MODERN HOLOGRAPHIC RECORDING AND ANALYSIS TECHNIQUES APPLIED TO EDGE–LIT HOLOGRAMS AND THEIR APPLICATIONS

Zane A. Coleman

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of

Doctor of Philosophy

in the Faculty of Pure and Applied Science

February 1997

DECLARATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified by acknowledgments or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a higher degree.

Signed.....

Dated.....

DEDICATION

To Mom,

You've always been there with your support, love, and assistance.

Thank You.

SYNOPSIS

The main aim of this research was to create and analyze a compact holographic optical system which is illuminated through the edge of the substrate. Edge–lit holograms are used in order to minimize the space requirements for various possible replay geometries. Many aspects of the recording, analyzing, and replay techniques were found to be very critical for creating an efficient optical system.

The edge–lit holograms were recorded in a photopolymer film developed by Du Pont. When recording an interference pattern in the film with a steep reference beam, the affects of attenuation, beam ratio, and refractive index matching on optical efficiency are much more pronounced than in traditional holography due to the effective film thickness and Fresnel reflections.

Many of the practical aspects involved when recording with photopolymers were carefully examined. The measured physical characteristics of the film and its packaging considerations proved very crucial in the choice of the optimum recording film and conditions.

Traditional recording geometries were explored and new recording methods were discovered and evaluated.

A new adaptation of the rigorous coupled wave diffraction theory was used to analyze edge–lit holograms and the regimes of validity were detailed for other approximate

theories. Using the rigorous theory, the diffraction characteristics of the edge–lit hologram were investigated. The theory is compared with the measured diffraction characteristics of the film before and after heat processing. The discrepancies between the rigorous theory and the actual recordings are explained and modifications to the theory are suggested.

For an edge–lit hologram, the considerations for replay are more complicated than those for recording. The influences of the replay source radiometry, directionality, wavelength, and choice of substrate on the optical characteristics are analyzed for different replay conditions.

Various possible applications are examined from an optical and manufacturing viewpoint. In particular, the applications of an edge–lit hologram as a fingerprint illuminator and a liquid crystal display illuminator are discussed and their benefits are weighed against traditional methods.

CONTENTS

PREFACE

The work for this thesis could not have been accomplished without the support of ImEdge Technology Inc., Loughborough University, Du Pont, and the Centre for Modern Optics at DeMontfort University. The significant progress made as a result of this research can hopefully justify the faith in the project those have had throughout the years.

During the past four and a half years I've learned much more than could ever be written in this thesis, and much of that I can attribute to Prof. N. J. Phillips. When I first set foot in the catacombs of the lab at Loughborough I had no idea how much I would learn about holography. Most of that is due to Nick's seemingly infinite supply of ideas and enthusiasm. I must thank Chris and Terry for welcoming me to the lab, showing me the ropes, and throughout the years offering advice and friendship. I would like to thank Bruce, Dave, and Adrian for being good friends and colleagues despite all of my complaining.

I would like to thank all of my family whom have had the patience to see me through furthering my education away from home and assisting wherever possible.

I would like to thank the staff at the physics department at Loughborough and in the SERCentre at De Montfort, as they have all been very friendly and helpful.

Finally, I would like to thank all my friends for their support and their encouragement which helped me to finish. A special thanks goes to Dave Armstrong for helping me take fresh breaks from sanity through our surrealistic conversations.

Chapter 1

INTRODUCTION AND BACKGROUND

Many people believe that holograms of the future can be made to create three dimensional walking and talking people “live in your living room.” This belief only serves to perpetuate the existing view that holograms do not live up to their potential. In many cases this is because other technology in the system is lacking. The main limitations of M.I.T.'s holographic video display (real–time hologram generation from computer data) for example is the processing speed of the computers being used and the data transfer rate [1.1]. Similarly, with many display holograms, the reason for the poor visible quality often lies with the light source.

Technical advances, such as the extremely rapid growth of computer power and the advent of compact and powerful laser diodes, will soon be able to complement the theoretical performance of holographic optical devices and three–dimensional displays.

One of the areas of technical advancement that has only begun to reach holography is that of miniaturizing the entire display system. Most holographic systems have always

been plagued with the awkward exterior lighting requirements. Methods for reducing the size of the holographic display system have been theorized, although very few extremely positive results have been demonstrated.

One method for reducing the size of the optical system is to position the replay light source at or near the edge of the substrate of the hologram. These holograms are commonly referred to as "edge–lit" or "edge–illuminated" holograms. Other benefits of edge–illuminated holograms include very little diffraction of external lighting and the ability to be replayed using a laser without the danger of stray light escaping the system.

Since most all of the compact geometries examined here are illuminated from the edge, they will be referred generally to as "edge–lit" holograms. The different techniques for recording holograms that are illuminated through the edge can significantly affect the replayed image. Therefore, a distinction must be made between the recording and replay process. The different techniques examined here will refer to recording only, such as steep referenced or waveguide hologram for example.

The evolution of the edge–lit hologram began with Lin's [1.2] first public reference to the edge illuminated hologram. The history from Lin to the beginning of the work presented here, is described by Henrion [1.3]. The main areas of work were done initially by Upatnieks [1.4–1.6], followed by Benton and colleagues [1.7–1.10] and most recently by Phillips, this author and colleagues [1.11–1.15].

In order to distinguish between the different types of edge–lit holograms, one must first look at how traditional holograms are recorded. In general, holograms are created by recording the fringes produced by the interference of a reference wavefront with a mutually coherent object wavefront scattered by the object being recorded. When illuminated using the reference wavefront the scattered light from the hologram replays the object wavefront so that the object still appears to be present. When recording a holographic optical element, the object is removed and the second

wavefront (referred to here as object wavefront only to distinguish the two wavefronts) is recorded directly.

In a thick recording medium the recorded holographic interference fringes form planes which are regions of constant index of refraction in a phase hologram. These planes subtend an angle with a line normal to the plane of the hologram. Holograms recorded whereby $0^\circ \leq |\text{fringe angle}| < 45^\circ$ are classed as transmission holograms and those recorded such that $45^\circ < |\text{fringe angle}| \leq 90^\circ$ are classed as reflection holograms, depending on the notation used. However, precisely whether a hologram is acting as reflection or transmission depends also on the replay angle. Holograms recorded by the edge–referenced method are one of the few types of holograms where the fringe angle falls very close to the transitional boundary at ±45°. A view of the typical fringe angle for different types of holograms with a signal beam incident normal to the film is given in Fig. 1.1.

In Fig. 1.1, the steep reference angles correspond to edge–referenced holograms commonly made in the laboratory. (Note: The fringe angles may also take on their negative values, although only the positive ones are shown here for clarity). The fringe angle divisions correspond to a total internal reflection (TIR) condition in the glass for a waveguide hologram and the maximum reference angles in the film for silver halide holograms.

In conventional silver halide holography where the reference beam enters the film directly from the large face, the fringe angles are limited to 0°–19° for transmission holograms, and 71°–90° for reflection holograms. This is due to the minimum reference angle of 142° for reflection holograms and a maximum reference angle of 38° for transmission holograms (from the surface normal) in the film. Similarly, for conventional photopolymer holograms, the fringe angle is limited to the range of 0°–20° and 70°–90° for transmission and reflection holograms, respectively.

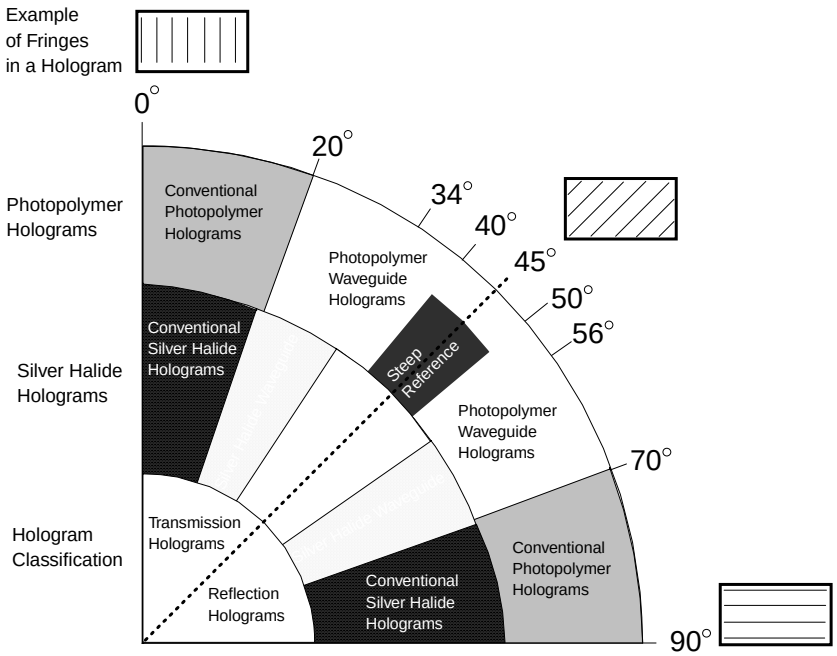


Figure 1.1 The fringe angles for various types of holograms with the signal angle incident normal to the film. The fringe angles associated with the silver halide holograms are calculated using refractive indices of 1.52 for the substrate and 1.64 for the emulsion. With photopolymers, the refractive indices can essentially be matched.

For the waveguide holograms depicted in Fig. 1.1, the criteria was determined by assuming the waveguide to be in a glass substrate (n_s = 1.52). The fringe angle range in silver halide is limited to 19°–34° and 56°–71° for waveguide holograms because of the refractive index difference between the substrate and the film (n_f = 1.64). The photopolymer is not as limited because the film can be made with the same refractive index as the glass, yielding a waveguide fringe angle range of 20° to 70°.

From Fig. 1.1, one can see the difficulty of recording slant angles near 45°. The steep referenced holograms (edge–referenced) and waveguide holograms in photopolymer are two of the more common regimes to obtain fringe angles approaching 45° with a normal incidence signal beam. Silver halide holograms in an index matching tank may also have fringe angles approaching 45°, although the tank and the problem of obtaining a reference conjugate with the glass substrate are two major problems inhibiting this method. These, along with the ease of processing and mass production possibilities, were the reasons photopolymers were investigated for edge–lit holograms in the research presented here.

Edge–lit holograms have many practical advantages over conventional holograms but to exploit them fully one needs to develop a rigorous theoretical understanding of how such holograms are recorded and replayed.

Henrion [1.10] and Birner [1.8] each have produced a masters thesis on edge–lit holography while Huang's doctoral thesis [1.16] focuses on substrate–guided wave holography. For the practical portion of their thesis, they all report on silver halide holograms, and most all of the theory is limited to silver halide holograms. Henrion and Birner assume Leith's theory [1.17] applied to the edge–lit hologram, while Huang redevelops Kogelnik's theory and expands it into two dimensions. Birner offers a look at a dispersion compensation for edge–lit holograms and Henrion models the effective exposure density (in terms of how much energy reaches the emulsion versus the developed optical density and efficiency) and looks at stray reflections.

Huang applies a Fourier method (used to describe two–dimensional diffraction in acousto–optics) to solve the two wave, two–dimensional coupled wave equations to model the diffraction characteristics of an edge–lit hologram. Huang's elegant analysis does not allow for multiple diffraction orders, however, and cannot be used for some real transmission hologram situations. More importantly, his theory does not account for diffraction at the boundaries, which always exists in real–time recording media. The regime where his, and Kogelnik's theory are valid is shown for the specific case of a silver halide hologram. Huang does not show, however, that the validity of his theory for transmission holograms is very dependent on the film thickness and refractive index modulation.

In comparison, the research presented here uses the rigorous coupled wave theory to model coupled wave diffraction from cosinusoidally modulated volume phase gratings. The theory applies to all of the edge–lit regimes except that of exactly 90° incidence and 90° diffraction. The theory presented here is also valid for high index modulations, very thick gratings, and accounts for boundary diffraction. While the practical aspects of the thesis concentrate on the recording of edge–lit holograms in

photopolymer, the theory is presented for cases of silver–halide and photopolymer and the results of the calculations can be translated to other recording media. Since the photopolymer offers the ability to record fringes from 0° to 90°, the rigorous theory is needed to examine the diffraction characteristics for all fringe angles.

Chapter 2 begins by examining the intensity profile of the interference pattern of a signal beam (normal to the film) and a steep angled reference beam. Taking into account absorption within the recording medium, the average fringe contrast is maximized for the thickness of the film for various fringe angles. The practical results are tested against theoretical predictions and are found to correlate very well.

The practical aspects of recording edge–lit holograms are examined in Chapter 3. The techniques for holding the sample, and the physical and optical properties of the photopolymers are examined. The importance of matching the refractive index between the substrate and the film is carefully examined. The real–time effects of recording with photopolymers are shown and their implications on efficiency are hypothesized.

Chapter 4 looks at different possible recording geometries for edge–lit holograms. In addition to the traditional transmission and reflection geometries, the evanescent and waveguide hologram are examined and discussed. Two new recording possibilities, alternative Bragg condition recording and direct fringe copying, are investigated.

In Chapter 5, holographic diffraction is modeled using the coupled wave equations. The rigorous solution is presented with new adaptations for thick holograms and high refractive index modulations. With the help of the rigorous theory, the approximate theories of Kogelnik and Vasnetsov are tested and the appropriate regimes for their use are described. Using the rigorous theory, the diffraction efficiency, wavelength sensitivity, and angular sensitivity of edge–lit holograms are predicted for the full range of replay angles. The effects of slanted fringes on these parameters are especially detailed.

Chapter 6 looks at techniques for measuring the physical and diffractive characteristics of recorded holograms. Results are presented for edge–lit holograms and compared with the rigorous theory developed in Chapter 5. The effects of heat processing on the physical and diffraction characteristics of photopolymers are discussed.

Chapter 7 examines the extremely critical replay considerations for edge–lit holograms. When the light source is brought in very close to the hologram, as in the edge–lit case, the lighting requirements become much more critical than with traditional holograms. The affects of radiometry, directionality, and wavelength of the hologram illumination are presented. The choice of replay substrate also affects the optical characteristics of the holographic system. The use of a coupling grating offers an alternative replay possibility with different light source requirements.

Different applications for edge–lit holograms are reviewed in Chapter 8. The edge–lit fingerprint illuminator and liquid crystal display illuminator are two important applications with large potential that are examined in detail. Other applications are described briefly and several traditional and new techniques for mass manufacture of edge–lit holograms are presented.

REFERENCES

1.1 St.–Hilaire, P., Benton, S., Lucente, M., Sutter, J., Plesniak, W., "Advances in holographic video," Proceedings from Practical Holography VII, SPIE vol. 1914, pp 188–196.
1.2 Lin, L.H.,"Edge illuminated holograms," Abstract of a talk from JOSA spring meeting, 60:714A, April 1970.
1.3 Henrion, M., "Diffraction and exposure characteristics of the edgelit hologram," Maters thesis, MIT, 1995.
1.4 Upatnieks, J., "Compact holographic sight," SPIE Vol. 883, Holographic Optics:Design and Applications, pp 171–176 (1988).
1.5 Upatnieks, J., "Edge–illuminated holograms," Applied Optics, Vol. 31, No. 8, pp1048–1052, (1992).
1.6 Upatnieks, J., "Compact hologram displays and method of making compact hologram," Pat. No. 5,151,800, Sept. 1992.
1.7 Benton, S.A., Birner, S., and Shirakura, A., "Edge–lit rainbow holograms," Practical Holography IV, SPIE vol. 1212, pp 149–157.

1.8 Birner, S., "Steep reference angle holography: Analysis and Applications," Masters thesis, MIT, 1989.
1.9 Farmer, W., "Edge–lit holographic stereograms," Masters thesis, MIT, 1991
1.10 Henrion, M., "Diffraction and exposure characteristics of the edgelit hologram," Maters thesis, MIT, 1995.
1.11 Phillips, N.J., Ce, Wang, "The recording and replay of true edge–lit holograms," IEE Conference Publication 342, pp 8–11, 1991.
1.12 Phillips, N. J., Coleman**,** Z.A.,and Wang,C.,"Holograms in the edge–illuminated geometry–new materials developments," Practical Holography VII:Imaging Materials, SPIE vol. 1914, pp 75–81 (1993).
1.13 Metz, M. , Flatow ,C., Coleman, Z.A., Phillips, N.J. ,"The use of edge–lit holograms for compact fingerprint capture," CardTec SecureTec Conference, April 10th, (1995).
1.14 Coleman, Z.A., Metz, M., Phillips, N.J., "Holograms in the extreme edge illumination geometry," Holographic Materials II, SPIE Proceedings vol. 2688, San Jose, CA (1996).
1.15 Metz, M., Phillips, N.J., Coleman, Z.A., Flatow, C., "Holographic optical element for compact fingerprint imaging system," Optical Security and Counterfeit Deterrence Techniques, SPIE Proceedings vol. 2660, San Jose, CA (1996).
1.16 Huang, Q., "Substrate guided wave holography: Analysis, experiments and applications," Doctoral thesis, University of Alabama in Huntsville, 1994
1.17 Leith, E., "Holographic Data Storage in Three Dimensional Media," Applied Optics, Vol. 5, pp 1303, (1966).

Chapter 2

RECORDING EDGE–LIT HOLOGRAMS

2.1 Introduction

A hologram is recorded in a light sensitive medium from the intensity variations of a interference pattern. When referring to volume holograms, the interference pattern is recorded throughout a film of substantial thickness. In a perfect system, the maximum contrast of the intensity pattern would not change throughout the thickness of the layer. However, since most recording mediums absorb some of the light during the exposure, the intensity of the contrast pattern can vary with depth if the distance the light has traveled at one point in the film is not the same for the two recording beams. Because of the steep recording angle involved, the distance traveled by the reference beam is much greater than the signal beam for edge–lit holograms. Since the distance traveled is greater, the effects of absorption are also greater.

One way of increasing the average fringe contrast would be to vary the amount of sensitizing dye throughout the volume of the recording medium. Unfortunately, this is very difficult to accomplish practically. A simpler way to improve the average

contrast is to adjust the relative intensities of the signal and reference beams. In order to understand these effects, one must first be able to model the contrast of the interference fringes created during the recording.

2.2 Interference and Contrast

A holographic grating is formed from the interference of two mutually coherent waves of light of similar frequencies in a recording medium. For simplicity, a view of the interference pattern formed from two plane waves is illustrated in Figure 2.1. For the illustrations here, only reflection phase holograms with the electric field normal to the plane of incidence (also known as TE, H–mode, or s polarization) will be shown. The equations and results are valid for transmission holograms and can be expanded to include amplitude or TM polarized holograms. The recording medium is in the x–z plane.

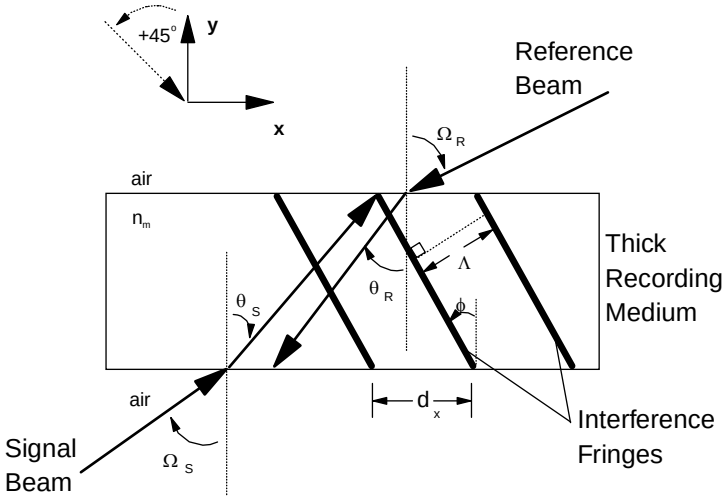


Figure 2.1. The interference fringes formed from a general two beam interference pattern.

2.2.1 General Wave Representation

Assuming that the signal and reference beams have no component in the z direction, the signal and reference waves are represented by

$$E_s = A_s e^{ik_m(x\sin\theta_S + y\cos\theta_S)} \quad \text{and} \quad E_R = A_r e^{-ik_m(x\sin\theta_R + y\cos\theta_R)}. \tag{2.1}$$

These equations represent the electric fields of the signal and reference waves respectively with A_s and A_r as the signal and reference amplitudes, k_m the wave propagation constant in the medium, and θ_R, θ_S the angles of the reference and signal beams from the surface normal inside the film.

2.2.2 Absorption and Interference

For the general case of recording without absorption, Eq. 2.1 would suffice. However, most holographic recording materials must absorb light in order to record the interference pattern. The absorption coefficient affects the recording of amplitude as well as phase holograms. If one were to consider the absorption of the reference and signal beams as they pass through the recording medium, the asymmetry of the intensity in most recordings is evident. For example, in a film with an absorption coefficient of α= 0.7133 μm^{-1} (such as Holographic Recording Film 352 from Du Pont) the intensity attenuation is represented by

$$\frac{I(y)}{I_0} = e^{\frac{-2\alpha y}{\cos\theta_R}} \tag{2.2}$$

where I(y) is the intensity at any point inside the hologram and I_0 is the initial intensity of the reference beam before entering the film (at y=0). At a reference beam angle of 88.7°, the light is at 10% of its original intensity at a depth of 3.66 μm. Thus, if the film were actually 20 μm thick, not much of the reference beam would penetrate even half of the film, creating a very thin and inefficient hologram. One can also use these equations to imagine the thickness which each beam 'sees' the film. With the signal incident normal to the surface of the hologram, it 'sees' the film as only 20 μm

thick with the normal effects of absorption. However, with the reference beam at 88.7°, it 'sees' the film as being

$$\frac{thickness}{\cos\theta_R} = \frac{20\mu m}{\cos(88.7)} = 881\mu m, \tag{2.3}$$

which, with the same absorption coefficient, reduces the intensity considerably when the beam reaches the other side of the film. This example roughly illustrates the importance of the absorption of the reference beam at steep angles. One notices that the exponent in Eq. 2.2 grows quickly as the angle approaches 90° as in the case of steep edge–referenced holograms. The effect of absorption is not as crucial when the angles of the signal and reference beam are more relaxed as in traditional face–lit holography.

Returning to the interference calculations, the effects of the absorption coefficient can be seen further. From Eq. 2.1 the electric fields inside the recording material are described by [2.1]

$$E_s = A_s e^{ik_m(x\sin\theta_S + y\cos\theta_S) - \frac{\alpha(y+t)}{\cos\theta_S}} \quad \text{and} \quad E_R = A_r e^{-ik_m(x\sin\theta_R + y\cos\theta_R) + \frac{\alpha y}{\cos\theta_R}}, \tag{2.4}$$

where α is the attenuation coefficient of the recording medium. The interference of these two waves is represented by the sum of the two electric fields inside the recording material. The total amplitude at any point in the hologram is the sum of these two fields E_S and E_R.

$$A_{tot} = E_s + E_R = A_s e^{ik_m(x\sin\theta_S + y\cos\theta_S) - \frac{\alpha(y+t)}{\cos\theta_S}} + A_r e^{-ik_m(x\sin\theta_R + y\cos\theta_R) + \frac{\alpha y}{\cos\theta_R}} \tag{2.5}$$

In the most common uses of edge–referenced holograms the signal is incident normal to surface of the recording medium ($\theta_S = 0$). Eq. 2.5 is then reduced to

$$A_{tot} = A_s e^{ik_m y - \alpha(y+t)} + A_r e^{-ik_m(x\sin\theta_R + y\cos\theta_R) + \frac{\alpha y}{\cos\theta_R}} \tag{2.6}$$

2.2.3 Intensity and Fringe Contrast Calculations

Holographic recording materials respond to variations in intensity inside the film and the intensity at any point within the film can be expressed as

$$I = A_{tot} \times A_{tot}^{*} \tag{2.7}$$

$$I = A_s^2 e^{-2\alpha(y+t)} + A_r^2 e^{\frac{2\alpha y}{\cos\theta_R}} + 2A_s A_r e^{-\alpha(y+t)+\frac{\alpha y}{\cos\theta_R}} \times \cos\left(k_m(y + y\cos\theta_R + x\sin\theta_R)\right). \tag{2.8}$$

For a phase hologram to be efficient it needs a high permittivity modulation, i.e. a high index of refraction modulation. Since the refractive index modulation of the recording material is directly related to the intensity variation, a high contrast interference pattern is desired to create a high index of refraction modulation. Ultimately, a high diffraction efficiency of the hologram is desired. The diffraction efficiency is directly related to the index of refraction modulation and is defined as the ratio of the intensity of the light diffracted in a defined order (usually the first order) to the incident intensity. To maximize diffraction efficiency one must maximize the average contrast. Thus, with contrast defined as

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{2.9}$$

one can deduce the fringe contrast for any point within the hologram. Since the absorption does not affect the relative contrast in the x–direction, one can determine the contrast in the y–direction by finding the maximum and minimum intensities from Eq. 2.8. The argument of the cosine term in Eq. 2.8 is rapidly varying (because k_m is very large) relative to the hyperbolic terms. Therefore, the contrast (as a function of y

is determined when the cosine term is at a maximum (+1) or a minimum (–1), yielding

$$C = \frac{2A_s A_r e^{-\alpha(y+t)+\frac{\alpha y}{\cos\theta_R}}}{A_s^2 e^{-2\alpha(y+t)} + A_r^2 e^{\frac{2\alpha y}{\cos\theta_R}}}. \tag{2.10}$$

This equation can be simplified by multiplying the numerator and denominator by $e^{\alpha y - \frac{\alpha y}{\cos\theta_R} + \alpha t}$ to yield the result

$$C = \frac{2A_s A_r}{A_s^2 e^{-\alpha y - \alpha t - \frac{\alpha y}{\cos\theta_R}} + A_r^2 e^{\alpha y + \alpha t + \frac{\alpha y}{\cos\theta_R}}}. \tag{2.11}$$

Eq. 2.11 defines the contrast at any point inside the recording medium. The highest diffraction efficiency would be produced from the highest average contrast inside the recording material. The average contrast is defined as

$$\langle C \rangle = \frac{1}{t}\int_{-t}^{0} C\, dy. \tag{2.12}$$

By letting

$$z = e^{\alpha y + \alpha t + \frac{\alpha y}{\cos\theta_R}} \text{ and } \frac{\partial z}{\partial y} = \left(\alpha + \frac{\alpha}{\cos\theta_R}\right) z \tag{2.13}$$

one can rewrite Eq. 2.12 using Eq. 2.11 as

$$\langle C \rangle = \frac{1}{\alpha t\left(1 + \frac{1}{\cos\theta_R}\right)} \int_{e^{\frac{-\alpha t}{\cos\theta_R}}}^{e^{\alpha t}} \frac{2A_s A_r}{A_s^2 + A_r^2 z^2}\, dz. \tag{2.14}$$

This can be further simplified by substituting

$$\xi = \frac{A_r}{A_S} z \text{ and } \frac{d\xi}{dz} = \frac{A_r}{A_s} z \text{ into Eq. 2.14, giving} \tag{2.15}$$

$$\langle C \rangle = \left(\frac{2}{\alpha t \left(1 + \frac{1}{\cos\theta_R} \right)} \right) \int_{\frac{A_r}{A_s} e^{\frac{-\alpha t}{\cos\theta_R}}}^{\frac{A_r}{A_s} e^{\alpha t}} \frac{1}{1+\xi^2} \, d\xi \tag{2.16}$$

Completing the integration of Eq. 2.16 results in the average fringe contrast of

$$\langle C \rangle = \frac{2}{\alpha t \left(1 + \frac{1}{\cos\theta_R} \right)} \left[\tan^{-1}\left(\frac{A_r}{A_s} e^{\alpha t} \right) - \tan^{-1}\left(\frac{A_r}{A_s} e^{\frac{-\alpha t}{\cos\theta_R}} \right) \right] \tag{2.17}$$

This result shows that the average fringe contrast is very dependent on the reference beam angle when it approaches 90º. The effect of the attenuation coefficient in the recording of holograms is grating non–uniformity. One can see from Eq. 2.17, that the contrast is very dependent on the ratio of the amplitudes of the signal and reference beams. By adjusting the measured intensity ratio (which is related to the amplitude ratio) one can optimize the average fringe contrast. In most practical holographic recording regimes only minor adjustments to the relative intensities of the reference or signal beams are needed to optimize the beam intensity ratio, $\frac{A_r^2}{A_S^2}$. However, for the edge–referenced hologram the $\frac{1}{\cos\theta_R}$ factor strongly affects the absorption of the reference beam. This result suggests that the beam ratio needs to be adjusted more than in the normal recording situations. However, before the optimum beam ratio can be calculated, the absorption coefficients must be measured.

2.3 Measuring the Absorption Coefficients

In order to accurately predict the contrast and diffraction efficiency of holograms, the absorption coefficients need to be measured directly for the various films at the specified laser wavelengths. Tests were attempted using standard spectrophotometers with very little success. The standard spectrophotometer uses non–collimated, non–polarized light and the measurement speeds are slow. Therefore, a device was constructed that measures the absorption coefficients correctly in real time (Fig. 2.2).

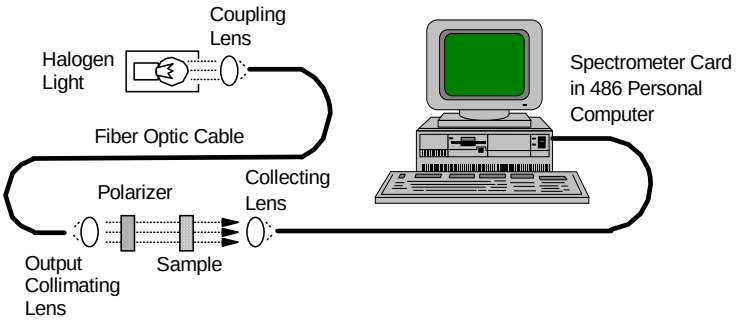


Figure 2.2. The spectrophotometer apparatus created to measure absorption coefficients and diffraction efficiencies.

The light from a halogen bulb was used as a reference spectrum for the wavelength range from 400 nm to 700 nm. The Ocean Optics PC1000™ spectrometer card in the computer used a blazed grating having 600/lines per millimeter giving a resolution of 0.5 nm. The optical fibers were used to steer the light and allowed for filtering and convenient collimation. A reference spectrum of the halogen bulb with a clear glass sample (to account for the Fresnel reflection when the sample was loaded) was taken as a baseline and the absorption of the sample was measured relative to this.

Transmission curves of several experimental Du Pont films (HRF 352, HRF 300x006, and HRF 700x071) were measured using this apparatus. Sensitization information and absorption coefficients of the various Du Pont films can be obtained from the transmission curves in Fig. 2.3.

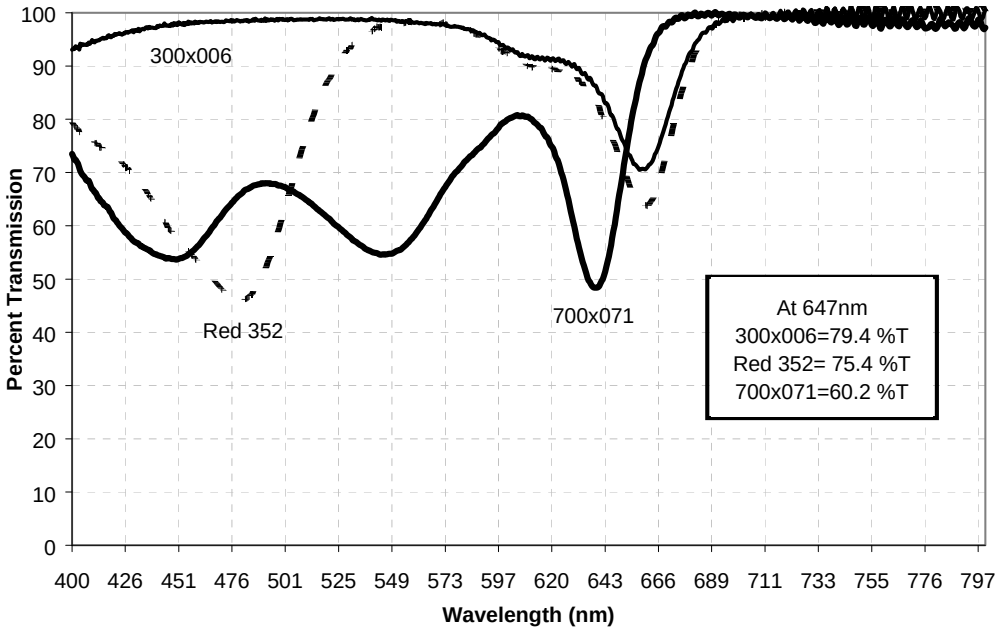


Figure 2.3. The transmission curves for three types of red–sensitive Du Pont Photopolymers using the spectrophotometer apparatus.

When 647 nm light is incident normal to HRF 700x071, 60% of the light is transmitted through the film. From this measurement, one can deduce from Eq. 2.2 that this 20 μm film has an absorption coefficient of $\alpha = 0.0127\ \mu m^{-1}$. Thus, with the absorption coefficient the average contrast can be predicted for a known recording angle.

With Millimask® silver halide recording film, the percent transmittance at 0° incidence is 78% at 647 nm and 34% at 514 nm. These equate to absorption coefficients of $\alpha = 0.018\ \mu m^{-1}$ and $\alpha = 0.077\ \mu m^{-1}$, respectively.

2.4 The Effects of the Beam Ratio on Hologram Efficiency

As mentioned in §2.2, the ratio of the incident intensities of the signal and reference beams can affect the intensity of the interference pattern. This intensity directly influences the refractive index modulation in the film. One must, however, maximize the average fringe contrast in order to maximize the refractive index modulation.

2.4.1 Maximizing the Average Contrast

The beam ratio that maximizes the average contrast can be determined by differentiating Eq. 2.17 as below:

$$\frac{d\langle C\rangle}{d\left(\frac{A_r}{A_s}\right)} = \frac{2}{\alpha t\left(1+\frac{1}{\cos\theta_R}\right)}\left[\frac{e^{\alpha t}}{1+\left(\frac{A_r}{A_s}e^{\alpha t}\right)^2} - \frac{e^{\frac{-\alpha t}{\cos\theta_R}}}{1+\left(\frac{A_r}{A_s}e^{\frac{-\alpha t}{\cos\theta_R}}\right)^2}\right]. \tag{2.18}$$

Setting $\frac{d\langle C\rangle}{d\left(\frac{A_r}{A_s}\right)} = 0$, gives a maximum average contrast at

$$\frac{A_r^2}{A_s^2} = \frac{e^{\alpha t\left(1+\frac{1}{\cos\theta_R}\right)}-1}{e^{2\alpha t} - e^{\alpha t\left(1-\frac{1}{\cos\theta_R}\right)}}. \tag{2.19}$$

One can note that $\lim_{\alpha\to 0}\frac{A_r^2}{A_s^2} = 1$ as expected for the case of no absorption . In the case of the pure Lippmann hologram (where $\theta_S = 0°$ and $\theta_R = 0°$) the beam ratio also converges to 1 as one would expect. Since $\frac{A_r^2}{A_s^2}$ is a direct indication of the measured intensity ratios, one can use this relationship to predict the optimum beam ratio for maximum average fringe contrast, thus maximum hologram efficiency.

2.4.2 Testing the Beam Ratio Calculations

To test Eq. 2.19 numerous holograms were recorded at a reference angle of $\theta_R = 78°$ and a signal angle of $\theta_s = 0°$ in Du Pont photopolymer with an absorption coefficient of $\alpha = 0.0127\ \mu m^{-1}$. This angle was chosen for a particular application, and is not as steep as the general steep reference applications. Using Eq. 2.19, one can deduce that

the optimal beam ratio is 2.6:1 (reference intensity to signal intensity), and that under those conditions, the average contrast will be 0.92.

The efficiency was maximized experimentally by adjusting the beam ratio to confirm these calculations. To determine the beam ratio, the intensities of the signal and reference beams were needed. The uniform signal intensity was measured directly. However, the reference beam is Gaussian and diverges from the cylindrical mirror (Fig. 3.1) and the intensity at the film can not be directly measured. The reference beam intensity approximation used was based on measuring the total power and knowing the input beam waist. The expansion of the reference beam is due to the angle of incidence in the film in the x direction and depends on the divergence from the mirror in the z direction. The angle of incidence is measured (as in §6.2.2) and the new beam waist could be calculated by knowing the divergence of a collimated beam on the mirror. Thus, the beam waist is not symmetrical and has a different size for the x and z directions. Knowing the beam waist in each direction, the new intensity could be approximated at a particular location.

The reference intensity was estimated at a particular location at which the efficiencies were measured on an optical fiber based spectrophotometer. Since one can not measure the fringe contrast directly, the diffraction efficiency must be measured and related to the beam ratio. The efficiencies of the hologram were measured prior to heating due to uncertainty of the effects of heat processing on the photopolymers. While the diffraction efficiency depends on the contrast of the fringes, it is also directly related to the refractive index modulation. Insufficient refractive index modulation will not result in 100% diffraction efficiency regardless of the fringe contrast. However, the contrast should be directly related to the efficiency. In order to see the effects of the fringe contrast (thus beam ratio), the maximum measured efficiency was adjusted so that it was near the maximum fringe contrast. The efficiency values were scaled by adding 0.52 to all of the efficiencies so that the curves would overlap. These scaled efficiencies are compared with the theoretical results of the average fringe contrast versus beam ratio (directly from Eq. 2.17) and are shown in Fig. 2.4. The shape and maxima of the curve and the data demonstrate that our data correlates well with the theory. The data follows the same shape as the

curve and the maxima are located at the same beam ratios. This is what one would expect assuming that average fringe contrast would be directly related to efficiency.

