverse) discrete Fourier transform (see Appendix B):

$$\Delta \hat{a}_{mn} = \frac{1}{N^2} \sum_{p,q=0}^{N-1} \Delta Q_{pq} \exp\left\{ \frac{i2\pi}{N} (mp + nq) \right\} \quad . \tag{5-8}$$

Inserting (5-7) into (5-8) and taking the expected value, yields

$$E\{\Delta\hat{a}_{mn}\} \simeq \frac{1}{N^2} \sum_{p,q=0}^{N-1} \exp\left\{\frac{i2\pi}{N}(mp+nq)\right\} \cdot E\left\{\Delta A_{pq}^t\right\} \cdot E\left\{\frac{K_0}{|A_{pq}|}\right\} \qquad (5\text{-}9)$$

where ΔA_{pq}^t and $|A_{pq}|^{-1}$ are assumed to be independent.

In order to calculate $E\{\Delta A_{pq}^t\}$, we further break up ΔA_{pq}^t into two orthogonal components. Referring to Figure 2.6b, the component of ΔA_{pq}^t that is parallel to ΔA_{pq} has a magnitude $|\Delta A_{pq}^t| \cdot |\sin \beta_{pq}|$, and the component that is orthogonal to ΔA_{pq} has a magnitude $|\Delta A_{pq}^t| \cdot |\cos \beta_{pq}|$. Since $|\Delta A_{pq}^t| = |\Delta A_{pq}| \cdot |\sin \beta_{pq}|$, then

$$\Delta A_{pq}^t = \Delta A_{pq} \sin^2\beta_{pq} + \left(\Delta A_{pq}\, e^{i(\pi/2)}\right) \sin \beta_{pq} \cdot \cos \beta_{pq} \qquad (5\text{-}10)$$

where the first and second terms are the components parallel to and orthogonal to ΔA_{pq}, respectively. The expected value of ΔA_{pq}^t is

$$E\left\{\Delta A_{pq}^t\right\} = \Delta A_{pq}\, E\left\{\sin^2\beta_{pq}\right\} + \left(\Delta A_{pq}\, e^{i(\pi/2)}\right) E\left\{\sin \beta_{pq} \cdot \cos \beta_{pq}\right\}. \qquad (5\text{-}11)$$

Assuming that β_{pq} is uniformly distributed on $[0,2\pi)$, then

$$E\left\{\sin^2\beta_{pq}\right\} = \frac{1}{2\pi} \int_0^{2\pi} \sin^2\beta_{pq}\, d\beta_{pq} = \tfrac{1}{2} \qquad (5\text{-}12a)$$

and

$$E\left\{\sin \beta_{pq} \cos \beta_{pq}\right\} = \frac{1}{2\pi} \int_0^{2\pi} \sin \beta_{pq} \cos \beta_{pq}\, d\beta_{pq} = 0 \,. \qquad (5\text{-}12b)$$

Therefore (5-11) reduces to

$$E\left\{\Delta A_{pq}^t\right\} = \tfrac{1}{2}\, \Delta A_{pq} \,. \qquad (5\text{-}13)$$

Inserting (5-13) into (5-9) and assuming that the $|A_{pq}|$ are identically distributed (i.e., $E\{K_0/|A_{pq}|\}$ is independent of (p,q)), we obtain the result

$$E\left\{\Delta\hat{a}_{mn}\right\} \simeq E\left\{\frac{K_0}{|A|}\right\}\frac{1}{N^2}\sum_{p,q=0}^{N-1}\exp\left\{\frac{i2\pi}{N}(mp+nq)\right\}\cdot\tfrac{1}{2}\,\Delta A_{pq} \qquad (5\text{-}14)$$

or

$$E\left\{\Delta\hat{a}_{mn}\right\} \simeq \tfrac{1}{2}\,\Delta a_{mn}\cdot E\left\{\frac{K_0}{|A|}\right\} \quad . \qquad (5\text{-}15)$$

From (5-15) we see that the expected change in the output is a fraction of the change in the input, that fraction depending on the statistics of the spectrum, $|A|$. If the spectrum is nearly level (that is, if $|A_{pq}| \simeq K_0$ for all values of (p,q)), then the expected change in the output is just half of the change in the input; that is, only half of the change Δa_{mn} makes it through the kinoform system.

Similarly, we compute the variance of $\Delta\hat{a}_{mn}$:

$$E\left\{|\Delta\hat{a}_{mn}-E\left\{\Delta\hat{a}_{mn}\right\}|^2\right\} = E\left\{|\Delta\hat{a}_{mn}|^2\right\} - |E\left\{\Delta\hat{a}_{mn}\right\}|^2 \quad . \qquad (5\text{-}16)$$

Taking the expected value of the squared modulus of (5-8), we obtain

$$E\left\{|\Delta\hat{a}_{mn}|^2\right\} = \frac{1}{N^4}\sum_{p,q=0}^{N-1}\sum_{r,s=0}^{N-1}\exp\left\{\frac{i2\pi}{N}\left[m(p-r)+n(q-s)\right]\right\}\cdot E\left\{\Delta Q_{pq}\Delta Q^*_{rs}\right\}. \qquad (5\text{-}17)$$

Using (5-7), we have

$$E\left\{\Delta Q_{pq}\Delta Q^*_{rs}\right\} \simeq E\left\{\frac{K_0^2}{|A_{pq}|\cdot|A_{rs}|}\right\}\cdot E\left\{\Delta A^t_{pq}\Delta A^{t*}_{rs}\right\} \qquad (5\text{-}18)$$

where again the ΔA^t_{pq} and $|A_{pq}|$ are assumed to be independent. Using (5-10), the second factor in (5-18) can be expressed as

$$\begin{aligned} E\left\{\Delta A^t_{pq}\Delta A^{t*}_{rs}\right\} = \Delta A_{pq}\cdot\Delta A^*_{rs}\cdot E\Big\{&\sin^2\beta_{pq}\sin^2\beta_{mn} + \\ &+ e^{-i(\pi/2)}\sin^2\beta_{pq}\sin\beta_{rs}\cos\beta_{rs} \\ &+ e^{i(\pi/2)}\sin\beta_{pq}\cos\beta_{pq}\sin^2\beta_{rs} \\ &+ \sin\beta_{pq}\cos\beta_{pq}\sin\beta_{rs}\cos\beta_{rs}\Big\}. \end{aligned} \qquad (5\text{-}19)$$

There are two cases to consider. The first case is for $(r,s) \neq (p,q)$; then, assuming that β_{pq} and β_{rs} are independent and using (5-12), we find that only the first term in (5-19) survives, leaving

$$E\left\{\Delta A^{t}_{pq}\Delta A^{t^*}_{rs}\right\} = \tfrac{1}{4}\,\Delta A_{pq}\Delta A^{*}_{rs} \qquad (r,s) \neq (p,q) \tag{5-20}$$

The second case is for $(r,s) = (p,q)$; then the pertinent integrals are

$$E\left\{\sin^4\beta_{pq}\right\} = \frac{1}{2\pi}\int_0^{2\pi} \sin^4\beta_{pq}\, d\beta_{pq} = 3/8 \tag{5-21a}$$

$$E\left\{\sin^3\beta_{pq}\cos\beta_{pq}\right\} = \frac{1}{2\pi}\int_0^{2\pi} \sin^3\beta_{pq}\cos\beta_{pq}\, d\beta_{pq} = 0 \tag{5-21b}$$

and

$$E\left\{\sin^2\beta_{pq}\cos^2\beta_{pq}\right\} = \frac{1}{2\pi}\int_0^{2\pi} \sin^2\beta_{pq}\cos^2\beta_{pq}\, d\beta_{pq} = 1/8 \ . \tag{5-21c}$$

Using (5-21), (5-19) reduces to

$$E\left\{\Delta A^{t}_{pq}\Delta A^{t^*}_{rs}\right\} = \tfrac{1}{2}\,\Delta A_{pq}\Delta A^{*}_{rs} \qquad (r,s) = (p,q) \ . \tag{5-22}$$

Combining (5-18), (5-20), and (5-22) with (5-17) results in

$$\begin{aligned} E\left\{|\Delta\hat{a}_{mn}|^2\right\} \simeq \tfrac{1}{4}\left(E\left\{\frac{K_0}{|A|}\right\}\right)^2 \cdot \frac{1}{N^2}\sum_{p,q=0}^{N-1} \exp\left\{\frac{i2\pi}{N}(mp+nq)\right\}\Delta A_{pq} \\ \cdot\frac{1}{N^2}\sum_{r,s=0}^{N-1} \exp\left\{\frac{-i2\pi}{N}(mp+ns)\right\}\Delta A^{*}_{rs} \\ + \frac{1}{N^4}\left[\tfrac{1}{2}E\left\{\frac{K_0^2}{|A|^2}\right\} - \tfrac{1}{4}\left(E\left\{\frac{K_0}{|A|}\right\}\right)^2\right]\sum_{p,q=0}^{N-1}|\Delta A_{pq}|^2 \end{aligned} \tag{5-23}$$

where $E\{(|A_{pq}|\cdot|A_{rs}|)^{-1}\} = E\{|A_{pq}|^{-1}\}\cdot E\{|A_{rs}|^{-1}\} = (E\{|A|^{-1}\})^2$, $(r,s) \neq (p,q)$ was used. Using (5-9) and Parseval's theorem, (5-23) reduces to

$$E\left\{|\Delta\hat{a}_{mn}|^2\right\} \simeq \tfrac{1}{4}|\Delta a_{mn}|^2\left(E\left\{\frac{K_0}{|A|}\right\}\right)^2 + \tfrac{1}{4}\left[2E\left\{\frac{K_0^2}{|A|^2}\right\} - \left(E\left\{\frac{K_0}{|A|}\right\}\right)^2\right]\frac{1}{N^2}\sum_{k,\ell=0}^{N-1}|\Delta a_{k\ell}|^2 \quad . \tag{5-24}$$

Combining (5-24) with (5-16) and (5-15), we obtain the result that the variance of the change in the output is

$$E\left\{|\Delta\hat{a}_{mn} - E\left\{\Delta\hat{a}_{mn}\right\}|^2\right\} \simeq \tfrac{1}{4}\left[2E\left\{\frac{K_0^2}{|A|^2}\right\} - \left(E\left\{\frac{K_0}{|A|}\right\}\right)^2\right] \cdot \frac{1}{N^2}\sum_{k,\ell=0}^{N-1}|\Delta a_{k\ell}|^2 \quad . \tag{2-25}$$

Equations (5-15) and (5-25) suggest how to go about manipulating the amplitude and phase in order to reduce quantization noise. If the computed output is something other than what is desired, then (5-15) suggests that if $\{\Delta a_d\}$ is the difference between the desired output and the actual output, then we should add to the input $\{\Delta a\} = 2(E\{K_0/|A|\})^{-1}\cdot\{\Delta a_d\}$. Then the expected value of the change in the output would be $\{\Delta a_d\}$, the desired change in the output. However, the actual change in the output would not be exactly $\{\Delta a_d\}$, due to the finite variance (5-25). The variance of the change in the output (5-25) can be considered to be a measure of additional noise (besides the original quantization noise) that is introduced into the output by the change in the input, and that noise prevents us from obtaining exactly the desired change in the output.

For the case of a kinoform of an object with a level (or nearly level) spectrum, (5-15) reduces to

$$E\{\Delta\hat{a}_{mn}\} \simeq \tfrac{1}{2}\Delta a_{mn} \qquad (5\text{-}26)$$

and (5-25) reduces to

$$E\left\{\left|\Delta\hat{a}_{mn}-E\{\Delta\hat{a}_{mn}\}\right|^2\right\} \simeq \frac{1}{4N^2}\sum_{k,\ell=0}^{N-1}\left|\Delta a_{k\ell}\right|^2 \quad . \qquad (5\text{-}27)$$

Equations (5-26) and (5-27) were tested by computer trials in which small changes were made in the input and the resulting changes in the output were computed. For small changes, (5-26) and (5-27) were found to be very accurate.

In order to obtain some indication of the accuracy with which the output can be changed, we will consider three cases. The first case is for $|\Delta a_{mn}| = |\Delta a_0|$ for all (m,n) ; that is, the modulus of the change in the input is the same for all points (m,n) . Then the standard deviation of the change in the output, given by the square root of (5-27), is equal to $|\Delta a_0|/2$. From (5-26), the modulus of the expected change in the output is $|E\{\Delta\hat{a}_{mn}\}| \simeq \tfrac{1}{2}|\Delta a_{mn}| = |\Delta a_0|/2$. In this case, the additional noise (introduced by the change in the input) is as great as the expected change in the output; therefore, changing the input in this case allows little control over the change in the output.

The second case we will consider is for $\Delta a_{mn} = \Delta a_0 \delta_{m,r}\delta_{n,s}$; that is, only the $(r,s)^{th}$ point in the input is changed. From (5-26), $E\{\Delta\hat{a}_{mn}\} \simeq \tfrac{1}{2}\Delta a_0 \delta_{m,r}\delta_{n,s}$. From (5-27), the standard deviation (a measure of the additional noise) of the change in the output is $|\Delta a_0|/(2N)$. The average (expected) value of the additional noise at a given point depends only on the total change in the input rather than on the change in the input at that particular point. For an image of modest complexity, N^2 is greater than 1000. Therefore, the standard deviation is less than

1/30 of the expected change at the $(r,s)^{th}$ point. That is, when only one point is changed, then the change at the corresponding output point can be controlled very accurately. Weak noise of average amplitude $|\Delta a_0|/(2N)$ is spread over the entire output as a result of the change of a single point in the input.

The third case we will consider is for $|\Delta a_{mn}|$ different for each point (m,n). This is the case of interest, since the desired change in the output is generally different for each point, depending on the error at each point. Once again, the average additional noise at each output point is independent of the change in the input at that particular point. Therefore, similar to the first case mentioned above, for those points with comparatively small values of $|E\{\Delta\hat{a}_{mn}\}|$, the actual change in the output, $\Delta\hat{a}_{mn}$, is heavily influenced by the additional noise and bears little resemblance to $E\{\Delta\hat{a}_{mn}\}$. However, similar to the second case mentioned above, for those points with comparatively large values of $|E\{\Delta\hat{a}_{mn}\}|$, Δa_{mn} is only slightly influenced by the additional noise, and is approximately equal to $E\{\Delta\hat{a}_{mn}\}$. That is, the output points which require the greatest change (i.e., have the greatest error) are the ones that can be changed most accurately. The output points with comparatively small error cannot be changed accurately (and may actually have greater error after the change than before); but there is less need to accurately change the points with small error if there are points with much larger error in the same output (more will be said about this later).

If the input is random phase coded, then the amplitudes $|A_{pq}|$ are identically Rayleigh distributed [Powers and Goodman, 1975]: the probability density of $|A|$ is

$$p(|A|) = (|A|/\sigma^2)e^{-|A|^2/2\sigma^2} \quad , \; 0 \leq |A| < \infty \tag{5-28}$$

where $2\sigma^2 = E\{|A|^2\}$ is the second moment. Using the identity $\int_0^\infty x^n e^{-ax^2} dx = \Gamma((n+1)/2)/2a^{(n+1)/2}$, we find that

$$E\{|A|\} = \sqrt{\frac{\pi}{2}}\,\sigma \tag{5-29}$$

and

$$E\{1/|A|\} = \sqrt{\pi/(2\sigma^2)} \quad . \tag{5-30}$$

Thus, for the random phase-coded case, (5-15) reduces to

$$E\{\Delta\hat{a}_{mn}\} \simeq \tfrac{1}{2}K_0\sqrt{\pi/(2\sigma^2)}\;\Delta a_{mn} \quad . \tag{5-31}$$

Using $K_0 = \sqrt{E\{|A|^2\}}$ in order to keep $\sum_{pq} K_0^2 = \sum_{pq}|A|^2$, (5-31) becomes

$$E\{\Delta\hat{a}_{mn}\} \simeq \frac{\sqrt{\pi}}{2}\,\Delta a_{mn} = 0.886\;\Delta a_{mn} \quad . \tag{5-32}$$

We also find that

$$E\{1/|A|^2\} = \int_0^\infty \frac{1}{|A|\sigma^2}\, e^{-|A|^2/2\sigma^2}\, d|A| \to \infty \quad . \tag{5-33}$$

Therefore, (5-25) diverges, due to the assumption that $|\Delta A_{pq}| << |A_{pq}|$ for all (p,q), an assumption that cannot be used in the case of Rayleigh distributed $|A|$. That is, (5-25) is invalid for any case that allows $|A_{pq}|$ to be very small. The accuracy of (5-15) is also reduced for those cases, making the prediction of $E\{K_0/|A|\}$, the value of which needed to control the change in the output, more difficult. In any case, though, the value of $E\{K_0/|A|\}$ can be determined experimentally, by making a change in one point of the input, calculating the resulting change in the corresponding point in the output, and using (5-15). In this way

$E\{K_0/|A|\}$ was found to be 0.90 for a random phase coded input, in close agreement with (5-32).

5.2.2. The General Case

While Section 5.2.1 dealt strictly with the kinoform system, we also found, by computer tests, similar behavior to (5-15) and (5-25) in computer-generated holograms with other types of quantization. In general, for all types of quantized computer-generated holograms, a change in the input results in a fraction of that change in the output, plus additional noise. That is,

$$E\{\Delta\hat{a}_{mn}\} = \alpha\Delta a_{mn} \qquad (5\text{-}34)$$

where the fraction α depends on the type of quantization and upon the statistics of $|A|$. The average intensity of the additional noise, which is equal to the variance of $\Delta\hat{a}_{mn}$, depends on the type of quantization, on the statistics of $|A|$, and on the amount of change in the input. (5-34) suggests that, in order to achieve a desired change $\{\Delta a_d\}$, the input should be changed by

$$\{\Delta a\} = \alpha^{-1}\{\Delta a_d\} \quad . \qquad (5\text{-}35)$$

5.2.3. A Second Method

An alternate method can also be used. Suppose that an output $\{\hat{a}\}$ is used as an input. Since $\{Q\}$, the transform of $\{\hat{a}\}$, is already quantized, then the quantization step will have no effect on it. Therefore, if the output $\{\hat{a}\}$ is used as an input, then its output will simply be itself. Thus, no matter what input actually resulted in the output $\{\hat{a}\}$, the output $\{\hat{a}\}$ can always be considered to have resulted from itself as an input. Consequently, the change given by (5-35) can be

added to the output, and the sum $\{\hat{a}\} + \alpha^{-1}\{\Delta a_d\}$ can be used as the new input. This alternate method (adding (5-35) to the output to form a new input) will be referred to as method II, and the previously discussed method of adding (5-35) to the input will be referred to as method I.

In the Fourier domain, the effect of method II is to use the sum $Q_{pq} + \Delta A_{pq}$ (rather than $A_{pq} + \Delta A_{pq}$) to determine the new quantized value Q'_{pq}. Then in the case of the kinoform system (see Fig.5.6) for method II,

$$Q_{pq} \simeq \Delta A^t_{pq} \tag{5-36}$$

replaces (5-7). For method II, we find that (5-15) is replaced by (5-26), and (5-25) is replaced by (5-27). That is, irrespective of the actual statistics of $|A|$, method II produces changes in the output that are the same as the changes obtained with method I for a perfectly level spectrum ($|A_{pq}| = K_0$). In fact, for method II, $\{\Delta\hat{a}\}$ is treated as though it did result from an input with a perfectly level spectrum, since $\Delta\hat{a}$ is treated as being the input, and $\Delta\hat{a}$ does have a perfectly level spectrum.

A disadvantage of method II is that $\alpha^{-1} = \frac{1}{2}$, whereas α^{-1} is generally greater than $\frac{1}{2}$ for method I, depending upon the statistics of $|A|$ (.89 for the case of a random phase coded input). Therefore, method II requires greater changes in the input (and a greater factor $\sum_{k\ell} |\Delta a_{k\ell}|^2$ in (5-26) contributing to the additional noise) than does method I, in order to achieve the same change in the output. However, depending on the statistics of $|A|$, the factor $\frac{1}{4}\left[2E\{K_0^2/|A|^2\} - (E\{K_0/|A|\})^2\right]$ in (5-25) may be considerably greater than the factor of $\frac{1}{4}$ in (5-27); therefore, in general there is no clear advantage of method II over

method I or vice versa. In practice the two methods work with comparable success.

Still other methods are possible, too. For example, the roles of the object domain and of the Fourier domain can be reversed. That is, the input and output could occur in the Fourier domain while the "system" consists of an inverse transform, a setting of the object domain amplitude equal to the desired amplitude $\{\tau\}$, and a forward Fourier transform. However, only methods I and II were investigated in any detail.

5.2.4. The Choice of the Desired Change

Recall that the desired output has amplitude $\{\tau\}$ and the phase of the output is unimportant. Then for each output point $\hat{a}_{mn}$, there are an infinite number of possible desired outputs, one for each possible value of phase, $[0,2\pi)$. Therefore, for a given output $\{\hat{a}\}$, there are an infinite number of possible choices of $\{\Delta a_d\}$, the desired change in the output. The particular choice of $\{\Delta a_d\}$ is very important, as it determines the success with which the method works. A few of the considerations that come into play in picking $\{\Delta a_d\}$ will be discussed in this section.

Desired Change Related to Noise

Consider the case for which the amplitude of the input, $\{|a|\}$, is $\{\tau\}$, the desired amplitude of the output. Then the quantization noise is

$$\{n\} = \{\hat{a}\} - \{a\} \quad . \tag{5-36}$$

Obviously a change in the output that would yield the desired output is $-\{n\}$. However, as will be shown, the change $-\{n\}$ is not effective in changing the output. In this case, the new input is

$$\{a'\} = \{a\} + \{\Delta a\} = \{a\} - \alpha^{-1}\{n\} \quad . \tag{5-37}$$

In the transform domain, the change in the transform is

$$\{\Delta A\} = -\alpha^{-1}\{N\} \quad . \tag{5-38}$$

In the case of the kinoform, Figure 5.1b shows that the phase of N_{pq} and the phase of A_{pq} differ by either 0 or π radians. Therefore, in this case the phase of $A_{pq} + \Delta A_{pq}$ differs from the phase of A_{pq} also by 0 or π radians. In particular, the phases of A_{pq} and $A_{pq} + \Delta A_{pq}$ are the same if $\alpha^{-1}|N_{pq}| < |A_{pq}|$ and differ by π radians if $\alpha^{-1}|N_{pq}| > |A_{pq}|$. Thus, the new quantized value in the Fourier domain is either Q_{pq} or $-Q_{pq}$, respectively. When $\alpha^{-1}|N_{pq}| < |A_{pq}|$ for all coefficients, the new quantized values do not differ from the old quantized values, and so the result is that no change occurs in the output at all. Similarly, no change whatsoever occurs in the output of a kinoform system if the input is changed by $\{n\}\cdot$const., if $0 \leq$ const. ≤ 1 . Note that const. = 1 is just the use of the output as the input. For the case of $\{\Delta a\}$ proportional to $\{n\}$, the results of Section 5.2.1 do not hold since the angle β_{pq} is then highly correlated with ξ_{pq} (that is, $\beta_{pq} = \pm \xi_{pq}$). Another case for which the results of Section 5.2.1 do not hold is for $\Delta a_{mn} = n_{mn} e^{i(\pi/2)}\cdot$const., again because β_{pq} is highly correlated with ξ_{pq} . In this case, $\Delta A^{t}_{pq} \simeq \Delta A_{pq}$, and the factor of ½ in the right hand side of (5-15) is replaced by unity. From these last two cases we see that changes in the input that tend to be orthogonal to $\{a\}$ and to $\{n\}$ tend to be much more effective in changing the output than are changes that tend to be parallel to $\{a\}$ and to $\{n\}$.

Other Considerations

Another important consideration is the magnitude of the desired change, $\{|\Delta a_d|\}$. Since the additional noise increases with increasing $\sum_{k\ell} |\Delta a_{k\ell}|^2$, it is desirable to minimize the magnitude of the desired change. The desired change that has the smallest magnitude is $\tau_{mn}(\hat{a}_{mn}/|\hat{a}_{mn}|) - \hat{a}_{mn}$, which has magnitude $|\tau_{mn} - |\hat{a}_{mn}||$.

Still another important consideration is the fact that the previous iterative phase-coding method is effective in reducing the dynamic range of the spectrum, which is highly desirable. In terms of the input-output point of view, that method uses a $\{\Delta a\}$ that rotates the phase angle of the input toward the angle of the output. Therefore, it is desirable to choose a $\{\Delta a_d\}$ that tends to rotate the angle of the input toward the angle of the output.

A Good Choice of the Desired Change

Taking all the aforementioned considerations into account, the optimal choice of $\{\Delta a_d\}$ is not clear. A number of different types of change were made in computer trials. In addition, various modifications were tried, such as changing the input only at the few points which contained the most errors. Rather than discuss the many possible types of changes and the modifications of the basic method that might be used, we will present here only one of the best methods found. A way of changing the input that worked particularly well is as follows. If an input a_{mn} results in an output $\hat{a}_{mn}$, and the desired amplitude of the output is τ_{mn}, then an effective choice of Δa_{mn} is one that has two components:

$$\Delta a_{mn} = \begin{cases} \left(\frac{\hat{a}_{mn}}{|\hat{a}_{mn}|} - \frac{a_{mn}}{|a_{mn}|} \right) \tau_{mn} + \left(\frac{\hat{a}_{mn}}{|\hat{a}_{mn}|} \tau_{mn} - \hat{a}_{mn} \right) & , \ \tau_{mn} \neq 0 \\ 0 & , \ \tau_{mn} = 0 \end{cases} \tag{5-39}$$

in which the first component (of the upper line) rotates the angle of the input toward the angle of the output, and the second component boosts (or shrinks) the amplitude of the output to the desired level. (For the images we worked with, $\tau_{mn} = 1$ or 0 only.)

It should be noted that the change (5-39) used with method I is equivalent to using the change

$$\Delta a_{mn} = \begin{cases} \left(2 \frac{\hat{a}_{mn} \tau_{mn}}{|\hat{a}_{mn}|} - \hat{a}_{mn} \right) & , \quad \tau_{mn} \neq 0 \\ -\hat{a}_{mn} & , \quad \tau_{mn} = 0 \end{cases} \tag{5-40}$$

with method II. From the point of view of method II, for which $\alpha = \frac{1}{2}$, the change (5-40) is expected to produce a new output with the desired amplitude τ_{mn} ($\tau_{mn} \neq 0$). Furthermore, the previous iterative phase-coding technique can be interpreted as being method II with

$$\Delta a_{mn} = \begin{cases} \left(\frac{\hat{a}_{mn} \tau_{mn}}{|\hat{a}_{mn}|} - \hat{a}_{mn} \right) & , \quad \tau_{mn} \neq 0 \\ -\hat{a}_{mn} & , \quad \tau_{mn} = 0 \end{cases} \tag{5-41}$$

which is only half the change in the input ($\tau_{mn} \neq 0$) needed to achieve the desired output. It can also be seen why the previous iterative phase-coding technique works its way into a situation in which the errors stabilize. After a few iterations the quantization noise tends to line up its phase with the phase of the input. Then the change (5-41) is

just $-n_{mn}$, and as discussed earlier, the result is no change in the output. This behavior was observed in computer trials. However, the new method never gets into that situation, since, even if the noise were to line up in phase with the input, then (5-40) would become $\Delta a_{mn} = -2n_{mn}$, $\tau_{mn} \neq 0$ and $\Delta a_{mn} = -n_{mn}$, $\tau = 0$, which results in ΔA_{pq} that is not lined up with A_{pq} in the Fourier domain, and so the change continues to be effective.

The exact nature of the change in the output due to a change in the input is only partially understood. Particularly important is the fact that there generally is some correlation between $\{\Delta a\}$ and $\{a\}$, so we would expect there to be some correlation between $\{\Delta A\}$ and $\{A\}$ as well, which would change the results of Section 5.2.1. However, the relationships are understood well enough at this time to make practical use of them.

5.2.5 Procedure

Figure 5.7 shows a block diagram of the procedure used to reduce quantization noise in computer-generated holograms. The procedure is started by random phase coding the input, which at the start has amplitude equal to the desired amplitude $\{\tau\}$. Then the input is transformed, the transform is quantized (according to the type of computer-generated hologram to be synthesized), then inverse transformed, thereby computing the output that would result from the quantized hologram, $\{Q\}$. Then the change Δa_{mn} (into which the α^{-1} factor is incorporated) is computed and added either to the previous input (for method I) or to the last output (for method II), to form a new input. The procedure is then continued for a fixed number of iterations or until an error criterion is met.