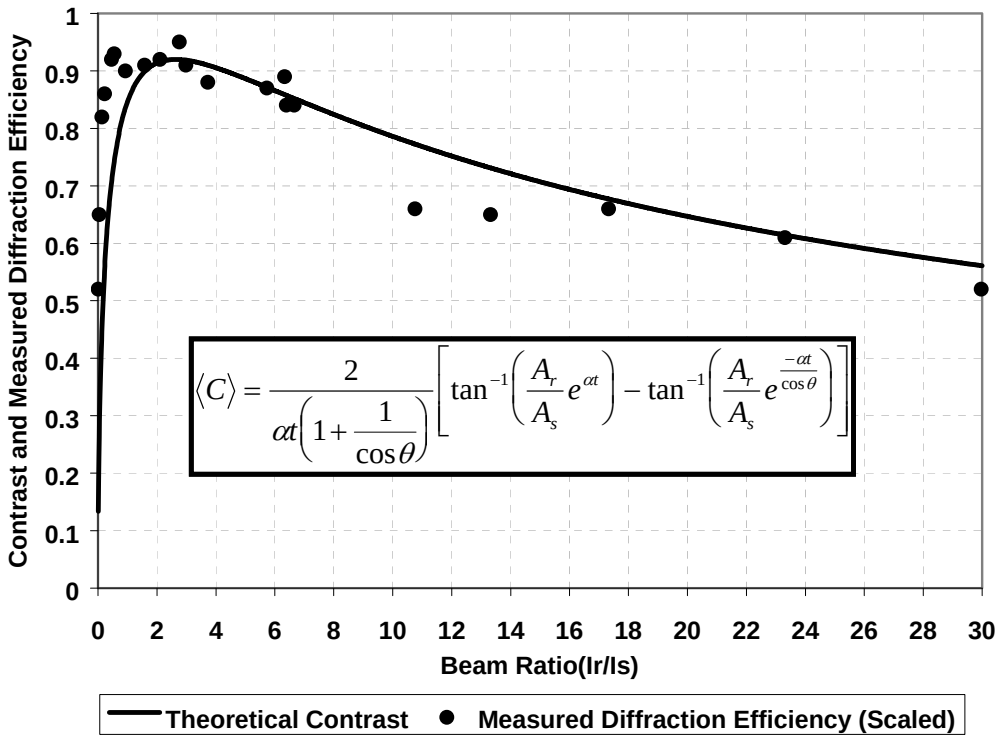


Figure 2.4. The measured and theoretical diffraction efficiency versus beam ratio based on the average fringe contrast model of Eq. 2.17. The holograms were made from the setup of Fig. 2 on HRF 700x071 film and exposed to a total of 2.18 mW/cm2 at 647 nm for 100 seconds followed by a UV cure.

2.4.3 Expanding the Beam Ratio Calculations

When calculating the optimum recording geometry or materials, it is useful to know how variables such as the recording angle and absorption will affect the final edge–lit hologram. From Eq. 2.17, the maximum average contrast of the hologram can be calculated for various absorption coefficients (characteristic to the film) and the recording angle (according to the setup). Some results are shown in Figures 2.5 and 2.6.

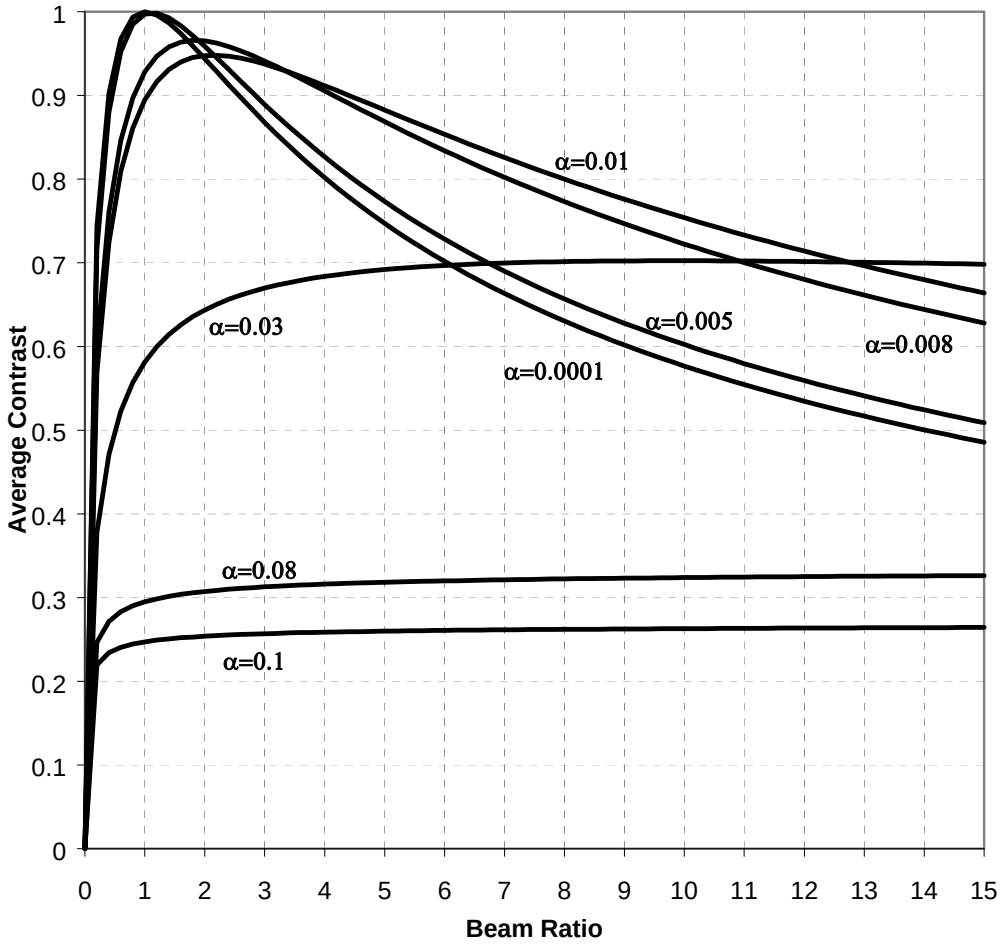


Figure 2.5 Average contrast versus the beam ratio (Reference:Signal) for various absorption coefficients. The data is calculated from Eq. 2.17 with $\theta_r = 78°$, $\theta_s = 0°$, and thickness = 20 μm. The values for the absorption coefficients are in μm^{-1}.

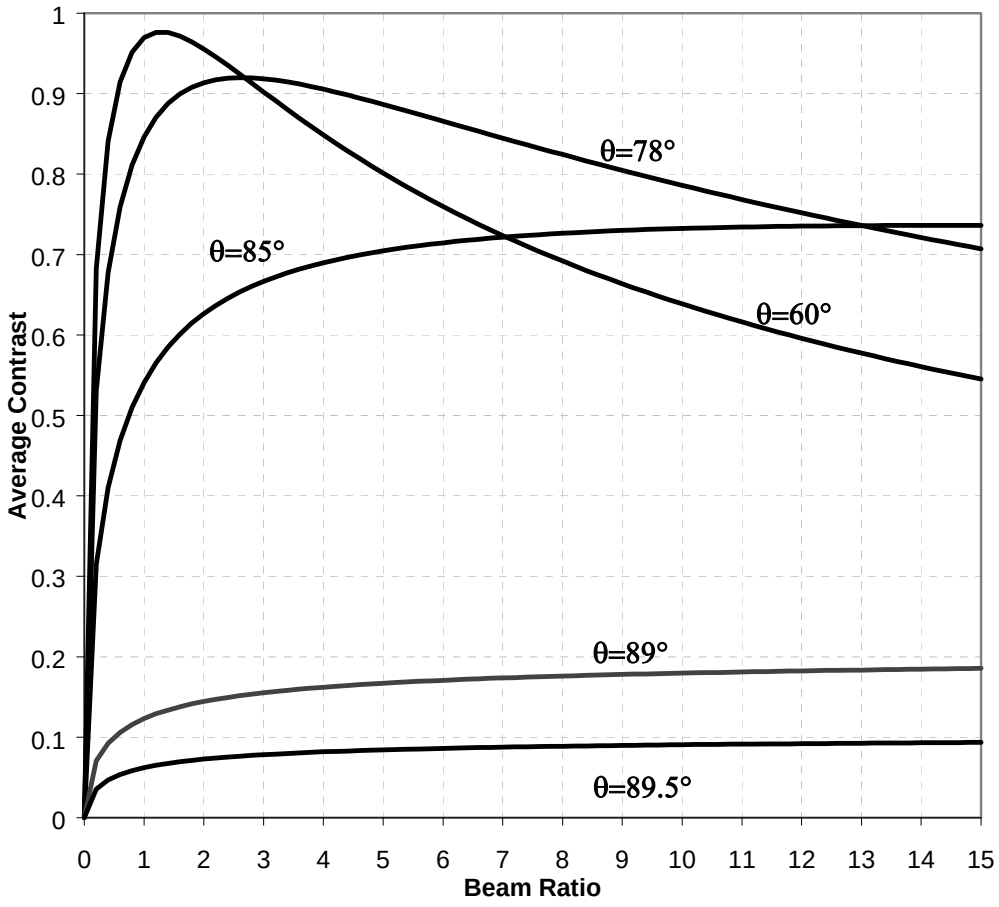


Figure 2.6 The average contrast versus the beam ratio (Reference:Signal) for various reference beam angles with θ_s=0°. The data is calculated from Eq. 2.17 with α = 0.0127 μm^{-1} and thickness = 20 μm.

The graphs in Figs. 2.5 and 2.6 are similar in shape. From Fig. 2.5, one can see that the beam ratio is more crucial when the absorption coefficient is low and a higher average contrast can be achieved. If a more absorptive film is used, the ratio is not very crucial, however, the maximum average contrast is lower.

Similarly, one can see that at a smaller angle, the beam ratio is more crucial and the average contrast is higher. As the angle approaches grazing, the maximum average contrast decreases and its dependence on beam ratio saturates.

These two graphs enforce a compromise in order to record a high contrast (high efficiency) edge–lit hologram. If one is going to directly record a very steep hologram, then one must have a low–absorption coefficient to get a reasonable

contrast and efficiency. One should also note that the absorption coefficient may need to be high enough to diminish the intensity of the reference beam so that the reflection back into the film does not reduce the dynamic range of the film.

The final efficiency of the hologram is very dependent on the linearity of the refractive index modulation versus incident intensity during recording. For example, in a non–linear recording medium the modulation increase resulting from a low light exposure (from a low contrast interference pattern) may be much greater, relatively speaking, than that of the dark area nearby. This may be exhibited if there is an exposure threshold with the photopolymer. Tests done with exposing the similar types of photopolymer through dark masks, as with the Microsharp® material have shown that there is a threshold of exposure energy. Until this threshold is reached, no appreciable refractive index modulation is seen in the final hologram.

The non–linearity of the heating process may also affect the efficiency. If one were to record a perfect sinusoidal interference pattern recorded in the medium, the index modulation profile would unlikely be a perfect sinusoid due to the limits of monomer and component diffusion [2.2].

As a result, with careful exposure considerations one is able to exploit the contrast calculations of Eq. 2.17, in combination with the exposure threshold and the heating effects, to achieve a higher refractive index modulation and efficiency. The effects of the absorption are strong, but they do not prohibit efficient recordings. Using Figs. 2.5 and 2.6 one could easily optimize the diffraction efficiency of a hologram by changing the recording beam ratio to suit the recording angle and absorption of the film.

REFERENCES

2.1 Coleman, Z.A., Metz, M., Phillips, N.J., “Holograms in the extreme edge illumination geometry”, Holographic Materials II, SPIE Proceedings vol. 2688, San Jose, CA (1996)

2.2 Colburn, W. and Haines, K., “Volume Hologram Formation in Photopolymer Materials,” Applied Optics, Vol. 10, No. 7, pp 1636–1641, July, (1971)

Chapter 3

Practical Hologram Recording Considerations

3.1 Introduction

There are many requirements that must be met when recording a hologram in a laboratory: if just one of these is not satisfied, then one is unlikely to record an efficient hologram. Often, the most difficult requirement to meet is stability. The laser, the table, all of the optical components, and the hologram itself must be very stable for effective recording. The requirements for the light sensitive recording medium are also very important. The choice of the correct film, in the case of edge–illuminated holography, is even more crucial than the stringent demands set for traditional face–lit holography. In addition to the normal film requirements, the refractive index is extremely crucial in an edge–lit hologram. Also important are the film packaging and the dynamic real–time effects when recording.

3.2 Laser and Stability Requirements

In order for a stable holographic interference pattern to be formed within a light sensitive medium, one requires quasi–monochromatic light with high coherence. A continuous wave laser operating in the lowest transverse mode (TEM_{00}) is spatially coherent. However, temporal coherence may vary among different types of lasers. A high temporal coherence is needed with holography, thus a significant coherence length is required. An argon ion gas laser with an etalon fitted, such as that used in the work presented here, has a coherence length of around 10 m. Low temporal coherence or poor stability of the optical components can reduce the fringe contrast during the recording of a hologram. If one object in one beam of a holographic setup moves relative to the other beam during recording, then the fringe contrast is reduced due to the phase difference introduced. Also, in order to achieve maximum fringe contrast, the electric fields of the two interfering beams must be parallel to each other. If they are perpendicular, then the fringe contrast falls to zero. [3.1, 3.2]

An argon ion laser was chosen for the research presented here because of the power requirements for exposing the Du Pont photopolymers (30 mJ/cm^2) over possibly large areas. In order to maintain stability of the setup and relative components, a large steel table on isolation tubes was used with magnetic mounts. Since movement of reflective optical components in the setup results in a larger phase change at a given point than transmissive components, their stability is much more crucial. Many standard optical component mounts are provided with spring mounts for positioning, which were found to be unacceptable because there was inevitable long–term creep or movement associated with them. Replacing these mounts with rigid, glued mounts provided high holographic stability and enabled long exposure times (over 3 minutes).

Any variation in temperature in a holographic setup can also reduce the contrast of an interference pattern because of the refractive index (thus phase) change introduced from the heat convection. For this reason, the setup was isolated from air currents by enclosing it in a cage with acrylic panels that could be moved on hinges.

Normally, in simple holographic arrangements mechanical switches are used as shutters. However, the shutter was found to be a critical component when the laser is working at high powers. The typical type of shutter involves something light–absorbing moving in and out to block the beam. However, at high powers, the absorbing element heats significantly and when moved out of the beam (usually about a centimeter at most) to commence exposure the heat convection from the heated element would introduce phase changes due to the change in the refractive index of the air. This destroyed the coherence of the exposure. To resolve this, the block was replaced with a mirror and the beam was 'dumped' to a block at a significant distance away such that the heat convection would not affect the beam. The results improved significantly.

When working with small area holograms, low laser powers were needed. Simply turning down the current sometimes caused mode–hopping in the laser. An in–line linear absorbing polarizer would reduce the intensity, however, the absorbing polarizers introduced linear bands into the beam profile. Therefore, a rotating a Glan–Thompson prism polarizer was sometimes used to remove a portion of the light without altering the rest of the setup.

3.3 Recording Setup

The setup used to record the edge–referenced holograms is depicted in Fig. 3.1 and a picture of the setup without the optical fiber can be found in Appendix A.

The polarized light output from the argon ion laser is first passed through a half–wave plate. The alignment of the half–wave plate is adjusted to rotate the polarization of the light entering the polarizing beamsplitter, thereby controlling the beam intensity ratio.

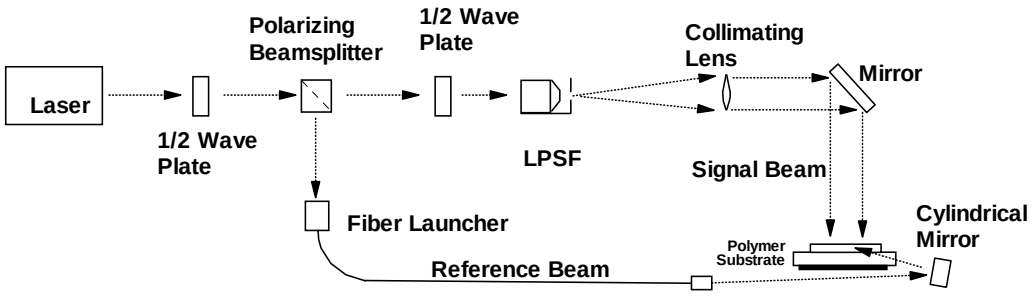


Figure 3.1 The holographic arrangement for recording a standard edge–referenced reflection hologram.

The signal beam passes through another half–wave plate to bring the polarization back to vertically linear polarization. The lens–pinhole spatial filter (LPSF) is used to filter the light yielding a homogeneous Gaussian beam profile. A further lens then collimates the light before reaching the mirror and then the film. It is important to use a light dump index matched to the back of the substrate of the film in order to eliminate unwanted reflections. The signal beam can be easily apodized using an anti–Gaussian filter. One may use an absorbing filter [3.3] or non–absorbing lens anti–Gaussian filter [3.4]. In order to use the absorbing filter, the $1/e^2$ beam waist of the signal beam has to match the characteristic diameter of the filter in order to properly remove the Gaussian characteristics of the beam to give a uniform intensity field. When a small anti–Gaussian filter was used, diffraction rings from the apertures created by the lenses and the filter in the setup shown in Fig. 3.2 caused unwanted rings in the beam. One can reduce the visibility of these rings by using larger lenses and a larger filter.

The reference beam may be launched into a fiber after the beamsplitter. The fiber serves two purposes: it allows easy positioning of the beam for entering into the substrate and more importantly, it filters the reference beam while maintaining its beam waist and divergence. The fiber should be polarization maintaining to keep the polarization matched with the signal beam, and it should be able to handle high powers. Unfortunately, there are inherent difficulties which may be encountered while using a fiber in one arm of a holographic setup. The length of the fiber may become an issue since in order to filter the beam well, several meters are required. However, the path length difference between the reference and signal beam must be

multiplied by the refractive index of the fiber and may contribute to a large difference in path length. Thus, a 3 meter fiber would result in a 4.5 meter path length difference between the two beams. In some arrangements this may be a problem. The fiber, being a reflective component, must also remain very stable during the exposure to prevent phase changes being introduced and destroying the contrast. The fiber was tacked down in the setup to prevent movement. Good results were obtained with and without the fiber.

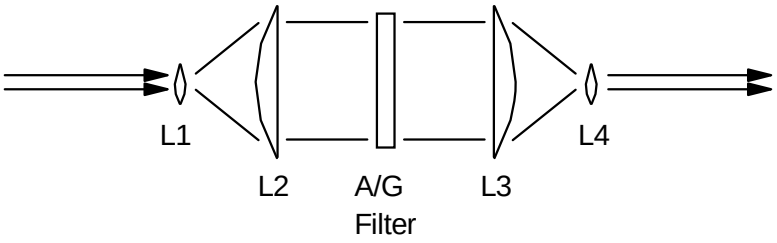


Figure 3.2 The anti–Gaussian arrangement for removing the Gaussian profile of the laser beam. Lens 1 diverges the light to the correct diameter depending on the size of the filter and Lens 2 collimates the light before entering the filter. Lens 3 then converges the light down to its original size while Lens 4 re–collimates the light. The lenses should all be A/R coated to prevent stray reflections.

The diverging optics for the reference beam are very crucial. The main application for the edge–lit reflection holograms required non–conjugate replay by a point light source. This means that the recording light needed to simulate a point source by being highly divergent. In the beginning, two cylindrical lenses were used in combination to diverge the light (in the z direction of Fig. 2.1) so that the reference beam would cover the film. This did not produce sufficient divergence. Later a small cylindrical mirror was used which had a higher divergence due to being a reflective optical element rather than a transmissive one. The light diverging in one direction enters the substrate through the narrow edge (typically 3 mm) before reaching the recording film. Ideally, there should be another substrate on the other side of the film which would allow the reference beam to be dumped out the other side. Fortunately, the high absorption at a steep angle helps alleviate the need for a light dump for the reference beam.

3.4 Sample Holding

The mount which holds the sample (consisting of an unexposed film on a substrate) must be very rigid and stable in order for high contrast fringes to be created. Several different types of plate holders were examined including vertical, vacuum held, and horizontal holders. The first holder positioned the sample vertically so that the glass substrate rested on a rod with a hard plastic holder on the top attached to a vertical arm. The hard plastic seemed to be to flexible and allowed slight movement of the substrate. The second attempt at sample holding involved creating a vacuum between the sample and an acrylic base inside an index matching tank as illustrated in Fig. 3.3.

An index matching tank was examined in order to try to remove stray reflection of the reference beam. With this method a "sandwich" geometry is created where the first substrate is the liquid, the second is the film, and the third is the film substrate.

The glass window at the edge of the tank allows the reference beam to enter where it then passes beneath the sample and onto the cylindrical mirror which is attached to steering controls that reach out of the tank. The light then reflects directly off the mirror onto the film. The light which is not absorbed passes through the substrate and is absorbed at the edge of the tank. The sample is held to the acrylic by a vacuum. Acrylic was used so that the holder could be fabricated as one piece and the signal could pass through the holder. Unfortunately, the flatness between the sample and the acrylic holder was never sufficient to allow a perfect vacuum, leading to the index matching fluid being sucked through the hose creating bubbles and obvious movement.

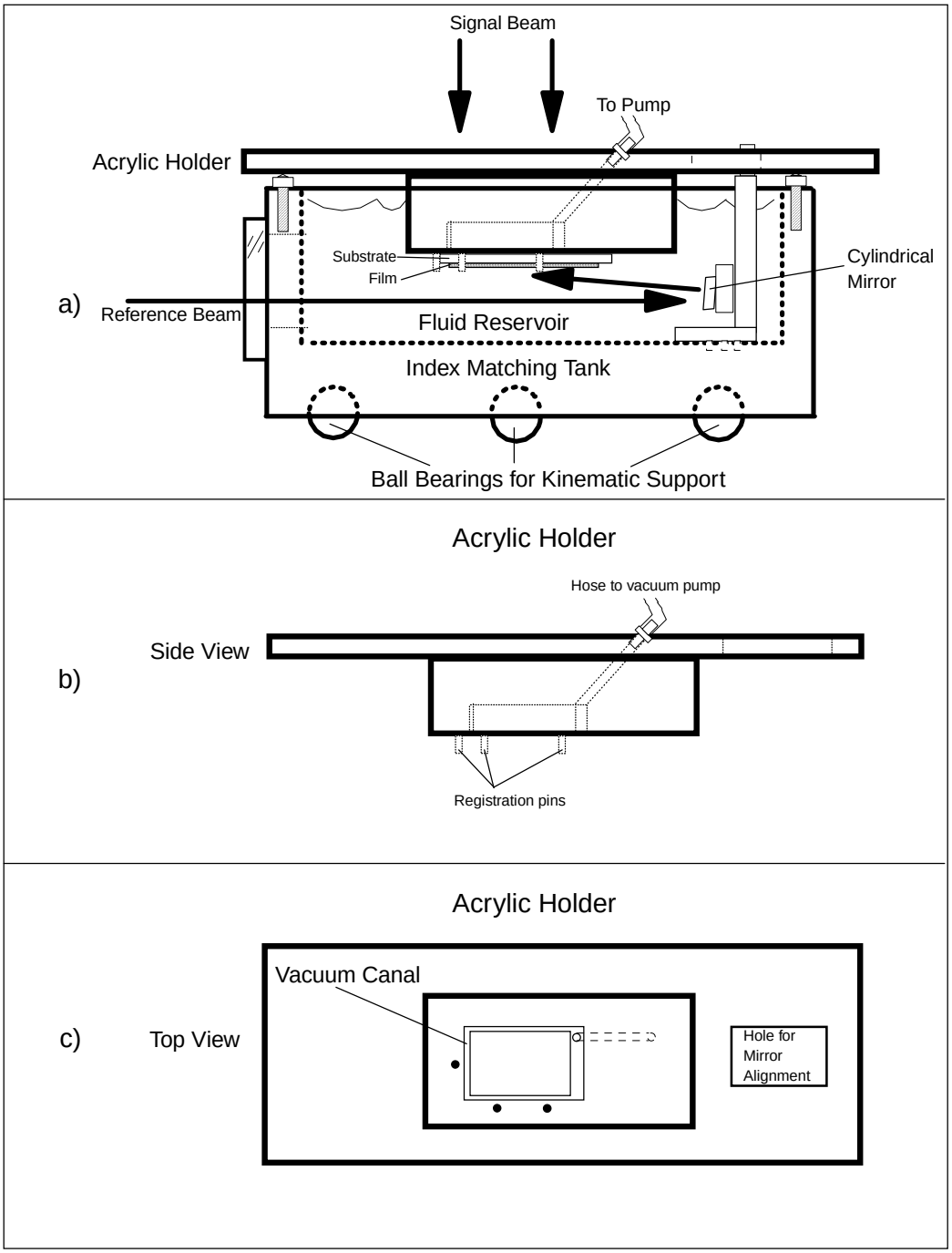


Figure 3.3 The mirror and sample holder for the vacuum method. a) The aluminum tank geometry and the acrylic holder in position for recording. b) The side view of the acrylic sample holder. c) The top view of the acrylic sample holder.

Another difficulty with the setup of Fig. 3.3 was the wait time required for the bubbles to settle after the holder was placed on the tank. The area between the sample

and the acrylic was originally a hollow area. This was modified to a channel around the sample with an island (Fig. 3.3 (c)) to reduce the bending forced on the sample/substrate combination (The substrate in these cases was acrylic so that the Mylar cover could be removed when the film bonded to the acrylic as described in §3.6.5).

The tank/vacuum method was abandoned due to low diffraction efficiency results attributed to the instability caused by leaks and the general difficulties involved in recording. A simpler technique, which used rigid kinematic support, was adopted with much better results (Fig. 3.4).

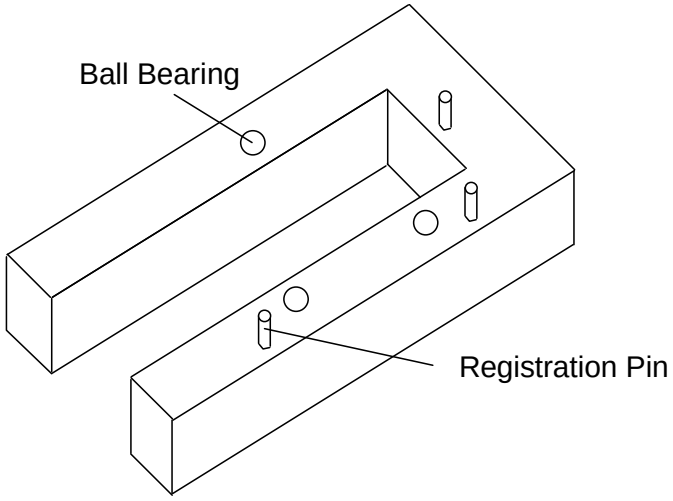


Figure 3.4 Gravity positioned, stable, sample holder

The horizontal sample holder adopted utilizes gravity to position and hold the sample as a constant force. The rectangular glass substrate rests on the three ball bearings kinematically. The sample holder is tilted slightly in the horizontal and vertical directions so that the sample rests against the pins. The pins also allow precise registration for repositioning the sample after measurement.

3.5 Choice of Recording Material

The demands on holographic film exceed those required by many other optical applications or recordings. Typically, up to 6000 lines per millimeter may be required to record a reflection hologram (or edge–lit hologram) while a simple photograph may only require 50 lines/mm. Two of the most common holographic films are silver halide emulsions (usually supplied by Agfa–Gevaert) and photopolymers (usually supplied by Du Pont). Silver halide emulsions may have typically between 3000 and 5000 lines/mm resolution (AGFA 10E75 and AGFA 8E75 recording films respectively). Photopolymers normally have resolution of 5000 lines/mm and the Omnidex® HRF 352 photopolymer supplied by Du Pont has a near constant refractive index modulation amplitude in the range from 3000–6000 lines/mm [3.19].

Recording edge–illuminated holograms has had limited success when using the traditional silver–halide materials [3.5–3.8]. One of the main limitation lies in the inherent refractive index difference between the substrate and the film. The silver–halide emulsion Holotest® (supplied by Agfa–Gevaert) has an average index of refraction of n=1.64 [3.6]. Compared with the usual glass substrate of n=1.52, the 0.12 refractive index difference creates spurious reflections, which are a significant problem. At steep angles, the intensity of the transmitted light is significantly reduced due to the reflection. This can be a significant problem if the available laser power is an issue.

Stray reflections can also create a standing "wood–grain" interference pattern within the film which is recorded and decreases intensity uniformity. While some of these cosmetic difficulties can be alleviated by using an index matching tank [3.6], the index tank introduces more complications (sample holding, alignment, etc.) and makes processing more difficult since one has to remove the index matching fluid. The silver halide recording films require multi–step wet processing and careful calibration in order to achieve acceptable efficiencies. Despite the low exposure requirement (typically around 25 $\mu J/cm^2$) relative to photopolymers, the unfavorable index difference between the film and substrate makes recording edge–lit holograms with silver halide material a very difficult process.

The Du Pont Omnidex® photopolymers offer many advantages over the silver halide films when used for edge-lit holograms. These photopolymers offer higher refractive index modulation possibilities which can, in some situations, lead to higher diffraction efficiencies than the silver halide film. The processing of the photopolymers is dry and a significant amount of real–time development occurs. The film is easier to process and a quick examination without any processing can reveal important information and save time. Another crucial benefit of the photpolymers supplied by Du Pont is that they can vary in many parameters such as absorption, index of refraction, and modulation vs. spatial frequency (i.e. films devoted to reflection or transmission holograms). Despite the high exposure requirement (typically 30 mJ/cm^2), these films have a much higher potential for success in creating an efficient edge–lit hologram.

3.6 Film Considerations

Determining and understanding the optical and dynamic characteristics of the photopolymer recording medium are crucial to the production and optimization of holograms. The photopolymerization that occurs during and after exposure is a complex process and many variables are still not precisely defined and understood. The fringes of higher refractive index resulting after polymerization are dependent on many variables such as the effective diffusion length, the exposure intensity and duration, etc.. In choosing a recording material, optical qualities such as sensitization to a desirable wavelength and the average refractive index and modulation are important factors. The packaging of the commercial film can also be a deciding or limiting factor.

When all of the optical qualities (such as the average refractive index) have been measured before recording, one must then compare them to the final optical qualities after processing to understand and optimize the effects of processing on the

photopolymer. Complications can arise in the case of photopolymers because they develop in real–time and the fluorescence of the absorbing dye may fog the recording.

3.6.1 Photopolymer Composition and Polymerization

A considerable amount of research has been conducted on the Du Pont photopolymers and other films similar to them. Much of the research has been conducted by Du Pont themselves [3.9–3.23] as well as other sources [3.24–3.34]. The Du Pont photopolymers are usually formulated according to the following formula: [3.18]

- Photosensitizing dye, 0.1–0.2%
- Initiator, 1–3%
- Chain transfer agent, 2–3%
- Plasticizer, 0–15%
- Acrylic monomer, 28–46%
- Polymeric binder, 45–65%

These percentages are with respect to the total composition weight and the solvents used to prepare the films were dichloromethane and methanol. The films are usually 25 μm thick web coated onto a 50 μm thick Mylar (polyethylene terephthalate) carrier with a 25 μm Mylar or polypropylene removable cover. The films are sensitized to various wavelengths and the absorption is usually centered close to the most popular laser wavelengths for the applications.

Understanding exactly what happens within the film when it is exposed to light is a complicated process and the details are not precisely known. The widely accepted model for understanding the index modulation created inside the film is the diffusion model presented first by Colburn and Haines [3.30] and continued by Wopschall and Pampalone [3.20] and Booth [3.13]. Small aspects of this model such as the initial decrease in index upon exposure and the inert characteristics of the binder have been criticized by Smothers et. al. [3.18]. Generally, the diffusion model is accepted and can be expanded to include the effects of heating in the processing of the Du Pont

photopolymers. This process for a general Du Pont photopolymer film can be summarized in four steps (Table 3.1):

STEP	PROCESS DETAILS
1. Initial exposure and polymerization	Expose film to 25 mJ/cm^2
3. Monomer diffusion	Wait (occurs almost instantaneously in real time)
3. Complete polymerization	UV cure >100 mJ/cm^2 at 365 nm
4. Final component diffusion	Heat at 100° C for 2 hours

Table 3.1 The four main steps for high refractive index modulation in Du Pont photopolymers.

In the first step, the chemical reaction occurring within the film is described as:

$$\text{Dye} \xrightarrow{h\nu} \text{Dye*}$$

$$\text{Dye*} + \text{Initiator} \longrightarrow \text{R}\bullet$$

$$\text{R}\bullet + \text{Transfer Agent} + \text{Monomer} \longrightarrow \text{Polymer}$$

When the film is exposed to light which is absorbed by the dye, the dye becomes excited and decomposes the initiator to form a free radical. This free radical then oxidizes the transfer agent to create another radical which when combined with the monomer begins the polymerization process.

The standing wave pattern created holographically within a film during recording has areas of constructive (light areas) and destructive (dark areas) interference. As the polymerization process occurs within the light areas, the monomer concentration is depleted because it is forming the polymer chain. Since the polymer is more dense than the monomer, the index of refraction is slightly higher in the regions of polymerization (the bright areas). Normally, this increase in index is connected with a volume change assuming the amount of all of the components is constant. However, the monomer concentration is thus different in the light and dark areas and the monomer diffuses into the light area to equalize the concentration. This diffusion

is where the main initial index modulation is attributed. This produces an initial index gradient which is visible in real time. As the exposure continues, more of the monomer is polymerized and the film continues to harden which prevents further migration. A significant amount of diffusion occurs in real time and simply waiting after an exposure can demonstrate a further increase in the initial (post–heating) diffraction efficiency. However, further tests need to be done before concluding whether this results in a higher final diffraction efficiency after heat processing.

After holographic exposure, the film is cured, or "fixed," with a uniform UV exposure which polymerizes most, if not all, of the remaining monomer. After this cure, the film is insensitive to any further exposure.

If the film is simply left alone, a further index modulation will occur. This was observed on holograms left overnight that were considerably brighter the next day. However, in the full process suggested by Du Pont, the film is uniformly heated at 100°C for two hours. The initial refractive index modulation can then be increased by up to a factor of four [3.14] with the heating. It is believed that this heating softens the components of the film and allows the components within the film to further migrate/diffuse between the photopolymer–rich and the binder–rich areas of the film. Without the heating step, only a small amount of migration can occur as monomer polymerization further plasticizes the film.

The modulation of the film is very dependent on the feature size [3.13, 3.40] because of the diffusion distance and rate of diffusion inherent in the photopolymer. Du Pont generally characterize their films based on this modulation relative to feature size as reflection or transmission films. The transmission films (such as HRF 600 and HRF 150) generally have a higher refractive index modulation for larger feature size while the reflection films (such as HRF 352, HRF 706, and HRF 700x031) have higher modulation for small feature sizes.

Observations in the lab suggest that in the edge–referenced holograms are subject to a failure of the reciprocity law. In general, the reciprocity law states that for a constant exposure the results are independent of the recording energy or recording time [3.35].

Failure of the reciprocity law means that a high intensity for a brief time may have a different effect on the recording than a low exposure for a long time. In photopolymers, the rate of polymerization is directly related to the intensity of the exposing light. Since the refractive index modulation is dependent on the rate of polymerization and the rate of diffusion, the modulation is dependent on the exposing intensity. If the rate of polymerization is much higher than the rate of diffusion, then less diffusion takes place and the refractive index modulation is lower. Thus, the lower the light level (above a threshold), the more monomer diffusion can take place before being polymerized. Du Pont suggest that the refractive index modulation decreases with higher exposure intensities for transmission holograms [3.10]. They have not noticed this dependence in reflection hologram tests. Lab results seem to suggest that the initial (pre-heat) refractive index modulation increases with low intensity exposures for slanted holograms. Since the edge–referenced hologram is essentially in the regime between transmission and reflection holograms, one might conclude that the refractive index modulation was intensity dependent. Laboratory results suggest that this is true. The maximum intensity for high modulation for the edge–referenced holograms is approximately 2 mW/cm^2. The results compare well with Du Pont's findings for transmission holograms [3.10, 3.13]. Lower efficiencies were noticed for higher intensities, while lower intensities were not very different in diffraction efficiency.