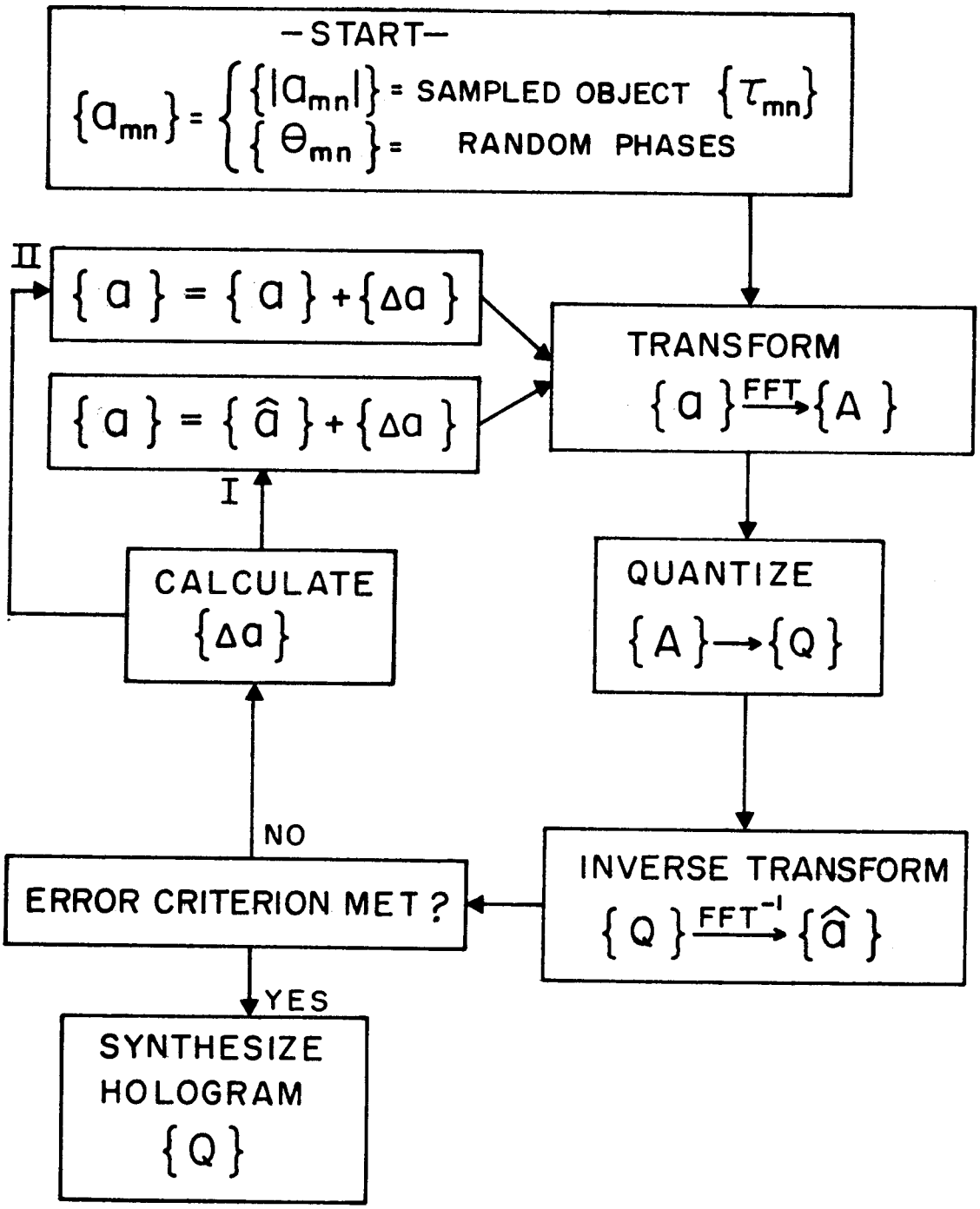


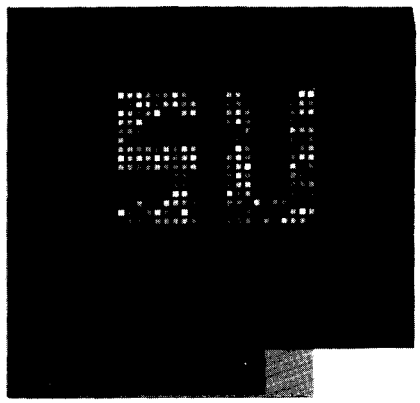
Fig.5.7 Iterative procedure based on the input-output approach.

5.3 Experimental Results

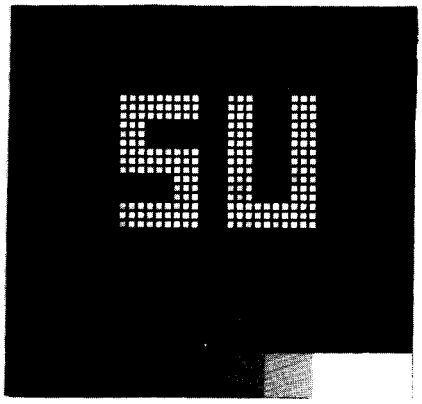
The results in this section are for binary objects ($\tau_{mn} = 0$ or 1), and the images shown are computer simulations of the reconstructions from quantized holograms (displayed on a CRT). Gray scales were included below each image in order to insure a valid comparison of different results.

The first example is for a hologram with four amplitude and four phase quantized levels, plus the zero level, as would be the case for a Lohmann binary detour-phase hologram using only 4×4 display resolution elements per Fourier coefficient. Figure 5.8a shows the intensity pattern that results when the object is random phase coded in the ordinary way. Figure 5.8b shows the result if the hologram is computed with 13 iterations of the procedure described in Section 5.2.5, using (5-39) with method I. A comparison of Figures 5.8a and 5.8b shows that the image is greatly improved by the iterative method.

When the application is to an optical memory, the quality criterion is the error rate. While the error rate has been computed for the random phase-coding case [Powers and Goodman, 1975], a similar study has not yet been performed that takes into account the iterative method. However, it is instructive to compare the intensities of the weakest one ($\tau_{mn} = 1$) bit and the strongest zero ($\tau_{mn} = 0$) bit in the image. When a one bit is less intense than a zero bit, then a wrong decision would be made in detecting the output. The graph of Figure 5.9 shows the weakest one bit and the strongest zero bit in the image, plotted against the number of iterations performed. Also plotted is the strongest one bit. Random phase coding (before any iterations) produced an image in which the weakest one bit is nearly equal to the strongest zero bit,



(a)



(b)

Fig. 5.8 Computer-simulated images from hologram with 4 amplitude and 4 phase quantized levels. (a) object random phase coded; (b) after 13 iterations of the present iterative method.

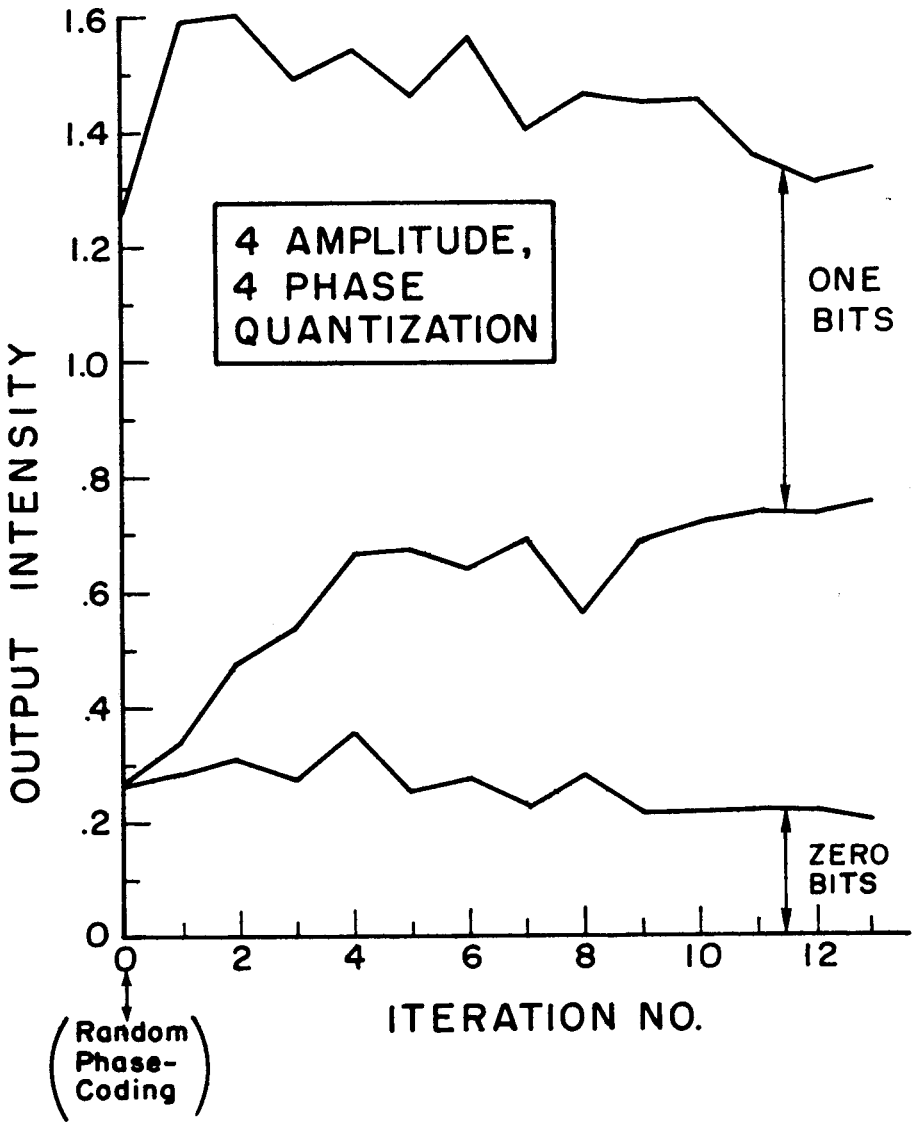
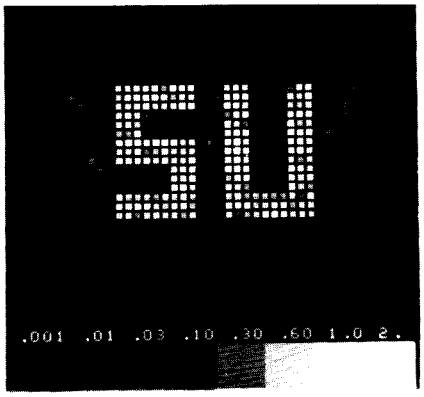


Fig.5.9 Range of output intensities vs. number of iterations, for 4 amplitude, 4 phase quantized hologram.

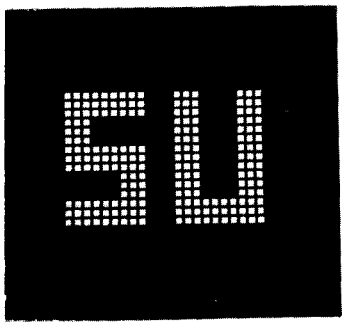
indicating a high error rate (perhaps on the order of 10^{-3}). After just a few iterations there is a comfortable gap between the weakest one bit and the strongest zero bit, indicating a very low error rate (a few orders of magnitude lower than 10^{-3}). Thus, using the iterative method, it is possible to achieve very low error rates, even with a severely quantized hologram.

The second example is for the kinoform, which has continuously controlled phase but only one amplitude level. Figure 5.10a shows the result of random phase coding in the ordinary way. Figure 5.10b shows the result after 8 iterations of method I using (5-39). For comparison, Figure 5.10c shows the result after 8 iterations of the previous iterative phase-coding method [Gallagher and Liu, 1973]. The image is greatly improved by both iterative methods. However, the variations in the intensity of the one bits is greater in the image shown in Figure 5.10c (previous iterative phase-coding method) than in the image shown in Figure 5.10b (new iterative method). For both iterative methods the weakest and strongest one bits and the strongest zero bit are plotted in Figure 5.11 for each iteration. As can be seen from Figure 5.11, the new iterative method of manipulating the amplitudes as well as the phases of the input allows better control over the errors in the image than the previous iterative phase coding method.

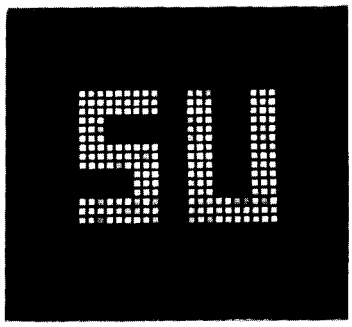
When comparing the images from different computer-generated holograms, the choice of the "best" image may depend on which quality criterion is used. For the memory application, the error rate is most appropriate, but is usually difficult to compute. Other common quality criteria are signal-to-noise ratio and mean squared error. Gallagher and Liu [1973]



(a)



(b)



(c)

Fig. 5.10 Computer-simulated images from kinoform. (a) object random phase coded; (b) after 8 iterations of the present iterative method; (c) after 8 iterations of the previous iterative phase coding method.

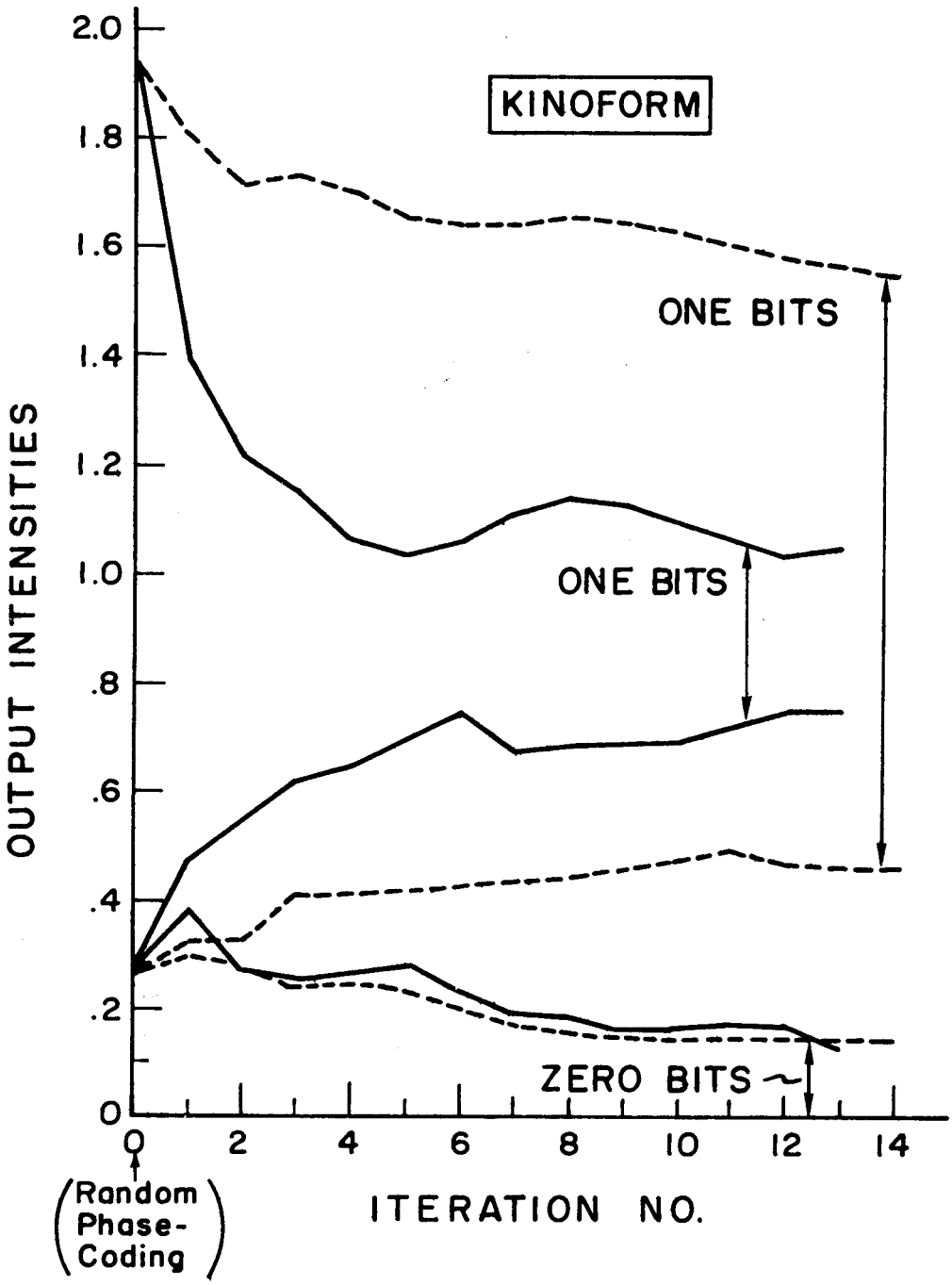


Fig.5.11 Range of output intensities vs. number of iterations for kinoform: solid lines for the present method, dashed lines for the previous iterative phase-coding method.

used the mean squared amplitude error,

$$\Phi_a = \frac{1}{N^2} \sum_{m,n=0}^{N-1} (|\hat{a}_{mn}| - \tau_{mn})^2 \qquad (5\text{-}42)$$

as a quality criterion. However, since the actual quantity measured is the intensity of the image, a more pertinent quality criterion would be the root-mean squared intensity error,

$$\Phi_i = \left[\frac{1}{N^2} \sum_{m,n=0}^{N-1} (|\hat{a}_{mn}|^2 - \tau_{mn}^2)^2 \right]^{\frac{1}{2}} . \qquad (5\text{-}43)$$

For the image for the random phase-coded case (Fig. 5.10a), $\Phi_a = .0381$ and $\Phi_i = .163$; for the image after 8 iterations of the previous iterative phase-coding method (Fig. 5.10c), $\Phi_a = .0195$ and $\Phi_i = .103$; for the image after 8 iterations of the present method (Fig. 5.10b), $\Phi_a = .0231$ and $\Phi_i = .0728$. Comparing the present method to the previous iterative phase-coding method, the previous method is better if judged by Φ_a , but the present method is better if judged by Φ_i . However, neither Φ_a nor Φ_i gives any direct indication of the error rate, since both Φ_a and Φ_i are the sum of the errors at all points in the image, whereas the error rate depends only on the error of the worst (one or few) points in the image. For example, if the error at nearly all the points in the image were very small, but the error at just one or a small number of points were large, then Φ_a and Φ_i would be small, but the error rate would be very high (an error rate of 10^{-3} is considered to be very high). On the other hand, if the error at all the points in the image were moderately large, but not enough to cause an incorrect decision, then Φ_a and Φ_i would be moderately large, but

the error rate would be very low. Therefore, a better indication of the error rate than either Φ_a or Φ_i would be the difference between (or possibly the ratio of) the weakest one bit and the strongest zero bit. Using this quality criterion, the present method would be judged to be better than the previous iterative method. Computer-generated holograms computed by the present method are expected to have error rates that are orders of magnitude better than holograms computed by the previous iterative method, for the following reason. The previous method does improve the mean squared error, but does nothing in particular to decrease the error at the individual points in the image where the error is greatest (and, as mentioned in Section 5.1, the worst errors tend to stabilize at high levels). However, the present method of manipulating the amplitude as well as the phase of the input works particularly well to reduce the error at the points with the worst error, as discussed in Section 5.2.1. Therefore, while the present method significantly reduces the mean squared error, it reduces the error rate to a far greater degree.