In many of the Du Pont photopolymers, the refractive index modulation generally saturates with increasing exposure [3.14]. Thus, a hologram can be exposed for long periods of time to ensure complete exposure. However, with some film thicknesses for transmission holograms, the refractive index modulation may be such that overmodulation occurs and the final diffraction efficiency may rise and fall again [3.19]. In these cases, the refractive index modulation still saturates, although the index modulation and film thickness combination may cause a decrease in diffraction efficiency. The total diffraction efficiency modulation depends on the thickness of the film. The film thickness is sometimes adjusted so that maximum diffraction efficiency will result for a particular fringe spacing.

3.6.2 Importance of Refractive Index in Edge–Illuminated Holograms

As explained by many authors, the index of refraction for films and substrates is very important in creating an edge–lit hologram [3.5–3.7, 3.36–3.38]. A large difference in refractive index between the substrate and the film results in spurious reflections which can diminish the quality of the hologram–if it can be recorded at all. In order to understand the possible reflections involved, a general model (Fig. 3.11) is applied to the various recording conditions (Table 3.2).

For the "block recording" silver halide configuration (the silver halide plate is index matched to a glass block), refractive index values of $n_1 = 1.52$, $n_2 = 1.65$, $n_3 = 1.497$, and $n_4 = 1.52$ [3.7] were used in the Fresnel equations (Eq. 3.1). For the "tank recording" method, refractive index values of $n_1 = 1.497$, $n_2 = 1.65$, $n_3 = 1.52$, and $n_4 = 1.497$ were used in the Fresnel equations. For the photopolymers HRF red 352 and 700x071, the values for Mylar ($n_D = 1.65$) and air ($n_D = 1$) were used for n_3 and n_4 respectively. The coefficients T0–T3 and R0–R3 in Fig. 3.11 and Table 3.2 are from the equations for the Fresnel coefficients below for S polarization [3.39]:

$$R_\perp = r_\perp^2 \text{ with } r_\perp = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}, \text{ and} \tag{3.1}$$

$$T_\perp = \left(\frac{n_t \cos\theta_t}{n_i \cos\theta_i}\right) t_\perp^2 \text{ with } t_\perp = \frac{2 \sin\theta_t \cos\theta_t}{\sin(\theta_t + \theta_i)}. \tag{3.2}$$

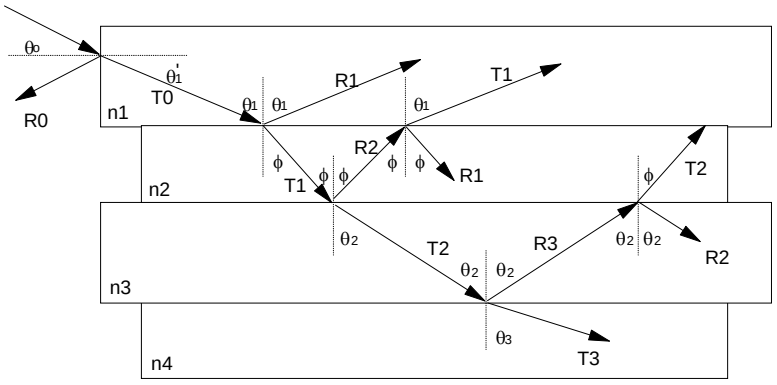


Figure 3.11 The ray diagram used to calculated the reflectances and the transmittances at the boundaries between layers of different indices of refraction.

Substrate n_D	Film n_2*	θ_1	R1	T1	R2	T2	R3	T3	ϕ	Notes
BK7 1.52	Ag Halide 1.64	80°	.19	.81	.76	.24	.51	.49	66°	BLOCK
Xylene 1.497	Ag Halide 1.64	87°	.60	.40	.17	.83	.27	.73	66°	TANK
BK7 1.52	red 352 1.504	87°	TIR	---	---	---	---	---	---	
BK10 1.4981	red 352 1.504	87°	.09	.91	.39	.61	TIR	---	84°	
Acrylic 1.491	red 352 1.504	87°	.18	.82	.29	.71	TIR	---	82°	
FK5 1.4876	red 352 1.504	87°	.22	.78	.25	.75	TIR	---	81°	
Fused Silica 1.4588	red 352 1.504	87°	.39	.61	.12	.88	TIR	---	76°	
BK7 1.52	700x071 1.4787	87°	TIR	---	---	---	---	---	---	
BK10 1.4981	700x071 1.4787	87°	TIR	---	---	---	---	---	---	
Acrylic 1.491	700x071 1.4787	87°	TIR	---	---	---	---	---	---	SIM
FK5 1.4876	700x071 1.4787	87°	TIR	---	---	---	---	---	---	SIM
Fused Silica 1.4588	700x071 1.4787	87°	.24	.76	.25	.76	TIR	---	80°	

* The refractive index range for red–sensitive 352 is from $n_D = 1.504$ to approximately 1.53. The index range for 700x071 is from $n_D = 1.4787$ to approximately 1.51. The Self–induced Index Matching (SIM) refers to coupling while the index of the film increases (§4.4.2).

Table 3.2 The values for the reflection and transmission coefficients for various recording situations. The coefficients for T and R refer to the transmittance and reflectance at the previous interface.

The data in Table 3.2 illustrates the relative intensities of the many reflections involved with recording edge–referenced holograms. With the silver halide block method (where the reference is coupled through the edge of a substrate), the R2 reflection occurring inside the film is 76% of the light first entering the film. This corresponds to multiple gratings being formed in the film and is unacceptable for recording efficient holograms.

Using the silver halide tank method one can record an edge–lit hologram. However, a large percentage of the incident reference light (R1=60%) is reflected away in recording as well as replay because of the high index of the recording film relative to the substrate. This can create problems in terms of light efficiency and laser powers during *recording and replay*.

With the refractive indices of the Du Pont photopolymers, simpler and more efficient recording can be made by using the block method with special substrates. The low index of the photopolymers relative to the substrates allows more of the reference beam to enter the film in *recording* and *replay*. The Total Internal Reflections (TIR) after the film are not as much of a problem as one might first suspect. Because of the low index of the film, much steeper reference angles are allowed within the film ($\phi \approx 80°$) so the inherent absorption diminishes the intensity of the wave as mentioned in § 2.2.2. This is actually similar in the photopolymer case to that of the silver halide hologram whose angle inside the film is essentially limited to $\phi \approx 66°$. For a 514 nm photopolymer recording, the percentage of the reference beam intensity at the reflection surface due to absorption is 8% (with $\alpha = 0.011 \mu m^{-1}$ from Fig. 2.3 and Eq. 2.2 for light incident to HRF 700x071 at 80°). With Millimask silver halide plates, the reflection is 7% of the incident intensity in the film (with $\alpha = 0.077 \mu m^{-1}$ and Eq. 2.2 for light incident to 7μm Millimask at 66°). Here, the high absorption in the thin emulsion has a low intensity reflection just as the thick photopolymer with low absorption.

The data in Table 3.2 illustrates the reflectance and transmittance for various practical recording substrates and films. Ideally, one would like a perfect index match between the photopolymer and the substrate. The difference between the refractive indices of

the film and the substrate determine the transmittance into the film. For very steep angles such as those encountered in edge–referenced holograms, the index difference is very crucial as illustrated in Fig. 3.12.

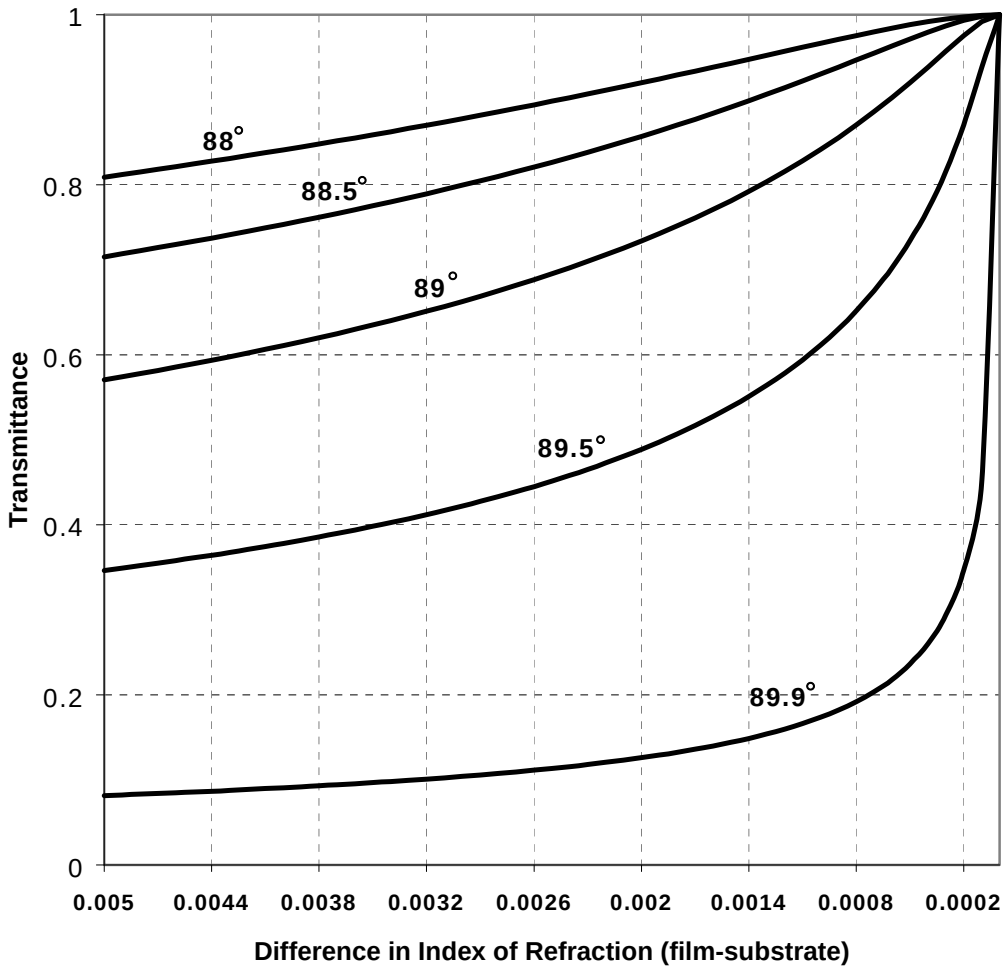


Figure 3.12 The transmittance across an interface with the index difference approaching zero for various steep reference angles. The graphs are calculated based on Eq. 3.2.

From Fig. 3.12, one can see that as the reference angle approaches 90°, the substrate refractive index must match the refractive index of the film very closely in order for the reference beam to penetrate into the film. Therefore, the refractive indices of the substrates and the films must be measured very accurately.

Another view of the transmission into the film can be made if one considers a constant index for the substrate and looks at the transmission for various film indices as in Fig. 3.13.

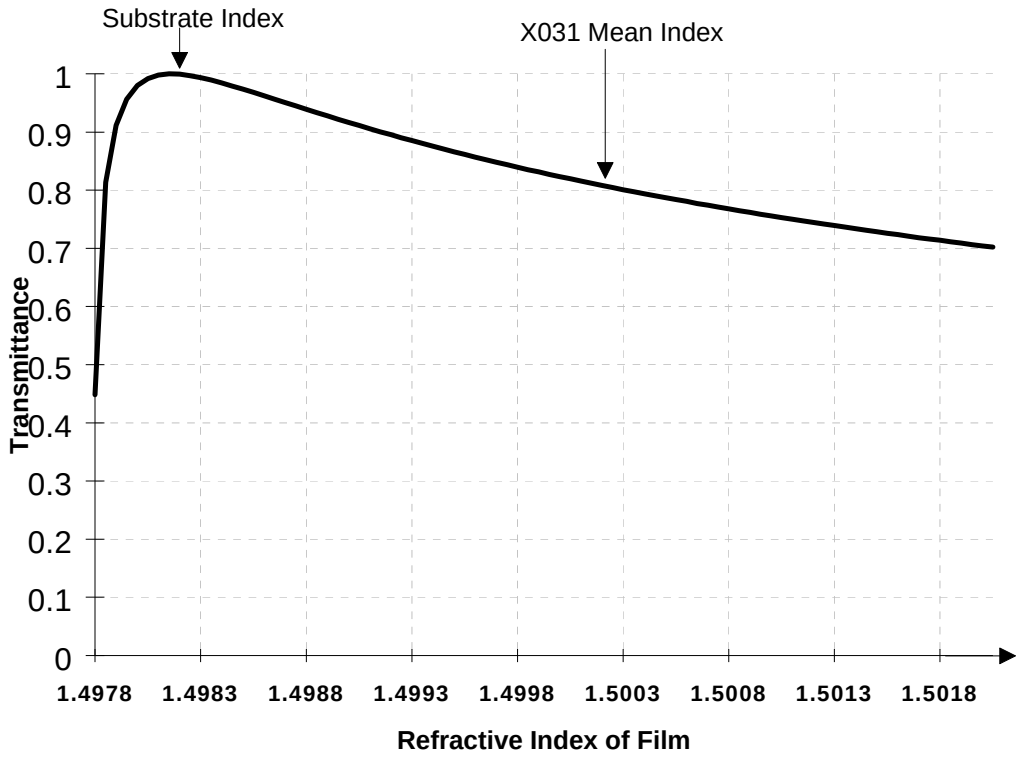


Figure 3.13 Showing the percent transmission into films with different indices of refraction (or a film whose refractive index changes) with a constant refractive index substrate. The values used were n_D(substrate) = 1.49811 and θ = 88.7° and HRF 700X031(lab coating) n_D = 1.5002.

One can look at Fig. 3.13 to determine which substrate or polymer to use. Given a particular substrate, one must consider the transmission into the film as dynamic because the index of the film in the bright areas of the interference pattern increases and in the dark areas, decreases. Therefore, the maximum transmission into the film would be at a refractive index equal to or just higher than the substrate. If the index was less than the substrate, the transmission falls off quickly as one approaches total internal reflection. One could imagine an effective gain in modulation if the indices of the substrate and the film precisely match. With equal indices of refraction, the areas of low intensity during recording would decrease in refractive index and even *less* light would reach the area from the Fresnel reflection (Fig. 3.13) This would

effectively increase the contrast of the interference pattern. This effect would also increase the non–linearity of recording.

One must be careful in measuring the refractive index using a device such as an Abbe refractometer as it can only measure the average refractive index. Therefore, since some areas of the emulsion may have a lower index (coating non–uniformity) which could create a localized TIR, it would be advisable to choose a photopolymer with an average starting index just above the substrate for a safe recording. For the same reasons, one might choose a recording substrate which is just slightly below that of the polymer.

3.6.3 Measuring the Photopolymer Refractive Index

A standard device for measuring the index of refraction for optical solids is the Abbe refractometer. This device works on the properties of total internal reflection from a prism to measure the precise angle (thus refractive index) where TIR occurs (Fig. 3.14). The angle is determined by aligning cross–hairs through a viewfinder on the dark band which signifies the TIR.

From Snell's law, one can see that the reference angle reaches 90° in the second medium at

$$\theta_c = Sin^{-1}\left(\frac{n_2}{n_1}\right), \tag{3.3}$$

thus signifying the transition between partial transmission and total reflection. The variables on which the Abbe refractometer depend are wavelength, temperature and precision of angular measurement. The temperature for the measurements remained constant and the angular measurement was very good because of the quality of the refractometer. The wavelength used was the sodium D–line. Attempts were made at using the precise laser wavelength of 514 nm, however, the speckle introduced made reading the angular measurement very difficult and imprecise.

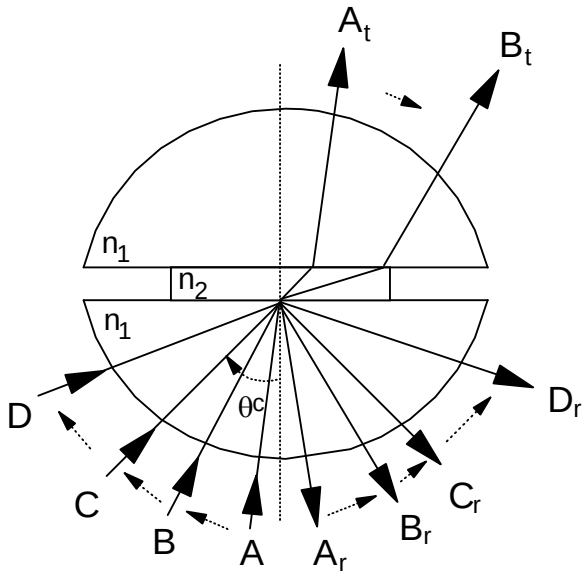


Figure 3.14 A diagram showing the principle of total internal reflection. Subscripts t and r represent the transmitted and reflected components respectively.

An inherent difficulty encountered when measuring optically sensitive films is that often the refractive index is desired at a wavelength at which the film is sensitive, thus the measurement is exposing the film. In the case of the Du Pont (real–time) photopolymers this was indicated by bands moving away from the main dark line of the TIR during measurement. The initial index of refraction was easily obtainable in most cases with quick measurements. In many cases, however, reading the bulk index of refraction of the film was difficult because of poorly coated or mixed 'experimental lab coatings' of the photopolymer. As the concentrations of the components varied, the index varied with it, thus there was more than one dark line over the viewing area signifying the TIR's. With a non–uniform film, measuring a smaller sample area (such as in the case of a ellipsometer) could yield false readings. Most of the non–uniformity was corrected on the commercial (non–experimental) coatings provided by Du Pont. The refractive index modulation of the photopolymers was very difficult to measure using the Abbe as the index modulation would occur in a small region, not the large sampling region. Faint lines were visible on the Abbe corresponding to higher refractive indices. Since the refractive index changes continuously across the pitch of the grating, many lines corresponding to specific indices of refraction are faintly visible.

3.6.4 Film Packaging

The packaging of the film is very critical in the case of edge–illuminated holograms. The Du Pont photopolymer film layer is usually 20 μm or 25 μm thick, although it can be coated from about 7 μm to 90 μm. The base, or film carrier, is a polyester (Mylar®) layer usually 50 μm thick. The removable cover layer is usually 25 μm thick and can be made of Mylar or poly–vinyl chloride (PVC). Thus, the film is usually in the layers indicated in Fig. 3.15.

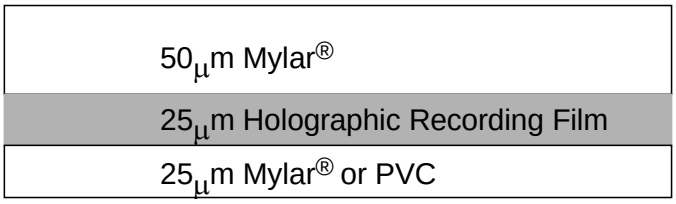


Figure 3.15 The standard packaging of Du Pont Holographic Recording Film.

The Mylar material, otherwise known as polyethylene terephthalate, is stress tri–refringent and thus effectively has three different indices of refraction [3.41]. Within the plane of support of the film, the index of refraction is n_α=1.50. Recording with light traveling in this direction is not feasible. In the direction normal to the surface, the refractive indices are n_γ=1.66 and n_β=1.65, depending on the polarization. However, all of these indices may change depending on the exact stress introduced during production. If the polarization of the incident laser is not aligned precisely with the either the γ–axis or the β–axis of the Mylar, then optical retardance of the polarization results. If a reflection hologram is made, then only one beam will pass through the Mylar before reaching the film (after the cover sheet is removed and it is laminated onto a substrate). Thus, one beam will have a different state of polarization than the other (unless the axis is aligned), resulting in a less efficient hologram. The alignment of the optical axis of the Mylar is not always constant on the film and is difficult to account for in advance. While in some situations (such as a vertically arranged single beam reflection hologram) the sample may be rotated in–situ to minimize the retardation, the fogging introduced by this alignment is generally unacceptable. For normal (non–edge referenced) holograms and most of the edge–

referenced ones, the decrease in efficiency due to the tri–refringence is very small based on observations.

Another disadvantage of using a Mylar carrier is its high refractive index. Inherent with the high refractive index is the stray Fresnel reflection generated at the polymer–Mylar interface. This reflection can be 12–39% in the examples presented in Table 3.2 (reflection R2). If the index were lower, this reflection would be minimized. If the Mylar were to be removed, the reference beam would reflect back at the air surface due to a TIR regardless (unless a tank was used). Unfortunately, it is very difficult to remove the Mylar carrier layer as the film usually adheres to it. It may be removed in certain circumstances by bonding the film to another substrate using an epoxy, however, the film must be covered during recording as exposure to oxygen will quench the reaction and result in a lower modulation [3.13]. Thus, laminating the film to a substrate and removing the Mylar is unacceptable because the film would be exposed to air. In many cases of the edge–referenced hologram, the Mylar is left on the film and most of the reference beam is absorbed before reaching the Mylar.

3.6.5 Photopolymer Preparation

With the standard form of the photopolymer (Fig. 3.15), the Mylar cover sheet is generally removed and the film is laminated onto a substrate (by using a rolling laminator for example). The initial tackiness caused by the solvents in the film allows for good lamination. Dust is a major issue and is more important in the edge–illuminated geometry. If there is dust present, there will be an air bubble around the dust in the film. In a face–referenced hologram, light would still reach the film around the bubble (although there may be a small reflection generated). However, in an edge–referenced hologram, the reference beam will not reach the film in that area at all because of a TIR from the substrate–air interface. This results in a dark spot on the final hologram. The film can be obtained in a liquid form which could be spin–coated onto a substrate. With this technique, the thickness is difficult to calibrate with many types of films and solvent concentrations over time. Also, the film would have

to be covered with a non–permeable medium that would not allow oxygen into the film. Since respectable efficiencies were obtained with the film coated on Mylar, spin coating was not investigated.

In many cases, the desired final substrate of the hologram is not the recording substrate. For most applications, the desirable final substrate would be of a very close index of refraction, cheap, and lightweight. Fortunately, acrylic plastic fits nicely into these categories. However, many attempts have been made at recording through an acrylic substrate with generally poor results. The optical quality of acrylic is generally not sufficient for edge–referenced holograms. The quality of the substrate is much more crucial with edge–referenced holograms because at a steep angle the reference beam effectively travels through the full width of the acrylic and any minor striae, stress, etc. will show up as non–uniform reference coverage. Another difficulty is the instability of the acrylic. As it is a lightweight, flexible material, it is much more subject to movement and bending than optical glass during a holographic exposure (which reduces the final diffraction efficiency).

When recording with such substrates as BK10, which is expensive and heavy, the film ideally needs to be transferred to acrylic. Transferring the film once it has been recorded is a very complicated issue, and thus three methods have been studied: dry transfer, HRF 150 transfer, and epoxy transfer.

3.6.5.1 Dry Transfer

In the dry transfer process no additional components are added when transferring the film to another substrate. In its original form, the photopolymer is very tacky, allowing good adhesion to the substrate. As the photopolymer is exposed, it gradually loses its tackiness. The UV cure after the exposure dramatically reduces the ability of the film to bond with another substrate. With a saturated holographic exposure or after the UV exposure, the film is not sufficiently tacky enough to re–laminate onto another substrate (the bond is temporary and air bubbles are trapped during the lamination).

Since the photopolymer exposure will saturate, it is possible to expose for the minimum time or perhaps slightly underexpose to allow sufficient tackiness (before UV) for bonding to a different substrate. Dust and stress are important factors in re–lamination. A second lamination is always likely to introduce more blemishes. Poor transfer can also result from stress induced on the film because of the handling difficulties. There is usually less film to handle since the excess usually needs to be trimmed off just before initial exposure. The stress on the film can result in a non–uniform grating (sometimes as if the fringes are stretched) or wrinkles and air bubbles. Dry transfer requires a tedious amount of care and needs a precise type of machine handling in a clean–room environment for successful transfer. Limited success has been achieved by hand and seems to only be useful for demonstrative purposes.

3.6.5.2 HRF 150 Transfer

Most commercially available adhesive films have a very low index of refraction (usually around n_D=1.47 for the glue layer). This low index will not permit steep angles for the illumination of the hologram because of a TIR at the substrate–glue layer. In the search for an optical adhesive for transferring the film to another substrate, it was noticed that the HRF 150 film (without a visible sensitizing dye) was

very near the required index of refraction, was sufficiently tacky, and would provide a permanent bond. Since the film itself is an acrylate photopolymer, its refractive index is quite similar to that of the recording photopolymer and the index of the acrylic substrate as well. However, since the unexposed HRF 150 film has monomers within it, it can swell a pre–recorded hologram just as the Du Pont Color Tuning Film [3.14]. First tests showed non–uniform swelling effects associated with using the HRF 150 as a laminate. Therefore, a process was developed which would enable transfer without significant wavelength shift. This process is described in Fig. 3.16.

1)	Laminate HRF 150 onto polished acrylic
2)	Trim
3)	Heat combination under IR lamps for 90 seconds (at 80°–100° C) to promote polymer adhesion to acrylic
4)	Let cool to room temperature
5)	Laminate HRF 700x071 onto recording substrate
6)	Expose
7)	UV cure 60 seconds (≈0.5 Joules) (need at least several hundred millijoules)
8)	Remove Mylar from HRF 150
9)	Quickly laminate HRF 700x071 onto HRF 150
10)	Wait 60 seconds for chemical bond
11)	UV cure for 3 seconds (app. 25 mJ)
12)	Trim

Figure 3.16 The steps for a workable transfer from a recording substrate to a new substrate using HRF 150 as a transfer film.

The HRF 150 will naturally bond to the acrylic, however, the combination is heated to expedite the bonding. In step 10 of Fig. 3.15, if the wait period is to long, then the solvents in the HRF 150 will react and effectively destroy or swell the fringes in the hologram (in the HRF 700x071). If the wait period is not long enough, then the hologram will not bond to the HRF 150. Also, if the UV cure in step 11 is to long, then the HRF 150 will become cloudy and diffuse because of non–uniform density variations possibly created from coating non–uniformities within the thick layer. The results with the HRF 150 transfer are generally acceptable, however, very careful control of the heating, UV, and dust is necessary. Since this method requires an

additional lamination layer (step 1), dust is more likely to contaminate the sample. Another disadvantage of this technique is that heating after the final lamination (after step 12) is not possible as this swells the fringes. This could inherently limit the final available refractive index modulation.

3.6.5.3 Epoxy Transfer

After considerable searching, a suitable one–step UV cure epoxy was found with an appropriate index of refraction [3.42]. It is very crucial that the solvents in the epoxy do not affect the film. After exposing and UV curing the sample, the epoxy was placed on the new substrate and the sample was rolled onto the epoxy. Then the combination was immediately UV cured according to the specifications of the epoxy. No adverse effects have been noticed after bonding. The results using this method appear to be very good and this method of transfer has been adopted when it is desired to change substrates.

3.6.6 Real–time Effects of Photopolymers

Holograms formed in the family of visibly sensitized Du Pont HRF photopolymers develop during the recording, and thus are susceptible to real–time effects inherent in the recording. Two of these effects, self–fluorescent fogging and real–time holographic diffraction, have been observed in the lab. In theory, these effects would reduce the diffraction efficiency of the resulting hologram.

3.6.6.1 Fluorescence Fogging During Recording

The sensitizing dyes used in general holography are necessary to make the recording medium sensitive to desirable recording laser wavelengths. The sensitizing dyes for many of the Du Pont photopolymers fluoresce at wavelengths at which they are sensitized. Thus, during the recording, the light from the fluorescence is fogging the recording material and reducing its dynamic range. In theory, this is also independent of the exposure conditions because the fluorescence is proportional to the intensity. Thus, the dark areas of a holographic interference pattern would have a fog level due to the fluorescence.

To verify the self–fogging characteristics, the fluorescence spectrum of HRF 705 panchromatic film was compared with the absorption spectrum. The overlap area is a fog level as shown in Fig. 3.17. To measure the fluorescence, the film was illuminated with 514 nm laser light and the fluorescence was focused into a spectrometer. The absorption was measured with a spectrophotometer.

The fluorescence from the photopolymer is not as dramatically visible in traditional recording as it is in the edge–referenced geometry. In normal holograms, only a small percentage of the incident light is actually absorbed by the hologram relative to the absorption as the reference beam approaches 90° incidence. Thus, at very steep angles, the reference beam is almost completely absorbed. This fluorescence is easily visible from either face of the hologram because none of the reference beam is escaping from the face. During the recording, the absorption due to the dye reduces because of the dye molecules breaking down (or bleaching). Fortunately, the intensity of the fluorescence does not appear to be sufficient enough to dramatically reduce the diffraction efficiency. Reasonably high efficiencies have still been obtained, however, despite the lack of a suitable non–fogging or non–fluorescent dye. The exact effects of the fogging on the film have not been determined conclusively.

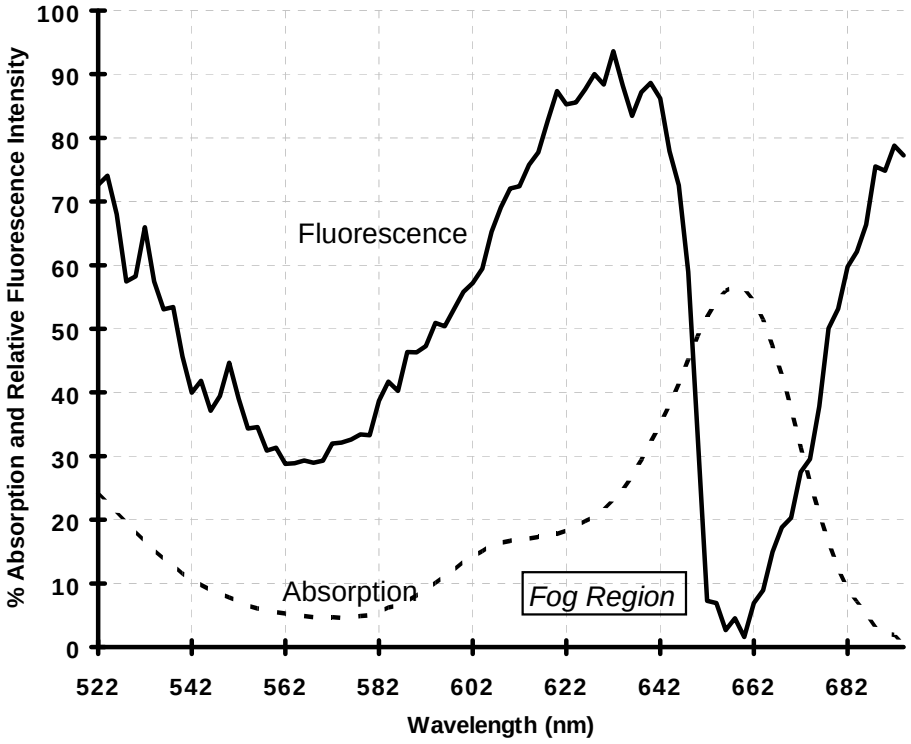


Figure 3.17 The fluorescence and absorption spectrums of HRF 705 panchromatic film. The overlap area is where fogging occurs.

3.6.6.2 Real–time Diffraction During Recording

Inherent with all real–time self–developing materials is the dynamic problem of diffraction of the signal and reference beams due to a grating forming during the recording. As photopolymers develop a refractive index modulation during recording, they are very susceptible to these effects. A dynamic theory of recording was first proposed by Nonomiya [3.43], and later modified by Magnusson *et al.* [3.44]. It was then applied to the photorefractive effect by Moharam *et al.* [3.45]. This theory is should be generally applicable to photopolymers with considerable adaptation.

Experimental, real–time observations of diffraction in photopolymers have been noted, however they are not often modeled [3.47]. Rush *et al.* [3.46] tried a different

model for the dynamic effects on holograms recorded in Du Pont photopolymers with limited success. Their theory only models correctly the first portion of exposure (not the entire recording period), and they defined crucial photopolymer variables arbitrarily to enable their results to fit the theory. Their photopolymer exposure variables were related to the speed of the film, representing an offset exposure related to the induction delay, and related the grating shrinkage to the index change.

These variables would obviously have a major effect on how the real–time diffraction affects the final hologram, however, they can not be simply chosen so that a model fits an experimental graph. They must be measured independently and related appropriately. Perhaps uncertainty in these values, and the extreme complication of the dynamic models explains the limited papers available in this area. An ideal model would perhaps contain some the coefficients suggested by Rush *et al.* and use the rigorous coupled wave adaptation for 3–dimensional, slanted, volume, phase holograms [3.48] with the dynamic model modifications [3.46].

Most of the practical recording complications with edge-lit holograms can be overcome to record holograms with high diffraction efficiencies. By carefully choosing and arranging the recording setup (including the recording film, recording and replay substrate, and sample holder) one can ensure that the edge-lit holograms produced have high diffraction efficiencies and can be incorporated into various optical systems.

REFERENCES

3.1 Collier, R., Burckhardt, C., and Lin, L., *Optical Holography,* Academic Press, New York, 1971

3.2 Syms, R., *Practical Volume Holography,* Clarendon Press, London, 1990

3.3 Optics for Research, Caldwell, NJ, USA, Apodizer, Catalog No. GC–25

3.4 Jiang, Wu, "Application of a laser beam profile reshaper to enhance performance of holographic projection systems," Doctoral thesis, University of Alabama, Birmingham, 1993

3.5 Henrion, M., "Diffraction and exposure characteristics of the edgelit hologram," Maters thesis, MIT, 1995

3.6 Birner, S., "Steep reference angle holography: Analysis and Applications," Masters thesis, MIT, 1989

3.7 Benton, S.A., Birner, S., and Shirakura, A., "Edge–lit rainbow holograms," Practical Holography IV, SPIE vol. 1212, pp 149–157
3.8 Coleman, Zane, Lab Report, Physics 4900, Georgia Institute of Technology, 1992
3.9 Armstrong, M. and Tipton, D., “Machine for continuous hologram production,” SPIE vol. 1914, pp 134–154
3.10 Gambogi, W., Gerstadt, W., Mackara, S., and Weber, A., “Holographic transmission elements using improved photopolymer films,” SPIE, International Symposium on Optical Science and Engineering, San Deigo, CA, (1991)
3.11 Trout, J., Gambogi, W., and Stevenson, S., “Photopolymer Materials for Color Holography,” Applications of Optical Holography, SPIE vol. 2577, pp 94–105, June, (1995)
3.12 Booth, B., “Recent photopolymer development for holographic and laser recording,” Holosphere, vol. 5, No. 5, pp 1–8, May, (1976)
3.13 Booth, B., “Photopolymer Material for Holography,” Applied Optics, Vol. 14, No. 3, pp 593–601, March (1975)
3.14 Gambogi, W., Smothers, W., and Steijn, K., “Color holography using DuPont holographic recording films,” SPIE vol. 2405, San Jose, CA, (1995)
3.15 Zager, S. and Weber, A., “Display holograms in Du Pont’s OmniDex films,” SPIE vol. 1461, San Jose, CA, March (1991)
3.16 Gambogi, W., Weber, A. and Trout, T.J., “Advances and applications of DuPont Holographic Photopolymers,” SPIE vol. 2043, Quebec, Canada, Aug. (1993)
3.17 Krongauz, V., Schmelzer and Yohannan, R., “Kinetics of anisotropic photopolymerization in polymer matrix,” Polymer, Vol. 32, No. 9, pp 1654–1662, (1991)
3.18 Smothers, W., Monroe, B., Weber, A. and Keys, D., “Photopolymers for holography,” SPIE vol. 1212, Practical Holography IV, pp 20–29, (1990)
3.19 Weber, A., Smothers, W., Trout, T.J. and Mickish, D., “Hologram recording in Du Pont’s new photopolymer materials,” SPIE vol. 1212, Practical Holography IV, pp 30–39, (1990)
3.20 Wopschall, R. and Pampalone, T., “Dry photopolymer film for recording holograms,” Applied Optics, Vol. 11, No. 9, pp 2096–2097, Sept. (1972)
3.21 Weber, A., “DuPont’s Omnidex holgoraphic recording materials,” Opto 7/Holographics ‘90 Conference, Nuremberg, Germany, Oct. 16–18, (1990)
3.22 Krongauz, V. and Yohannan, R., “Polymerization kinetics in photopolymer films,” SPIE vol. 1213, Photopolymer Device Physics, Chemistry, and Applications, pp 174–183, (1990)
3.23 Krongauz, V. and Yohannan, R., “Kinetics of Photopolymerization in polymer matrix,” Mol. Cryst. Liq., vol. 183, pp 495–503, (1990)
3.24 Lessard, R. and Manivannan, G., editors, *Selected papers on photopolymers:physics, chemistry, and applications*, SPIE Optical Engineering Press MS 114, 1995
3.25 Jenney, J., "Holographic recording with photopolymers," JOSA Vol. 60, No. 9, pp 1155–1161 (1970)
3.26 Ingwall, R., and Fielding, "Hologram recording with new photopolymer system," Optical Engineering, Vol. 24, No. 5, pp 808–811 (1985)

3.27 Rush, D., Vance, J., Goldhar, J. and Burdge, G., “High resolution reflection holography using DuPont photopolymer holographic recording film,” SPIE vol. 2688, Photonics West, (1996)
3.28 Rhee, U., Caulfield, H.J., Vikram, C. and Shamir, J., “Dynamics of hologram recording in DuPont photopolymer,” Applied Optics, Vol, 34, No. 5, pp 846–853, Feb. (1995)
3.29 Colburn, W. and Haines, K., “Volume Hologram Formation in Photopolymer Materials,” Applied Optics, Vol. 10, No. 7, pp 1636–1641, July, (1971)
3.30 Curtis, K. and Psaltis, D., “Characterization of the DuPont photopolymer for three–dimensional holographic storage,” Applied Optics, Vol. 33, No. 23, pp 5396–5399, Aug. (1994)
3.31 Rhee, U., Caulfield, H.J., Vikram, C., Shamir, J. and Mirsalehi, M., “Characteristics of the DuPont photopolymer for angularly multiplexed page–oriented holographic memories,” Optical Engineering, Vol. 32, No. 8, pp 1839–1847, Aug. (1993)
3.32 Kloosterboer, J., *Advances in Polymer Science*, Springer–Verlag, pp 4–61, (1988)
3.33 Waldman, D.A., Ingwall, R.T.,et. al., "Cationic ring–opening photopolymerization methods for volume hologram recording," SPIE Vol. 2689, pp 127–141
3.34 Simova, E. and Kavehrad, M.,"Light shaping diffusers for indoor wireless infrared communications via a holographic approach," SPIE, Vol. 2689, pp 284–291, (1996)
3.35 Bjelkhagen, H.I.,*Silver–Halide Recording Materials*, Springer–Verlag, Berlin, 1993
3.36 Upatnieks, J., "Compact holographic sight," SPIE Vol. 883, Holographic Optics:Design and Applications, pp 171–176 (1988)
3.37 Upatnieks, J., "Compact hologram displays and method of making compact hologram," Pat. No. 5,151,800, Sept. 1992
3.38 Upatnieks, J., "Edge–illuminated holograms," Applied Optics, Vol. 31, No. 8, pp 1048–1052, (1992)
3.39 Hecht, E., and Zajac, A., *Optics*, Addison–Wesley, Massachusetts, 1974
3.40 Barnett, C.A., "Modern Lithographic techniques applied to stereographic imaging,"Doctoral thesis, Loughborough University, 1993
3.41 James, T. H., *Theory of the photographic process*, Fourth Edition, McMillan Publishing Co. Inc., New York, p580, (1966)
3.42 Photobond LS–706, UV epoxy from Stonehill Chemicals, Ottawa, Ontario, CA
3.43 Ninomiya, Y., JOSA , vol. 63, pp 1124, (1973)
3.44 Magnusson, R. and Gaylord, T., “Use of dynamic theory to describe experimental results from volume holography,” Journal of Applied Physics, vol. 47, pp 190–199, (1976)
3.45 Moharam, M. and Young, L., ”Hologram writing by the photorefraction effect,” Journal of Applied Physics, vol. 48, pp 3230–3236, (1977)
3.46 Rush, D., Vance, J., Goldhar, J. and Burdge, G., “High resolution reflection holography using DuPont photopolymer holographic recording film,” SPIE vol. 2688, Photonics West, (1996)
3.47 Rhee, U., Caulfield, H.J., Vikram, C. and Shamir, J., “Dynamics of hologram recording in Du Pont photopolymer,” Applied Optics, Vol, 34, No. 5, pp 846–853, Feb. (1995)

3.48 Moharam, M. and Gaylord, T., “Three–dimensional vector coupled–wave analysis of planar–grating diffraction,”, JOSA A, Vol 73, No. 9, pp 1105–1112, Sept. (1983)

Chapter 4

Recording Methods

4.1 Introduction

There are two main classifications for holograms: reflection and transmission. Waveguide holograms can be classified as either of these, however evanescent wave holograms must be classed differently because they are created by the interference of at least one inhomogeneous wave. Each of these types of holograms has important factors to be examined for recording. Since there is more than one way to record the same hologram, all of the possible recording methods must be reviewed. First, the traditional transmission hologram will be examined.

4.2 Transmission Holograms

Transmission holograms are normally thought of as those which are recorded with signal and reference beam arriving in the recording medium from the same side of the substrate. For replay the light is incident from one side and the main diffracted light is transmitted through to the other side. With edge–referenced holograms various factors for the recording and replay beams must be considered. Edge–referenced transmission holograms have many additional restrictions for recording than the traditional face–lit transmission holograms.

4.2.1 Transmission Recording

A typical recording setup for an edge–referenced transmission hologram is shown in Fig. 4.1.

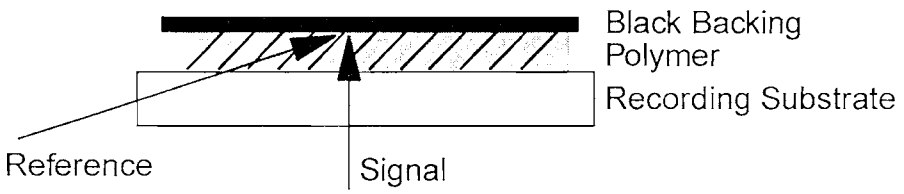


Figure 4.1 The recording geometry for an edge–referenced transmission hologram in photopolymer.

In a transmission hologram, it is normal procedure to black back the film so that the reference and signal reflections do not create spurious gratings. Just as mentioned in §2.2.2, the reference beam intensity is very small when it reaches the full thickness of the film for very steep angles. Hence, this reflection is not very detrimental. However, the signal beam reflection does create a significant stray grating. In order to successfully remove these reflections, an appropriate absorbing light dump must be used. Removing the Mylar is not necessary for recording as the light from the signal

beam may be absorbed after traveling through it. However, as will be seen in many replay considerations, it should optimally be removed.

4.2.2 Transmission Replay

Typically, there are two types of transmission replay geometries for a hologram. In terms of normal holographic images, these are referred to as the 'real' or 'virtual' images. For references to Holographic Optical Elements (HOE's) they will be referred to here as conjugate or non–conjugate replay corresponding to the type of illumination required. As detailed in Fig. 4.2 (a) and (b), the two typical replay geometries are invoked with (a) corresponding to the non–conjugate (virtual) and (b) corresponding to the conjugate (real) replay geometries for a transmission hologram recorded as in Fig. 4.1.

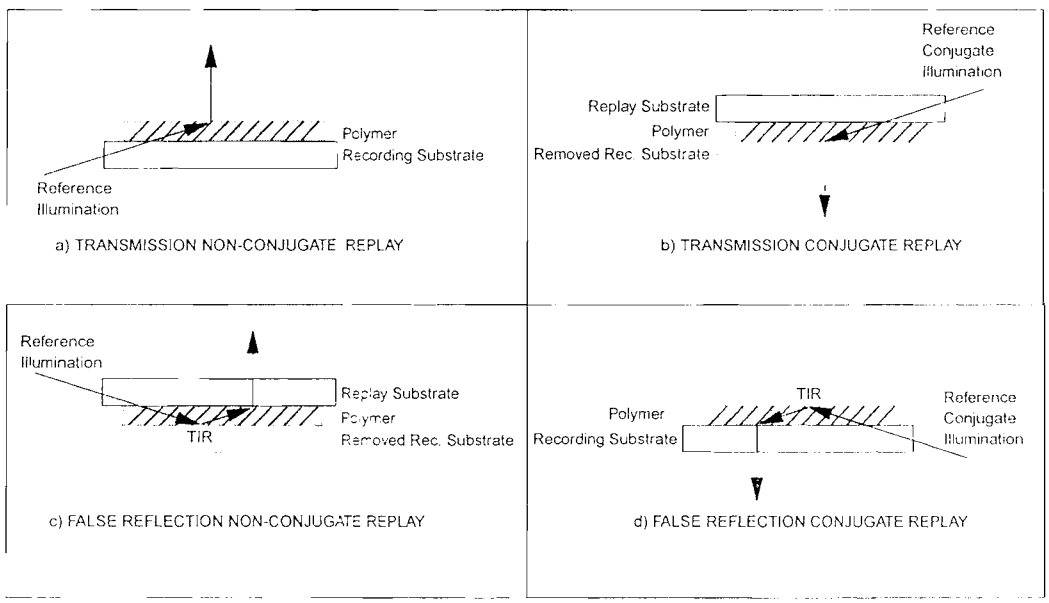


Figure 4.2 The replay possibilities of an edge–referenced transmission hologram. (a) and (b) are the traditional conjugate and non–conjugate replay, respectively. (c) and (d) are the new 'false reflection' non–conjugate and conjugate replay, respectively, which incorporate a TIR within the film.

4.2.3 False Reflection Replay

Two unique additional replay geometries are possible with the edge–referenced hologram. As a result of the steep angles, a TIR can be generated within the polymer which allows a transmission hologram to be replayed in a seemingly reflection geometry. This geometry seems to be unique to edge–lit holograms because the TIR actually occurs within the film as opposed to reflecting from another element. While Fresnel reflections can generate these replay situations for angles not as steep, the intensities are not normally near the 100% as in a TIR.

In order for the hologram to be replayed through the original recording substrate as in Fig. 4.2 (a) and (d), the black light dump must be removed. If the light dump were not removed in (a), then the light would not exit, and in (d), the TIR would not occur. No additional substrate lamination is necessary. The Mylar may be removed if desired, however, the signal reflections caused by the presence of Mylar during recording are minimal. Successful results have been achieved for these two types of replay geometries.

For Fig. 4.2 (b) and (c), the recording substrate may be left in contact with the film with no effect in (b) and a TIR shifted to the substrate–air boundary in (c). In (b) and (c) the Mylar should be removed as the reflection from a replay substrate–Mylar boundary is 63% of the incident intensity which would result in inefficient replay. The Mylar may be removed after recording and then the replay substrate epoxied to the film.

4.2.4 Removing the Mylar

A major difficulty with transmission recordings is the packaging of the Du Pont photopolymers. If the film is laminated onto a *glass* recording substrate, then before recording it is extremely difficult to remove the Mylar from the film without again

removing the film. A new technique for removing the Mylar has been developed. This involves lamination of the film onto an *acrylic* substrate. Once on the substrate, the film should be left at room temperature for a day or heated for about 10 minutes at 60° C and left to cool. In both methods, it is necessary that the sample remain unexposed to light. If the sample is left any longer in the oven, then the acrylic substrate is likely to soften and distort. The chemical permanent bond which forms between the photopolymer film and the acrylic appears to be related to the solvents within the film. The photopolymer does not appear to have any obvious reduced performance after this transfer.

Thus, in order to remove the Mylar, the film must be laminated to a clear acrylic substrate for recording. Unfortunately, the difficulties of recording through an acrylic substrate as mentioned in §3.6.5 prevented definitive results for edge–referenced transmission hologram replayed as in Fig. 4.2 (b) and (c). Future possibilities may arise if a silane (a chemical used to promote bonding to a glass) is added to the film composition. Du Pont offer a film, HRF 750, with higher adhesion to glass, although this film had a refractive index to low for edge–referenced recording when used with the original glass substrates.

4.3 Reflection Holograms

Reflection holograms are those recorded with the incident light beams arriving from opposite sides of the film. On efficient replay, most of the diffracted light leaves the film on the same side which it entered. As with transmission holograms, the recording and replay conditions must be carefully considered when recording edge–referenced reflection holograms.

4.3.1 Reflection Recording

A typical recording setup for an edge–referenced reflection hologram is depicted in Fig. 4.3.

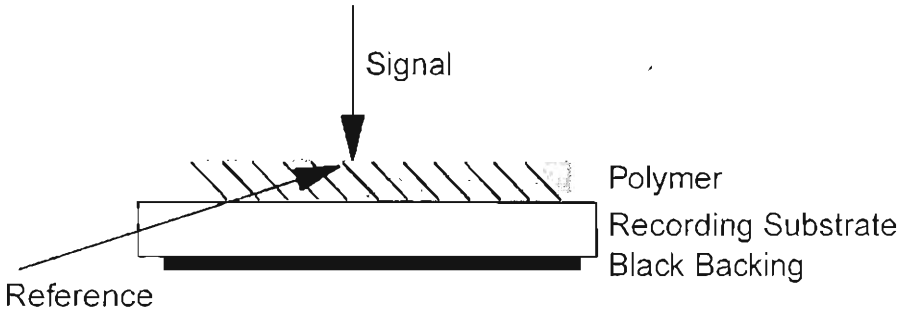


Figure 4.3 The recording geometry for an edge -referenced reflection hologram.

The black backing is essential for a light dump for the signal beam as even the 4% reflection from the substrate–air interface creates a stray reflection which significantly affects the performance of the hologram in some applications. Again, with the difficulty of recording through an acrylic substrate, the Mylar is left on the polymer and the effects appear to be minimal on the final hologram performance. The possibilities for replaying the hologram recorded in Fig. 4.3 are shown in Fig. 4.4.

4.3.2 Reflection Replay

Replaying a non -conjugate reflection hologram is simple to do on the recording setup by removing the black backing. The real–time refractive index modulation from photopolymerization allows for quick determination of a successful recording. If the signal beam is blocked after the backing is removed, then the diffraction efficiency can be judged by the diffracted beam. If the signal beam is left unblocked, then a real -time interferometer is created from the recorded hologram and the real–time interference of the signal and reference beams. This interferometer helps determine

an efficient recording and allows for accurate stability checks on the optical components (to help determine creep and damping factors associated with the components). Due to the fluorescence, the real–time fringes can even be seen inside the photopolymer (Appendix B). These fringes also verify correct repositioning if the hologram has moved, for example, when the black backing is removed. The non–conjugate reflection holograms from Fig. 4.4 (a) have proven to yield excellent HOE's for edge–lit fingerprint illuminators (§8.2).

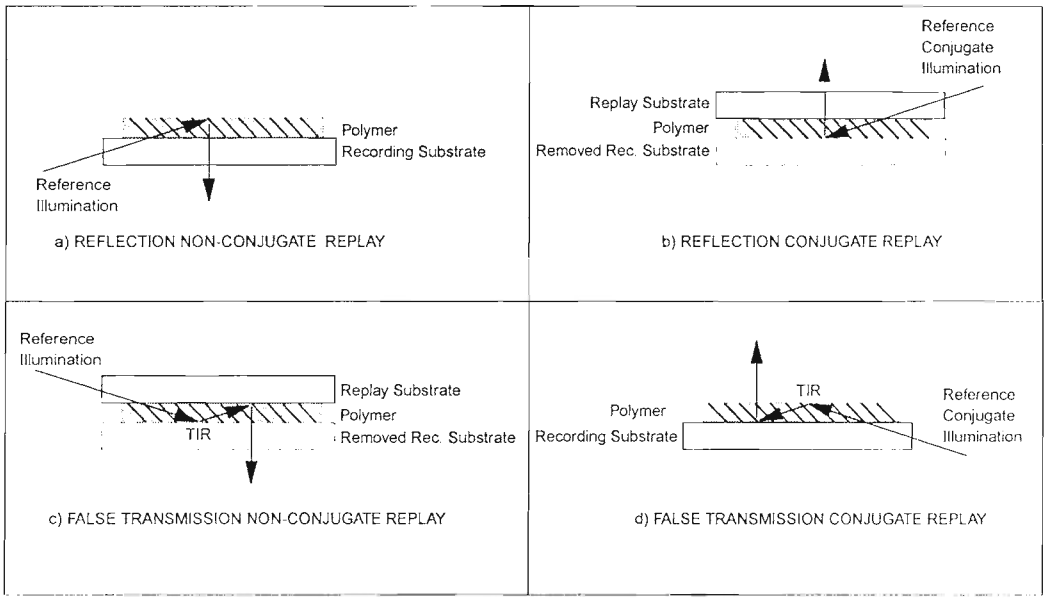


Figure 4.4 The replay possibilities of an edge–referenced reflection hologram. (a) and (b) are the traditional conjugate and non–conjugate replay, respectively. (c) and (d) are the new 'false transmission' non–conjugate and conjugate replay geometries, respectively, which incorporate a TIR within the film.

The same difficulties which arise with transmission holograms for adding a new replay substrate also arise in reflection holograms. In theory, the photopolymer can be transferred to a replay substrate as in Fig. 4.4 (b) and (c), however, this transfer is very difficult to accomplish. This transfer is not necessary in most of the applications studied because it is possible to replay in the non–conjugate reflection mode.

4.3.3 False Transmission Replay

Replaying a reflection hologram in a transmission mode has many attractive benefits. The dispersion normally attributed to a transmission hologram is not as severe as in a reflection hologram. Also, transmission holograms are generally preferred in optical systems because of the ability of placing the next or previous element in close proximity without blocking the replaying light.

For example, in an edge–lit LCD illuminator, the hologram would need to be as close as possible to the LCD to precisely illuminate the correct pixels. In reflection replay, the LCD would have to be at least the thickness of the substrate away. In a false transmission replay geometry, the LCD could be essentially in contact with the hologram. Reflection edge–lit LCD illuminators and display holograms have been made which replay very well in the false transmission geometry of Fig. 4.4 (d).

4.4 Evanescent Holograms and SIM

Traditionally, holograms are recorded with homogeneous waves (waves whose planes of constant amplitude coincide with those of constant phase). Stetson [4.1] first proposed that holograms could be made from inhomogeneous waves (a wave whose planes of constant amplitude cross planes of constant phase at an angle) [4.2]. Since then, many applications involving evanescent waves have been studied [4.2–4.19]. Evanescent waves can be created from edge–referenced holograms depending on the refractive indices of the substrate and film. In general, there are three possible recording regimes for the edge–referenced reflection hologram depending on the refractive indices as indicated in Fig. 4.5.

When the substrate refractive index is less than the photopolymer (a), reference beam coupling into the film is possible. However, depending on the angle and index

difference, there can be a Fresnel reflection of significant intensity (§3.6.2). The most effective recording regime is that of Fig. 4.5 (b) because is couples the full incident intensity of the reference beam into an appreciable depth of the recording medium.

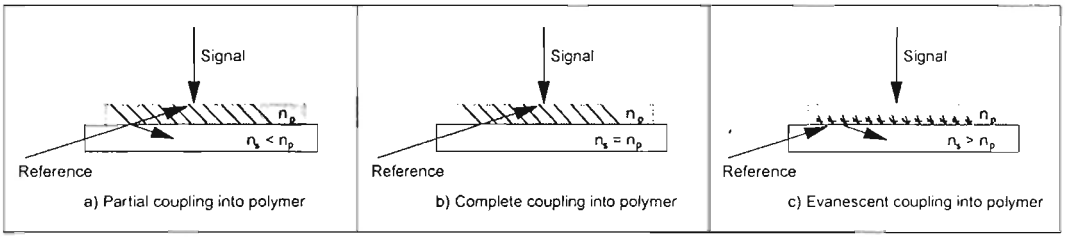


Figure 4.5 The three possible recording regimes for edge–referenced holograms which depend on the substrate index.

4.4.1 Evanescent Holograms

When the index of the substrate is higher than that of the polymer as in Fig. 4.5 (c), a steep reference beam at an angle greater than the critical angle is totally internally reflected at the substrate–polymer boundary. In a static hologram (one which does not develop in real time), this would create an evanescent hologram (Fig. 4.6).

The amount of light which enters the rarer medium depends on the angle and the indices of refraction of the two mediums. The depth, λ_a is derived from rigorous treatment of Maxwell's equations: [4.2]

$$\lambda_a = \frac{\lambda_0}{\sqrt{n_s^2 \sin^2\theta - n_p^2}}, \text{ and the beam displacement is} \tag{4.1}$$

$$d = \frac{\lambda_0 \sin\theta}{n_s \pi \sqrt{\sin^2\theta - \left(\frac{n_p}{n_s}\right)^2}}. \tag{4.2}$$

where λ_0 is the wavelength in a vacuum, θ is the angle inside the denser medium, n_s and n_p are the indices of refraction of the denser medium (substrate) and rarer medium (polymer) respectively. The evanescent wave amplitude falls to 1/e of its incident value at a depth in the rarer medium of $\lambda_a/2\pi$. The lateral beam displacement, d , is commonly known as the Goos–Hänchen effect and here is defined for the case of TE polarized light. The depth and the displacement become very significant when the refractive index of the polymer is such that the angle of incidence for the reference beam is at the critical angle. This phenomenon can occur in a dynamic recording medium such as the photopolymers.

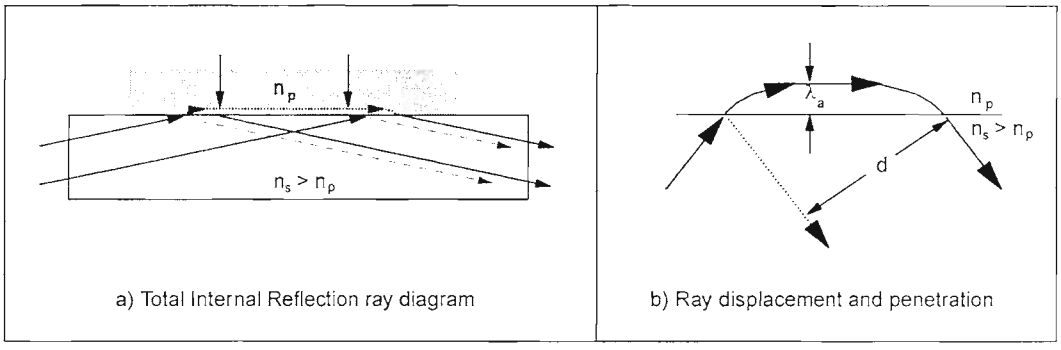


Figure 4.6 The representation of the reflected beam path in an evanescent hologram (a). The reflection is displaced by a distance d, and the beam penetrates to a depth of λ_a as shown in (b). The arrows in the rarer medium represent the direction of energy flow. [4.2]

4.4.2 SIM Holograms

Recording an edge–referenced hologram with a dynamic recording medium (such as with the Du Pont photopolymers) can result in a unique phenomenon. In a TIR situation at the substrate–polymer interface, an evanescent hologram is recorded and in this area the index increases. Since the index is increasing (and the index difference between the substrate and the film is decreasing), the penetration depth of the reference beam increases and continuous evanescent holograms are recorded (Fig.

4.7). This process continues until the modulation is insufficient to allow further penetration or the light traverses the depth of the film. This phenomenon is referred to as Self–induced Index Matching.

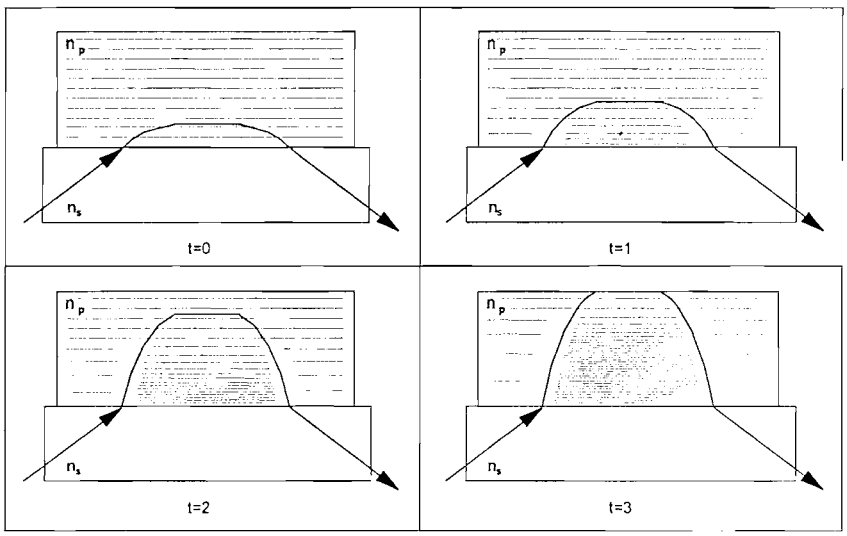


Figure 4.7 A simplified model for the continuous recording of an evanescent hologram in a medium with a time–varying index of refraction. The horizontal lines represent density of photopolymer.

4.4.2.1 Penetration Depth and Lateral Displacement

In Fig. 4.7, when the recording begins (t=0), an evanescent wave is penetrating into the film. The gradual index increase in this exposed area allows more coupling of light into the higher index areas while simultaneously increasing the penetration depth (Fig. 4.8). The depth would increase until it reaches the edge of the film, however, most of the light is absorbed before it reaches the opposite side of the film due to the steep angle involved and the absorption of the film. When the polymer index reaches the 'critical index' corresponding to the critical angle, the evanescent coupling converts to direct coupling and is depicted in the asymptotes in Fig. 4.8.

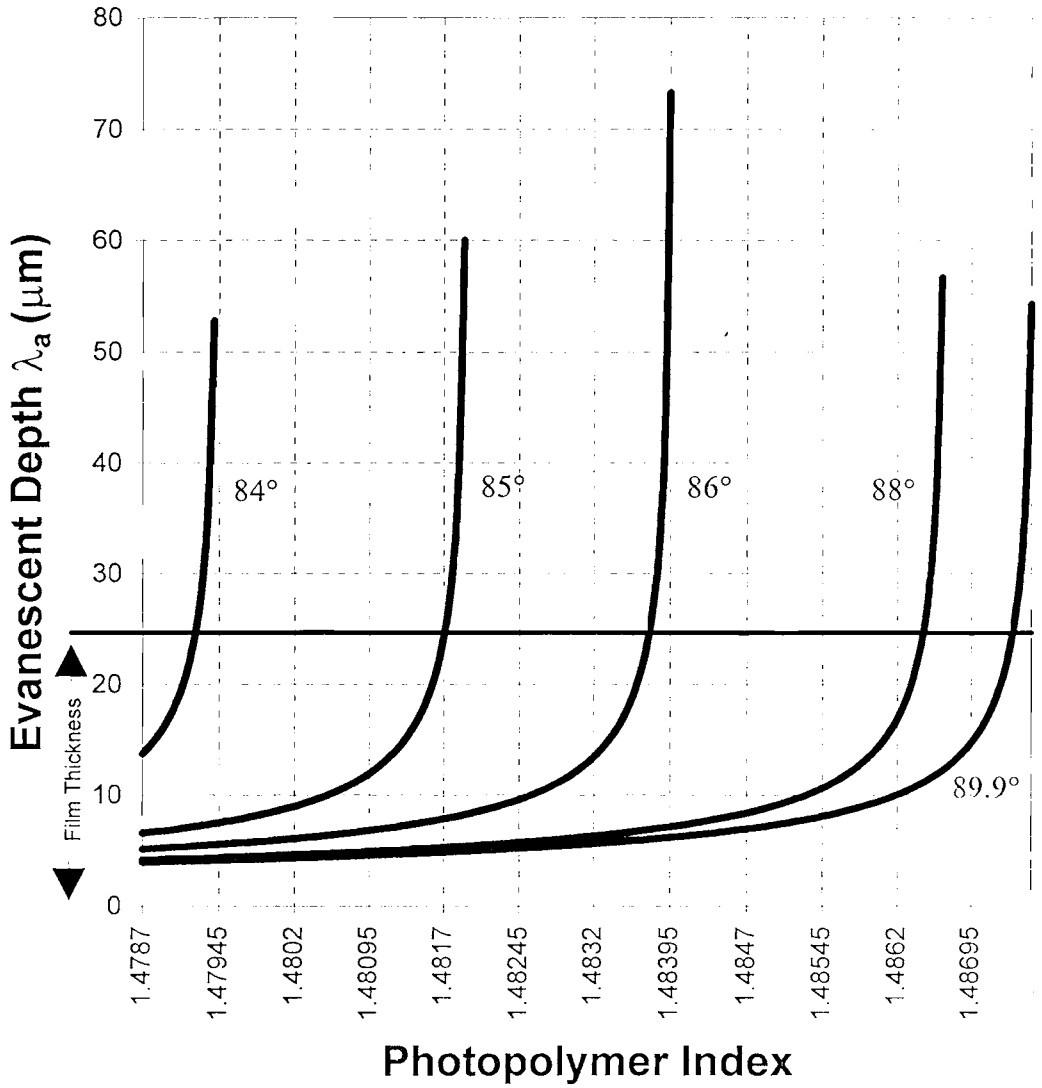


Figure 4.8 The evanescent penetration depth as the photopolymer index of refraction increases during exposure for various steep reference angles for a substrate with $n_s = 1.4876$.

The values obtained for Figs. 4.8 and 4.9 are from Eqs. 4.1 and 4.2 where $n_s = 1.4876$, $\lambda = 0.647$ μm, and the photopolymer is HRF 700x071. From Fig. 4.8, one can see that the closer the angle of incidence is to the initial critical angle (83.7°), the larger the initial penetration depth. The evanescent depth is at least 5 μm into the film for the steep angles between 84°–89°. Thus, with the 1/e depth of $\lambda_a/2\pi$ (which is in this case at least 0.8μm) there is sufficient amount of exposure area to initiate polymerization. This shows that in this steep referenced system with closely matched indices, any angle above the critical angle will record a SIM evanescent hologram.

Since the signal beam is exposing the area where the reference beam has not yet reached throughout the exposure, it is inherently fogging the material and reducing the dynamic range of the material. This reduction inhibits the further penetration of the evanescent beam and reduces the refractive index modulation available in the areas yet unexposed to the reference beam. One would therefore want to minimize this fogging by reducing the amount of refractive index modulation needed for full reference beam penetration. This would result in a thicker hologram with higher modulation. In Fig. 4.8, for example, one would choose a recording angle of 84° as opposed to 85° so that the penetration would reach the film before the signal exposed (fogged) the area significantly. Therefore, it is more desirable to work near the critical angle when recording edge referenced evanescent or SIM holograms. Fig. 4.8 also shows the high angular dependence on the refractive index modulation required for full film penetration. In other words, when recording with a reference beam angle of 84°, a refractive index modulation near 0.0008 is required for full film penetration while at 88°, a refractive index modulation near 0.008 is required. Therefore, recording a volume SIM hologram (using the full film thickness) requires ten times more index modulation at 88° than at 84°. This example shows the small angular range within which to record an evanescent hologram of moderate thickness and efficiency.

From Eqs. 4.1 and 4.2, one can see that the lateral displacement as well as the penetration depth also maximizes at the critical angle. A very large lateral displacement may cause spurious reflections if it is larger than the recording area. As can be see from Fig. 4.9, the displacements grow towards infinity when they approach the critical angle (or in this case when the polymer approaches the critical index).

Fortunately, as the polymer approaches the critical index, the penetration depth is also increased and most all of the reference light is absorbed and does not cause spurious reflections of significant intensity.

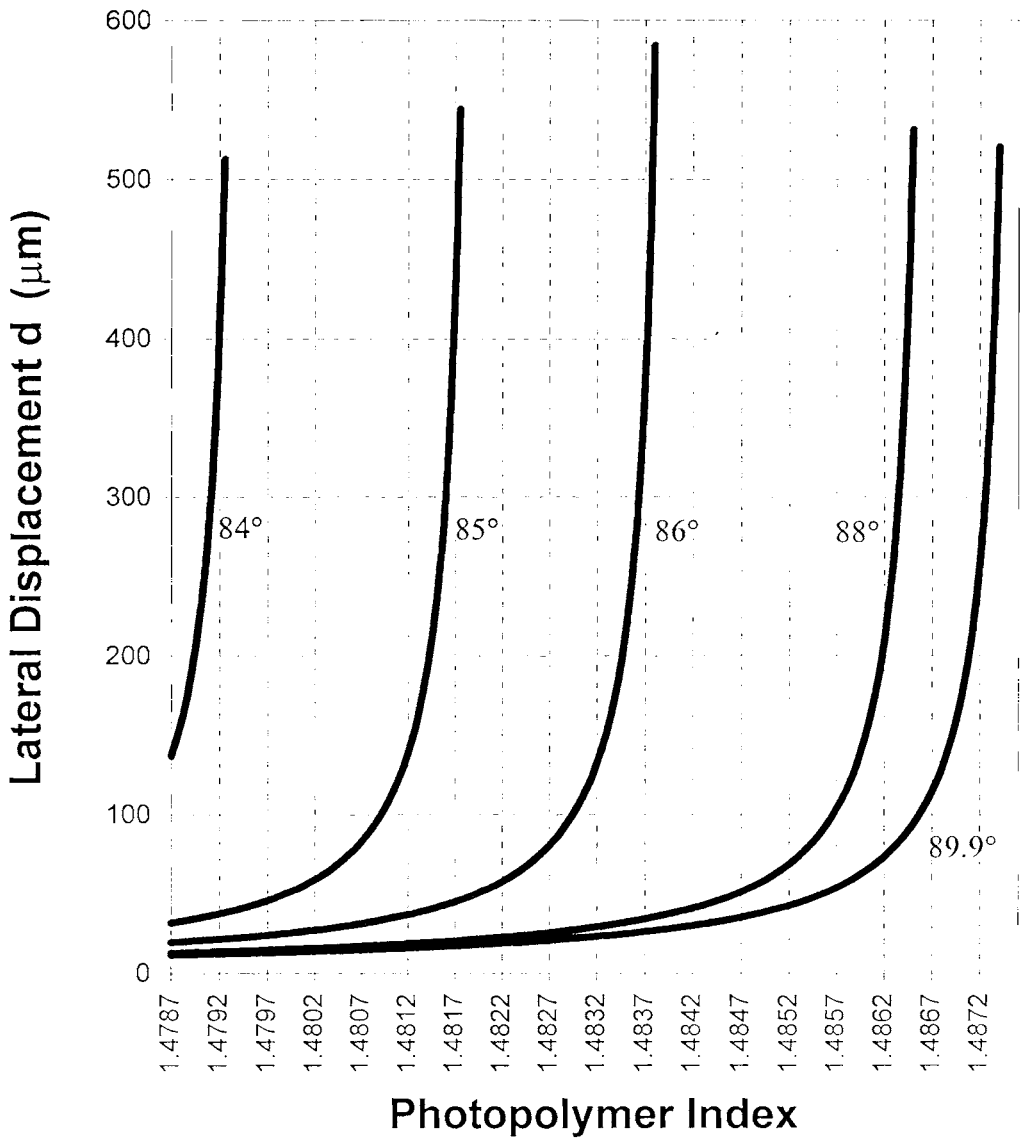


Figure 4.9 The lateral displacement of the reflected beam as the photopolymer index increases for various steep angles in a substrate with $n_s = 1.4876$.

4.4.2.2 Increased fluorescence

Since the film absorbs light at the recording wavelength, the reflection in an evanescent or SIM hologram is an Attenuated Total Internal Reflection (ATIR). The reflection is not a Frustrated Total Internal Reflection (FTIR) because direct coupling does not occur. Since the average index of the entire film does not increase enough to

provide complete coupling, an ATIR exists throughout the recording. Some of this absorbed light is re–emitted as fluorescence from the sensitizing dye. As the penetration depth increases, more of the sensitizing dye is exposed and the fluorescence increases. This increase in fluorescence was the first indicator of the occurrence of the SIM phenomenon and an example is shown in Fig. 4.10 [4.20].

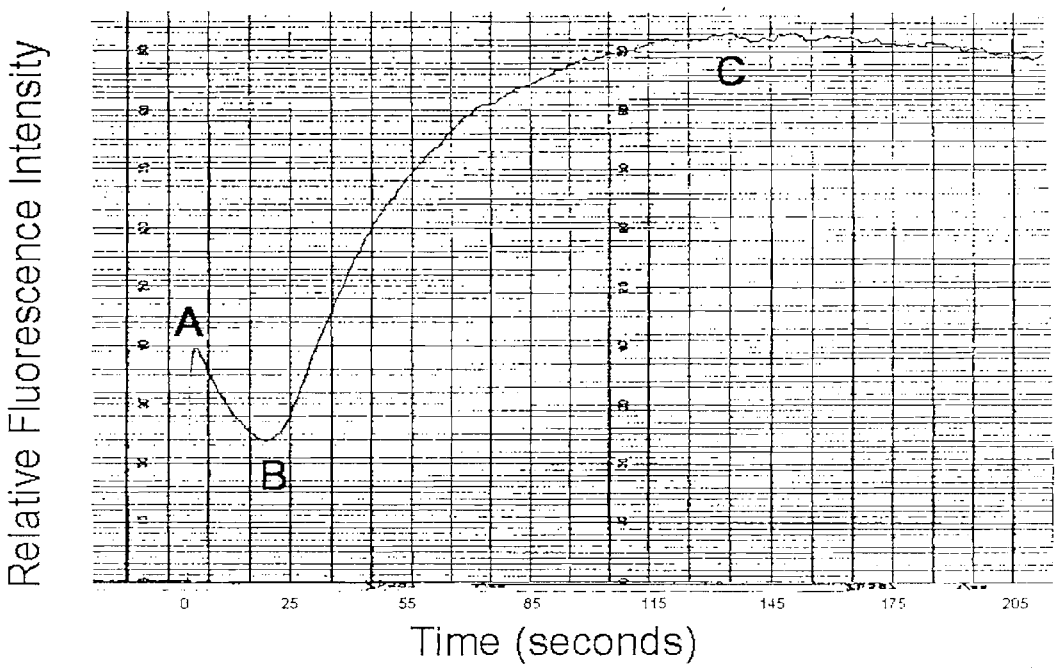


Figure 4.10 The increase in fluorescence from a constant intensity reference beam for 215 seconds in the setup of Fig. 3.1 (without the signal beam) with the substrate refractive index higher than the index of the photopolymer.