5.4 Further Applications

The iterative method outlined in Section 5.2.5 is extremely versatile and has applications far removed from computer holography. The iterative method (or a modified version of it) can be used to solve any problem of the following type: given a set of constraints placed on the object (or image) domain and another set of constraints placed on the Fourier domain, find a Fourier transform pair (i.e., an object or image and its Fourier transform) that satisfies both sets of constraints. The reduction of quantization noise is this type of problem: the image-domain constraint is that the amplitude pattern of the image must have a certain desired shape, and the Fourier-domain constraint is that the

Fourier transform is quantized according to the quantization levels of a particular type of computer-generated hologram.

Problems that can be solved by the iterative method fall into two categories. The first category includes problems in which the sets of constraints arise from physical measurements or conditions, and an exact Fourier transform pair exists between the physical quantities of concern. For the first category, the uniqueness of the Fourier transform pair is an important question. The second category includes problems, such as those in computer holography, in which the constraints are such that an exact Fourier transform pair may or may not exist. The problem is considered to be solved if a Fourier transform pair is found that closely approximates the constraints. The basic method is the same for both categories.

Both the iterative method based on the input-output approach and the previous iterative phase-coding method can be used to solve this type of problem. Which method works best depends upon the specific problem. For example, the iterative phase-coding method was unsuccessful for the 4-amplitude, 4-phase quantized hologram example of Section 5.3. In practice, a combination of methods (i.e., one method for a few iterations, then the other method for a few more iterations) can work better than either method alone.

Two problems were previously solved by methods similar to the iterative phase-coding method. The first problem is the following: given the intensity (or amplitude) of an object both in its image plane and its Fourier plane, determine the phase in either plane [Gerchberg and Saxton, 1972]. The second problem is the following: given the finite

width of an object and given the object's Fourier transform for low spatial freauencies, determine (analytically continue) the Fourier transform to higher spatial frequencies [Gerchberg, 1974]. These problems can also be solved by the present iterative method.

5.4.1 Spectrum Shaping

Analagous to the first problem solved by Gerchberg and Saxton [1972] is the following problem in computer holography: given a specified image intensity pattern, compute the hologram in such a way that the hologram itself looks like a second specified pattern. For example, as illustrated in Figure 5.12, we might desire the image of a fish to be reconstructed from a hologram that is itself a picture of a bird. Figure 5.13 shows the actual bird and fish patterns used for the computer trial. The fish object was random phase coded and Fourier transformed. The transform amplitude was set equal to the bird pattern, and then inversed transformed, thereby computing the image (shown in Fig. 14a) that would result from the bird hologram. The iterative procedure of Section 5.2.5 was then used for seven iterations, setting the Fourier amplitudes equal to the bird pattern at each iteration, resulting in the improved image shown in Figure 5.14. Increasing the number of iterations resulted in a further increase in the quality of the image.

5.4.2 Phase Retrieval Problem

The phase retrieval problem can be stated as follows: given the fact that an object function is real, nonnegative, and spatially bounded, and given the modulus (amplitude) of its Fourier transform, determine the phase of the Fourier transform (or, equivalently, determine the object function). This problem arises (for example) in speckle inter-

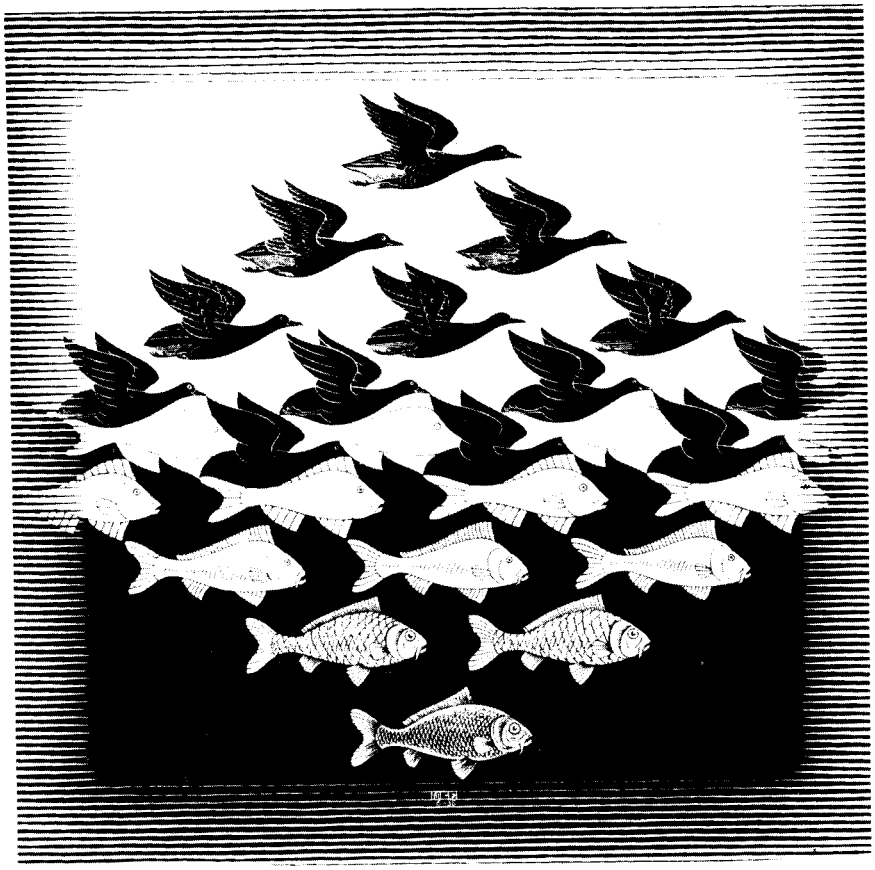


Fig. 5.12 Bird transforms into fish.

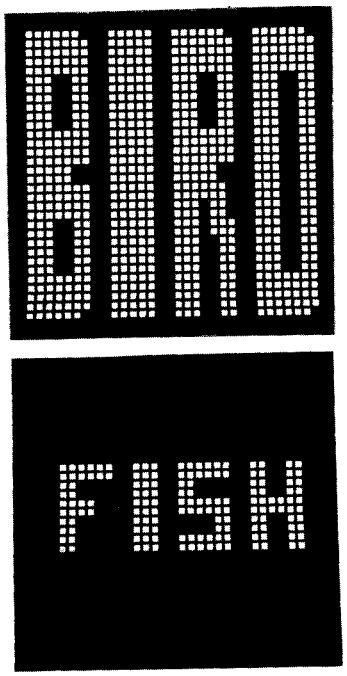
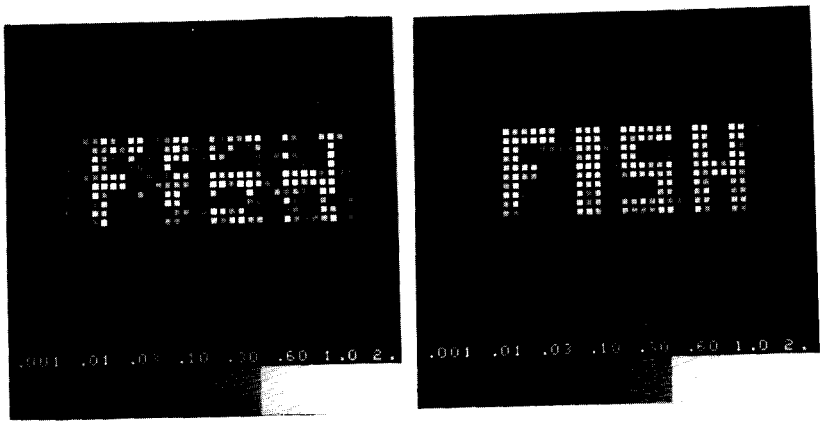


Fig.5.13 Bird hologram and desired fish image.



(a) (b)

Fig.5.14 Fish image: (a) random phase coding; (b) after 7 iterations.

ferometry [Labeyrie, 1970]. Ordinarily, if an astronomical object is viewed or photographed through a very large telescope, atmospheric turbulence limits the resolution of detail to an upper spatial frequency that is far below the diffraction limit of the telescope. However, by using the speckle interferometer of Labeyrie, the modulus of the Fourier transform of the intensity pattern of the object can be measured out to the diffraction limit of the telescope. The transform of the modulus squared of the object's Fourier transform is the autocorrelation of the object [Bracewell, 1965]. The autocorrelation is sufficient to determine only very limited information about the object, such as its width.

The phase problem was attacked with a modified version of method II of the iterative procedure of Section 5.2.5. The known modulus was the only Fourier-domain constraint used. The input-output domain was taken to be the object domain. The object-domain constraints were as follows: that the output be real and nonnegative and that it be restricted to the width of the actual object (this width can be determined from the autocorrelation of the object). Since the object is real rather than complex, the considerations in choosing the desired change are different than those discussed in Section 5.2.4. Rather than discuss the many different ways that the change can be chosen to achieve the desired output, we will present here only two of the best methods investigated. The first method consists of choosing the new input to be equal to zero if the output does not satisfy the constraints, and equal to the output if the output does satisfy the constraints. It can be shown that the mean squared error of the modulus of the Fourier transform of the input can only decrease when this first method is used. The second method

consists of choosing the new input to be equal to the difference between the previous input and the output if the output does not satisfy the constraints, and equal to the output if the output does satisfy the constraints. The second method does not necessarily decrease the mean squared error of the modulus of the Fourier transform of the input, but in practice it usually converges more quickly than the first method.

Figure 5.15 shows the nonzero portion of the one-dimensional object used (only ¼ of the object field is shown - the object is zero outside this interval). The Fourier transform of the object is computed, and only the modulus information is kept. The initial input is chosen to be a sequence of random numbers (uniformly distributed on $[0,1)$) over the known width of the object. The mean squared error of the modulus of the Fourier transform of the input is found to decrease very rapidly with each iteration and level off only after that error is less than 10^{-3} (after 20 or 30 iterations). At that point the solution is considered to be found, since it would agree with the sets of constraints in both domains to well within the expected accuracy of the measurement of the modulus. Different solutions are found, depending upon the initial value of the input, because the solution to the phase problem is not unique (see, for example, [Roman and Marathay, 1963]). However, despite the non-uniqueness of the solution, some useful additional information about the object can be extracted. Figure 5.16 shows, superimposed on the object, the output solutions resulting from 15 different random initial inputs. As can be seen from Figure 5.16, there is a fair degree of correlation between each of the solutions and the object. Therefore, while it is not possible to absolutely determine the object, the iterative

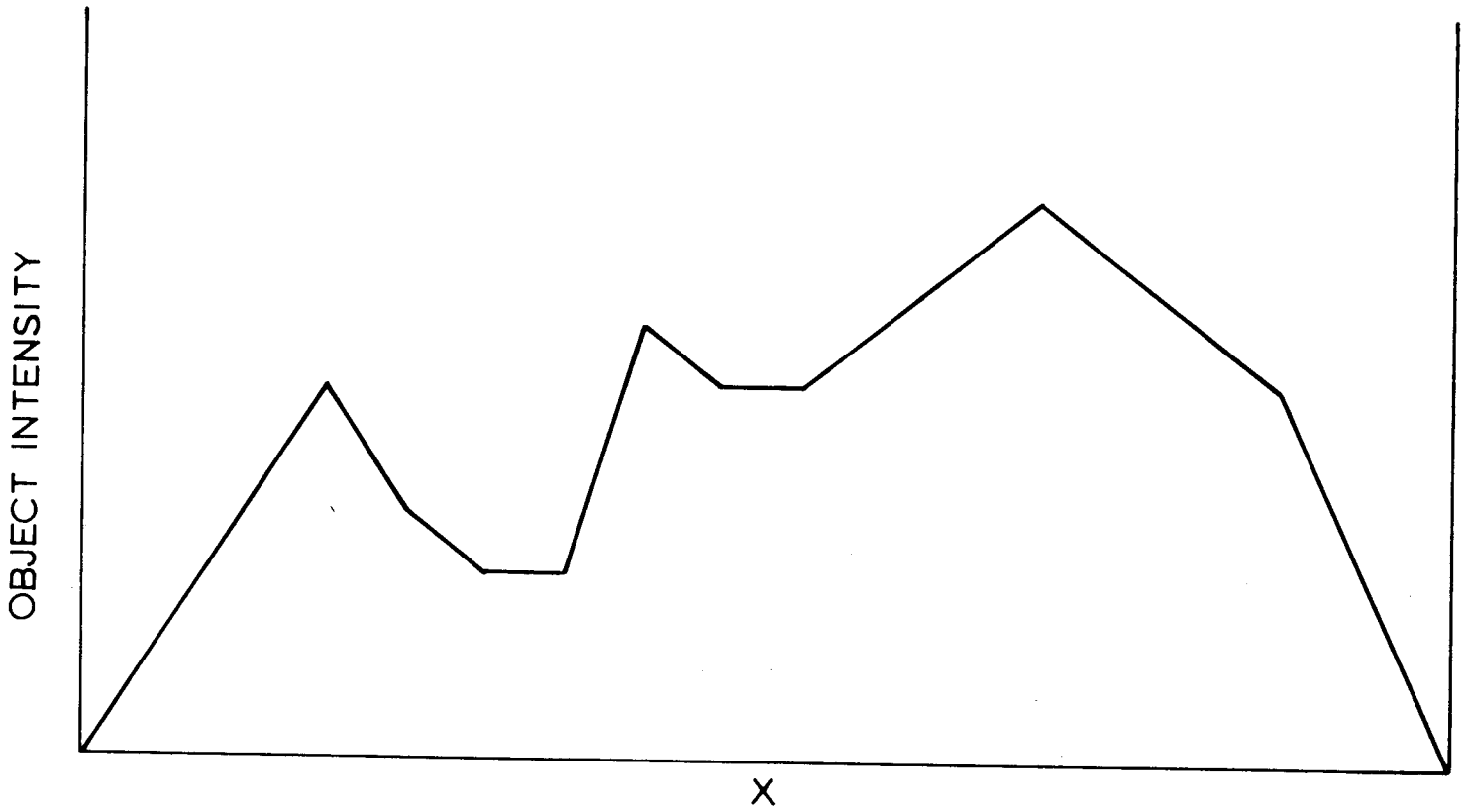


Fig.5.15 Test object for phase problem.