For the measurement of Fig. 4.10, no signal beam is used and the detector is placed behind the sample instead of the black light dump in Fig. 3.1. The exposure begins at point A, and most of the diffusion begins at point B where the reference begins to penetrate through the film until point C where the fluorescence begins to saturate and gradually fall–off due to bleaching. When the reference beam is turned on, the fluorescence emission is constant. The fall–off from point A to B in Fig. 4.10 is due to bleaching of the dye. During the polymerization process, the sensitizing dye is broken down and thus there is less of it to fluoresce. This effect is quicker than the diffusion process which explains the slight decrease in fluorescence intensity from points A to B. When the diffusion begins and the monomer moves to the exposed regions, the

local index of refraction increases. The reference penetrates further in these areas, reaching more unexposed dye and thus the total fluorescence increases. The fluorescence intensity reaches a saturation value at C which indicates it has penetrated as far into the film as possible with the available refractive index modulation. This saturation occurs during the direct recording of an edge–illuminated hologram and indicates that the reference beam has penetrated as far into the film as possible.

Since at point A, the reference beam may have penetrated approximately 5 μm before any diffusion takes place (based on the calculations of Fig. 4.8), one can assume that an evanescent hologram has been formed in this region. This portion of the hologram is formed from an evanescent wave which is inhomogeneous. Later, as the index of refraction increases in this area, the coupling becomes more direct, and the wave becomes more homogeneous closer to the boundary. Since the refractive index is changing, the path and angle at which the reference beam travels also changes. Thus, the real–time effect of these recordings would inevitably create dynamic fringes with slightly varying angles and spacing. This would inherently reduce the efficiency of the hologram. Lab results have confirmed the reduced efficiency which could be attributed to this effect in combination with the fogging of the signal wave.

The SIM phenomenon is only specific to a situation of TIR where the lower refractive index increases to allow penetration and could apply to non–edge–illuminated geometries as well. The low diffraction efficiency of SIM holograms does not make it a preferred method, however, monitoring the fluorescence in the TIR situation may also be useful in measuring other characteristics of the photopolymer such as diffusion.

4.5 Alternative Bragg Condition

In the basic recording geometry for edge–lit holograms (Figs. 4.1, 4.3) the angular restriction imposed on the reference beam because of the thin recording substrate

limits the light coverage on the surface of the film. The thin substrate also imposes strict intensity uniformity since the small area of the reference beam is spread over a much larger area in the photopolymer. Diffraction effects at the edges of the substrate can prevent a uniform reference intensity. In order to alleviate some of these difficulties, Bragg's Law [4.21] was examined in order to determine other possible recording regimes:

$$2\Lambda \sin(\phi - \theta_0) = \frac{L\lambda_0}{n} \tag{4.3}$$

where Λ is the absolute fringe spacing, ϕ is the fringe angle, θ_0 is the Bragg angle, L is the Bragg order, λ_0 is the Bragg wavelength and n is the index of refraction of the film. Thus, looking at different first order Bragg conditions (L=1) for a particular grating (Λ, ϕ, and n are constant) yields an infinite number of wavelength and angle combinations which will result in maximum diffraction. (Normally the first order of diffraction has a higher efficiency than other orders for cosinusoidal volume phase gratings). As a result, one may record a specific grating at other Bragg angle and wavelength combinations that satisfies Eq. 4.3.

Therefore, with a given grating structure (periodic with constant fringe angle and spacing), one can visualize that the same grating (or interference pattern) could be created by a large number of combinations of wavelength and recording angles. One could also view this by noting that there are always different Bragg conditions for replay of a hologram. It is possible to record at one Bragg condition (defined by the fringe spacing and two recording angles) and replay at another Bragg condition (imposing a different combination of replay angles and wavelength). This hypothesis led to an alternative recording geometry to generate an edge-lit hologram which would have more relaxed recording angles (thus making it easier to align, filter, and measure the recording beams.).

4.5.1 Alternative Bragg Condition Recording

Using the diagram in Figure 2.1, one can deduce the equation for the absolute fringe spacing, Λ, and the equation for the fringe spacing in the x–direction, d_x, from Bragg's Law:

$$d_x = \frac{\lambda_m}{|\sin\theta_S - \sin\theta_R|}. \tag{4.4}$$

Using the term for the wavelength in air, one can rewrite this using $\lambda_m = \frac{\lambda_a}{n}$ which yields,

$$d_x = \frac{\lambda_a}{n(|\sin\theta_S - \sin\theta_R|)}. \tag{4.5}$$

From this, an equation for the absolute fringe spacing, $\Lambda = d_x \cos\phi$, is deduced:

$$\Lambda = \frac{\lambda_a \cos\phi}{n(|\sin\theta_S - \sin\theta_R|)} \text{ where} \tag{4.6}$$

$$\phi = \frac{\theta_S + \theta_R}{2} \text{ is the fringe angle.} \tag{4.7}$$

Using the trigonometric identity

$$\sin\alpha - \sin\beta = 2\cos\frac{1}{2}(\alpha+\beta)\sin\frac{1}{2}(\alpha-\beta), \tag{4.8}$$

it can also be shown that the fringe spacing can be written in terms of the wavelength in air and the two recording angles as

$$\Lambda = \frac{\lambda_a}{2n\left|\sin\left[\frac{1}{2}(\theta_S - \theta_R)\right]\right|} \tag{4.9}$$

One can think of a simple volume phase hologram formed from two collimated wavefronts as simply a set of fringes of refractive index modulation with a specific angle and spacing. With this view, different recording regimes are possible which could create the same set of fringes, thus the same hologram.

Equations 4.7 and 4.9 give the defining characteristics of the fringes which make the hologram. To make the same hologram through a different technique, it would have to have the same fringe spacing and fringe angle. The two methods are denoted by subscripts 1 and 2. Thus, $\Lambda_1 = \Lambda_2$, and $\phi_1 = \phi_2$. For two different recording methods Eq. 4.9 becomes

$$\frac{\lambda_{a1}\cos\phi_1}{n(\sin\theta_{S1} - \sin\theta_{R1})} = \frac{\lambda_{a2}\cos\phi_2}{n(\sin\theta_{S2} - \sin\theta_{R2})}. \tag{4.10}$$

Since the fringe angles must be the same, this reduces to

$$\lambda_{a2} = \left(\frac{\sin\theta_{S2} - \sin\theta_{R2}}{\sin\theta_{S1} - \sin\theta_{R1}}\right)\lambda_{a1}. \tag{4.11}$$

Thus, it is possible to use two or more different recording geometries to achieve the same grating structure.

4.5.2 Dual Prism Method

As an alternative to recording a slanted fringe hologram with a steep reference beam, an alternative Bragg condition recording method was attempted which involved two

prisms and a different recording wavelength. The recording geometry using the prisms is denoted with the subscript 1, and the steep edge–referenced recording is denoted with the subscript 2 as indicated in Fig. 4.11 and 4.12.

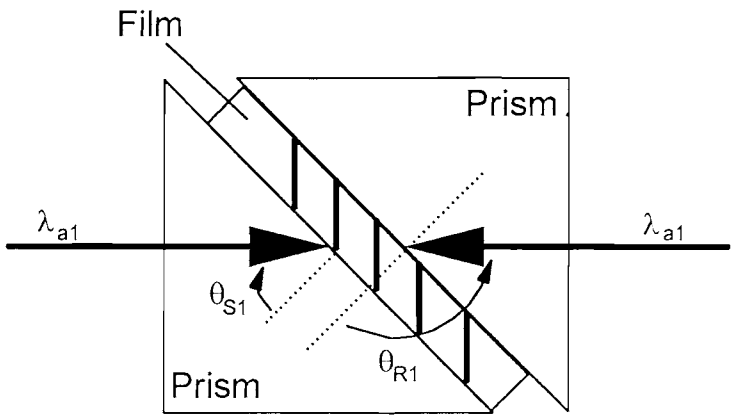


Figure 4.11 The slanted fringe hologram recorded using two prisms.

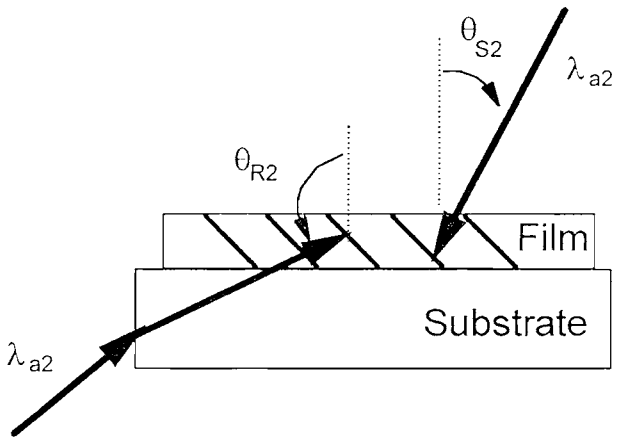


Figure 4.12 The slanted fringe hologram recorded using a steep–angled reference beam through the edge.

In order to eliminate edge–referencing, the signal and reference beam angles had to be significantly altered while keeping the fringe angle constant. Therefore, the prism geometry of Fig. 4.11 was conceived so that the reference angle would not be as steep and the uniformity and coupling restrictions would not be as severe. As a result, the recording wavelength was adjusted according to Eq. 4.11. Thus, a blue edge–referenced hologram recorded as in Fig. 4.12 could also be recorded using a red

wavelength and the geometry of Fig. 4.11. The main drawback of this technique is the requirement that the interference fringes have a constant fringe angle over the entire recording area (i.e. parallel fringes). This forces the recording and replay beams to be collimated.

A hologram was made using a single beam and the two prism geometry as illustrated in Fig. 4.13 with a mirror reflecting light back at a specific angle. The single beam method was used for simplicity and higher vibration stability because of the lack of beam separation. However the single beam method introduced an undesirable beam ratio due to the absorption of the film. The hologram was not heated to be certain that a Bragg wavelength shift would not be introduced.

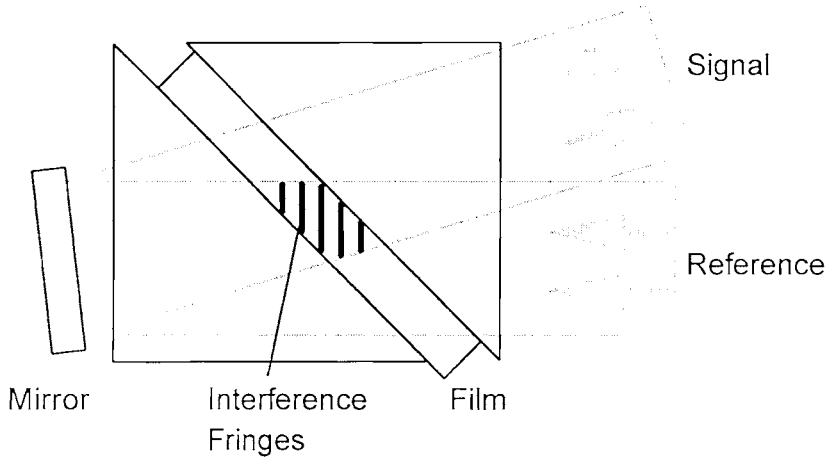


Figure 4.13 The recording setup for a single–beam slanted fringe hologram using two prisms and a mirror.

Another drawback of the single beam method is smaller overlap area of the two beams. This can be seen from Fig. 4.13 where the recorded fringes occupy a smaller area than the beam exposing the film. For large areas, this is a major drawback as the prisms and optics would need to be much larger than the final hologram size. In the dual prism method of recording, the prisms must be mounted kinematically while allowing for easy sample retrieval. A dual prism holder was constructed as illustrated in Fig. 4.14.

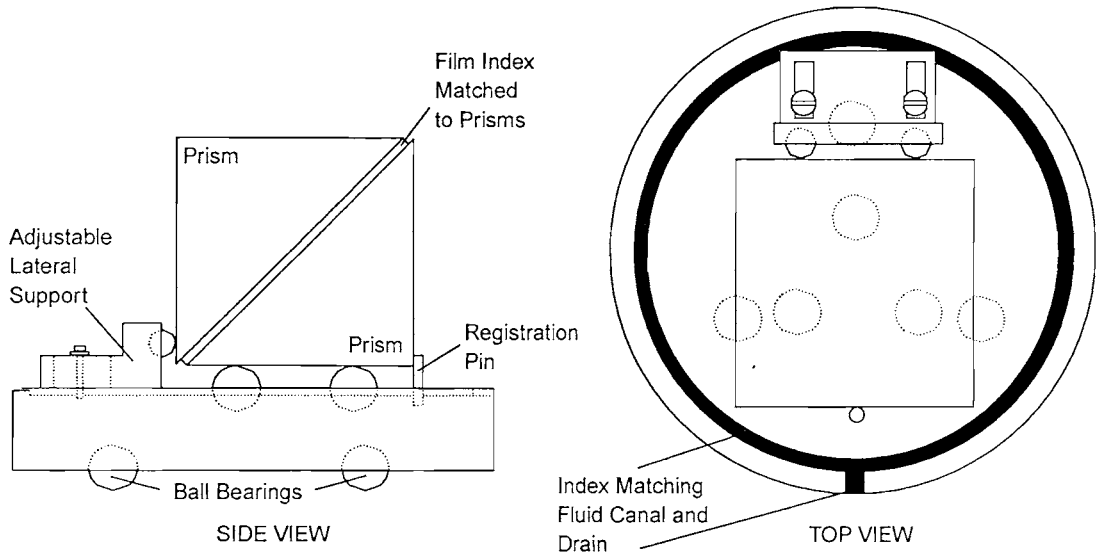


Figure 4.14 The dual prism kinematic holder. The ball bearings support the holder as well as the prisms. The dark area on the top view is a channel for the index matching fluid to drain. The film is laminated to a substrate and the combination is index matched to both prisms.

4.5.3 Prism Method Results

For the application of the edge–illuminated holograms examined in this research, the geometry restricted the angles to $\theta_{S2} = 0°$ and $\theta_{R2} = 98.3°$ for replay. These angles define the fringe angle to be (according to Eq. 4.7)

$$\phi_2 = \frac{0°+(98.3°)}{2} = 49.15°. \tag{4.12}$$

Therefore, if the same hologram is to be made using the prism method, the criterion $\phi_1 = \phi_2$ must be met. For convenience, two 45° prisms were used with the signal and reference beams at angles of $\theta_{S1} = -45°$ and $\theta_{R1} = 143.3°$ which gave the same fringe angle as the edge–illuminated hologram. These angles can be substituted into Eq. 4.11 which yields the relationship

$$\lambda_{a2} = \frac{\sin(0°) - \sin(98.3°)}{\sin(-45°) - \sin(143.3°)} \lambda_{a1} = 0.758\lambda_{a1}. \tag{4.13}$$

If one were to record using the prism geometry and a wavelength of $\lambda_{a1} = 647\text{nm}$, this hologram could be replayed in an edge–illuminated geometry at

$$\lambda_{a2} = (0.758)(647nm) = 490nm. \tag{4.14}$$

A hologram recorded using the dual prism method at 647 nm can be replayed as an edge–lit hologram at 490 nm. The beam uniformity is easier to control in the alternative (prism) method in contrast to that of the direct edge–referenced recording. Recording using the prism method is also of significant importance when one considers the difficulty in obtaining high powered blue lasers (if one wanted to record directly using the edge–referenced technique). One could extend these results to recording with the dual prism with a visible blue wavelength to generate a UV edge–lit hologram, although for replay, the UV absorption in the photopolymer is much higher.

The transmission spectrum of the prism recorded hologram of Fig. 4.13 is measured in Fig. 4.15 at normal incidence (replaying as $\theta_{S2} = 0°$). The transmission minimum for the hologram (corresponding to a reflection of light from the hologram at $\theta_{R2} = 98.3°$ out of the edge of the substrate) is at 491 nm which is just as expected from Eq. 4.14 (allowing for possible angular measurement errors).

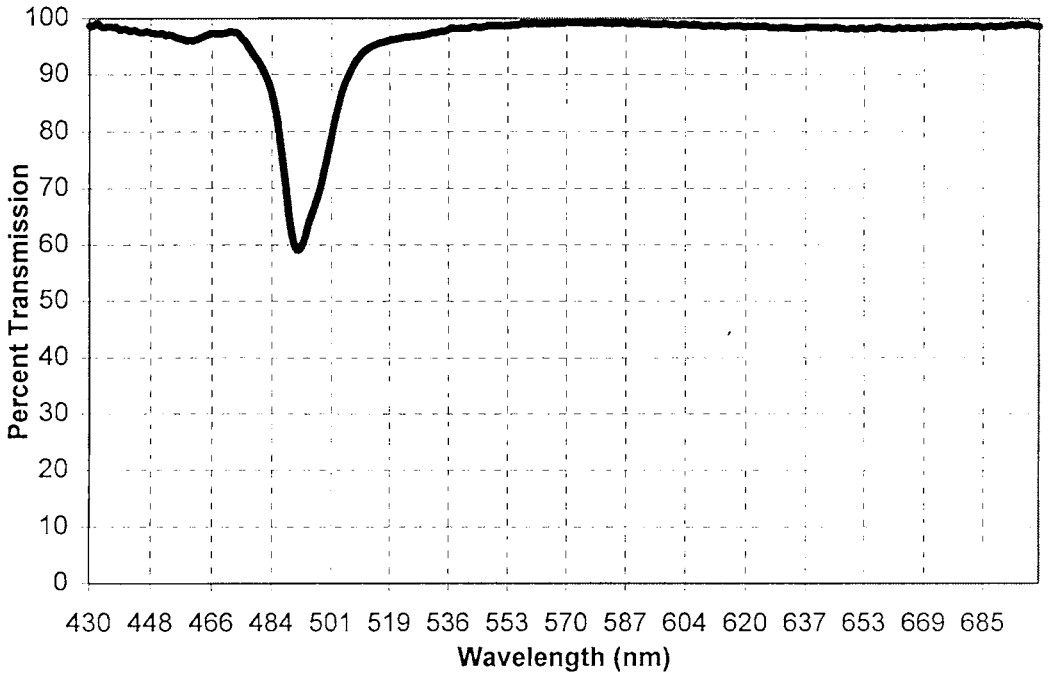


Figure 4.15 The zero-order transmission spectrum of a dual prism method hologram illuminated normal to the film. The recording wavelength was 647 nm with the signal and reference angles at $\theta_s = -45°$ and $\theta_r = 143.3°$.

4.5.4 Other Bragg Conditions

Using the rigorous coupled wave theory (described in §5.3), one can also analyze other possible Bragg conditions. With a particular grating, one can see all of the possible first order Bragg conditions by looking at the diffraction efficiencies of the first diffracted order. All of the possible wavelength and replay angle combinations for the Bragg condition occur when the first order diffraction efficiency is maximized. These are the same conditions that result from Eq. 4.11. The example given in the previous section is analyzed using the rigorous coupled wave theory and the results are graphed in Fig. 4.16.

The tail-like structure in Fig. 4.16 (a) shows the recording and replay possibilities for a hologram produced by recording using the dual prism method while (b) illustrates the recording and replay possibilities of an edge-referenced hologram. The second

Bragg condition is usually much weaker in diffraction efficiency for volume phase sinusoidal holograms. The second Bragg condition is shown in (b) as the faint mark near 90° and 270 nm. In Fig. 4.16, a vertical line drawn at a particular wavelength (one which will cross the curve) will intersect the curve at two angles, thus describing the two possible recording angles and a wavelength for recording the same hologram (Note that since these are reflection holograms, one of the *recording* angles would be $\theta_r=\theta_{graph(replay)}-180°$).

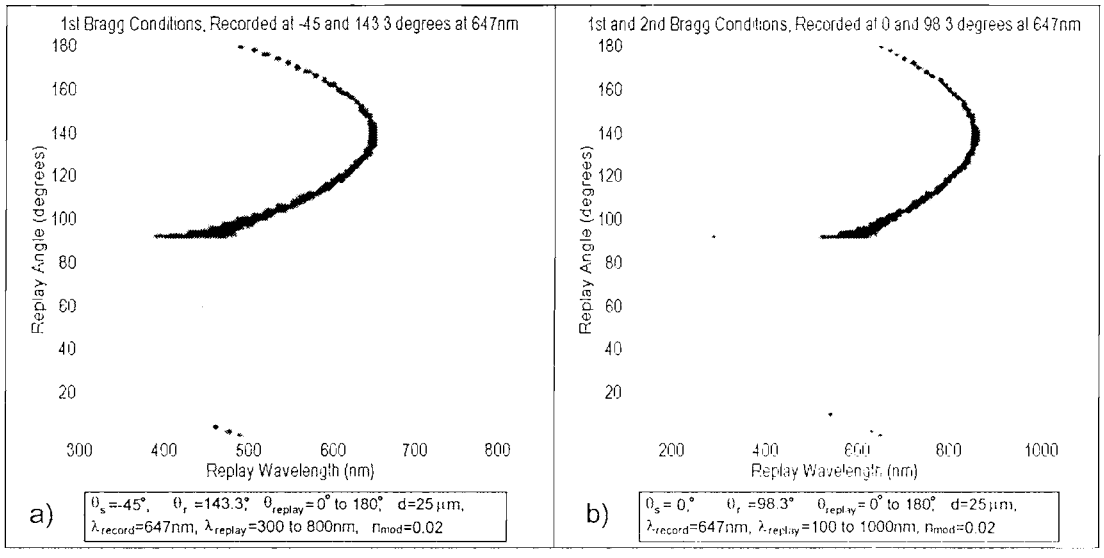


Figure 4.16 The possible Bragg replay (or recording regimes) for two slanted fringe holograms.

An advantage of calculating the possible Bragg conditions from Eq. 4.11 is that it is quick and simple (while the rigorous program can take an hour to calculate 90 x 90 data points on an IBM PC compatible 486–66DX). However, the rigorous method also shows the angular and wavelength bandwidth at all of the possible Bragg replay conditions. For example, in Fig. 4.16 (a), the hologram has its highest *angular* bandwidth near the 647 nm recording angle and a high *wavelength* bandwidth closer to the 90° limit.

4.5.5 Advantages of Alternative Bragg Condition Recording

Many of the difficulties associated with recording edge–illuminated holograms can be overcome by recording at a different Bragg condition. By using the dual prism method, the angles are much less restrictive. With this method, the beam alignment and uniformity are much simpler to maintain. Since the size of the beams can be larger than in the edge–referenced method, the anti–Gaussian filtering is much easier to accomplish. The refractive index matching of the substrate and film for recording is virtually eliminated because the angles of incidence are much smaller (for edge–lit replay, the restrictions still apply). The attenuation of the reference beam is much less, so the profile of the recorded grating is much less likely to be tapered. One of the few disadvantages is that the grating fringes must be parallel, thus one could not use diverging or converging wavefronts efficiently on replay.

4.6 Waveguide Holograms

Whereas in a steep reference hologram the light reaches the film directly, in a waveguide hologram the reference beam is bounced through a total internal reflection inside a substrate before it reaches the recording film. Many authors have reported on waveguide holography, beginning with Suhara [4.22] in 1976. Since then "waveguide holography" referred to here has also been called "guided–wave holography" [4.25], variations of "gratings in waveguides" [4.26, 4.23], "waveguide holograms" [4.27, 4.28, 4.22, 4.29, 4.32, 4.33, 4.34, 4.35, 4.37], "substrate–mode holograms" [4.30, 4.31], "substrate guided–wave holograms" [4.36], "guided–mode holograms" [4.38], and others [4.39].

As a result of the bouncing, the main differences between waveguide holograms and steep reference holograms are the *non–uniform reference coverage* and the *formation of unwanted gratings*.

4.6.1 Non–uniform Reference Coverage

When the reference beam in a waveguide hologram reaches the film, the beam either overlaps itself from several modes, or has a discontinuous coverage leaving unexposed areas. In recording the waveguide hologram, two types of conventional reference wavefronts can be used–collimated or diverging (as shown in Fig. 4.17). Converging wavefronts are rarely used because the beam would focus and would result in a diverging wavefront after the focus. This would create serious non–uniformities near the focus. Waveguide holograms with collimated wavefronts result in linear patches of light while waveguide holograms which have diverging wavefronts result in multi–moding patch overlap.

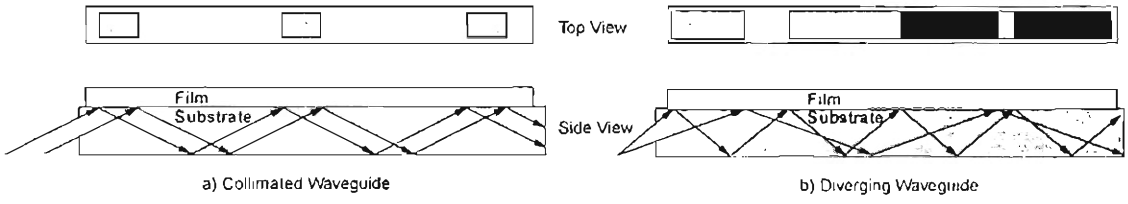


Figure 4.17 The reference coverage through a substrate onto a film from a collimated waveguide (a) and a diverging waveguide (b).

The linear patches created from a collimated waveguide result in dark areas where no hologram is formed at all. Therefore, the collimated wavefront is usually unacceptable cosmetically. In the diverging waveguide, the areas of overlap have higher intensities resulting in a non–uniform intensity pattern which affects the refractive index modulation uniformity.

The inherent absorption of the holographic recording material also affects the uniformity of the reference beam. As the beam passes in and out of the film, it is gradually absorbed. The angle the beam travels in the film for a waveguide hologram is steep and the absorption depends on the cosine of the angle (as in Eq. 2.2). This results in a highly attenuated reference beam. The attenuation is *repeated for every reflection* in a waveguide hologram. Thus, the intensity of the reference beam at the

last reflection is much less than at the first reflection. This unidirectional non-uniformity alters the effective beam ratio at various places in the film and therefore reduces the diffraction efficiency as discussed in § 2.4.3.

4.6.2 Formation of Unwanted Gratings

After a reference beam with an angle of incidence of θ reaches the surface of the film in a waveguide, the light reflects back into the substrate at an angle, θ_R, with the value $\theta_R=180°-\theta$ for a reflection hologram, or $\theta_R=\theta+90°$ for a transmission hologram using the angular conventions of Fig. 2.1. The area where two reference beams overlap is represented as the gray patches in the top view (corresponding to the film) in Fig. 4.17b. When this bouncing reference beam interferes with the signal beam, three gratings (three sets of fringes) are produced as in a Stetson TIR hologram [4.40]. This significantly reduces the dynamic range (the amplitude of the refractive index modulation) available to the desired fringes in this area of overlap, and therefore, the diffraction efficiency is reduced. In the black areas of Fig. 4.17, there are two reference beams incident on the surface, both of which reflect. Thus, there are four reference waves in these areas interfering together with one signal wave which produces *ten different interference patterns* or sets of fringes in the volume hologram.

Not only is the efficiency of the hologram reduced, but in replay the light reaching the desired set of fringes is reduced. On replay, the multiple angles in the overlap areas arriving at an unwanted set of fringes may diffract out of the film in unwanted directions. This corresponds to stray light in HOE's or false images in display holography. Therefore, the overlap areas create fringes with lower refractive index modulation in recording and unwanted diffracted light in replay which dramatically reduce the overall efficiency of the system.

4.6.3 Comparisons with Steep Reference Holograms

Steep referenced holograms may have collimated, diverging, or converging reference wavefronts in recording just as waveguide holograms. The difficulty of matching the replay source is also a factor in steep reference holograms. However, the existence of only one set of fringes and the possibility of a collimated reference in a steep–referenced hologram increases the system efficiency and replay flexibility.

As the thickness of the substrate decreases in a waveguide hologram, the maximum possible divergence decreases as well. As the divergence decreases and the direction of the *reference* beam approaches 90°, the angle of the *reflected* beam approaches 90° as well. Thus, as the angular difference in the incident and reflected beams in the overlap areas decreases, the fringes approach 45° (with a 0° signal angle). This type of waveguide hologram is essentially the same as an edge reference hologram recorded with a reference beam angle of 90°. One might think that this is irrelevant since one can not practically record at angles extremely close or at 90° anyway. However, 45° fringes with the correct spacing can be created by other methods, such as alternative Bragg condition recording (§4.4) or possibly using the direct fringe copying method.

4.7 Direct Fringe Copy Holograms

Advances in Microsharp™ diffuser technology developed in the laboratory led to the idea that one could copy an amplitude hologram into photopolymer using incoherent UV light [4.41]. To attempt this type of recording, the photopolymer was laminated onto a rainbow amplitude H2 hologram recorded in silver halide. When the photopolymer was exposed to broadband UV light, the amplitude hologram could be transferred to a volume phase hologram within the film. Two experimental models and arrangements were developed for the copying as illustrated in Fig. 4.18.

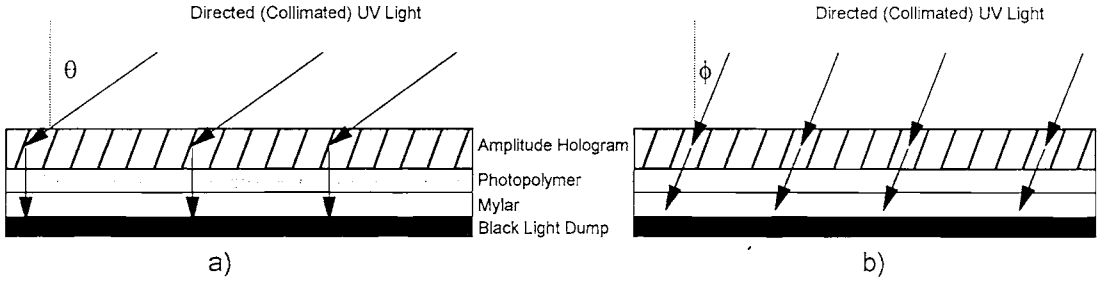


Figure 4.18 The two models for copying the amplitude hologram into a phase hologram. (a) uses a Bragg reconstruction model where θ is the Bragg angle. (b) uses the 'light tunnel' model where ϕ is the fringe angle.

Contact copying amplitude holograms with the incident light normal to the surface was attempted. This yielded very poor results. Then, attempts were made at copying the holograms at an angle similar to the Bragg angle, θ, as when copying a reflection hologram with a laser (Fig. 4.18a). These seemed to record partially with aberrations. Later it was theorized that a slightly steeper angle, ϕ, was needed which would enable the light to run parallel to the fringes (channeling it through at the correct angle). These results seemed to work better, yielding brighter samples with better field of view.

This method worked better because the light is not in an interference recording mode as in reflection contact copying. If the copy is recorded as in 4.18a, then in order to get a phase hologram equivalent to the original amplitude hologram, the transmitted first order diffracted light must interfere with the zero order light. Due to the extremely low temporal coherence of the UV source, any interference pattern would be virtually negligible. An image was visible, although this is probably due to copying the *volume* amplitude hologram into a *plane* phase hologram at the surface of the polymer. With the recording angle as in Fig. 4.18b, the fringes are copied directly, virtually irrelevant of the temporal coherence, and only specific to the spatial coherence which is related to the degree of collimation of the incident beam.

The period of the fringes in the transmission amplitude master was near 4 μm (with an aperture of 2 μm). For the equivalent hologram to be recorded as a reflection hologram the fringe spacing will be considerably smaller and diffraction from the edges may become an issue. However, the self–modifying characteristics of the photopolymer will increase the local refractive index in the regions between the amplitude fringes (inside the 'light tunnels') which may help counteract the spreading of the light due to diffraction (which would increase the spacing of the fringes and reduce the diffraction efficiency).

In early experiments, Du Pont recording films HRF 150 and HRF 600 (both 100 microns thick) were used for the copies, however, the images seemed to be over–modulated. The holograms had remarkable efficiency (more efficiency than the master), however they seemed to have a lensing effect on replay. This was attributed to a formation of a very thick hologram and possibly poor collimation of the UV source. Results were more efficient and had better angle of view when copied into thinner emulsions (such as 25 micron HRF 352). Due to the high contrast of the amplitude copy holograms, the amplitude modulation is almost binary. This results in an extremely good mask for copying into the photopolymer, best described by Phillips et al.[4.41]:

> "The unusual aspect of this type of contact copy method lies in its use of high contrast (non–sinusoidal) patterns of interference. Thus the driving force behind monomer migration is greater than that present in conventional holographic recordings where the optical patterns involve at best $\cos^2$ fringes. Once the monomer diffuses, then it contributes to a linearization of the high gamma silver recording. Were the copy hologram to retain the effects of such high gamma, it would lose efficiency due to the need for a high level of beam ratio consistent with image linearity."

The amplitude holograms were made using a special low–noise process. The optimum process was found to be a 4:1 reference to signal ratio and a 500 $\mu J/cm^2$ exposure of an Agfa Millimask® silver halide film which is developed in 4–Amino Phenol developer [4.41] for 2 minutes. This yielded a high contrast amplitude hologram with low noise. Some of the speckle noise was reduced by adjusting the slit width used in the rainbow copy hologram to a ratio of 50 cm object distance: 1 cm slit width.

For the exposure of the photopolymer through the amplitude hologram, it was necessary to use *low intensity* UV (broadband) light in order for the rate of diffusion to be greater than the rate of polymerization so that the photopolymer would not 'freeze' and polymerize everywhere without the diffusion (which is needed for a high index of refraction modulation and diffraction efficiency). Good results were obtained using a 4 mW/cm^2 broadband UV light source (Fig. 4.19) and a 5 minute exposure (1.2 $Joules/cm^2$).

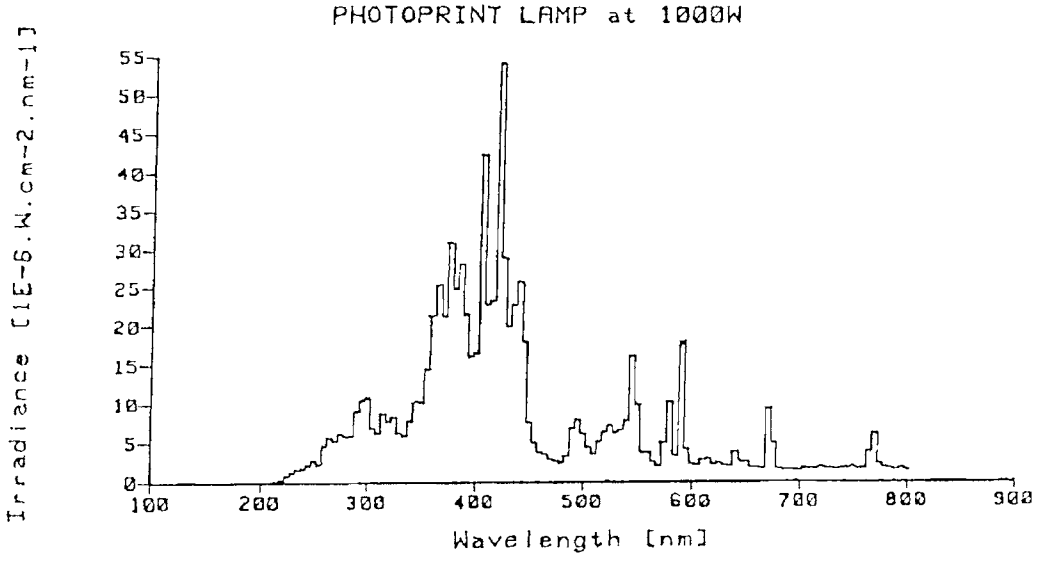


Figure 4.19 The broadband spectral output of the mercury discharge lamp used for the direct fringe copying of the amplitude masters into photopolymer.

When the copy geometry was altered to replay at the angle ϕ (Fig. 4.18) higher efficiencies and viewing angles were noted. Unfortunately, the illumination wavefront used was not at an optimum as it was later realized that since the fringes of the amplitude *rainbow* hologram are not parallel, the UV light needs to be diverging similar with the fringes for the light to travel down the 'tunnels'. Thus, much better results should be obtainable with a diverging UV recording instead of collimated.