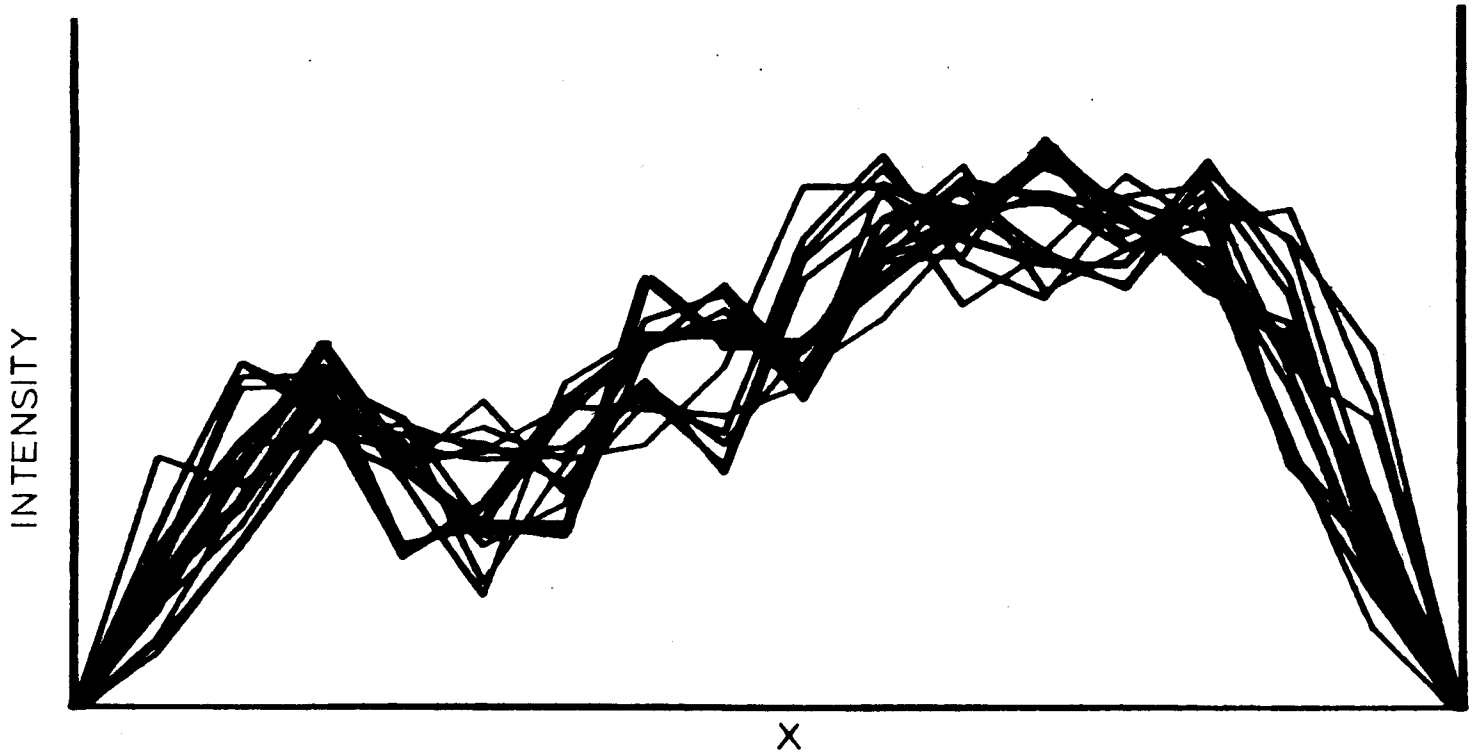


Fig.5.16 Test object and 15 solutions to the phase problem.

procedure can be used as an easy method of generating possible solutions, from which an idea of the general appearance of the object can be inferred.

5.4.3 Further Proposed Applications

Due to the general nature of the iterative method, it can probably be used to solve a number of other problems. Two problems that might be solved by the iterative method are discussed in this section.

One problem in computer holography to which it might be applied is as follows. As discussed in Section 3.6, the diffraction efficiency of an ideal ROACH, which is given by (3-32), is less than 100% because it is necessary for the hologram to absorb light in order to control the amplitude of the wavefront. Neither do kinoforms have 100% diffraction efficiency in the general case, since a substantial fraction of the light intensity goes into noise rather than into the desired image. A method of making computer holograms having a diffraction efficiency of 100% (irrespective of the dynamic range of the Fourier amplitude) is to modulate the wavefront by two different kinoforms. The first kinoform would be placed in the object plane of Figure A.1 and would be illuminated by a plane wave. The function of the first kinoform would be to produce the desired Fourier amplitude pattern in the hologram plane (Figure A.1). The second kinoform would be placed in the hologram plane and would modulate only the phase of the wavefront (which already has the desired amplitude pattern). The second kinoform would have a phase transmittance equal to the difference between the phase of the desired wavefront and the phase of the incident wavefront. By the combined effects of the two kinoforms, a desired wavefront $F(u,v)$ could be produced with no loss in intensity. The problem is to find a phase-only function (the first

kinoform) that has a Fourier transform of amplitude $|F(u,v)|$. The iterative procedure of Section 5.2.5 was designed for that purpose. Unfortunately, it is not generally possible to find a phase-only function that has a Fourier transform of amplitude exactly equal to $|F(u,v)|$. However, $|F(u,v)|$ can be manipulated by assigning a phase code to the desired image $f(x,y)$. Therefore, we can instead try to solve an alternative problem: find a phase-only function (the first kinoform) and a phase code for $f(x,y)$ so that the phase-only function and $f(x,y)$ have Fourier transforms with the same modulus. This problem probably has an exact solution (or many exact solutions) whereas the previously mentioned one probably does not. The second approach would require a more complicated procedure, requiring four Fourier transforms per iteration (to go from the image plane to the plane of the first kinoform and back again to the image plane).

A second problem for which the iterative method might be used is the "hidden line" problem. Suppose a computer-generated hologram is made in the ordinary way for a three-dimensional object that consists of two planar objects separated by a certain distance. Upon reconstruction of the hologram it will be found that the image in the first plane is degraded by the out-of-focus image from the second plane. This degradation could be reduced by phase-coding the object in the second plane so that the amplitude of its near-field diffraction pattern in the first plane is approximately the same as the amplitude pattern of the object in the first plane. If a sampled hologram is used to produce the image, then the solution cannot be obtained simply by substituting a Fresnel transform for the Fourier transform in the iterative method of Section 5.2.5 and working between the two image planes. Instead, four Fourier

transforms are required per iteration, to go from the second image plane to the hologram plane, then to the first image plane, then back again.

CHAPTER 6

CONCLUDING REMARKS

The most important contribution of this thesis is the ROACH, a new computer-generated hologram based on the simultaneous and independent control of the amplitude transmittance and the phase transmittance of a material. ROACHs have been successfully synthesized on Kodachrome II and on Kodachrome 25, using ordinary commercial processing. The properties of Kodachrome film as a holographic material were studied with sufficient depth to enable us to exert a considerable degree of control over the material and to understand its limitations.

The materials presently available for the ROACH were designed for other purposes. It is safe to say that the present technology of photographic emulsions would allow the production of a ROACH material that is far superior to Kodachrome II. Only two layers would be necessary, except when the film itself is used to reject false color images. The development of superior ROACH materials would be most welcome.

Practical methods of producing full-color images from computer-generated holograms were formulated and successfully employed. Methods for eliminating false color images were demonstrated.

A new iterative procedure for reducing quantization noise in computer-generated holograms was developed. The new method, based on an

input-output point of view, was found to be far superior to previous methods in reducing the error rate (due to quantization noise) in computer-generated holographic memories. Other applications of the iterative method were found. Particularly exciting is its use to find solutions to the phase retrieval problem. Since the phase of the Fourier transform can often be measured at low spatial frequencies, the low-frequency phase is one more constraint that can be imposed on the transform domain and might possibly enable us to extract still more information about the object by narrowing the set of possible solutions. It would also be interesting to apply the iterative method to other problems, such as those suggested in Section 5.4.3. How the iterative method works is only partially understood. A better understanding of the relationship between the change in the input and the change in the output might lead us to an improved iterative method.

A problem briefly mentioned in Appendix B that deserves attention is the problem of speckle in the images from computer-generated holograms. We eliminated the speckle by repeating the hologram to give the image a dot structure. However, since the speckle is an effect of phase coding, it should be possible to eliminate it by an appropriate phase-code, without repeating the hologram. One such phase code might be to assign every other point to be real (± 1) and interlacing with them every other point to be pure imaginary $(\pm i)$, in a checkerboard fashion, where the sign (+ or -) is chosen quasi-randomly. Other ways of creatively manipulating the phase code to produce various effects on the microstructure of the image should be possible.

The contributions of this thesis can be briefly summarized as follows:

1. The development of a method to control the complex transmittance

of a material.

2. The measurement of pertinent characteristics of Kodachrome film.

3. The development of a versatile iterative procedure and applying it to reducing quantization noise in computer-generated holograms, spectrum shaping, and the phase retrieval problem.

4. The experimental realization of full-color images from computer-generated holograms and verification of the phase-null effect.

APPENDIX A

FOURIER TRANSFORM HOLOGRAMS

In this appendix we will discuss the nature of the Fourier transform geometry and the properties that make it ideal for the holographic memory application. This discussion is included for the sake of completeness and is not an original contribution (see, for example, [Chu, 1974]).

Referring to Figure A.1, if an object transparency with complex transmittance $f_0(x_0,y_0)$ is placed in the object plane and is illuminated by a plane wave, then by (1-1), the wavefront produced in the hologram plane is

$$F(u,v) = \frac{1}{\lambda f}\int\int_{-\infty}^{\infty} f_0(x_0,y_0)e^{-i(2\pi/\lambda f)(ux_0+vy_0)}dx_0dy_0 \quad . \qquad (A\text{-}1)$$

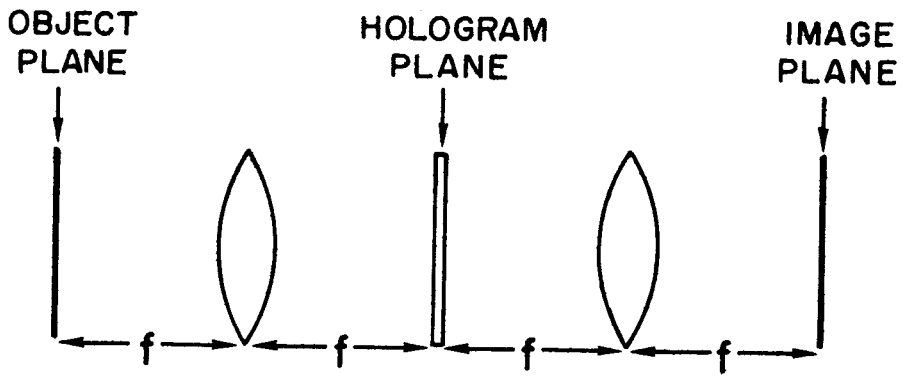


Fig.A.1. Fourier transform geometry. Reconstruction of the hologram is accomplished by placing a monochromatic point source on axis in the object plane.

If no hologram were present, allowing the undisturbed wavefront to propagate to the image plane, the wavefront in the image plane would be (using (1-1) again)

$$f_i(x_i,y_i) = \frac{1}{\lambda f}\iint_{-\infty}^{\infty} F(u,v)e^{-i(2\pi/\lambda f)(ux_i+vy_i)}\,dudv \quad . \qquad \text{(A-2)}$$

Inserting (A-1) into (A-2), we find that

$$f_i(x_i,y_i) = \frac{1}{(\lambda f)^2}\iint_{-\infty}^{\infty} f_0(x_0,y_0) \cdot \underbrace{\left[\iint_{-\infty}^{\infty} \exp\left\{\frac{-i2\pi}{\lambda f}\left[u(x_0+x_i) + v(y_0+y_i)\right]\right\} dudv\right]}_{(\lambda f)^2\delta(x_0+x_i)\delta(y_0+y_i)} dx_0dy_0$$

$$= f_0(-x_i,-y_i) \quad . \qquad \text{(A-3)}$$

That is, if no hologram is present, the wavefront in the image plane is equal to the wavefront in the object plane reflected through the origin (optical axis). Similarly, if (A-1) is used to compute the hologram, and if the hologram (which is assumed to be a ROACH) is placed in the hologram plane and is illuminated with a plane wave (which could be obtained by having a point source in place of the object), then the reconstructed image is the same as the object reflected through the origin. In order to obtain the correct orientatation of the image, one of three things can be done: either the holograms can be rotated 180° about the optical axis, or the hologram can be computed by the inverse Fourier transform, or the coordinate system of the image plane can be redefined. This question of the orientation of the image is sometimes a cause for confusion. In particular, for the analysis of Chapter 5, the output from the

hologram is computed by the inverse Fourier transform rather than the forward Fourier transform; this is done so that the input and output have the same orientation. An actual image produced by a hologram computed by the iterative procedure of Chapter 5 is reflected through the origin compared to the computed output. As a matter of convenience, we occasionally use the terms "object" and "image" interchangeably, referring to both as $f(x,y)$, even though they may have different orientations.

An extremely useful property of the Fourier transform geometry is the shift-invariance (translation-invariance) property. That is, the hologram can be translated without affecting the intensity pattern of the image. This property can be shown by replacing $F(u,v)$ by $F(u+u_0,v+v_0)$ in (A-2). Then the new image is

$$\begin{aligned} f_i'(x_i,y_i) &= \frac{1}{\lambda f}\int_{-\infty}^{\infty}\int F(u+u_0,v+v_0)e^{-i(2\pi/\lambda f)(ux_i+vy_i)}dudv \\ &= \frac{1}{\lambda f}\int_{-\infty}^{\infty}\int F(u,v)\exp\left\{-i\left(\frac{2\pi}{\lambda f}\right)\left[(u-u_0)x_i + (v-v_0)y_i\right]\right\}dudv \qquad \text{(A-4)} \\ &= \exp\left\{i\left(\frac{2\pi}{\lambda f}\right)(u_0x_i+v_0y_i)\right\}f_i(x_i,y_i) \quad . \end{aligned}$$

Since the image from the translated hologram differs from the original image only by a phase factor, the intensity patterns (which are actually detected) of the two images are the same. Therefore, holograms in the Fourier transform geometry do not require accurate x-y positioning in order for the image to appear in the correct location.

A second extremely useful property of the Fourier transform geometry is that the effect of small localized defects, dust particles, or scratches on the hologram result not in localized defects in the image, but in a slight general loss of fidelity of the entire image. This property can be shown by adding to $F(u,v)$ a term representing a localized defect at

(u_0,v_0) , $a\cdot\delta(u-u_0,v-v_0)$. Inserting this term into (A-2), the result is $f_i(x_i,y_i)+(a/\lambda f)\exp\{-(i2\pi/\lambda f)\cdot(u_0x_i+v_0y_i)\}$. The second term, which is due to the defect in the hologram, is spread uniformly over the image plane.

A third useful property of the Fourier transform geometry is that the hologram can be repeated a number of times, side by side, thereby increasing the hologram area and redundancy, without requiring additional computation. The only effect of the repetition is to give the image a dot structure, which is desirable for the binary data of a holographic memory. This property is discussed further in Appendix B and is demonstrated by the image shown in Figure 2.13, which resulted from a ROACH repeated in a 4×4 array.

APPENDIX B

THE DISCRETE FOURIER TRANSFORM AND SAMPLED HOLOGRAMS

In this appendix, we will discuss the effect on the image of using the discrete Fourier transform to compute a sampled hologram.

We use the notation

$$\text{comb}(x/d) = \sum_{n=-\infty}^{\infty} \delta(x/d-n) \tag{B-1}$$

$$\text{rect}(x,d) = \begin{cases} 0 , & |x/d| > \frac{1}{2} \\ \frac{1}{2} , & |x/d| = \frac{1}{2} \\ 1 , & |x/d| < \frac{1}{2} \end{cases} \tag{B-2}$$

$$\text{sinc}(x/d) = \frac{\sin(\pi x/d)}{\pi x/d} \quad . \tag{B-3}$$

The symbol $\supset$ denotes the Fourier transform relationship and $*$ denotes convolution [Bracewell, 1965].

If the object is considered to be an array of point sources on a square lattice (which may be used to approximate a sampled object), then it can be represented by the summation (N×N sampled points)

$$f(x,y) = \sum_{m,n=0}^{N-1} f_{mn}\delta(x-ms,y-ns) \tag{B-4}$$

where s is the spacing of the object points. The case of a rectangular lattice with different spacings in the x and y directions is a trivial extension of this analysis. Inverse transforming (B-4), we find

that

$$F(u,v) = \frac{1}{\lambda f} \sum_{m,n=0}^{N-1} f_{mn} \exp\{i(2\pi/\lambda f)\cdot(ums+vns)\} \quad . \qquad (B\text{-}5)$$

Since $f(x,y)$ is sampled, $F(u,v)$ is periodic, with period $\lambda f/s$, which can be seen from (B-5). If a hologram were made of the form (B-5), then the reconstruction would consist of a single image of the form (B-4). However, when calculating (B-5) on a digital computer and displaying that information on a CRT or plotter, it is necessary to use only samples of $F(u,v)$ rather than the continuous complex function. We use only the sampled values (Fourier coefficients)

$$F_{pq} \equiv F(pd,qd) = \frac{1}{\lambda f} \sum_{m,n=0}^{N-1} f_{mn} \exp\{(i2\pi/\lambda f)(mpsd+nqsd)\} \qquad (B\text{-}6)$$

where d is the spacing of the samples in the hologram. If d is chosen so that there are also $N\times N$ samples of F_{pq} in one period, that is, $d = \lambda f/Ns$, then (B-6) reduces to

$$F_{pq} = \sum_{m,n=0}^{N-1} f_{mn} \exp\{(i2\pi/N)(mp+nq)\} \qquad (B\text{-}7)$$

where the constant factor $(\lambda f)^{-1}$ is dropped. The inverse of (B-7) is

$$f_{mn} = \frac{1}{N^2} \sum_{p,q=0}^{N-1} F_{pq} \exp\{(-i2\pi/N)(mp+nq)\} \quad . \qquad (B\text{-}8)$$

The relations of (B-7) and (B-8) are the discrete Fourier transform relations ordinarily used in computer holography. The (inverse) discrete Fourier transform (B-7) yields sampled values of the (inverse) Fourier transform of a sampled object. A particular advantage of the discrete Fourier transform is the existence of a fast algorithm (fast Fourier

transform) for its computation [Cooley and Tukey, 1965].

The ROACH consists of an array of square cells of width d and spacing d, the $(p,q)^{th}$ cell having a complex transmittance F_{pq} uniform over the area of the cell. That is, the complex transmittance of the ROACH of $f(x,y)$ is (omitting a scaling factor)

$$H(u,v) = \text{rect}\left(\frac{u}{d},\frac{v}{d}\right) * \left[F(u,v)\cdot\text{comb}\left(\frac{u}{d},\frac{v}{d}\right)\right] \qquad \text{(B-9)}$$

where

$$F(u,v)\cdot\text{comb}\left(\frac{u}{d},\frac{v}{d}\right) = \sum_{p,q=-\infty}^{\infty} F_{pq}\,\delta(u/d-p,v/d-q) \quad . \qquad \text{(B-10)}$$

The image produced by a ROACH is given by the Fourier transform (continuous, not discrete) of (B-9):

$$h(x,y) = \mathcal{F}\{H(u,v)\}$$

$$= \text{sinc}\left(\frac{dx}{\lambda f}\right)\text{sinc}\left(\frac{dy}{\lambda f}\right)\cdot\left[f(x,y)*\text{comb}\left(\frac{dx}{\lambda f},\frac{dy}{\lambda f}\right)\right]. \qquad \text{(B-11)}$$

Using the substitution $d/\lambda f = 1/Ns$, (B-11) becomes

$$h(x,y) = \text{sinc}\left(\frac{x}{Ns}\right)\text{sinc}\left(\frac{y}{Ns}\right)\cdot\left[f(x,y)*\text{comb}\left(\frac{x}{Ns},\frac{y}{Ns}\right)\right] \qquad \text{(B-12)}$$

where Ns is the width of the object (B-4). The convolution of $f(x,y)$ with the comb function causes $f(x,y)$ to be repeated in an array of spacing Ns, due to the sampling of the hologram. The sinc factors, due to the finite cell size, cause the repeated off-axis images to be suppressed. Except for the partial attenuation of the edges of the on-axis image due to the sinc factors, the on-axis image term in (B-12) is identical to the object (B-4). The attenuation at the edges can be compensated for by multiplying (B-4) by $\text{sinc}^{-1}(x/Ns)\text{sinc}^{-1}(y/Ns)$ before computing the hologram.

If only M periods of the periodic discrete Fourier transform are used, then (B-9) becomes

$$H(u,v) = \left\{ \text{rect}\left(\frac{u}{d},\frac{v}{d}\right) * \left[F(u,v)\cdot \text{comb}\left(\frac{u}{d},\frac{v}{d}\right)\right]\right\}\cdot \text{rect}\left(\frac{u}{MNd},\frac{v}{MNd}\right) \qquad \text{(B-13)}$$

which results in the image

$$h(x,y) = \left\{\text{sinc}\left(\frac{x}{Ns}\right)\text{sinc}\left(\frac{y}{Ns}\right)\left[f(x,y)*\text{comb}\left(\frac{x}{Ns},\frac{y}{Ns}\right)\right]\right\}*\left[\text{sinc}\left(\frac{Mx}{s}\right)\text{sinc}\left(\frac{My}{s}\right)\right]. \qquad \text{(B-14)}$$

That is, each point in $f(x,y)$ is convolved (in both dimensions) with a sinc function of width equal to s/M, $1/M$ times the spacing of the sampling points. If M is an integer, then the function $\text{sinc}(Mx/s)$ is zero for $x = ns$, $n = \pm 1, \pm 2, \ldots$. Therefore, if we look only at the locations of the sampling points in the image, we find that

$$h(ms,ns) = \text{sinc}\left(\frac{n}{N}\right)\text{sinc}\left(\frac{m}{N}\right)\cdot f_{mn} \quad . \qquad \text{(B-15)}$$

And if, as previously mentioned, the sinc factors are compensated for, then the image at the sampling points is given by

$$h(ms,ns) = f_{mn} \quad . \qquad \text{(B-16)}$$

That is, exactly the desired image is reconstructed <u>at the sampling points</u> [Chu, 1974].

If only one period of the discrete Fourier transform is used ($M=1$), then the sinc functions with which $f(x,y)$ is convolved in (B-14) have width s, the spacing of the sampling points. In between the sampling points, the value of the image is influenced by all nearby sampling points, due to the convolution with the sinc functions. Adjoining sampling points may have complex values opposite in phase, and $h(x,y)$ must vary continuously in the complex plane between the sampled values. Therefore,

even if all the amplitudes of $h(ms,ns)$ in a region of the image are the same, $|h(x,y)|$ may vary considerably in that region, and in some locations may even be zero. The result is a "speckle" pattern in the image from a computer-generated hologram of a random phase-coded object. Phase codes other than random produce other characteristic patterns, depending on the particular form of the phase code. By an appropriate choice of phase code it is possible to obtain various kinds of patterns or even to reduce the variations in $|h(x,y)|$ between points. For example, if the phase difference between two adjoining sampling points of equal amplitude is 78°, and if those two points are isolated, then the value of $|h(x,y)|$, for (x,y) between those two sampling points, varies by less than 2%. Alternatively, by using $M > 1$, the width of the $\text{sinc}(Mx/s)\text{sinc}(My/s)$ term in (B-14) is decreased, and the result in the reconstructed image is a dot-structure that can be seen in the image shown in Figure 2.13, which resulted from a ROACH with $M = 4$. For $M = 2$ or greater, the dot-structure dominates, and the particular form of the phase code no longer significantly affects the appearance of the image.

APPENDIX C

PRODUCTION OF SINUSOIDAL INTENSITY PATTERNS OF ARBITRARY MODULATION AND SPATIAL FREQUENCY

In order to obtain the intensity pattern of (3-21), it is necessary to produce two coherent plane waves of intensities A^2 and B^2, respectively, and cause them to intersect at an angle ϕ given by

$$\sin\phi = 2\pi\lambda/k = \nu\lambda \qquad (C\text{-}1)$$

where ν is the spatial frequency. If two plane waves (expanded and collimated laser beams) were made to intersect at a beamsplitter, then the angle between one transmitted beam and one reflected beam could be easily varied by changing the tilt of the beamsplitter. Unfortunately, this method would require a beamsplitter flat to within a fraction of a wavelength over the area of each beam; and such a beamsplitter was not available to us. Therefore, it was necessary to vary the angle in a less direct way, and a method suggested by Tichenor [1974] was employed.

A holographic diffraction grating is made by recording the interference of two plane waves A and B at an angle ϕ_0 (Fig. C.1a). After processing, the grating is replaced in its original position and is illuminated by both A and B. A number of plane waves are produced corresponding to the various orders of diffraction from each of the two plane waves A and B. Referring to Figure C.1a, we will concentrate

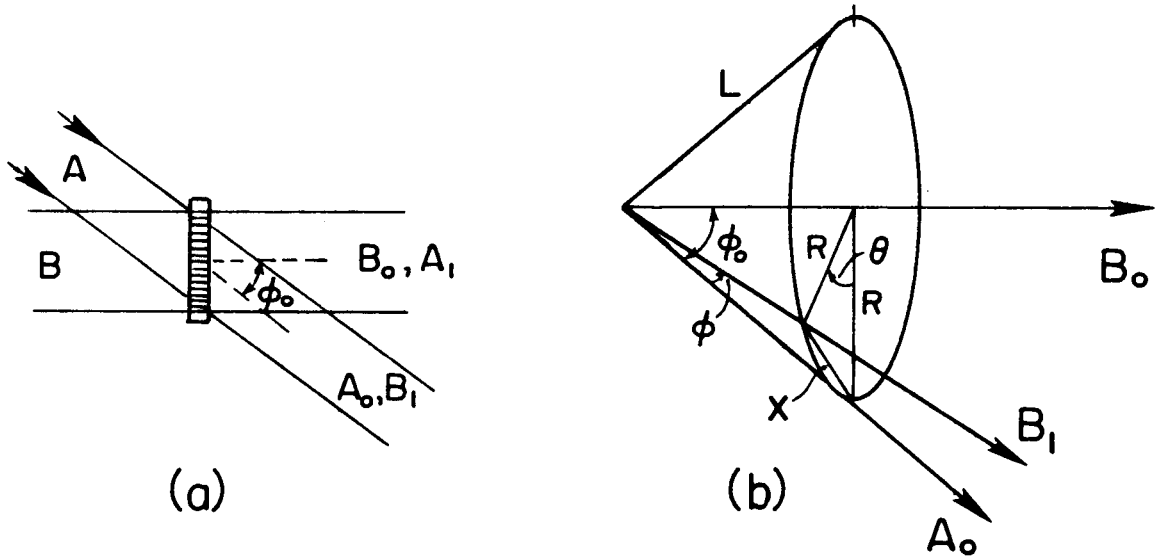


Fig.C.1 (a) Beams A and B that form the grating are also used to illuminate the grating to produce beams A_0 and B_1. (b) Geometry for determining ϕ , the angle between B_1 and A_0, as a function of Θ , the angle of rotation of the grating.

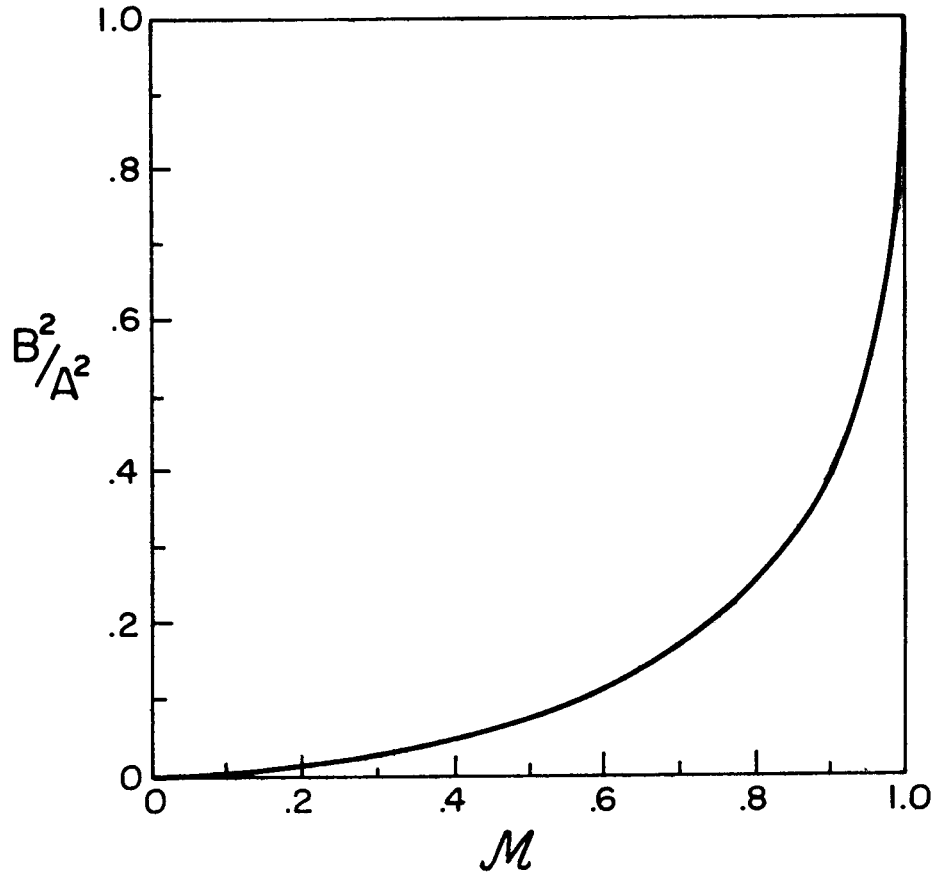


Fig.C.2 Beam ratio vs. modulation

only on two waves: A_0 , the undiffracted component of A , and B_1 , the first order diffracted wave arising from B . B_1 can be thought of as being the reconstruction of the wavefront A from a hologram that had B as the reference beam and A as the object beam. Therefore B_1 and A_0 coincide exactly (as do the similar pair A_1 and B_0). If the grating is rotated about its axis (about $B\text{-}B_0$) then A_0 is unaffected, but B_1 is rotated (out of the page in Fig. C.1). The rotation of B_1 introduces an angle ϕ , between A_0 and B_1 (which overlap) that can be easily varied by the amount of rotation of the grating. In this manner the spatial frequency of the sinusoidal intensity pattern caused by the interference of A_0 with B_1 can be easily varied without the need for a high quality beamsplitter.