REFERENCES

4.1 Stetson, K., Laser Focus, Vol. 3, pp 25, (1967)

4.2 Bryngdahl, O. "Evanescent Waves in Optical Imaging," *Progress in Optics* XI, E. Wolf, North–Holland, pp 167–221, (1973)

4.3 Fellers, R. and Taylor, J.,"Internal Reflections in Dielectric Prisms," IEEE Transactions on Microwave Theory and Techniques, pp 584–587, Nov. (1964)

4.4 Harrrick, N.J., "Use of frustrated total internal reflection to measure film thicknesses and surface releifs," Journal of Applied Physics, Vol. 13, No. 9, pp 2774–2775

4.5 Chakravarty, B. N., "On the diffraction of light incident at nearly the critical angle on the boundary between two media," pp 503–511??(1921)

4.6 Harrick, N. J., "A continuously variable optical beam splitter and intensity controller," Applied Optics, Vol. 2, No. 11, pp 1203–1204

4.7 Harrick, N. J.,"Electric Field Strengths at Totally Reflecting Interfaces," JOSA, Vol. 55, No. 7, pp 851–857, (1965)

4.8 Harrick, N. J., "Determination of refractive index and film thickness from interference fringes," Applied Optics, Vol. 10, No. 10, pp 2344–2349.(1971)

4.9 Harrick, N. J., and Carlson, A. I., "Internal reflection spectroscopy: Validity of effective thickness equations," Applied Optics, Vol. 10, No. 1, pp 19–23, (1971)

4.10 Harrick, N.J. and du Pre, F.K., "Effective thickness of bulk materials and of think films for internal reflection spectroscopy," Applied Optics, Vol. 5, No. 11,pp, 1739–1743 (1966),

4.11 Burghardt, T. and Thompson, N., "Evanescent intensity of a focused Gaussian light beam undergoing total internal reflection in a prism," Optical Engineering, Vol. 23, No. 1, pp 62–67, (1984)

4.12 Edwards, J., Ausserre, D., Hervet, H. and Rondelez, F., "Quantitative studies of evanescent wave intensity profiles using optical fluorescence," Applied Optics, Vol. 28, No. 10, pp 1881–1884, May (1989)

4.13 Renard, R., "Total Reflection: A new evaluation of the Goos–Hanchen shift," JOSA Vol. 54, No. 10, pp 1190–1197, Oct. (1964)

4.14 Lee, E., Benner, R., Fenn, J. and Chang, R., "Angular distribution of fluorescence from liquids and monodispersed spheres by evanescent wave excitation," Applied Optics, Vol. 18, No. 6, pp 862–868, (March 1979)
4.15 Allain, C., Ausserre, D., Rondelez, "Direct optical observation of interfacial depletion layers in polymer solutions," Physical Review Letters, Vol. 49, no. 23, pp 1694–1697, Dec. (1982)
4.16 Hall, E., "The penetration of totally reflected light into the rarer medium," Phys. Rev. , Vol. 15, pp 73–106, (1902)
4.17 Nassenstein, H, "Superresolution by diffraction of subwaves," Optics Communications, Vol. 2, No. 5, pp 231–234, Oct. (1970)
4.18 Carniglia, C. and Mandel, L., "Absorption and emission of evanescent photons," JOSA, vol. 62, No. 4, pp 479–486, April (1972)
4.19 Zhu, S., Yu, A., Hawley, D. and Roy, R., "Frustrated total internal reflection:a demonstration and review," Am. J. Phys., vol. 54, No. 7, pp 601–607, July, (1986)
4.20 Coleman, Z.A., Metz, M., Phillips, N.J., "Holograms in the extreme edge illumination geometry", Holographic Materials II, SPIE Proceedings vol. 2688, San Jose, CA (1996)
4.21 Bragg, W. L., "The diffraction of short electromagnetic waves by a crystal," Proc. Camb. Phil. Soc., Vol. 17, pp 43–57, (1912)
4.22 Suhara, T., Nishihara, H., Koyama, J., "Waveguide holograms: a new approach to hologram integration," Optics Communications, Vol. 19(3),pp 353–358, (1976)
4.23 Wood, V., Hartman, N., Verber, C., Kenan, R., "Holographic formations for gratings in optical waveguiding layers," Journal of Applied Physics, Vol. 46(3), pp 1214–1215, (1975)
4.24 Platte, W., "Effect of surface recombination on microwave performance of laser–induced DBR gratings in silicon coplanar waveguides," IEE Proceedings, Vol. 139(6), pp 399–401, (1992)
4.25 Saarinen, J. et al., "Computer–generated guided–wave holography:application to beam splitting," Optics Letters, Vol. 17(4), pp 300–302, (1992)
4.26 Syms, R., *Practical Volume Holography,* Clarendon Press, London, 1990
4.27 Chen, T. et. al., "Guided–wave planar optical interconnects using highly multiplexed polymer waveguide holograms," Journal of Lightwave Technology:a Joint IEEE/OSA Publication, Vol. 10(7), pp 888, (1992)
4.28 Saarinen, J. et. al.,"Computer–generated waveguide holograms by double–ion exchange processes in glass," Electronics Letters, Vol. 28(9), pp 876–877, (1992)
4.29 Bablumyan, A., Morozov, V., Putilin, A. and Shermergor, T., "Waveguide holograms in communication, storage, and information processing systems," Proceedings of the Lebedev Physics Institue, Academy of Sciences of the USSR, Vol. 185, pp 248–290, (1987)
4.30 Kostuk, R. et. al., "Reducing alignment and chromatic sensitivity of holographic optical interconnects with substrate–mode holograms," Applied Optics, Vol. 28(22), pp 4939–4944, (1990)
4.31 Kostuk, R., Kato, M. and Huang, Y.,"Polarization properties of substrate–mode holographic interconnects," Applied Optics, Vol. 29(26), (1990)

4.32 Xu, W., et al., "The dual twin image effect from a waveguide hologram," Optics and Lasers in Engineering, vol. 23, pp 137–143 (1995)
4.33 Jones, M., Kenan, R., and Verber, C., "Rectangular characteristic gratings for waveguide input and output coupling," Applied Optics, Vol. 34, No. 20 (1995)
4.34 Caulfield, H., Huang, Q., and Shamir, J., "Wide field of view transmission holography," Optics Communications, Vol 86, pp 487–490 (1991)
4.35 Putilin, A., Morozov, V., Huang, Q., Caulfield, H.,"Waveguide holograms, with white light illumination," Optical Engineering, Vol. 30 No. 10, pp 1615–1619 (1991)
4.36 Huang, Q., and Gilbert, J., "Diffraction properties of substrate guided–wave holograms," Optical Engineering, Vol. 34, No. 10, pp 2891–2899
4.37 Huang, Q., and Caulfied, H., "Waveguide holography and its applications," Practical Holography V, SPIE vol. 1461, pp 303–312 (1991)
4.38 Singher,L., and Shamir, J., "Waveguide holographic elements recorded by guided modes," Applied Optics, vol. 33 No. 7, pp 1180–1186 (1994)
4.39 Henrion, M., "Diffraction and exposure characteristics of the edgelit hologram," Maters thesis, MIT, 1995
4.40 Stetson, K.,"An analysis of the properties of total internal reflection holograms," Optik, Vol. 29(5), pp 520–536 (1969)
4.41 Phillips, N.J., Barnett, C., Wang, C., Coleman, Z.A., "Links between holography and lithography," SPIE vol. 2333, pp 206–214 (1995)

Chapter 5

Diffraction Theories

5.1 Introduction

This chapter describes the various theories used to model diffraction from a hologram. In particular, the coupled wave diffraction theory is analyzed, and the assumptions made in Kogelnik and Vasnetsov theories are evaluated for the regime of heavily slanted holograms. In order to verify or invalidate these theories, the rigorous form of couple wave diffraction of Moharam and Gaylord [5.4] is solved and a modification of the theory is made in order to rigorously compute the diffraction intensities for thick holograms with many possible space harmonics (pairs of diffracted orders).

Over the years, many people have attempted to describe the diffraction effects when a plane wave is incident on a dielectric medium with a sinusoidal variation in permittivity, such as the case in holography. These theories are all derived from Maxwell's equations and have been classed into two categories: 'dynamic' and 'kinematic' theories. The kinematic theory has important flaws–the main one being that power conservation is not satisfied. Thus, this theory is not usually used, and will

be disregarded here. All of the dynamic theories can be classified into three main groups: coupled wave, modal, and perturbation theory [5.1]. These theories vary according to how the multiple scattering within the media is described. The coupled wave theory uses differential equations, the modal theory uses simultaneous linear equations, and the perturbation theory uses integral equations. These three theories have been proven to be equivalent [5.2, 5.3]. Since the coupled wave theory is the most commonly used and is easy to associate with experimental results, it is the one examined here.

5.2 Coupled Wave Equations

First, conditions for which the diffraction will be modeled must be defined. In the work presented here, the light is TE polarized (H mode or S polarized) from a medium with dielectric constant, ε_1, incident on a slab phase grating with a dielectric permittivity which can be represented by

$$\varepsilon(x,z) = \varepsilon_2 + \varepsilon_m \cos[K(x\sin\phi + z\cos\phi)], \tag{5.1}$$

or in terms of the refractive index,

$$n(x,z) = n_2 + n_m \cos[K(x\sin\phi + z\cos\phi)] \text{ where} \tag{5.2}$$

$$\frac{\varepsilon_m}{\varepsilon_2} = \frac{2n_m}{n_2} \text{ and the commonly used term } \Delta n \text{ is defined by} \tag{5.3}$$

$$\Delta n = 2n_m. \tag{5.4}$$

K is the magnitude of the grating vector at an angle ϕ with the z–axis and is defined by $K = \frac{2\pi}{\Lambda}$ where Λ is the fringe spacing (distance between regions of constant permittivity) as indicated in Fig. 5.1.

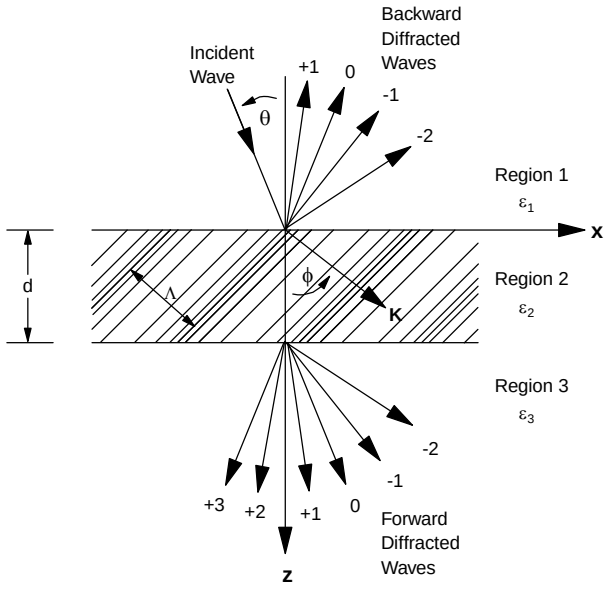


Figure 5.1 The geometry for diffraction from a slanted planar grating. [5.5]

The light leaves the opposite side of the grating through a medium with a dielectric constant ε_3. The grating does not vary in the y–direction and all three regions are assumed to have the permeability of free space. The following derivation follows closely that of Moharam and Gaylord [5.5] and uses their notation wherever possible.

The total electric field in region 1 can be thought of as a sum of the incident wave, plus the sum of all of the reflected (backward diffracted) waves. The backward diffracted waves in region 1 are as a result of grating (bulk) diffraction as well as boundary diffraction. In region 2, the electric field is a supposition of all of the forward and backward diffracted waves. In region 3, the electric field is simply the sum of all of the forward diffracted waves. These waves may be normalized and expressed by

$$E_1 = \exp\left[-j\left(\beta_0 x + \xi_{10} z\right)\right] + \sum_i R_i \exp\left[-j\left(\beta_i x - \xi_{1i} z\right)\right] \tag{5.5}$$

$$E_2 = \sum_i \hat{S}_i(z) \exp\left[-j\left(\beta_i x + \xi_{2i} z\right)\right] \tag{5.6}$$

$$E_3 = \sum_i T_i \exp\left[-j\left(\beta_i x + \xi_{3i}\left(z-d\right)\right)\right] \tag{5.7}$$

with

$$\beta_i = k_1 \sin\theta - iK\sin\phi\,, \tag{5.8}$$

$$k_l = \frac{2\pi\sqrt{\varepsilon_l}}{\lambda}, \tag{5.9}$$

$$\xi_{li}^{\ 2} = k_l^{\ 2} - \beta_i^{\ 2}, \text{ for regions 1 and 3, and} \tag{5.10}$$

$$\xi_{2i} = k_2 \cos\theta' - iK\cos\phi \text{ for region 2,} \tag{5.11}$$

where l =1, 2, 3 is the region index; i is the wave index; R_i and T_i are the normalized amplitudes of the ith reflected and transmitted wave; S_i is the amplitude of the ith wave in region 2; λ is the free space wavelength of the incident light; θ is the angle of incidence in region 1; and θ' is the angle of incidence inside region 2. The waves represented by Eqs. (5.5), (5.6), and (5.7) are phase matched at the two interfaces and derived with the assumption that the modulation inside region 2 is periodic. However, if the slant angle, ϕ, equals zero, the modulation is no longer periodic and there are a finite number of cycles. Thus, the Floquet theorem which was used to describe the waves [5.4] is no longer valid, and the rigorous chain–matrix method solution should be used [5.6], and not the "adaptation" of the rigorous coupled wave theory suggested by Zylberberg [5.7] as indicated by Moharam [5.8].

In order to accurately predict the diffraction behavior, the amplitudes of the waves are needed. These can be obtained by solving Maxwell's wave equation for the modulated region:

$$\nabla^2 E_2 + (2\pi/\lambda)^2 \varepsilon(x,z) E_2 = 0\,. \tag{5.12}$$

Thus, by substituting Eqs. (5.6) and (5.1) into (5.12), the differentiation yields the infinite set of coupled wave equations *without* the usual approximations:

$$\frac{\varepsilon_m}{8\varepsilon_2}\frac{d^2S_i(u)}{du^2} = (\cos\theta' - i\mu\cos\phi)\frac{dS_i(u)}{du} - \rho i(i-B)S_i(u) + S_{i+1}(u) + S_{i-1}(u) \tag{5.13}$$

with

$$S_i(u) = \hat{S}_i(z),\ u = \frac{j\pi\varepsilon_m}{2\lambda\sqrt{\varepsilon_2}}z = j\kappa z\,,\ \mu = \frac{\lambda}{\Lambda\sqrt{\varepsilon_2}}\,,\ \rho = \frac{2\lambda^2}{\Lambda^2\varepsilon_m}\,,\ \text{and} \tag{5.14}$$

$$\mathrm{B} = \frac{2\Lambda\sqrt{\varepsilon_2}}{\lambda}\cos(\phi - \theta')\,. \tag{5.15}$$

This version of the coupled wave equations is as stated in [5.4] and is simply derived as in Appendix 2 of Syms [5.1]. Eq. (5.13) is an infinite set of second order coupled differential equations. These equations do not have an exact solution. However, one may compute an approach to the solution by letting *i* reach very high values, and solving the resulting equations at the desired accuracy.

An approximate theory is much simpler to use and program than the rigorous theory. Some of the approximate theories can be ruled out for the slanted fringe regime without even the need for comparison with the rigorous results (such as Kogelnik [5.9]). Unfortunately some approximate methods (such as Vasnetsov [5.10]) must be compared with the rigorous solution to test their validity. By first solving the rigorous equations, one can then verify whether or not an approximate theory is valid for a certain diffraction regime (such as a hologram with fringes near 45°).

5.3 Rigorous Solution of the Coupled Wave Equations

The computational solution of the rigorous coupled wave (RCW) theory begins by following Moharam [5.5] and converting the equations to unforced state equations. The infinite set of second order coupled wave equations is represented by two first order infinite state equations. By representing them in a state–space, a solution can be obtained from computation of the eigenvalues and eigenvectors of the state–space coefficient matrix such as in Liu [5.11].

5.3.1 Solution of the Coupled Wave Equations

By substituting column vectors of S, S', and S" for S_i, dS_i/du, and d^2S_i/du^2, Eq. (5.13) can be re–written in state space as

$$\begin{bmatrix} S' \\ S'' \end{bmatrix} = \begin{bmatrix} \\ b_{rs} \\ \\ \end{bmatrix} \begin{bmatrix} S \\ S' \end{bmatrix}, \tag{5.16}$$

where the matrix b_{rs} is determined from the substitutions:

$$S_{1,i} = S_i(u), \tag{5.17}$$

$$S_{2,i} = \dot{S}_i(u). \tag{5.18}$$

By substituting Eqs. (5.17) and (5.18) into Eq. (5.13), the state equations for Eq. (5.13) can be written as

$$\dot{S}_{1,i} = S_{2,i}, \tag{5.19}$$

$$\dot{S}_{2,i} = aS_{1,i+1} + b_i S_{1,i} + aS_{1,i-1} + c_i S_{2,i} \text{ where} \tag{5.20}$$

$$a = \frac{8\varepsilon_2}{\varepsilon_m},\ b_i = -a\rho i(i - B)\text{, and } c_i = a(\cos\theta' - i\mu\cos\phi)\,. \tag{5.21}$$

One can rewrite Eqs. (5.19) and (5.20) in a matrix form as

$$\begin{bmatrix} \vdots \\ \dot{S}_{1,2} \\ \dot{S}_{1,1} \\ \dot{S}_{1,0} \\ \dot{S}_{1,-1} \\ \dot{S}_{1,-2} \\ \vdots \\ \\ \dot{S}_{2,2} \\ \dot{S}_{2,1} \\ \dot{S}_{2,0} \\ \dot{S}_{2,-1} \\ \dot{S}_{2,-2} \\ \vdots \end{bmatrix} = \begin{bmatrix} & & & \vdots & & & & & \vdots & & & \\ & 0 & 0 & 0 & 0 & 0 & & 1 & 0 & 0 & 0 & 0 & \\ & 0 & 0 & 0 & 0 & 0 & & 0 & 1 & 0 & 0 & 0 & \\ \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 & \cdots \\ & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 1 & 0 & \\ & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 1 & \\ & & & \vdots & & & & & \vdots & & & & \\ & & & & & & & & & & & & \\ & b_2 & a & 0 & 0 & 0 & & c_2 & 0 & 0 & 0 & 0 & \\ & a & b_1 & a & 0 & 0 & & 0 & c_1 & 0 & 0 & 0 & \\ \cdots & 0 & a & b_0 & a & 0 & \cdots & 0 & 0 & c_0 & 0 & 0 & \cdots \\ & 0 & 0 & a & b_{-1} & a & & 0 & 0 & 0 & c_{-1} & 0 & \\ & 0 & 0 & 0 & a & b_{-2} & & 0 & 0 & 0 & 0 & c_{-2} & \\ & & & \vdots & & & & & \vdots & & & & \end{bmatrix} \begin{bmatrix} \vdots \\ S_{1,2} \\ S_{1,1} \\ S_{1,0} \\ S_{1,-1} \\ S_{1,-2} \\ \vdots \\ \\ S_{2,2} \\ S_{2,1} \\ S_{2,0} \\ S_{2,-1} \\ S_{2,-2} \\ \vdots \end{bmatrix}. \tag{5.22}$$

Thus, Eq. (5.16) is represented in full matrix form in Eq. (5.22). The coefficients of the matrix b_{rs} are as defined above and the solution to the equation can be found in terms of the eigenvectors and eigenvalues of b_{rs}. Moharam and Gaylord [5.4] use a solution expressed in the form

$$S_i(u) = \sum_m C_m w_{i,m} \exp(q_m u)\,, \tag{5.23}$$

where w_{im} is the mth element (i.e. column m) of the row in the matrix [w] consisting of the eigenvectors (the eigenvectors are in the rows) where i is the ith row of the matrix corresponding to the ith order and q_m is the mth eigenvalue. The coefficients C_m are to be determined from the boundary conditions. Eq. (5.13) may also be written in terms of z as

$$\hat{S}_i(z) = \sum_m C_m w_{i,m} \exp(jq_m \kappa z) \tag{5.24}$$

With this expression for $S_i(z)$, one can note that the exponential term has the potential to reach extremely high values if the eigenvalues (q_m) have large, negative imaginary components. This is computationally very difficult for a computer to handle because the large values obtained are placed within the boundary conditions and the resulting system of equations must be solved without numerical overflow. When the *very large* values of the equations are placed with the other *very small* values, calculations are much more difficult for any computer and possible errors are generated. For the case of very deep gratings or when one is considering many harmonics (diffractive orders), some of the eigenvalues have large, negative imaginary components.

This difficulty occurs more often in the case of binary gratings and Moharam addresses the computational difficulties by exploiting the symmetry of binary gratings [5.12] and redefining the solution (such as redefining Eq. (5.24)). The symmetry involved with binary gratings allowed him to use only the positive eigenvalues. However, the symmetry of binary gratings is not present in slanted fringe gratings. The eigenvalues generated from slanted gratings do exhibit half with positive imaginary components, and half with negative imaginary components, though they are not necessarily equal nor complex conjugates. A solution to this type of computational problem in diffraction has not been published. Therefore, as suggested through conversation with Dr. Moharam, a new solution is created by replacing Eq. (5.24) with

$$\hat{S}_i(z) = \sum_m \left[P_{m+} w_{i,m+} \exp(jq_{m+} kz) + Q_{m-} w_{i,m-} \exp\left[jq_{m-} kz(z-d) \right] \right] \tag{5.25}$$

where P_{m+} and $w_{i,m+}$ are the coefficients and eigenvectors corresponding to the *positive imaginary eigenvalues* q_{m+}; and Q_{m-} and $w_{i,m-}$ are the coefficients and eigenvectors corresponding to the *negative imaginary eigenvalues* q_{m-}. Thus, one essentially changes the single wave solution of Eq. (5.24) into the supposition of two

separate waves traveling in opposite directions (the +z and +(z–d) directions). By separating the q_m values into q_{m+} and q_{m-} at the boundary conditions of z = 0 and z = d, the arguments of the exponential of Eq. 5.25 will always be zero or negative. This will preempt numerical overflow and computational difficulties.

In order to determine the coefficients P_{m+} and Q_{m-}, Eq. (5.25) must be placed back into Eq. (5.6), and then the waves must be solved at the boundary conditions of z=0 and z = d. At the boundary conditions (surfaces 1–2 and 2–3) the tangential component of the electric fields must be continuous as well as the tangential component of the magnetic fields. Since with H mode polarization, the component of the electric field is only in the y–direction, the tangential component is simply the electric field itself. The tangential component of the magnetic field, however, is in the x–direction and is determined by Maxwell's Equation

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t}, \text{ thus the tangential component for the magnetic field is} \quad (5.26)$$

$$H_x = \left(\frac{-j}{\omega\mu}\right)\frac{\partial E_y}{\partial z}. \quad (5.27)$$

By matching the electric and magnetic fields at the boundary conditions at z = 0 and z = d from Eqs. (5.5), (5.6), (5.7), (5.26), and (5.27), one forms the equations:

Tangential E at z = 0:

$$\delta_{i0} + R_i = E_2(z=0), \quad (5.28)$$

Tangential H at z = 0:

$$\xi_{1i}(R_i - \delta_{i0}) = \frac{dE_2}{dz}(z=0), \quad (5.29)$$

Tangential E at z = d:

$$T_i = E_2(z = d)\,, \tag{5.30}$$

Tangential H at z = d:

$$-\xi_{3i} T_i = \frac{dE_2}{dz}(z = d)\,, \tag{5.31}$$

where δ_{i0} is the Kroneker delta function and is zero unless i=0. Putting Eqs. (5.25) and (5.6) into Eqs. (5.28) to (5.31) yields

Tangential E at z = 0:

$$\delta_{i0} + R_i = \sum_m P_{m+} w_{i,m+} + Q_{m-} w_{i,m-} X_{m-}\,, \tag{5.28}$$

Tangential H at z = 0:

$$\xi_{1i}(R_i - \delta_{i0}) = \sum_m P_{m+} w_{i,m+} \left(q_{m+}\kappa - \xi_{2,i}\right) + Q_{m-} w_{i,m-} \left(q_{m-}\kappa - \xi_{2,i}\right) X_{m-}\,, \tag{5.29}$$

Tangential E at z = d:

$$T_i = \sum_m P_{m+} w_{i,m+} X_{m+} Y_i + Q_{m-} w_{i,m-} Y_i\,, \tag{5.30}$$

Tangential H at z = d:

$$-\xi_{3i} T_i = \sum_m P_{m+} w_{i,m+} X_{m+} Y_i \left(q_{m+}\kappa - \xi_{2,i}\right) + Q_{m-} w_{i,m-} Y_i \left(q_{m-}\kappa - \xi_{2,i}\right)\,, \tag{5.31}$$

where

$$X_{m+} = \exp(jq_{m+}\kappa d)\,,\ X_{m-} = \exp(-jq_{m-}\kappa d)\,, \text{ and } Y_i = \exp(-j\xi_{2,i}d)\,. \tag{5.32}$$

Eliminating T_i and R_i from Eqs. (5.28) through (5.31) gives

$$\sum_m P_{m+}\left[\left(\xi_{1,i} + \xi_{2,i} - q_{m+}\kappa\right)w_{i,m+}\right] + Q_{m-}\left[\left(\xi_{1,i} + \xi_{2,i} - q_{m-}\kappa\right)w_{i,m-}X_{m-}\right] = 2\xi_{1,i}\delta_{i0}\,, \tag{5.33}$$

$$\sum_m P_{m+}\left[\left(\xi_{2,i} - \xi_{3,i} - q_{m+}\kappa\right)w_{i,m+}X_{m+}\right] + Q_{m-}\left[\left(\xi_{2,i} - \xi_{3,i} - q_{m-}\kappa\right)w_{i,m-}\right] = 0\,. \tag{5.34}$$

In matrix form, Eqs. (5.33) and (5.34) can be written as

$$\begin{bmatrix} W_+G_1 - V_+W_+ & X_-W_-G_1 - X_-V_-W_- \\ X_+W_+G_3 - X_+V_+W_+ & W_-G_3 - V_-W_- \end{bmatrix}\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} D \\ Z \end{bmatrix}, \text{ where} \tag{5.35}$$

n is the number of space harmonics retained in the calculations and

W_+ is the (n x n) eigenvector matrix with the eigenvectors of matrix b_{rs} corresponding to the eigenvalues with the *positive* imaginary components in the *columns*,

W_- is the (n x n) eigenvector matrix with the eigenvectors of matrix b_{rs} corresponding to the eigenvalues with the *negative* imaginary components in the *columns*,

V_+ is the (n x n) diagonal matrix of eigenvalues with *positive* imaginary components obtained from matrix b_{rs},

V_- is the (n x n) diagonal matrix of eigenvalues with *negative* imaginary components obtained from matrix b_{rs},

X_+ is the (n x n) diagonal matrix with the diagonal elements equal to $\exp(jq_{m+}\kappa d)$ where q_{m+} are the eigenvalues with the *positive* imaginary components,

X_- is the (n x n) diagonal matrix with the diagonal elements equal to $\exp(-jq_{m-}\kappa d)$ where q_{m-} are the eigenvalues with the *positive* imaginary components,

G_1 is the (n x n) diagonal matrix with the *i*th diagonal equal to $\xi_{1,i} + \xi_{2,i}$,

G_3 is the (n x n) diagonal matrix with the *i*th diagonal equal to $\xi_{2,i} - \xi_{3,i}$,

P is the (n x 1) matrix with the coefficients corresponding to P_{m+},

Q is the (n x 1) matrix with the coefficients corresponding to Q_{m-},

D is the (n x 1) matrix with all of the matrix elements zero except the element corresponding to *i*=0, which has the value of $2\xi_{1,0}$,

Z is the (n x 1) null matrix.

The system of linear equations represented by Eq. (5.35) can be solved for *P* and *Q* by various techniques such as Gaussian elimination. One may note that there are *2n* systems of equations with *2n* constants to be determined (*n* for P_{m+} and *n* for Q_{m-}).

Once P_{m+} and Q_{m-} have been determined, these can be put back into Eqs. (5.28) and (5.30) to solve for the amplitudes of the reflected and transmitted waves. The diffraction efficiencies of these waves are then given by

$$DE_{1,i} = \mathrm{Re}\left(\frac{\xi_{1,i}}{\xi_{1,0}}\right) R_i R_i^* , \tag{5.36}$$

$$DE_{3,i} = \mathrm{Re}\left(\frac{\xi_{3,i}}{\xi_{1,0}}\right) T_i T_i^* , \tag{5.37}$$

where the quantity $\mathrm{Re}\left(\frac{\xi_{l,i}}{\xi_{1,0}}\right)$ is the ratio of cosines of the *i*th wave to the zero order wave inside the *l*th medium. These results are verified by the law of conservation of power such that

$$\sum_i \left(DE_{1,i} + DE_{3,i}\right) = 1 \tag{5.38}$$

The results of the rigorous coupled wave theory are shown in §5.5 and were computed from a program written to calculate the results in the matrix based computational software MATLAB® (Appendix C).

5.3.2 Limitations of the Rigorous Coupled Wave Solutions

There are two main difficulties with the current format of the rigorous coupled wave equations. The setup of the equations *does not allow for angles of incidence or diffraction at exactly 90°*. The solutions are not valid in these regimes. The Floquet condition, upon which this rigorous solution is based, is valid only for periodic gratings. To define a boundary for the diffracted (or incident wave) at 90°, one must have a finite number of fringes. Thus, the Floquet condition would not be valid. This limitation surprisingly arose when comparing the rigorous coupled wave theory with the Vasnetsov two wave solution (§5.4.2). The limitation occurred when the angle of diffraction of the *second* order was at 90°. This discrepancy shows that the rigorous coupled wave solution is invalid in all regimes where *any diffraction order is at 90°*. The rigorous theory seems to be continuous and is deemed to be valid at reference or replay angles approaching 90° from either side.

While the calculations presented here are only for cosinusoidal index modulated phase gratings, the input into the rigorous method can be adapted to include any type of modulation phase or amplitude grating. This can be inputted by breaking down any grating profile into its Fourier harmonics to any level of desired accuracy. The rigorous theory has been calculated for both TE, TM or arbitrary degree of polarization [5.15] although only TE mode has been examined here.

5.4 Approximate Theories

Because many researchers have not been interested in the full rigorous solutions of the coupled wave equation, there have been many approximate solutions to suit the more common regimes or those in which they were working. As a result there have been many proposed theories or solutions to the equations–some of which are equivalent, and some of which are only applicable to very few situations. Gaylord and Moharam depicted the differences between the various theories in [5.5] as indicated in Fig. 5.2.

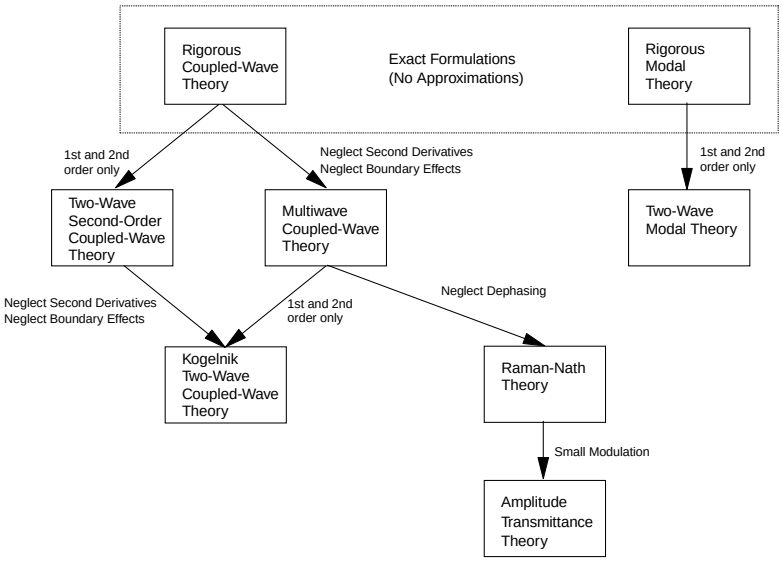


Figure 5.2 The planar diffraction theory hierarchy (Gaylord and Moharam [5.5]).

The method most referred to for analyzing diffraction from a hologram is Kogelnik's coupled wave theory [5.9]. This theory, while simple and easy to use for some gratings, is only an approximate theory. Unfortunately, Kogelnik's theory is not valid for steep–reference holograms, mainly because it leaves out the second order derivatives. This will be shown in the next section, as his theory can be shown to be invalid for steep slanted or highly modulated holograms without the need to compare with the rigorous coupled wave theory.

5.4.1 Kogelnik's Two–Wave Coupled Wave Theory

Kogelnik's coupled wave equations [5.9] do not involve the second order derivatives, boundary diffraction effects or contain information regarding higher orders. His second order equations are reduced to simple first order equation for phase gratings:

$$R'\cos\theta = -j\kappa S \tag{5.39}$$

$$\left(\cos\theta - \frac{K}{\beta}\cos\phi\right)S' + j\vartheta S = -j\kappa R \tag{5.40}$$

where R, S, R', S' are the reference wave, signal wave and their first derivatives respectively, θ is the angle of incidence, α is the average absorption constant, j is $\sqrt{-1}$, κ is the coupling constant, K is the grating vector, β is the propagation constant, ϕ is the grating slant angle, and ϑ is the dephasing measure.

The effect of the second order derivatives is dependent on the grating *slant angle*, ϕ, and the modulation. One may hypothesize with Kogelnik [5.9] that the second order components do not significantly affect the equation. Kogelnik assumes that "the energy interchange between S and R is slow and that the energy is absorbed slowly, if at all." However, when the slant angle approaches 45°, as in the regime of steep

referenced holograms, or when the modulation reaches high values (such as $n_m = 0.06$ with some of Du Pont's photopolymers), this may not be true. A slant angle of 45° is the boundary between reflection and transmission holograms. (One may also refer to the *fringe angle* when classifying a hologram because it is simply half of the sum of the two recording angles. The fringe angle is the angle between the K vector and the *x–axis* as in Fig. 5.1 and is simply 90–ϕ where ϕ is the slant angle defined as the angle between the K vector and the *z–axis*).

In order to determine if the second order derivatives are valid, the grating must first be defined, specifically the fringe angle. In the classification of steep referenced holograms here, the fringe angle is usually between 40° and 50°. In waveguide holograms, the fringe angle is much more relaxed than steep reference holograms and one might hypothesize as Huang [5.13] that the second order derivatives might be ignored.

In a paper on the diffraction properties of substrate guided–wave holograms, Huang [5.13] limited the second–order derivatives to ten times less than the first order derivatives. With this restriction, silver–halide waveguide holograms can use the first–order derivative solution for fringe angles from 0°–41.25° and 45.15°–90° for the parameters given in the paper. These results essentially suggest that reflection silver halide waveguide holograms can use the Kogelnik first–derivative solution for the coupled wave equations, and transmission waveguide holograms can use Kogelnik's solution unless the reference angle is such that the fringe angle is greater than 41°. Huang's results must be taken one step further to show that the validity of the theory is still very dependent on the *index modulation and thickness* when combined with the heavily slanted fringes.

In order to verify the range for which the first–order solution is valid for slanted holograms, one can first assume that the first order solution holds true. Then, with this solution, the first order derivatives must be compared with the second order derivatives. If the second order derivatives are, for example, one tenth of the value of the first order derivatives, then the approximation which ignores the second order derivatives is not valid. Mathematically, Huang [5.13] represented this by

$$\left|\frac{\partial^2 \Psi_e}{\partial z^2}\right| << \left|\frac{\partial \Psi_e}{\partial z}\right| \quad \text{where} \tag{5.41}$$

Ψ_e is the complex envelope of the waves propagating inside the grating which after substituting in the first order solutions becomes

$$\nu|\tan \nu| << k_2 d \cos\theta_0 \quad \text{for reflection and} \tag{5.42}$$

$$\nu|\tanh \nu| << k_2 d \cos\theta_0 \quad \text{for transmission holograms} \tag{5.43}$$

where

$$\nu = \frac{d k_2 n_m}{2 n_2 \sqrt{\pm \cos\theta_r \cos\theta_s}}, \text{ and } k_2 = \frac{2\pi n_2}{\lambda}, \tag{5.44}$$

θ_r , θ_s and θ_0 are the reference, signal, and replay angles from the film surface normal, respectively, and the other parameters are as defined in the previous section with the assumption of replay at the first Bragg angle, $\theta_0=\theta_r$. One may suggest as Huang that the second derivatives become influential when they are one–tenth of the value of the first derivatives. Huang illustrated for *one simple case* of silver–halide holograms where the first order derivative approximation was valid. For his case of silver halide with a 7μm thickness and a modulation of $n_m = 0.01$, the approximation was valid for reference angles between 0° and 82.5° for transmission holograms and between 90.3° and 180° for reflection holograms. However, he did not expand his results to include different modulations or other recording materials. His results are expanded here (showing the regimes where the Kogelnik approximation is *not valid*) for variations in modulation and thickness for various recording angles in silver halide and photopolymer holograms (Fig. 5.3).

Huang shows that for his particular recording case of substrate guided wave holograms, he can use Kogelnik's theory. However, by expanding the calculations,

one can see that if the thickness increases, Fig. 5.3(d), or the modulation increases, Fig. 5.3(b), one can not use Kogelnik's theory for silver halide holograms. For transmission silver halide holograms, one can not effectively use Kogelnik's theory at all for thicknesses of 32 μm. While many of the commercial film producers such as AGFA produce at 7 μm, some holographers coat emulsions much thicker. For silver halide reflection holograms, one is essentially safe using Kogelnik's approximation unless recording an *extremely* steep referenced hologram (or substrate guided hologram). The index modulation does not usually affect the use of Kogelnik's theory for reflection holograms because the modulation would not be extremely high (such as the $n_m = 0.08$ required for $\theta_r \cong 95°$) to invalidate Kogelnik's theory.

Huang uses Kogelnik's theory to make claims regarding the angular and wavelength bandwidths. Huang's angular and wavelength analysis of edge–referenced reflection holograms illuminated from the edge does show the correct trends when compared to the rigorous theory of §5.5, however, the results are still valid only for some regimes. As a result, Kogelnik's theory is not applicable for analyzing steep referenced holograms and should only be used in perhaps a limiting single case and should not be generally applied.

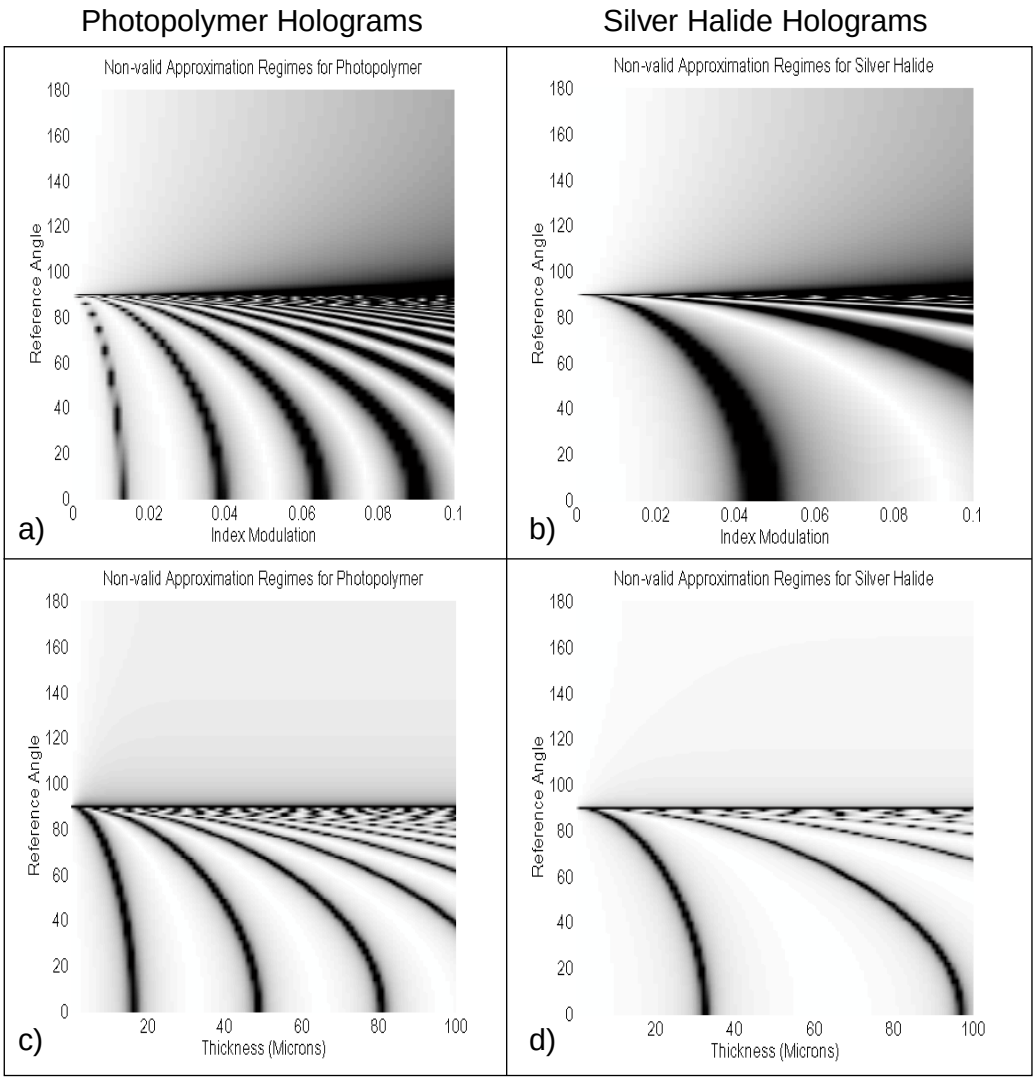


Figure 5.3 The regimes where the Kogelnik approximation of ignoring the second derivatives is not valid. The black areas (the invalid regimes) represent where $\nu|\tan\nu| \geq (0.1)k_2 d\cos\theta_0$ for reflection or $\nu|\tanh\nu| \geq (0.1)k_2 d\cos\theta_0$ for transmission holograms. The gray shading areas represent values proportionately less than 0.1. (b) and (d) are for silver halide holograms where n_2=1.63, λ=647 nm, with n_m = 0.01 for (d) and d=7 μm for (b). (a) and (c) are for photopolymer holograms where n_2=1.5, λ=647 nm, with n_m = 0.02 for (c) and d=25 μm for (a).

Huang only analyzes holograms with a fixed signal angle at 0°, and does not examine the efficiency, wavelength or angular bandwidth with respect to signal angle illumination. In other words, Huang does not examine grazing diffraction. Another approximate theory suggested specifically for very steep referenced holograms and grazing diffraction was described by Vasnetsov et. al. [5.10].

5.4.2 Vasnetsov's Two Wave Second Order Coupled Wave Theory

In 1977, Kong [5.14], expanded Kogelnik's work to include the second order derivatives and analyzed grazing diffraction for *unslanted* holograms. Vasnetsov et al. [5.10], further expanded this theory for grazing diffraction of heavily *slanted* fringe holograms. Vasnetsov's solution takes the two wave coupled wave equation and omits the terms corresponding to the -1 order because they are evanescent (Vasnetsov refers to them as "non–existent") in the case of grazing diffraction from a heavily slanted fringe grating (44.45° in the case examined). Also, in Vasnetsov's solution, one can not account for higher orders, such as those which occur in an unslanted hologram. Thus, one can rule out Vasnetsov's solution for slightly slanted or very thin transmission holograms which normally exhibit efficiencies in the higher harmonics.

In order to determine the exact regimes where Vasnetsov's theory can be used, the results for Vasnetsov's calculations were compared with results from the rigorous coupled wave theory and are shown in Fig. 5.4. The dark areas correspond to regimes where the difference in first order diffraction efficiencies between the rigorous method and Vasnetsov's method are greater than or equal to 5%. The gray areas correspond to values proportionately close to 5%. In other words, the black areas of Fig. 5.4 are regimes where one can not use Vasnetsov's theory.

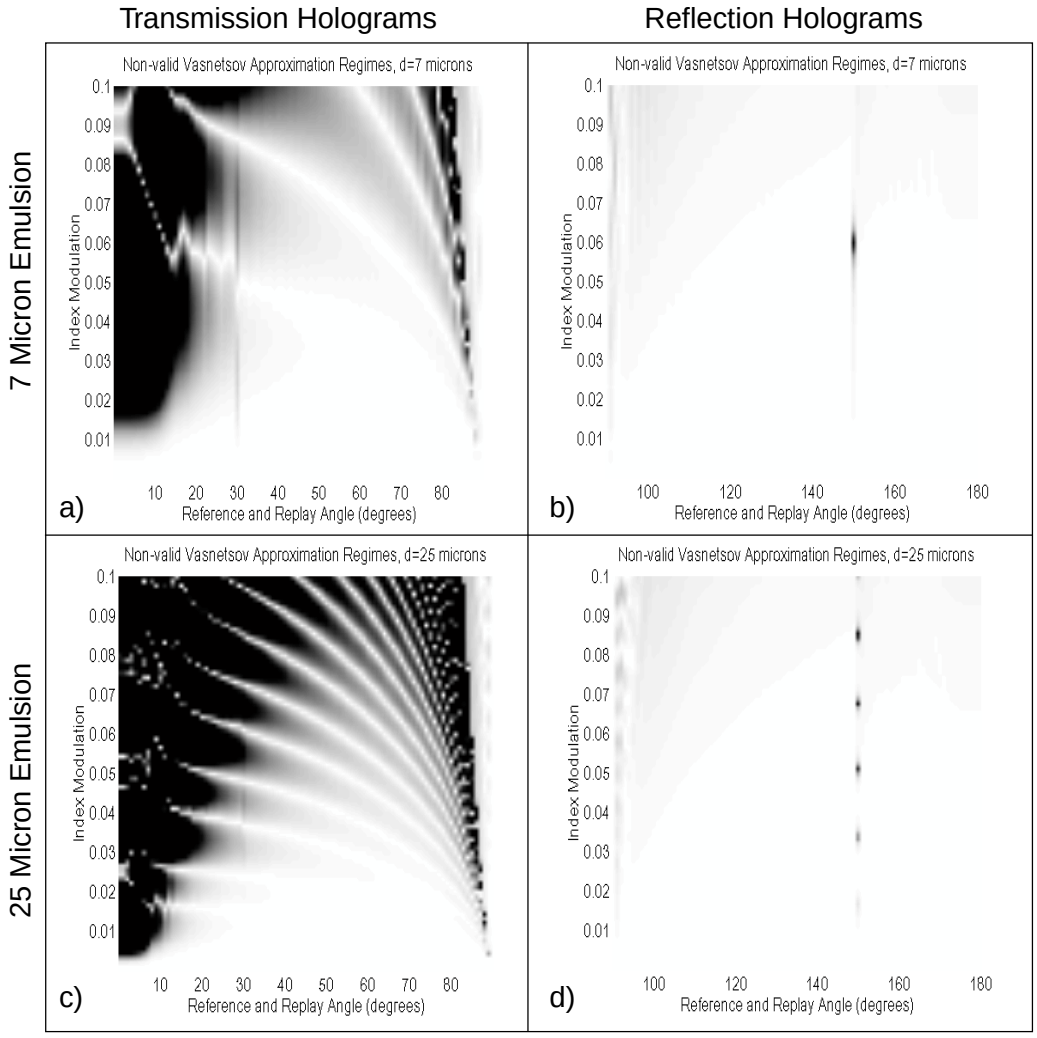


Figure 5.4 The regimes where Vasnetsov's two–wave coupled wave theory is not valid. The black areas correspond to a 5% or more error in the computed diffraction efficiency of the first order. The faint lines or marks at 30° for transmission holograms and 150° for reflection holograms correspond to regimes where the second order of diffraction is at 90° and the rigorous coupled wave theory is not valid. The holograms were recorded with a 0° signal angle and the reference and replay angle varied for transmission or reflection cases. The recording and replay wavelength was 647 nm in air and the indices of refraction are $n_1 = n_2 = n_3 = 1.5$. In (a) and (b), the emulsion is 7 μm thick, while in (c) and (d) the emulsion is 25 μm thick.

As can be seen from Fig. 5.4, Vasnetsov's theory can essentially be used for reflection holograms. However, in transmission holograms, a high refractive index modulation severely limits the applicable regimes. Also, as the fringe angle approaches 45° (the reference angle approaches 90°), the refractive index modulation at which the theory

becomes inapplicable decreases. Interestingly, the regime just before 45° degrees *is valid* for most index modulations. The example examined by Vasnetsov falls into this regime and is therefore valid. However, the theory is not valid at most refractive index modulation values when the fringe angle is near 43° (this invalid regime is larger with thicker emulsions as can be seen in Fig. 5.4 (a) and (c)). Therefore, like Kogelnik's theory, the range of Vasnetsov's theory is very limited and should not be used in the general case of steep reference holograms.

The faint lines on Fig. 5.4 at 30° and 150° are actually areas where the *rigorous coupled wave theory is not valid*. As explained in §5.3.1, the rigorous theory is not valid where the angle of diffraction for any order is 90°. Thus, in these very limited cases where the 2nd order of diffraction is at exactly 90°, the Vasnetsov calculations are actually closer to being rigorously correct.

5.5 Theoretical Diffraction Characteristics using RCW theory

Holograms inherently have many important qualities which enable them to be useful in many optical applications. Some of the most important qualities include diffraction efficiency, angular sensitivity, wavelength sensitivity, and polarization sensitivity. All of these characteristics are crucial to the performance of a hologram as an optical element and can be modeled using the RCW theory.

Using the RCW theory one can draw conclusions about the characteristics that are not analytically obvious from the coupled wave equations. Thus, for example, one could predict how the wavelength bandwidth will vary with slant angle or index modulation. In this section the angular sensitivity, wavelength sensitivity and diffraction efficiency will be analyzed with regards to the normal variables of slant angle, index modulation, and thickness. Polarization sensitivity is not examined here due to the additional complication involved with writing the computer program, although basic RCW theory can be expanded to include arbitrary polarization [5.15]. As most of the

traditional holographic recordings are done with TE polarized light, this is the polarization examined here.

In all of the graphs in this section, θ_s is the recording signal angle and is always equal to 0°, θ_r is the recording reference angle, θ_{replay} is the replay angle, n_{mod} is the amplitude of the refractive index modulation, d is the film thickness, λ_{record} and λ_{replay} are the recording and replay wavelengths in air respectively. Therefore, the fringe angle of the hologram is always half of the reference angle. To analyze only the diffractive effects within the medium (thus ignoring boundary diffraction), the hologram is surrounded by a medium with the same average index of refraction (i.e. $n_1 = n_2 = n_3$). All of the data for the graphs was created while retaining at least 3 harmonics (+3 order to -3 order).

The RCW theory can analyze the diffraction efficiency in any order desired, while most common orders examined are those corresponding to 0, +1, and -1 for the forward and backward diffracted waves. The graphs in this section are representations of 3–D graphs where the darker the area, the higher the diffraction efficiency. Therefore, conclusions about the diffraction efficiency can be made from all of them. Since angular and wavelength bandwidth are usually defined based on the diffraction efficiency, the effects of certain physical parameters on diffraction efficiency will be examined first.

5.5.1 Diffraction Efficiency

The diffraction efficiency of the first diffracted order of a hologram will vary depending on the fringe angle, index modulation, and the thickness (Fig. 5.5).

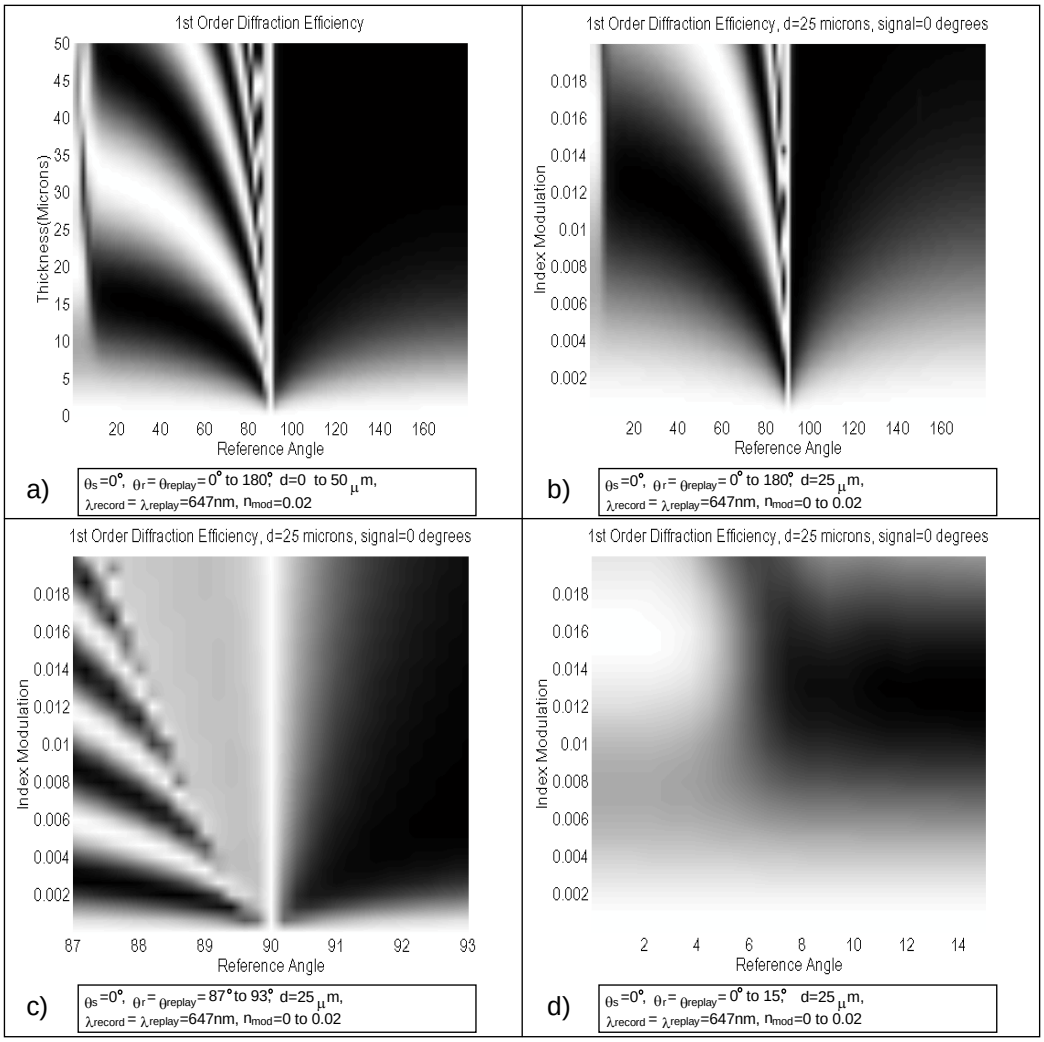


Figure 5.5 The diffraction efficiency of a hologram for varying thicknesses, fringe angles, and index modulations. The black areas represent diffraction efficiency near 100% with the gray areas proportionately less. (a) shows the effect of varying the thickness on the efficiency of a hologram with various fringe angles. In (b) the effect of the refractive index modulation on diffraction efficiency is shown for all fringe angles. (c) and (d) are enlargements of (b) showing the regions near 90° and 0° respectively.

From Fig. 5.5, one can see the expected modulation of the diffraction efficiency for transmission holograms as the refractive index modulation or thickness is increased. For reflection holograms, one can see the saturation of diffraction efficiency when the refractive index modulation or the thickness is increased. As the fringe angle approaches 45°, the frequency of the modulation of diffraction efficiency for transmission holograms seems to increase (Fig. 5.5(a)). However, as the fringe angle becomes *very close* to 45° for a transmission hologram, the diffraction efficiency does

not vary with index, and only one index of refraction modulation will yield the maximum diffraction efficiency (Fig. 5.5(c) at 89.5° for example).

Upon examining Fig. 5.5(a) and (b), one can also see that the minimum refractive index modulation required for maximum diffraction efficiency is reduced the closer the fringe angle is to 45° for a reflection or transmission hologram. This has important implications if the holographic recording medium has an inherent low refractive index modulation.

5.5.2 Wavelength Selectivity

The Bragg condition defines the possible angle and wavelength combinations for efficient diffraction. When replaying at a Bragg angle with a different wavelength, the diffraction efficiency will decrease. The rate at which the diffraction efficiency falls off for deviations from the Bragg wavelength is described as the hologram's wavelength selectivity. The wavelength bandwidth of the hologram is a measured value used to describe the wavelength selectivity. Wavelength bandwidth is usually defined as the wavelength range at which the diffraction efficiency is above half of the maximum value (otherwise known as Full Width at Half Maximum, FWHM).

The wavelength selectivity of holograms has been analyzed by many different methods, the most common of which is based on Kogelnik's theory [5.9]. In general, transmission hologram are associated with low wavelength selectivity (high wavelength bandwidth) and reflection holograms are associated with high wavelength selectivity (low wavelength bandwidth). Most theories suggest a continuous decrease in wavelength bandwidth as the fringe angle increases from a transmission hologram to a reflection hologram. Leith et. al. [5.16] in 1966 used a kinematic theory to model the wavelength and angular selectivity. Unfortunately, his kinematic theory can not account for interaction between the zero–order wave and the diffracted first order wave. This, in addition to the fact that he states within the paper that his theory is inapplicable at angles approaching 90°, would rule out this theory as a basis for

determining the angular or wavelength bandwidth for steep reference angle holograms. However, Birner [5.17] and then Henrion [5.18] each used the assumptions made by Leith for bandwidth analysis in their thesis on edge–lit holograms.

The RCW theory supports the diffraction characteristics of examples seen in the lab and it suggests very different wavelength bandwidth results from those of Leith. which. The diffraction efficiency of the first order when light is incident at the reference angle is equivalent to that when the light is incident at the signal angle (for all lossless, uniform, volume, phase holograms surrounded by a medium with the same average refractive index). The RCW theory predicts this as well as most other theories. However, one might not realize that the angular and wavelength bandwidths are going to differ very significantly (in the case of a slanted volume hologram). The asymmetry introduced by a slanted grating introduces a strong asymmetry in the bandwidth characteristics as well. A simple way to see this asymmetry is to realize that when a replay beam is incident at a steep reference angle, the effective thickness of the film is much larger than when the light is incident at a 0° signal angle. The results of this asymmetry in wavelength bandwidth can be seen clearly in Fig. 5.6

In Fig. 5.6, (a) and (b) represent the first order diffraction characteristics versus replay wavelength of a hologram with a refractive index modulation of $n_{mod} = 0.003$ as the fringe angle varies from 0° to 90°. Similarly, (c) and (d) with $n_{mod} = 0.02$ and (e) and (f) with $n_{mod} = 0.04$ show the wavelength bandwidth characteristics where the replay light is incident at the reference angle (0° to 180°) in the left column and the light is incident at the signal angle (0°) in the right column.

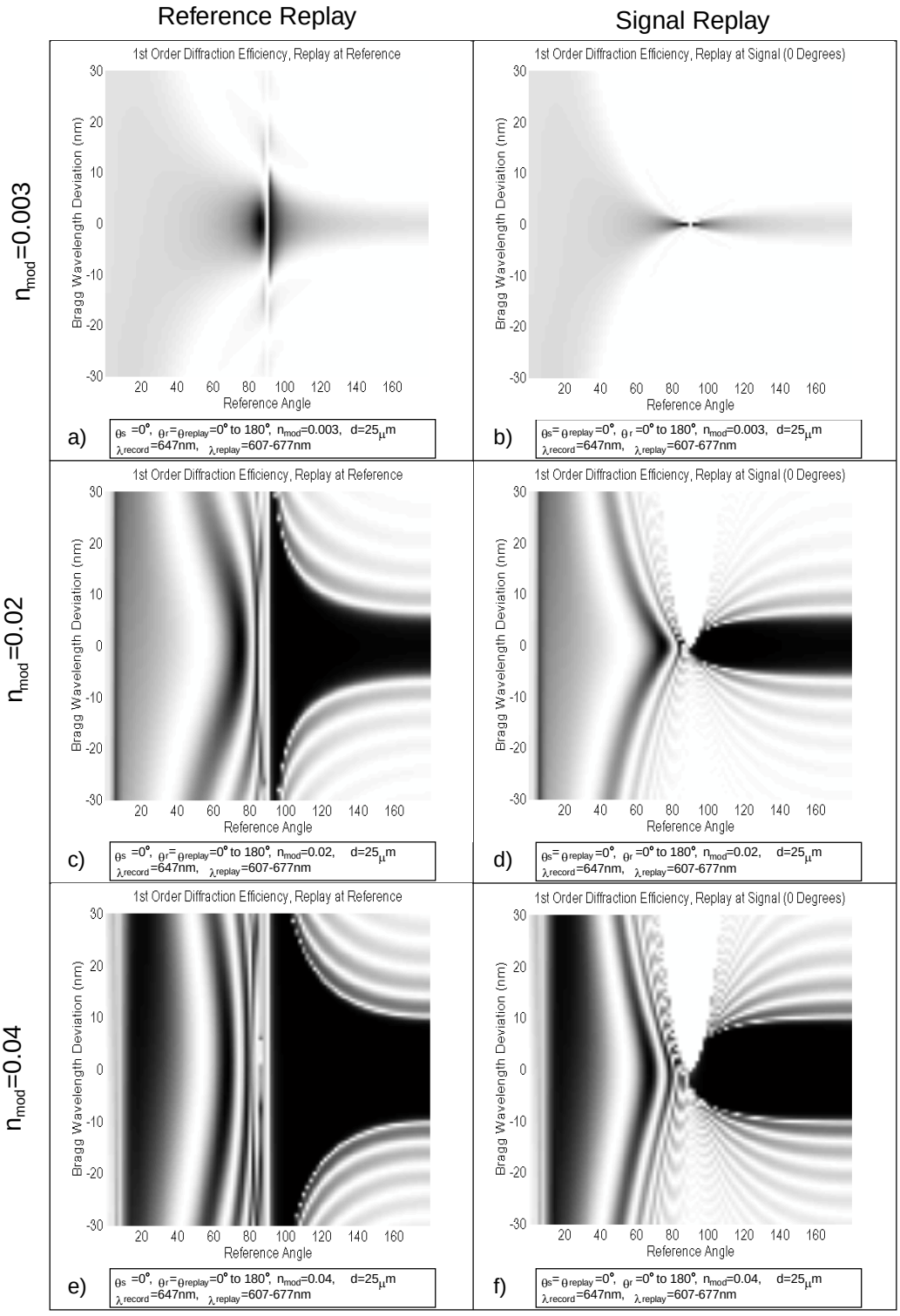


Figure 5.6 The dependence of wavelength selectivity on fringe angle and replay angle for three refractive index modulations for volume phase holograms. The black areas represent diffraction efficiency near 100% with the gray areas proportionately less.

Firstly, one can easily see from Fig. 5.6 that the wavelength bandwidth is smaller with signal replay than with reference replay. The vertical linear bands on the left sides of the graphs represent the large bandwidth associated with transmission holograms. One can see that these bands begin to curve (or bend) the closer the fringe angle is to 45° as seen easily from (c) and (d). This suggests that the wavelength bandwidth decreases with increasing fringe angle as suggested by most diffraction theories. Also, when looking at a constant reference angle (choosing one which corresponds to a high diffraction efficiency) the transmission bands seem to bend as the replay is changed from the reference angle to the signal angle. This corresponds to a reduced wavelength bandwidth. This is noticed mostly moving from (c) to (d). The bending is not as significant from (e) to (f) because of the effects of the high index modulation.

At first look, the diffraction characteristics of the transmission holograms in Fig. 5.6 seem to generally agree with accepted theory. However, upon close inspection, one can see that from (b) that an efficient transmission hologram can be made with a very small wavelength bandwidth. If the reference angle is just less than 90° and the refractive index modulation is low, the resulting bandwidth is very low (at a reference angle of 85° in Fig. 5.6 (b), the efficiency is 89% and the wavelength bandwidth (FWHM) is near *1.25 nm* with very small sidelobes).

Reflection gratings with a steep reference angle also seem to have diffraction characteristics not predicted by previous theories. While for *reference* replay the wavelength bandwidth appears to be continuously decreasing from the transmission case through the reflection case, the wavelength bandwidth for *signal* replay appears to decrease as the reference beam angle moves from 0° to 90°, then increases as the reference angle approaches 180°. From Fig. 5.6 (e), one can see that the wavelength bandwidth of a reflection hologram can be high as well. This can be viewed another way with regards to the asymmetry. As the reference angle approaches 90° for a reflection hologram the wavelength bandwidth increases if replayed at the reference angle and the wavelength bandwidth decreases if replayed at the signal angle.

One should note that the results in Fig. 5.6 are self–consistent. The values on the far left side of the graphs in the left column approach those on the far left side in the right

column, and similarly with the right sides. These results are as expected when the reference angle approaches the signal angle.

The three rows in Fig. 5.6 indicate that in general, the bandwidth increases with increasing index modulation for transmission and reflection holograms (assuming one looks at the next high diffraction peak in a transmission hologram). This effect is seen more clearly in Fig. 5.7.

The white lines at exactly 90° in Fig. 5.6 are due to the fact that the rigorous coupled wave theory is not valid at 90° incidence (left column) or diffraction (right column). Also, the white areas of parabolic shape in the top middle of (b), (d), and (f) are the off–Bragg conditions of zero efficiency which result when the first order diffracted beam is evanescent.

For the transmission hologram of Fig. 5.7 (a) and (c) one can see that moving to the next diffraction efficiency maximum (dark area) when increasing the refractive index modulation will increase the bandwidth. A vertical line drawn at $n_{mod} = 0.015$ in (c) will essentially cross more black area than a line at $n_{mod} = 0.005$. One must be careful to choose a line at the peak diffraction efficiency for these conclusions to be valid. The peak diffraction efficiency here is represented by a dark area, and it could be assumed for these illustrations that the peak efficiency was at an index modulation in the middle of the dark patch on the horizontal axis. Therefore, one could conclude from Fig. 5.7 that the wavelength bandwidth increases with increasing index modulation for transmission and reflection holograms.

The conclusion drawn from Fig. 5.6 that the wavelength bandwidth is smaller when the replay angle is at the signal angle as opposed to the reference angle is seen much more clearly in Fig. 5.7 for both reflection and transmission holograms.

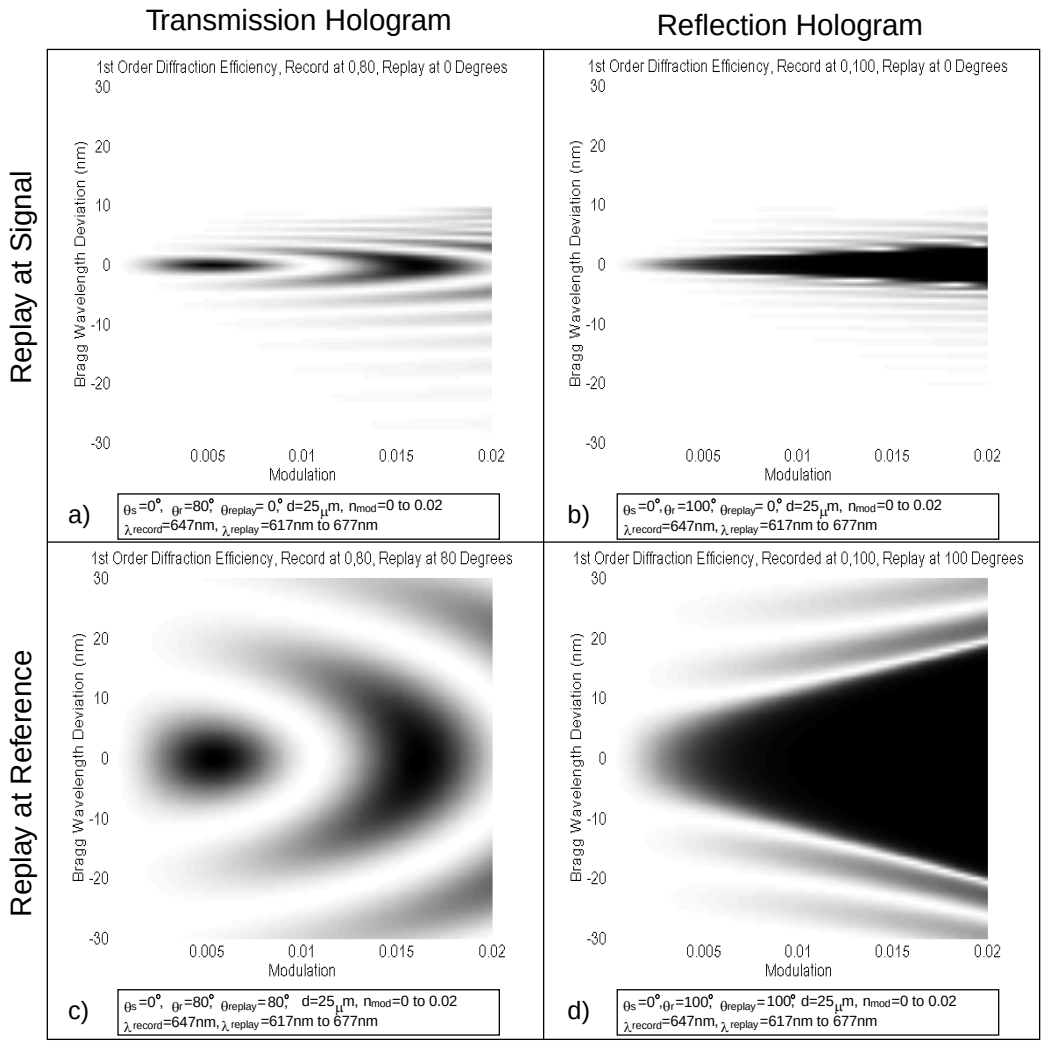


Figure 5.7 The dependence of wavelength bandwidth on index modulation for a transmission and reflection hologram. The wavelength bandwidth is shown for a transmission hologram in (a) and (c) and a reflection hologram in (b) and (d) for replay at the signal and reference angles.

The data in Fig. 5.7 is consistent with the accepted theory that the diffraction efficiency at the Bragg condition (wavelength deviation of 0°) is the same whether the replay angle is the signal or reference. This can be seen on the graphs by a horizontal line at a wavelength deviation of 0° passing through bright and dark areas at the same index modulations for (a) and (c) and then again for (b) and (d).

The thickness of a hologram can also affect the wavelength bandwidth as is seen from Fig. 5.8.

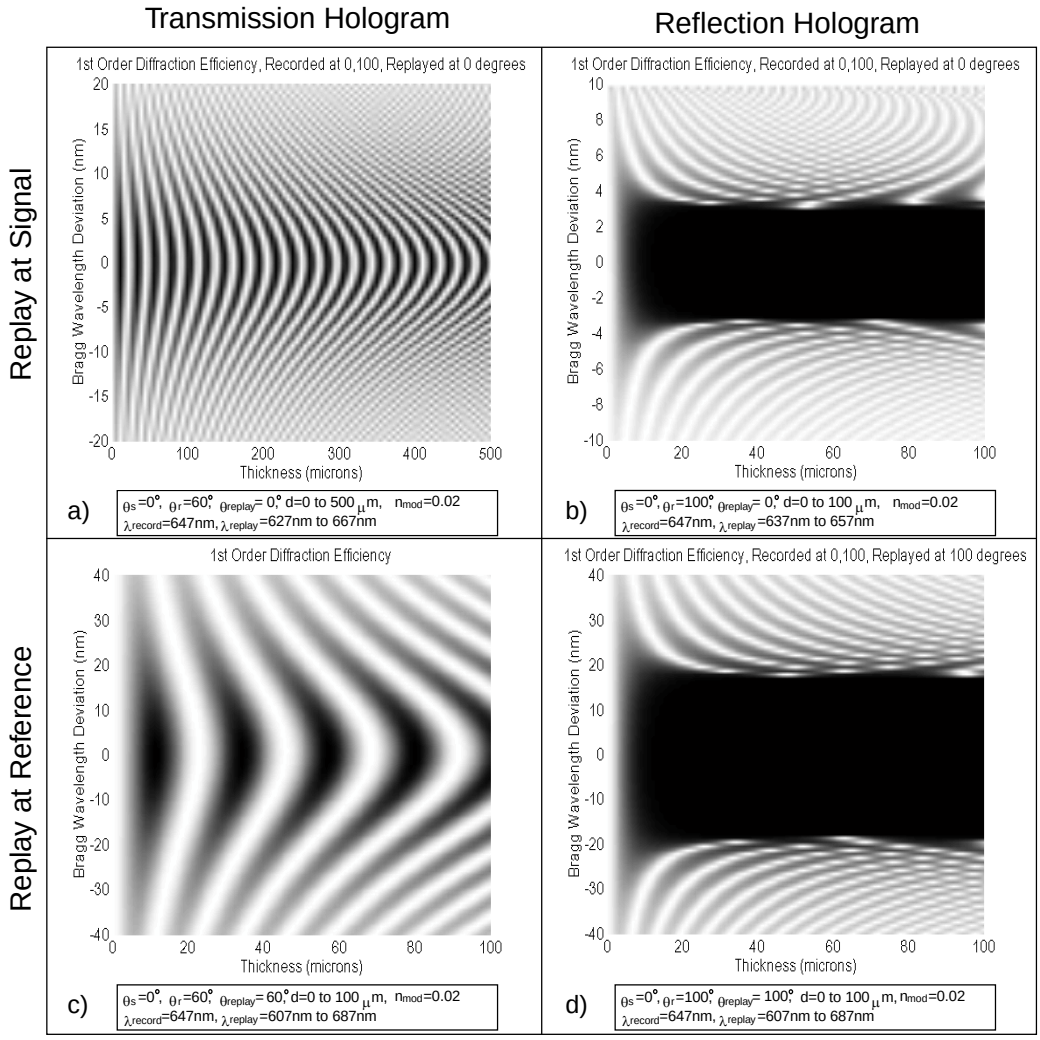


Figure 5.8 The dependence of wavelength bandwidth on film thickness for a transmission and reflection hologram. The wavelength bandwidth is shown for transmission hologram (a) and (c) and a reflection hologram (b) and (d) for replay at the signal and reference angles.

Once high diffraction efficiency has been obtained in a reflection hologram, the wavelength bandwidth is essentially independent of film thickness as can be seen from Fig. 5.8 (b) and (d). For transmission holograms, an increase in film thickness decreases the wavelength bandwidth. The increase in thickness increases the number of side–lobes and increases the proximity (in terms of wavelength proximity) of the side lobe peaks with respect to the on–Bragg peak. When considering only the main peak at the Bragg condition and not the sidelobes, the wavelength bandwidth diminishes as the thickness increases for transmission holograms.

5.5.3 Angular Selectivity

Similarly with wavelength deviations from the Bragg condition, angular deviations from the Bragg condition result in a decrease in diffraction efficiency. This angular selectivity is measured by the angular bandwidth which is usually defined at the FWHM of the diffraction efficiency versus replay angle curve.

In addition to small wavelength bandwidths, a transmission hologram can have small angular bandwidths. The angular bandwidth for reference replay is very different than for signal replay in the regime of heavily slanted transmission and reflection holograms. This was first noticed by Vasnetsov [5.10] for the case of heavily slanted (fringe angles near 45°) transmission holograms. He noted that for a 44.45° fringe angle in a transmission hologram, the angular bandwidth (FWHM) for the reference replay was approximately 50 times that of the signal replay. Vasnetsov did not expand these calculations to include reflection holograms or wavelength sensitivity for the regime of a steep referenced hologram. These regimes can be seen using the rigorous coupled wave theory as shown in Fig. 5.9. Vasnetsov's results are verified for the angular bandwidth as his theory is valid in his specific case hologram as mention in §5.4.2.