Referring to Figure C.1b, the angle between A_0 and B_0 is ϕ_0 , and the rotation of the grating through an angle of 360° causes B_1 to sweep out a cone about B_0 . By the geometry of Figure C.1b, the angle ϕ , between A_0 and B_1 is given by $(X/2)/L = \sin(\phi/2)$. Similarly, $R/L = \sin\phi_0$ and $(X/2)/R = \sin(\Theta/2)$, where Θ is the angle of rotation of the grating. Combining these expressions, we obtain

$$\sin(\phi/2) = \sin(\Theta/2)\cdot\sin\phi_0 \quad . \qquad \text{(C-2)}$$

(which reduces to $\phi \simeq \Theta\sin\phi_0$ for $\Theta << 1$). Since $\sin 2x = 2\sin x\cdot\cos x$,, we have $\sin\phi = 2\sin(\phi/2)\cos(\phi/2) = 2\sin(\phi/2)\sqrt{1-\sin^2(\phi/2)}$. Combining this expression with (C-1) and (C-2), we find that

$$\nu = \frac{\sin\phi}{\lambda} = 2\sin(\Theta/2)\cdot\frac{\sin\phi_0}{\lambda}\sqrt{1-\sin^2(\Theta/2)\sin^2\phi_0} \quad . \qquad \text{(C-3)}$$

A grating was produced by exposing 649-F plate to the interference of two plane waves of wavelength 501.7 nm from an argon-ion laser, with a beam ratio of .38, and a mean exposure of 25 $\mu J/cm^2$. The plate was prehardened, developed in D-19 for five minutes, and bleached [Lehmann, Lauer, and Goodman, 1970]. The angle, ϕ_0, at which the two beams intersected was 8.66°, or $\sin\phi_0 = .15$, making the spatial frequency of the grating $(\sin\phi_0)/\lambda = 300$ l/mm. Since only spatial frequencies up to 170 l/mm were desired, the maximum angle of rotation of the grating needed was 33°; and for those small angles, (C-3) can be approximated by

$$\nu = \frac{\sin\phi}{\lambda} = 2\sin(\Theta/2)\cdot\frac{\sin\phi_0}{\lambda} \qquad \text{(C-4)}$$

which was used to determine the angle of rotation of the grating, Θ, required to produce the desired spatial frequency, .

Let A^2 be the intensity of beam A_0 and B^2 be the intensity of beam B_1. A^2 (or B^2) is measured by blocking off beam B (or beam A). The modulation, $\mathfrak{m}$, in (3-21) is given by

$$\mathfrak{m} = 2AB/(A^2+B^2) = \frac{2B/A}{1+(B/A)^2} \quad . \qquad \text{(C-5)}$$

Solving (C-5) for (B/A), assuming $B^2 < A^2$, yields

$$B/A = (1 - \sqrt{1-\mathfrak{m}^2})/\mathfrak{m} \quad . \qquad \text{(C-6)}$$

Squaring, we find that the ratio of the intensities of the two beams needed to produce a given $\mathfrak{m}$ is

$$\begin{aligned} B^2/A^2 &= (1-\sqrt{1-\mathfrak{m}^2})^2/\mathfrak{m}^2 = (2-\mathfrak{m}^2-2\sqrt{1-\mathfrak{m}^2})/\mathfrak{m}^2 \\ &\simeq (\mathfrak{m}/2)^2 \quad \text{for } \mathfrak{m} \lesssim .3 \quad . \end{aligned} \qquad \text{(C-7)}$$

Figure C.2 shows a plot of B^2/A^2 vs. $\mathfrak{m}$. In order to produce a given

modulation, m, neutral density filters are placed in the paths of beams A and B (Fig. C.1) until the beam ratio B^2/A^2 determined from (C-7) is produced.

APPENDIX D

PHASE-NULL MATHEMATICS

In this appendix we derive the actual image produced by the phase-null method. The same notation is used as in Appendix B, in which it was shown that a hologram of the form

$$H(u,v) = \text{rect}\left(\frac{u}{d},\frac{v}{d}\right) * \left[F(u,v)\cdot\text{comb}\left(\frac{u}{d},\frac{v}{d}\right)\right] \qquad \text{(B-9)}$$

produces an image

$$\begin{aligned} h(x,y) &= \mathfrak{F}\{H(u,v)\} \\ &= \text{sinc}\left(\frac{x}{Ns}\right)\text{sinc}\left(\frac{y}{Ns}\right)\cdot\left[f(x,y)*\text{comb}\left(\frac{x}{Ns},\frac{y}{Ns}\right)\right] \qquad \text{(B-12)} \end{aligned}$$

where $Ns = \lambda f/d$ is the width of the sampled object $f(x,y)$. Whether $H(u,v)$ contains many or only one period of the discrete Fourier transform does not matter for the purpose of this analysis, since that affects only the microstructure of the image.

(B-9) is the form of the ordinary ROACH, which consists of an array of square cells of side d. Each cell has uniform transmittance over its area.

Now suppose that each hologram cell is divided into two halves along the u-direction. The right half has the same complex transmittance as before, and the left half has that complex transmittance multiplied by A, a complex constant (Fig.D.1b). Then we have

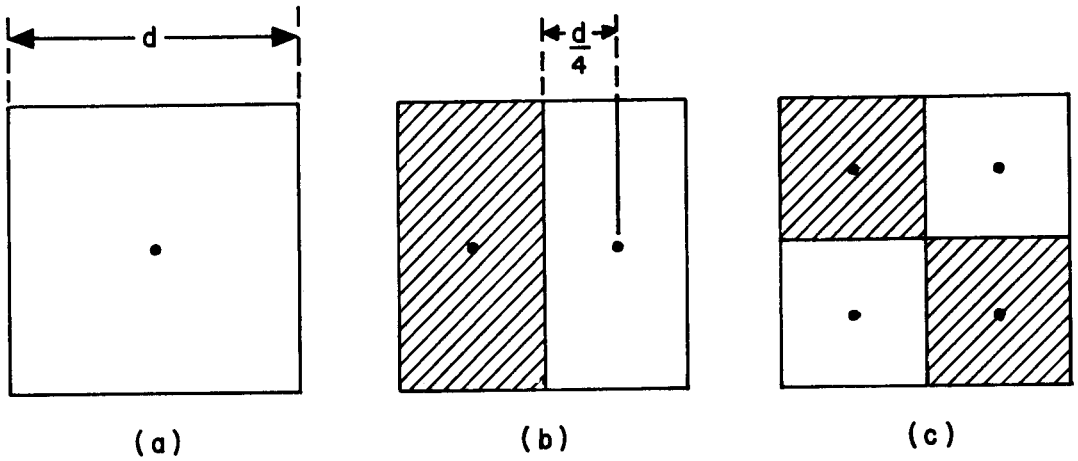


Fig.D.1 Dividing the Fourier coefficient cell for the phase-null method. Shaded areas have complex transmittance equal to A times the transmittance of the unshaded areas. (a) regular cell, (b) phase-null method in one dimension, (c) phase-null method in both dimensions.

$$H(u,v) = \left[\text{rect}\left(\frac{2u}{d} - \tfrac{1}{2},\frac{v}{d}\right) + A\cdot\text{rect}\left(\frac{2u}{d} + \tfrac{1}{2},\frac{v}{d}\right)\right] * \left[F(u,v)\cdot\text{comb}\left(\frac{u}{d},\frac{v}{d}\right)\right] \supset$$

$$\supset h(x,y) = \tfrac{1}{2}\left[\exp\left(-i2\pi\,\frac{x}{4Ns}\right) + A\cdot\exp\left(+i2\pi\,\frac{x}{4Ns}\right)\right]\text{sinc}\left(\frac{x}{2Ns}\right)\cdot$$

$$\cdot\,\text{sinc}\left(\frac{y}{Ns}\right)\cdot\left[f(x,y) * \text{comb}\left(\frac{x}{Ns},\frac{y}{Ns}\right)\right] \quad , \qquad \text{(D-1)}$$

where the relation $\lambda f/d = Ns$ is used.

To the desired color, the left half of the cell is opaque, that is, $|A| \cong 0$. Then the reconstruction becomes

$$h(x,y) \cong \tfrac{1}{2}\exp\left(-i2\pi\,\frac{x}{4Ns}\right)\cdot\text{sinc}\left(\frac{x}{2Ns}\right)\cdot\text{sinc}\left(\frac{y}{Ns}\right)\cdot$$

$$\cdot\left[f(x,y) * \text{comb}\left(\frac{x}{Ns},\frac{y}{Ns}\right)\right] \quad . \qquad \text{(D-2)}$$

This is the same as our original reconstruction (B-12) except that the sinc factor falloff is more gentle in the x-direction and the overall image is less intense. The phase factor is the result of shifting the centers of the hologram cells by $\frac{1}{4}d$ (Fig. D.1a,b).

To the undesired color, the left half of the cell is π radians

out of phase with the right half, that is, $A = -1$. Then the reconstruction is

$$h(x,y) = -i\cdot\sin\left(\frac{\pi x}{2Ns}\right)\cdot\mathrm{sinc}\left(\frac{x}{2Ns}\right)\cdot\mathrm{sinc}\left(\frac{y}{Ns}\right)\cdot \left[f(x,y)*\mathrm{comb}\left(\frac{x}{Ns},\frac{y}{Ns}\right)\right] \quad . \qquad \text{(D-3)}$$

Due to the $\sin(\pi x/2Ns)$ factor, this undesired image is attenuated near the center.

Figure D.2 shows the squares of the factors attenuating the various image amplitudes, the intensities of which are actually observed. We see that even with the phase-null effect, the false image will reconstruct with one half its previous intensity at the edge of the pattern, but that the degree of attenuation increases toward the center. Thus, by imbedding the image in a sufficiently large field of zeros, the false image can be attenuated to any degree desired. A much greater attenuation of the false image can be achieved by dividing the cell still further, as in Figure D.1c, for example, in which the phase-null effect is used in both the u and v-directions. Then the attenuating factor is $\sin^2(\pi x/2Ns)\cdot \sin^2(\pi y/2Ns)$. In this case, if the image is imbedded to half the size in both the x and y-directions, then the false image position in the corner will reconstruct with only 0.0215 times its original intensity, and the interior positions are attenuated still further. Thus the phase-null method can be very effective in eliminating an undesired color.

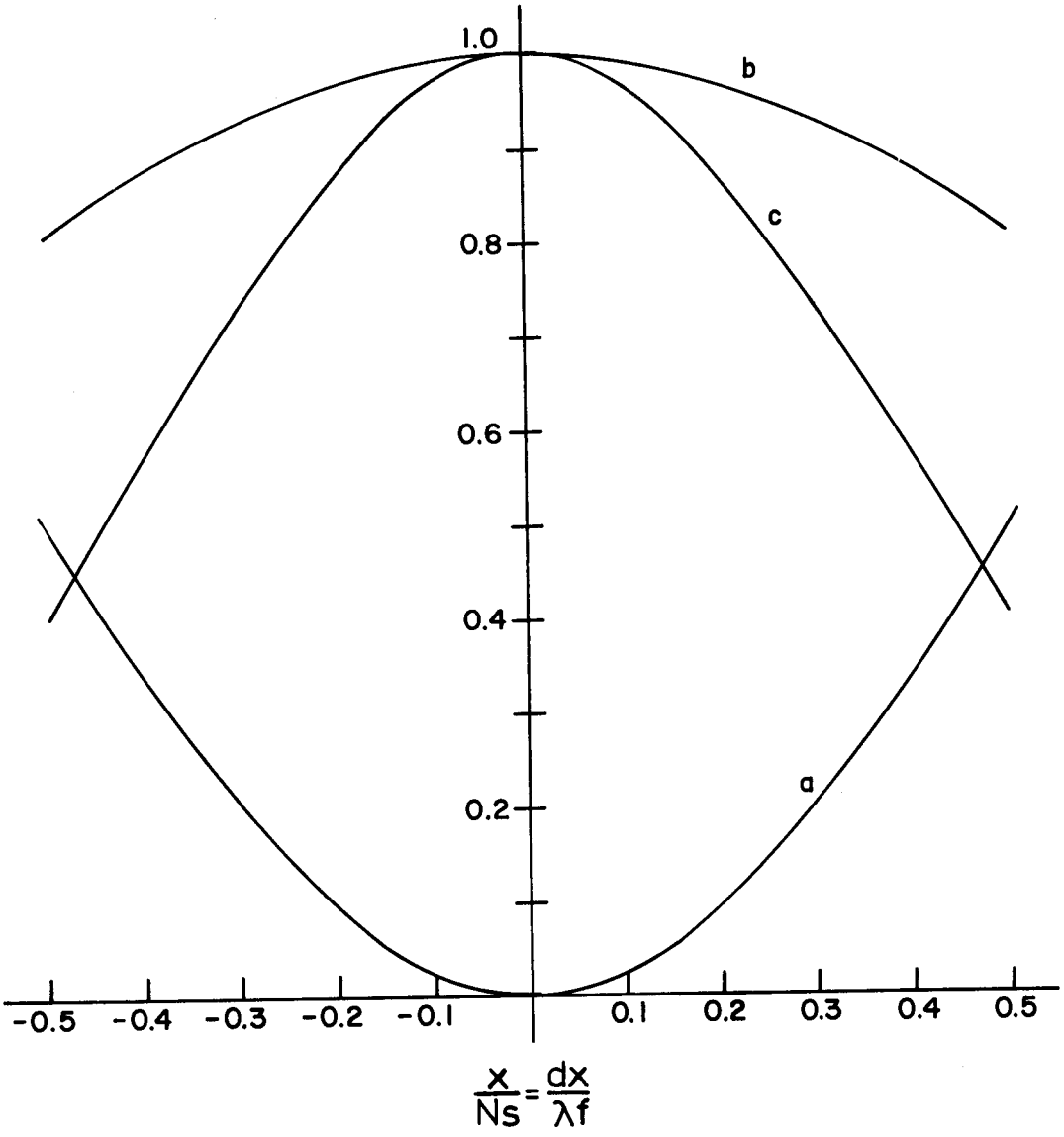


Fig.D.2 Phase-null effect. (a) $\sin^2(\pi x/2Ns)$ factor attenuating the undesired color; (b) $\text{sinc}^2(x/2Ns)$ factor attenuating the desired color; (c) $\text{sinc}^2(x/Ns)$ factor that would attenuate the image if the phase-null method were not used.

REFERENCES

Akahori, H., "Comparison of Deterministic Phase Coding with Random Phase Coding in Terms of Dynamic Range", Appl.Opt. 12, 2336 (1973).

Alferness, R., "Analysis of Optical Propagation in Thick Holographic Gratings", Appl.Phys. 7, 29 (1975).

Bartolini, R.A., "Characteristics of Relief Phase Holograms Recorded in Photoresists", Appl. Opt. 13, 129 (1974).

Bell, R.A., "Principles of Cathode-ray Tubes, Phosphors, and High-speed Oscillography", Application Note 115, H.P. Oscilloscope Systems (Hewlett-Packard Co., Colorado Springs, Colo., 1970).

Biederman, K., "The Scattered Flux Spectrum of Photographic Materials for Holography", Optik 31, 367 (1970).

Birch, K.G. and F.J. Green, "The Application of Computer-Generated Holograms to Testing Optical Elements", J. Phys.D. Applied Physics 5, 1982 (1972).

Bracewell, R., The Fourier Transform and its Applications (McGraw-Hill, New York, 1965).

Brandt, G.B., "Coherent Optical Power Spectra of Photographic Materials", Appl. Opt. 9, 1424 (1970).

Brown, B.R. and A.W. Lohmann, "Complex Spatial Filtering with Binary Masks", Appl.Opt. 5, 967 (1966).

Brown, B.R. and A.W. Lohmann, "Computer-Generated Binary Holograms", IBM J.Res.Develop. 13, 160 (1969).

Bryngdahl, O., "Radial- and Circular-Fringe Interferograms", J.Opt.Soc. Am. 63, 1098 (1973).

Bryngdahl, O., "Geometrical Transformations in Optics", J.Opt.Soc.Am. 64, 1092 (1974).

Bryngdahl, O. and W.H. Lee, "Shearing Interferometry in Polar Coordinates", J.Opt.Soc.Am. 64, 1606 (1974).

Burch, J.J., "A Computer Algorithm for the Synthesis of Spatial Frequency Filters", Proc. IEEE 55, 599 (1967).

Burckhardt, C.B.,"A Simplification of Lee's Method of Generating Holograms by Computer", Appl.Opt. 9, 1949 (1970).

Chu, D.C. and J.W. Goodman, "Spectrum Shaping with Parity Sequences", Appl.Opt. 11, 1716 (1972).

Chu, D.C., J.R. Fienup, and J.W. Goodman, "Multiemulsion On-Axis Computer Generated Hologram", Appl.Opt. 12, 1386 (1973).

Chu, D.C., private communication, 1973.

Chu, D.C., "Spectrum Shaping for Computer Generated Holograms", Ph.D. Thesis, Stanford University, (1974).

Chu, D.C. and J.R. Fienup, "Recent Approaches to Computer-Generated Holograms", Opt.Eng. 13, 189 (1974).

Clair, J.J., "New Method to Synthesize Kinoforms", Opt.Comm. 6, 135 (1972).

Collier, R.J., C.B. Burckhardt, and L.H. Lin, Optical Holography (Academic Press, New York, 1971).

Cooley, J.W. and J.W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series", Math. Comp. 19, 297 (1965).

Damman, H., "Computer Generated Quarternary Phase-Only Holograms", Phys.Lett. 29A, 301 (1969).

Dallas, W.J., Y. Ichioka, and A.W. Lohmann, "Computer-Generated Color Holograms", J.Opt.Soc.Am. 62, 739(A) (1972).

Evans, R.M., W.T. Hansen, and W.L. Brewer, Principles of Color Photography, (J. Wiley and Sons, Inc., New York, 1953).

Fienup, J.R., (a) "Color Images from Computer-Generated Holograms", Proc. S.P.I.E. 48, Acquisition and Analysis of Pictorial Data, (August, 1974).

Fienup, J.R., (b) "Reduction of Quantization Noise in Kinoforms and Computer-Generated Holograms", J.Opt.Soc.Am. 64, 1395A (1974).

Fienup, J.R. and J.W. Goodman, "New Ways to Make Computer-Generated Color Holograms", Nouv.Rev.Optique 5, 269 (1974).

Fillmore, G.L., "Kinoform Viewed as a Random Number Generator", Appl. Opt. 11, 2193 (1972).

Frieden, B.R., "On Arbitrarily Perfect Imagery with a Finite Aperture", Optica Acta 16, 794 (1969).

Frieden, B.R., "The Extrapolating Pupil, Image Synthesis, and Some Thought Applications", Appl.Opt. 9, 2489 (1970).

Gabel, R.A. and B. Liu, "Minimization of Reconstruction Errors with Computer Generated Binary Holograms", Appl.Opt. 9, 1180 (1970).

Gallagher, N.C. and B. Liu, "Method for Computing Kinoforms that Reduces Image Reconstruction error", Appl.Opt. 12, 2328 (1973).

Gabor, D., "A New Microscope Principle", Nature 161, 777 (1948).

Gabor, D., "Microscopy by Reconstructed Wave-fronts", Proc.Roy.Soc. (London) A197, 454 (1949).

Gerchberg, R.W. and W.O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures", Optik 35, 237 (1972).

Gerchberg, R.W., "Super-Resolution Through Error Energy Reduction", Optica Acta 21, 709 (1974).

Goodman, J.W., "Film-Grain Noise in Wavefront-Reconstruction Imaging", J.Opt.Soc.Am., 57, 493 (1967).

Goodman, J W., Introduction to Fourier Optics (McGraw-Hill, New York, (1968).

Goodman, J.W. and A.M. Silvestri, "Some Effects of Fourier-Domain Phase Quantization", IBM J.Res.Develop. 14, 478 (1970).

Goodman, J.W. and R.M. Gray (principal investigators), Annual Report to the National Science Foundation, NSF Grant GK-31606, November 15, 1972.

Gradshteyn, I.S. and I.M. Ryzhik, Table of Integrals, Series and Products (Academic Press, New York, 1965).

Haskell, R.E. and B.C. Culver, "New Coding Technique for Computer-Generated Holograms", Appl.Opt. 11, 2712 (1972).

Haskell, R.E., "Computer-Generated Holograms with Minimum Quantization", J.Opt.Soc.Am. 63, 504 (1973).

Haskell, R.E. and P. Tamura, "Another Look at Computer-Generated Binary Holograms", Proc. S.P.I.E. 52, Coherent Optical Processing (August, 1974).

Hirsch, P.M. et al., U.S. Patent No. 3,619,022 (November 9, 1971).

Holladay, T.M. and J.P. Gallatin, "Phase Control by Polarization in Coherent Spatial Filtering", J.Opt.Soc.Am. 56, 869 (1966).

Huang, T.S. and B. Prasada, "Considerations on the Generation and Processing of Holograms by Digital Computers", MIT Res.Lab. of Electr.,Quart. Prog.Rep. No. 81, 199 (1966).

Huang, T.S., "Digital Holography", Proc. IEEE 59, 1335 (1971).

Hurter, F. and V.C. Driffield, "Photo-Chemical Investigations and a New Method of Determination of the Sensitiveness of Photographic Plates", Soc.Chem.Ind., London 9, 455 (1890).

Jones, L.A., "Photographic Sensitometry, Part I", J.Soc.Mot.Pict.Engrs. (JSMPTE) 17, 695 (1931).

Kermisch, D., "Image Reconstruction from Phase Information Only", J.Opt.Soc.Am. 60, 15 (1970).

Kirk, J.P. and A.L. Jones, "Phase-Only Complex Valued Spatial Filter", J.Opt.Soc.Am. 61, 1023 (1971).

Kodak Data: Characteristic curves for Kodak films can be obtained on request from the Eastman Kodak Co., 343 State Street, Rochester, N.Y. 14650.

Kodak, Kodak Films for the Amateur, Kodak Information Book AF-1, (Eastman Kodak Co., Rochester, N.Y. 1971), p. D-5.

Kodak, Kodak Filters for Scientific and Technical Uses, Kodak Data Book B-3, (Eastman Kodak Co., Rochester, N.Y. 1970).

Kodak, Color as Seen and Photographed, Kodak Data Book E-74 (Eastman Kodak Co., Rochester, N.Y., 1966).

Kogelnik, H.,"Reconstructing Response and Efficiency of Hologram Gratings", Proc.Symp.Modern Optics, Polytechnic Inst., Brooklyn, March, 1967, 605.

Kogelnik, H., "Coupled Wave Theory for Thick Hologram Gratings", Bell Sys.Tech.J. 48, 2909 (1969).

Kozma, A., "Effects of Film-Grain Noise in Holography", J.Opt.Soc.Am. 58, 436 (1968).

Kozma, A., W.-H. Lee, and P.J. Peters, "Holographic Recording and Retrieval System", IEEE/OSA Conference on Laser Engineering and Applications (June 2-4, 1971).

Kozma, A. et al. Holographic Storage/Readout Techniques, Tech.Report RADC-TR-71-54, Rome Air Development Center, AFSC, Griffiss AFB, N.Y., June 1971.

Kozma, A. and D.L. Kelly, "Spatial Filtering for Detection of Signals Submerged in Noise, Appl. Opt. 4, 387 (1965).

Labeyrie, A., "Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images", Astron.Astrophys. 6, 85 (1970).

Lamberts, R.L., "Characterization of a Bleached Photographic Material", Appl.Opt. 11, 33 (1972).

Lamberts, R.L., F.C. Eisen, and F.G. Kaspar, "Measurement of the Optical-Path-Length Variation of a Photographic Grating for Determination of Complex Amplitude Transmittance", J.Opt.Soc.Am. 63, 480 (A) (1973).

Langford, M.J., Advanced Photography (Focal Press, New York, 1974).

Latta, J.N.,"Analysis of Multiple Hologram Optical Elements with Low Dispersion and Low Aberrations", Appl.Opt. 11, 1686 (1972).

Lee, W.-H., "Sampled Fourier Transform Hologram Generated by Computer", Appl.Opt. 9, 639 (1970).

Lee, W.-H., "Effect of Film-Grain Noise on the Performance of Holographic Memory", J.Opt.Soc.Am. 62, 797 (1972).

Lee, W.-H., "Binary Synthetic Hologram", Appl.Opt. 13, 1677 (1974).

Lehmann, M., J.P. Lauer, and J.W. Goodman, "High Efficiency, Low Noise, and Suppression of Photochromic Effects in Bleached Silver Halide Holography', Appl.Opt. 9, 1948 (1970).

Leith, E.N. and J. Upatnieks, "Reconstructed Wavefronts and Communication Theory", J.Opt.Soc.Am. 52, 1123 (1962).

Leith, E.N. and J. Upatnieks, "Wavefront Reconstruction with Diffused Illumination and Three-Dimensional Objects", J.Opt.Soc.Am. 54, 1295 (1964).

Lesem, L.B., P.M. Hirsch, and J.A.Jordan, Jr., "Holographic Display of Digital Images", AFIPS Conference Proceedings, Fall Joint Computer Conference 41 (Spartan Books, Inc., New York) 41 (1967).

Lesem, L.B., P.M. Hirsch, and J.A. Jordan, Jr., "Computer Synthesis of Holograms for 3-D Display", Comm. of the ACM 11, 661 (1968).

Lesem, L.B., P.M. Hirsch, and J.A. Jordan, Jr., "The Kinoform: A New Wavefront Reconstruction Device", IBM J.Res.Develop. 13, 150 (1969).

Lohmann, A.W., D.P. Paris, and H.W. Werlich, "A Computer Generated Spatial Filter, Applied to Code Translation", Appl.Opt. 6, 1139 (1967).

Lohmann, A.W. and D.P. Paris, "Binary Fraunhofer Holograms, Generated by Computer", Appl.Opt. 6, 1739 (1967).

Lohmann, A.W. and D.P. Paris, "Computer Generated Spatial Filters for Coherent Optical Data Processing", Appl.Opt. 7, 651 (1968).

Lohmann, A.W. and H.W. Werlich, "Incoherent Matched Filtering with Fourier Holograms", Appl.Opt. 10, 670 (1971).

MacAnally, R.B., "Single-Emulsion Phase and Amplitude Transparency", J.Opt. Soc.Am. 64, 563 (A) (1974).

Marathay, A.S., "Realization of Complex Spatial Filters with Polarized Light", J.Opt.Soc.Am. 59, 748 (1969).

Matthews, J. and R.L. Walker, Mathematical Methods of Physics (Benjamin, New York, 1970), p.110.

Mees, C.E.K. and T.H. James, editors, The Theory of the Photographic Process (Macmillan Co., New York, 1966).

Mezrich, R.S., "Magnetic Holography", Appl.Opt. 9, 2275 (1970).

Neblette, C.B., Photography: Its Materials and Processes (Van Nostrand, New York, 1962).

Pearlman, W.A., "Quantization Error Bounds on Computer Generated Holograms", Ph. D. Thesis and Information Systems Lab. Tech. Report No. 6503-1, Stanford University (1974).

Pennington, K.S., "Holographic Parameters and Recording Materials", Ch. 21 in CRC Handbook of Lasers with Selected Data on Optical Technology (Chemical Rubber Co., Cleveland, Ohio, 1971).

Powers, R.S., "Quantization Errors in Computer-Generated Holograms", Ph. D. Thesis, Stanford University (1975).

Powers, R.S. and J.W. Goodman, "Error Rates in Computer-Generated Holographic Memories", to appear in Applied Optics (1975).

Powers, S.A. and O.E. Miller, "Pitfalls of Color Densitometry", Phot.Sci. Eng. 7, 59 (1963).

Ransom, P.L., "Synthesis of Complex Optical Wavefronts", Appl.Opt. 11, 2554 (1972).

Roman, P. and A.S. Marathay, "Analyticity and Phase Retrieval", Il Nuovo Cim. 30, 1452 (1963).

Severcan, M., "Computer Generation of Coherent Optical Filters with High Light Efficiency and Large Dynamic Range", Ph. D. thesis and Information Systems Lab. Tech. Report No. 6415-6, Stanford University (1973).

Smith, H.M., "Photographic Relief Images", J.Opt.Soc.Am. 58, 533 (1968).

Smith, H.M., "Production of Photographic Relief Images with Arbitrary Profile", J.Opt.Soc.Am. 59, 1492 (1969).

Thomas, C.E., "Film Characteristics Pertinent to Coherent Optical Data Processing Systems", Appl.Opt. 11, 1756 (1972).

Tichenor, D.A., "Extended Range Spatial Filters for Image Deblurring", Ph. D. thesis and Information Systems Lab. Tech. Report No. 6418-3, Stanford University (1974); and private communication.

Tsujiuchi, J., "Correction of Optical Images by Compensation of Aberrations and by Spatial Frequency Filtering", in Progress in Optics, Vol. II (North Holland, Amsterdam, 1963) pp. 131-180.

Turner, A.E., Jr., "I/O Systems Optimization for the Digital Processing of Photographic Data", Information Systems Lab. Tech. Report No. 6419-9, Stanford University (1972).

Vander Lugt, A., "Signal Detection by Complex Spatial Filtering", IEEE Trans.Inform.Theory, IT-10, 139 (1964).

Voglesong, W.F., "Color Sensitometry", in Color: Theory and Imaging Systems, R.A. Eynard, ed. (Society of Photographic Scientists and Engineers, Washington, D.C., 1973).