In Fig. 5.9, (a) and (b) represent the first order diffraction characteristics versus replay angle for a hologram with a refractive index modulation of $n_{mod} = 0.003$ as the fringe angle varies from 0° to 90°. Similarly, (c) and (d) with $n_{mod} = 0.02$, and (e) and (f) with $n_{mod} = 0.04$ show the angular bandwidth characteristics when the replay light is incident near the reference angle (θ_r–10° to θ_r+10°) in the left column and the light is incident near the signal angle (-10° to 10°) in the right column.

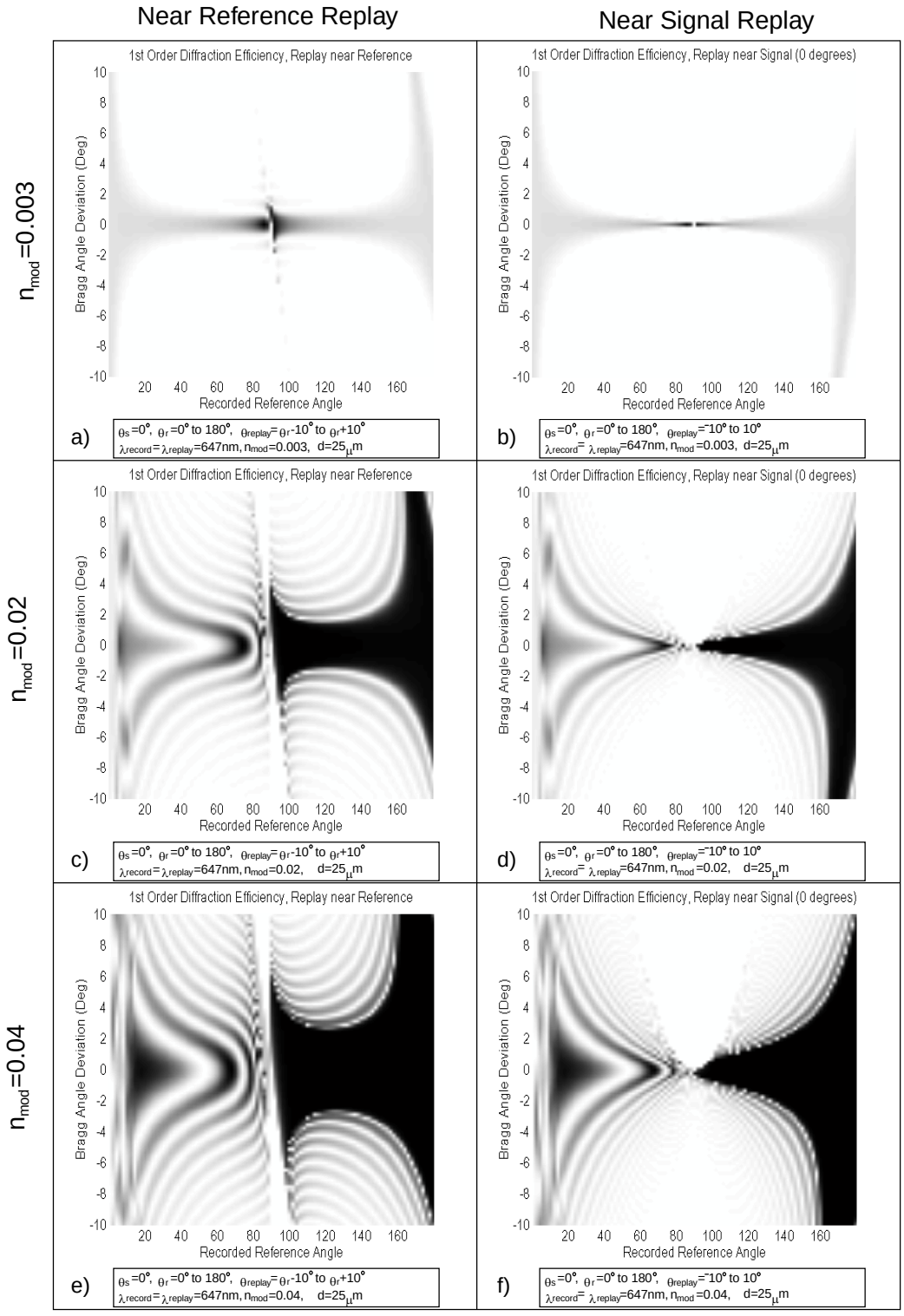


Figure 5.9 The dependence of angular selectivity on fringe angle and replay angle for three index modulations for volume phase holograms. The black areas represent diffraction efficiency near 100% with the gray areas proportionately less.

One can see from Fig. 5.9 that the angular bandwidth is smaller when the replay angle is near the signal angle as opposed to near the reference angle. One can also see from the left side of all the graphs in Fig. 5.9 that for a transmission hologram the angular bandwidth relationship to the fringe angle is not as easily quantified. Here, the angular bandwidth will be defined at the peak diffraction efficiency with the successive angular bandwidth at the next peak and the side–lobe characteristic of the transmission hologram will be ignored. On the graphs of Fig. 5.9, this could be represented by a vertical line in the dark band which enables the bandwidth to be calculated. With these definitions for the angular bandwidth of a transmission hologram, one could infer from the graphs that angular bandwidth decreases the closer the fringe angle is to 45° from 0°. This is seen most easily in Fig. 5.9 (f).

The angular bandwidth of reflection holograms exhibit unexpected dependencies on the fringe angle. When replaying near the signal angle, the angular bandwidth decreases the closer the fringe angle is to 45° from 90° (Fig. 5.9 (d) and (f)). However, when the replay angle is near the reference angle, the angular bandwidth initially decreases as the fringe angle approaches about 60° from 90°, then it begins to increase to high angular bandwidths as it approaches 45°.

The white lines at exactly 90° in Fig. 5.9 are where the rigorous coupled wave theory is not valid. In (a), (c), and (e), the white area line is slanted because the replay angle is changing, therefore, the recorded reference angle which would result in a 90° diffraction angle is also changing. The lines have a white area associated with them just as the upper parts of (b), (d), and (f) have white parabolic areas above them. These areas represent the off–Bragg condition where the first order of diffraction is evanescent just as in the wavelength bandwidth analysis.

The angular bandwidth increases with increasing refractive index modulation for both transmission and reflection holograms, as seen from the three rows of Fig. 5.9. Often the effects of refractive index modulation and hologram thickness on diffraction characteristics are similar. The effect of increasing the film thickness can be seen in Fig. 5.10.

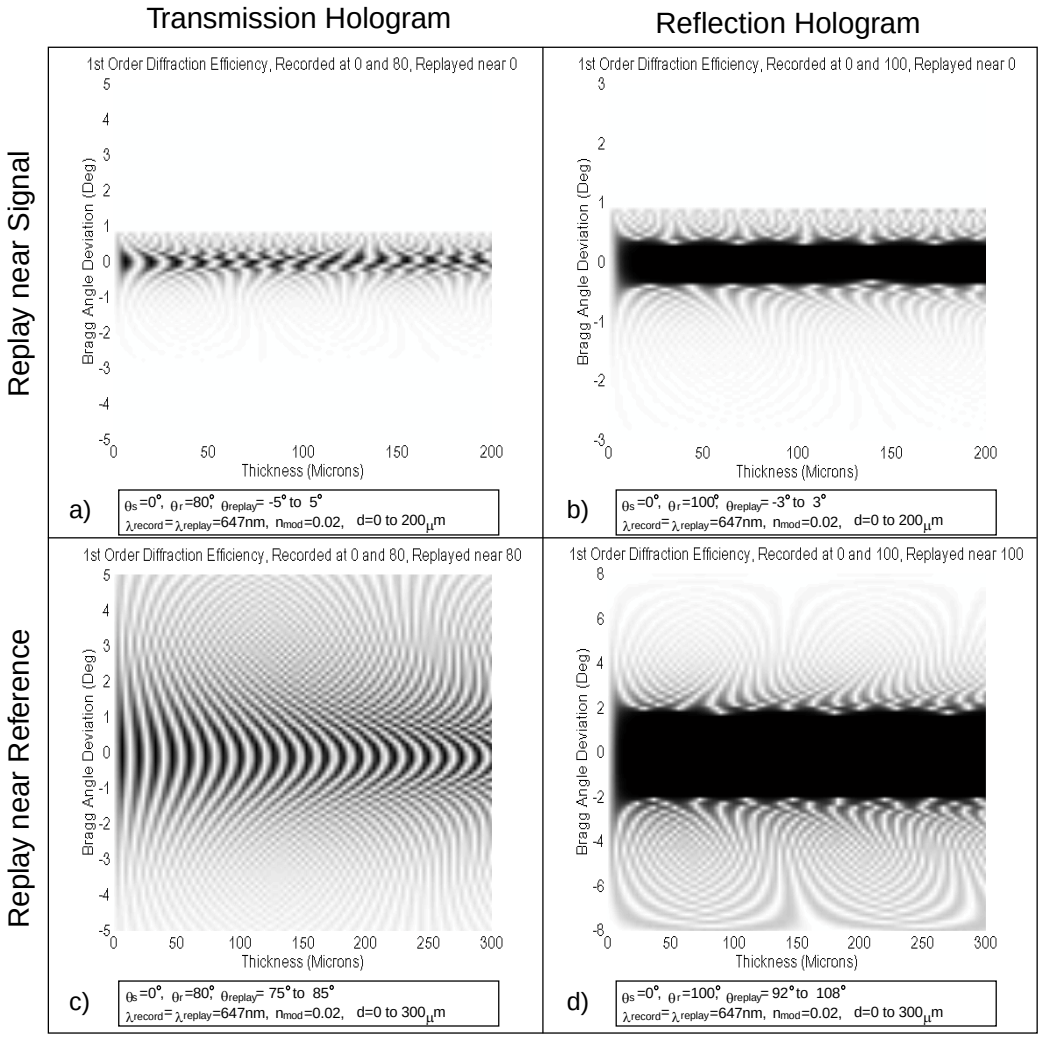


Figure 5.10 The dependence of angular bandwidth on film thickness for a transmission and reflection hologram. The angular bandwidth is shown for transmission hologram (a) and (c) and a reflection hologram (b) and (d) for replay at the signal and reference angles.

The effects of the film thickness on the angular bandwidth are as traditionally expected from approximate theories such as Kogelnik's. For transmission holograms, the angular bandwidth decreases with increasing thickness as indicated from Fig. 5.10. The number and proximity of side–lobes increases as the thickness increases just as with the wavelength bandwidth. Additionally, as in the wavelength bandwidth analysis, the angular bandwidth for a reflection hologram is essentially independent of thickness once a minimum thickness has been obtained.

5.5.4 Theoretical Results Summary

The results of the calculations on volume holograms can be summarized in Table 5.1 where 'inv' refers to inversely proportional and 'prop' refers to proportional and unless indicated, bandwidth refers to both wavelength and angular bandwidth.

Transmission Holograms

Slant

(inv) The larger the fringe angle (closer to 45°), the smaller the bandwidth.

Replay Angle

(prop) The larger the replay angle (closer to 45°), the larger the bandwidth.

Index Modulation

(prop) The larger the index modulation, the larger the bandwidth.

Thickness

(inv) The larger the thickness, the smaller the bandwidth.

Reflection Holograms

Slant

Replay Reference (inv, varies) The smaller the fringe angle (closer to 45°) the larger the *wavelength* bandwidth. The smaller the fringe angle, the *angular* bandwidth decreases then increases.

Replay Signal (prop) The smaller the fringe angle (closer to 45°) the smaller the bandwidth.

Replay Angle

(prop) The smaller the replay angle (closer to 0°), the smaller the bandwidth.

Index Modulation

(prop) The larger the index modulation, the larger the bandwidth.

Thickness

(ind) Once the minimum thickness for high efficiency has been obtained, the bandwidth is independent of thickness.

Diffraction Efficiency

The closer to 45° the fringes become, the index modulation and thickness required for maximum diffraction efficiency are both lower for transmission and reflection holograms.

Table 5.1 A summary of the effects of physical parameters on diffraction characteristics

REFERENCES

5.1 Syms, R., *Practical Volume Holography*, Clarendon Press, London, 1990

5.2 Syms, R. A."Path integral formulation of multiple scattering problems in integrated optics," Applied Optics, Vol. 25, pp 4402–12.(1987)

5.3 Syms, R. A."Path integral formulation of nonuniform optical coupled wave problems," Applied Optics, Vol. 26, pp 4220–30.(1987)

5.4 Moharam, M. and Gaylord, T., “Rigorous coupled–wave analysis of planar–grating diffraction,” JOSA A, Vol. 71, No. 7, pp 811–818, July (1981)

5.5 Gaylord, T. and Moharam, M., “Planar Dielectric Grating Diffraction Theories,” Applied Physics B, Vol. 28, pp 1–14, (1982)

5.6 Moharam, M., and Gaylord, T., "Chain–Matrix analysis of arbitrary–thickness dielectric reflection gratings," JOSA vol. 72, pp 187–190 (1982)

5.7 Zylberberg, Z. and Marom, E., “Rigorous coupled–wave analysis of pure reflection gratings,” JOSA, Vol. 73, No. 3, pp 392–398, March, (1983)

5.8 Moharam, M. and Gaylord, T., “Comments on analysis of reflection gratings,” JOSA A, Vol. 73, No. 3, pp 399–401, March (1983)

5.9 Kogelnik, H., "Coupled wave theory for thick hologram gratings," Bell System Tech. Journal, Vol. 48, No. 9, pp 2909–2947, (1969)

5.10 Vasnetsov, M., Soskin, M. et al., “Grazing diffraction by volume phase gratings,” Opt. Acta, vol. 32, pp 891–9

5.11 Liu, C. and Liu, J., *Linear Systems Analysis*, McGraw–Hill, New York, (1975)

5.12 Moharam, M., Pommet, D., Grann, E., and Gaylord, T.,”Formulation for stable and efficient implementation of the rigorous coupled–wave analysis of binary gratings,” JOSA A, vol. 12, No. 5, pp 1068–1076 (1995)

5.13 Huang, Q., and Gilbert, J., "Diffraction properties of substrate guided–wave holograms," Optical Engineering, Vol. 34, No. 10, pp 2891–2899

5.14 Kong, J., JOSA, vol. 67, p 825 (1977)

5.15 Moharam, M. and Gaylord, T., “Three–dimensional vector coupled–wave analysis of planar–grating diffraction,”, JOSA A, Vol 73, No. 9, pp 1105–1112, Sept. (1983)

5.16 Leith, E., "Holographic Data Storage in Three Dimensional Media", Applied Optics, Vol. 5, pp 1303, (1966)

5.17 Birner, S., "Steep reference angle holography: Analysis and Applications," Masters thesis, MIT, 1989

5.18 Henrion, M., "Diffraction and exposure characteristics of the edgelit hologram," Maters thesis, MIT, 1995

Chapter 6

MEASUREMENT AND ANALYSIS OF PHYSICAL AND DIFFRACTION CHARACTERISTICS

6.1 Introduction

In order to fully understand the diffraction characteristics of a hologram, one must first be able to measure or accurately predict the physical characteristics of the hologram. The difficulty in measuring the physical characteristics lies in the fact that the physical 'parts' of the hologram are just patterns of gradual refractive index differences. These differences are not finite, and are difficult to measure. Also, the size of the features is very small. For edge–lit holograms, the pitch of the features can be near 300 nm, which limits the type of measuring device which can be used. Many people use different diffraction theories to try to model the some of these characteristics instead of physically measuring them.

With information on the physical characteristics, one still must have accurate techniques for measuring the diffraction characteristics of the hologram. The diffraction characteristics are much simpler to measure. Most of the diffraction characteristics can be based on wavelength, polarization, and intensity and angle of diffracted light. All of these characteristics are relatively easy to measure. This is why diffraction characteristics are sometimes used to model the complicated physical characteristics.

6.2 Physical Characteristics

The recorded fringes of modulated index of refraction in a volume hologram are the distinguishing physical features. Many aspects of these fringes are crucial in defining the diffraction characteristics. The amplitude of the refractive index modulation and the thickness (or depth profile) are very crucial in determining the diffraction efficiency of a hologram. The physical spacing of the fringes determines the wavelength of diffracted light, while the angle of the fringes in a volume hologram determines the angle of diffraction.

6.2.1 Refractive Index Modulation

A sufficient refractive index modulation is necessary for a hologram to have high diffraction efficiency. However, a lower diffraction efficiency may be acceptable if the holograms is sufficiently thick. Measuring the modulation is critical, as the modulation may be too low or too high (transmission holograms) resulting in a low efficiency hologram.

The most common theory for determining the index modulation is Kogelnik's coupled wave theory [6.1]. Kogelnik's theory is the basis for Du Pont's predictions on refractive index modulation, or Δn [6.2]. Kogelnik's theory has been proven to be quite valid [6.3] for explaining the diffraction effects of non–slanted (0° or 90° fringe

angles) reflection or transmission holograms. Du Pont base their measurements on the transmission of the zero order undiffracted beam through a hologram and use this transmission for the first order diffraction efficiency in Kogelnik's calculations. A very important consideration here is that Du Pont [6.2] quote Kogelnik's theory incorrectly by replacing n_1 with Δn. This is not correct because n_1 is the amplitude of the modulation factor as

$$n(x) = n_0 + n_1 \cos(Kx) \quad \text{and with} \tag{6.1}$$

$$\Delta n = n_{\max} - n_{\min}, \text{ therefore} \tag{6.2}$$

$$\Delta n = 2n_1. \tag{6.3}$$

The Du Pont values for their predicted Δn are actually half of the value that they should be reporting. This is taken into account whenever the predicted refractive index modulation is analyzed. For Du Pont photopolymers the amplitude of the refractive index modulation crucially depends on the fringe spacing. The amplitude of the modulation also varies with different types of photopolymers and this is how Du Pont differentiate between their films for reflection and transmission use. Thus, in order to predict the refractive index modulation for steep referenced holograms, one must look at the refractive index modulation in previous calculations corresponding to a similar spatial frequency.

One might imagine measuring the refractive index modulation using a non–holographic method such as an Abbe refractometer. Using an Abbe refractometer for measuring the refractive index of photopolymers has been partially successful as mentioned in §3.6.3. One might imagine that when recording in photopolymer, the bright areas (constructive interference) result in a higher refractive index and the dark areas (destructive interference) result in a lower refractive index and that the average index of refraction remains the same. Du Pont report a slight shrinkage in the film after the various processing stages [6.15], although this is based on the Kogelnik

theory and does not take into account any non–linear fringe recording aspects of the film.

If the maximum value for the area of higher refractive index was constant over a region, then this would show up on the Abbe as a fairly dark line, with other faint lines beneath it corresponding to the maximum TIR angle. As seen with Eq. 2.11, the contrast varies throughout the thickness of the film due to absorption. However, at the y=0 plane (the surface of the film in Fig. 2.1), the maximum should remain constant over the surface (the x–z plane, assuming uniform, collimated reference and signal coverage during recording) and this is the plane which the Abbe would measure.

Unfortunately, due to the need for strict uniformity over the large sampling area, tests on the Abbe were not useful in determining the amplitude of the index modulation. Some thin lines were visible and a faint dark area appeared in the range of the index modulation, although the edges of this area were not distinctive enough for measurement.

6.2.2 Fringe Angle

The fringe angle is much simpler to predict than the refractive index modulation. One can simply predict the fringe angle based on the angular bisector of the reference and signal beams as in Eq. 4.6. When recording an edge–referenced hologram with the signal normal to the plane, the reference beam angle is sometimes difficult to accurately measure. The performance of the hologram is very critical as the angle approaches 90° as shown in Fig. 5.9 and must be accurately measured. This can be done with reasonable accuracy from the depiction of Fig. 6.1

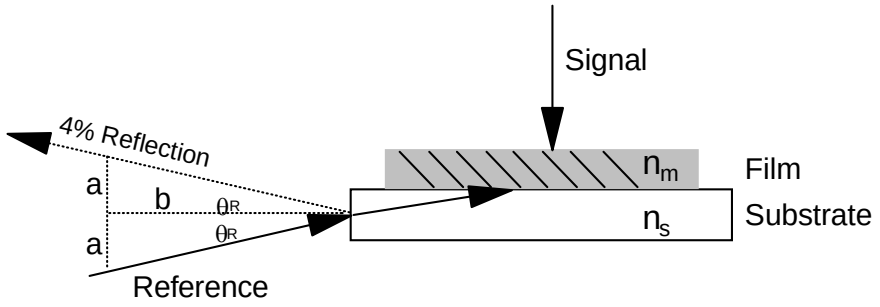


Figure 6.1 The technique used for measuring the reference beam angle in air for an edge referenced hologram.

From the law of reflection, one can measure the reference beam angle in air as

$$\theta_R = \tan^{-1}\left(\frac{a}{b}\right), \tag{6.4}$$

thus, using Snell's law the reference angle in the recording film, n_m, is

$$\theta_{R(film)} = \sin^{-1}\left(\frac{n_s \sin\left[90 - \sin^{-1}\left[\frac{\sin\theta_R}{n_s}\right]\right]}{n_m}\right). \tag{6.5}$$

One may also measure the fringe angle by looking at the diffracted light from an unaltered laser beam at the recording wavelength and measuring the angles. Although this requires a measurement similar to the one above for the diffracted or incident light depending on which replay geometry was used. If one is looking at the possibilities of film shrinkage, then the angle of diffraction should be examined.

Another method for measurement of the fringe angle as well as the fringe spacing is the Scanning Electron Microscope (SEM).

6.2.3 Scanning Electron Microscope Measurement

In a reflection SEM electrons from a field–emission source are focused into a probe (of a size near 0.2–0.5 nm) which can be scanned across a sample [6.4]. Electrons which are reflected and scattered by each point in the specimen are collected by an electron collector and the energies are recorded and can be displayed by a computer in real–time. Reflection SEM's are usually used for *topographical contrast* and show the reflections from the surface profile.

In order to 'see' the fringes recorded in the photopolymer, the sample had to be freeze–cracked to see a cross section showing the slant of the fringes. A significant amount of experimental sample preparation was needed to enable detection of the fringe structure. In general, four requirements should be met for electron microscopy of significant detail [6.4]:

1) Avoiding Specimen Collapse

Since specimens must be measured in a vacuum, collapsing of the structure must not occur. When a photopolymer sample is first recorded and UV cured, the sample is still evolving solvent vapors. These vapors emit a slight odor and will evolve until after a long heating process. The higher solvent content before heating promotes the bonding to the Mylar carrier of the film. The Mylar film can not be easily removed until after heating. This would not be a significant problem, however, the samples needed to be freeze–fractured in order to have a 'clean' break across the film. The Mylar carrier does not freeze–fracture well in liquid nitrogen and no results were obtained with the Mylar. This means that the fringe pattern on unheated samples were not obtained using an SEM.

2) Sufficiently Thin Sample

In transmission electron microscopy, the object specimen must be sufficiently thin so that multiple scattering does not occur from within the sample and confuse the image. In attempts to create a sample for TEM, freeze–microtoning was tried with very poor success with Mylar, and better, though not usable, success without the Mylar. Sample preparation was attempted with an epoxy used for acrylic fibers. However, no measurements were obtainable because the photopolymer appeared to be slightly soluble to the solvents in the epoxy. While the TEM can examine detail down to near atomic resolution, the fringe spacing is on the order of 300 nm. The SEM which has a resolution of a maximum around 2 nm is much more ideal. With a reflection SEM, one can measure the average fringe angle [Appendix D] because a fringe (≈ 0.15 μm) can just barely be seen relative to the thickness of the film (20 μm). The entire film thickness could not be seen on the TEM. Thus, due to the microtoning problems and the size limitations, the main focus for electron measurement was reflection SEM.

Thick samples are not a problem in reflection microscopy since normally the electrons are reflected from the specimen surface. However, the photopolymer sample must be less than 1 cm in length and about 5 mm in depth to be supported in the SEM apparatus.

3) Electron Beam Damage Avoidance

If the current and voltage on the SEM are sufficiently high, they can cause damage to the sample. This was noticed in the photopolymer samples at high voltage. The samples were heated and the surface profile was altered due to the applied voltage. This limited the effective operating range, but the range was still sufficient for measurement.

4) Maximizing Image Contrast

In order to promote electron reflection and improve contrast, a standard technique for organic materials is to sputter coat them with a thin metal coating (usually about 10 nm of gold/palladium). This was attempted with the photopolymer samples. Unfortunately, sputter coating was not possible because the heat required for the glow discharge completely melted the photopolymer.

6.2.4 Fringe Spacing

The fringe spacing may be predicted as in Eq. 4.6. This is verified if the diffraction efficiency is at a maximum at the recording angle and wavelength with the proper diffraction angle. The fringe spacing may also be verified with the SEM. By freeze fracturing the sample, the fringes may be seen [Appendix D], although the position of the index maximum/minimum is vague. An average of the spacing over many periods offers the closest approximation. In conventional transmission holography, a standard phase microscope can measure the fringe spacing because the sample does not need to be freeze–cracked to look along the plane of the hologram (no fringe slant) and the spacing is usually larger. Due to the slant and size of the edge–lit hologram, an SEM measurement is required.

6.2.5 Depth Profile

Predicting the refractive index profile throughout the thickness of the film for a steep referenced hologram is a difficult task. One can estimate the contrast throughout the medium from Eq. 2.11, although a steep reference hologram has a varying fringe contrast with depth, and therefore the refractive index modulation with depth also varies. Since the exact relationship between the contrast and the final index modulation in three dimensions is not known, one can not accurately predict the

modulation variation or the effective thickness. Another factor inhibiting this prediction is the fact that the sensitizing dye actually bleaches in real time, thus changing the absorption constant during recording.

Using the rigorous coupled–wave theory for diffraction (§5.3), one might be able to model the refractive index modulation profile by breaking down the thickness into discrete sections of constant index modulation. This could be compared with the measured angular selectivity and other diffraction characteristics to confirm the predictions. In a simpler model, one may predict an average modulation and compare that with the diffraction characteristics.

Exact measurement of the refractive index profile is very difficult as well. A rough measurement can be made from the SEM pictures as in Appendix D. One could perhaps see chirping, tapering [6.3], shrinking, or even multiple sets of fringes through the pictures, although this also relies on the uniformity of freeze–cracking. The distortion in the fringes in Appendix E at the top surface could be caused from a spurious reflection at the surface causing three sets of fringes to be formed in that area. These fringes would strengthen the photopolymer in that area and perhaps make the break in the freeze–cracking non–uniform.

6.3 Diffraction Characteristics

The important diffraction characteristics generally related to HOE's are diffraction efficiency (normally 1st order), angular sensitivity, wavelength sensitivity, and polarization sensitivity. The desired values for each of these characteristics depends on the application for the HOE. Measurement of these can be done with spectrophotometers and rotation stages with the proper arrangements and considerations (allowing for Fresnel reflections, polarization, etc.).

6.3.1 Spectrophotometer

A system has been developed which will measure all of these characteristics quickly with the data recorded in a computer (Fig. 6.2).

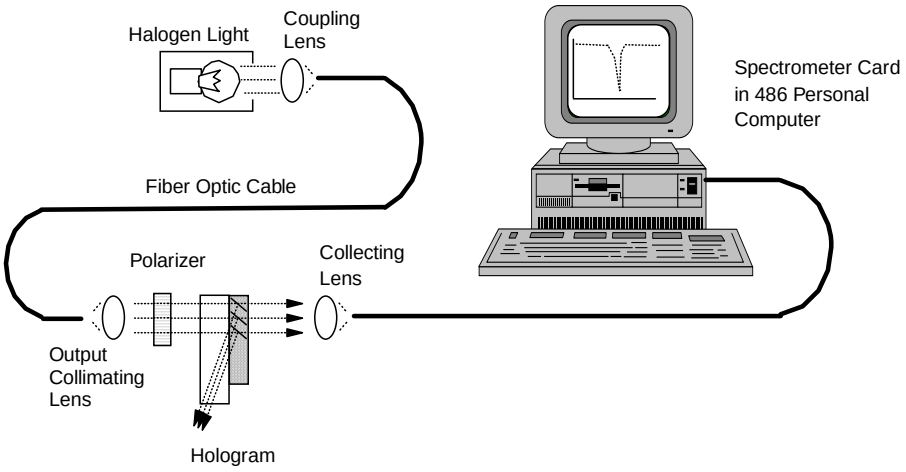


Figure 6.2 The spectrophotometer setup for measuring diffraction efficiency, angular sensitivity, wavelength sensitivity, and polarization sensitivity. (The rotation stage for the sample and the polarizer are not shown).

When using the conjugate signal replay of an edge–referenced reflection hologram, the diffracted light leaves the substrate through the edge and the zero–order transmitted light passes through the hologram. Assuming that the light not transmitted through the hologram is due only to first order diffraction (reflected order), one thus measures the diffraction efficiency of the first order. However, in the case of steep reference holograms, both the +1 forward and backward diffracted orders may have significant intensities. Therefore, one must keep in mind that in the above setup, only the zero order transmittance is measured and the *exact* diffraction efficiency of the first order light (reflected or transmitted) can not be determined.

The samples were examined using the spectrophotometer setup in Fig. 6.2. One can relate the term 'diffraction efficiency' to any order by considering the percent of light diffracted relative to the incident light. Since the spectrometer measures the intensity (relative to a reference which is an unexposed portion of the film) for a wavelength range, one immediately knows the diffraction efficiency and wavelength sensitivity.

By rotating the sample one can measure the angular sensitivity, and by rotating the polarizer, the polarization sensitivity is determined.

If one were measuring all of the different diffracted orders, such as in the case of unslanted transmission holograms, one could use a goniophotometer to measure the intensity and angle of each diffracted order.

6.3.2 Diffraction Efficiency Analysis

One of the most crucial factors which determines the diffraction efficiency is the refractive index modulation. When using photopolymers, the index modulation depends on the diffusion of the components as mentioned in §3.6.1. The final location of the chemical components determines the density and refractive index profile within the photopolymer. Since the diffusion length is limited, photopolymers can be considered to be non–linear recording materials. This diffusion length is different for each composition of photopolymer and usually distinguishes the transmission films (low spatial frequency or large fringe spacing) from the reflection films (high spatial frequency or small fringe spacing). Since the fringe spacing changes with wavelength and recording angles, to obtain accurate information regarding diffraction efficiency the angles and wavelength must be held constant (as when measuring the effects of beam ratio on diffraction efficiency in §2.4.2). The possible effects of the non–linear recordings on diffraction efficiency of holograms are briefly discussed by Bjelkhagen [6.5] with references to other work.

The value for the refractive index modulation can also change due to absorption in the recording process. This effect is known as taper [6.3]. As mentioned in §2.2, the fringe contrast can vary dramatically throughout the volume of a steep referenced hologram. Thus, the value for Δn is actually a function of the depth in the film.

Another critical factor for determining the diffraction efficiency is the actual refractive index profile. One might assume that the interference profile is exactly

sinusoidal, however, one can definitely not assume that this is true after the exposure. For a particular film with a specific diffusion length, there may be a precise, unique, fringe spacing which would result in a perfect sinusoidal profile, however, this is not usually the case. In order to assume that the index profile was completely sinusoidal, one would have to say that the resulting index modulation was linearly proportional to the intensity of the recording interference pattern.

In order to accurately predict the diffraction efficiency, one should represent the refractive index profile as

$$n(x,z) = n_0 + n_1(z)c_1(\Lambda, x)\cos(Kx) \quad \text{where} \tag{6.6}$$

n_1 varies in the z–direction as a result of the absorption during recording, and c_1 is a function for the photopolymer related to the diffusion length which is related to fringe spacing and position in the x–direction. The affects of non–linear recording could also be represented by the addition of cosinusoidal terms representing further harmonics instead of the function c_1. Due to the additional complexity involved with programming these variables in a computational program, the 'real' recording effects of taper and non–linear recording were not introduced to the rigorous coupled wave calculations for diffraction efficiency.

Due to the non–linear recording characteristics of the photopolymer, different films can easily result in different index profiles as well as varying amplitudes of refractive index modulations. For example, when making reflection edge–referenced holograms, by simply changing experimental photopolymer films, the maximum diffraction efficiency increased by 54% before heating.

While examining different photopolymers for reflection edge–referenced holograms, different maximum diffraction efficiencies were obtained just after recording (Table 6.1).

Experimental Film	Maximum Diffraction Efficiency
Red Sensitized HRF 352	8%
HRF 300x006	8%
HRF 700x071	62%

Table 6.1 Maximum first order diffraction efficiencies obtained for steep referenced holograms just after recording for three different Du Pont photopolymers.

Therefore, the successive tests on steep referenced holograms were done using the HRF 700x071 film. In order for the diffraction efficiency to be accurately predicted, one must know the amplitude of the refractive index modulation. Unfortunately, since an acceptable technique for measuring the index modulation in photopolymers was not found (§6.2.1), the index modulation is usually predicted from the diffraction efficiency results.

6.3.3 Wavelength Bandwidth Analysis

Steep referenced reflection holograms were recorded using HRF 700x071 laminated onto fused silica. The recording angles used were 0° for the signal and 103° for the reference. The resulting bandwidth before heating is illustrated in Fig. 6.3.

In Fig. 6.3, the rigorous coupled wave calculations were based on a refractive index modulation which would yield the same diffraction efficiency at the Bragg condition. The FWHM wavelength bandwidth measured is approximately 7 nm compared to the 2 nm predicted value. The discrepancy here is believed to be due from tapering effects and a deviation from an exact sinusoidal profile. There are many papers which describe the affects of taper on the angular and wavelength bandwidth [6.6–6.10]. Even though these papers do not use a rigorous form of the coupled wave theory, their results and experimental verifications show that the angular and wavelength bandwidths *increase* for non–uniform holograms recorded with increasing values of the absorption coefficient. Since the affect of absorption is more critical in steep

referenced holograms, one could conclude that this is possibly a significant reason for the experimental results having a larger bandwidth than predicted.

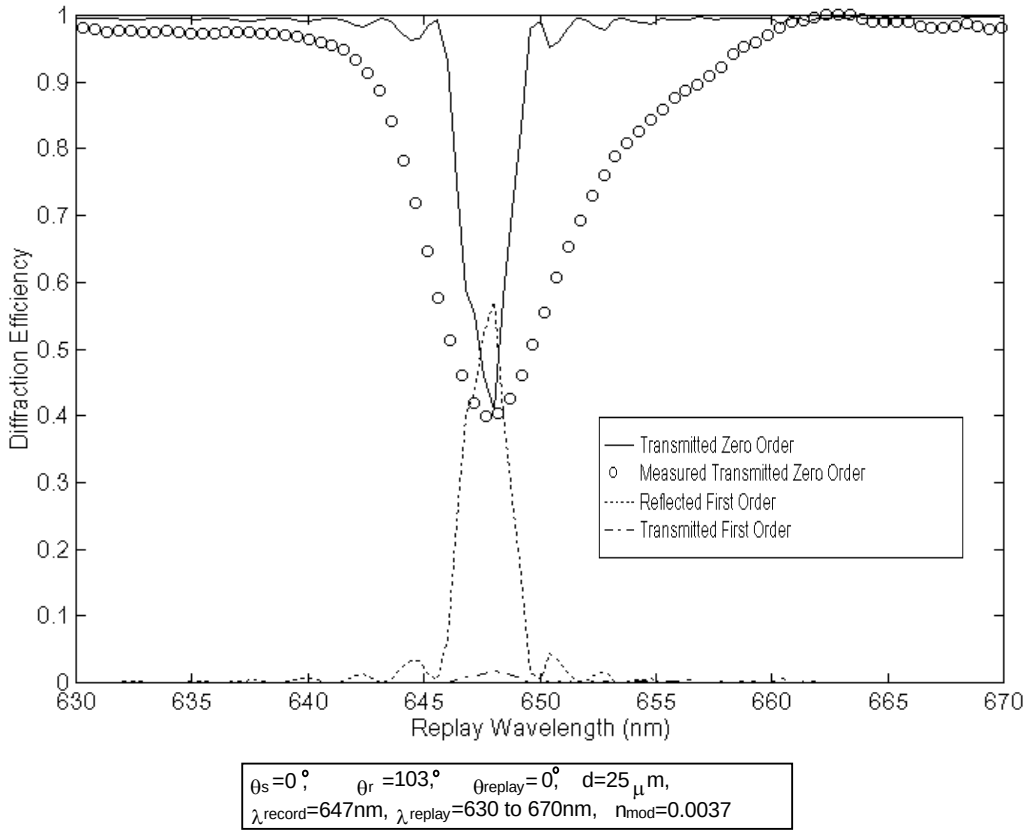


Figure 6.3 The measured zero–order transmitted diffraction efficiency versus replay wavelength compared with the predicted diffraction efficiencies using the rigorous coupled wave theory.

6.3.4 Angular Bandwidth Analysis

The angular bandwidth was measured for the same sample as in Fig. 6.3. The results are displayed in Fig. 6.4.

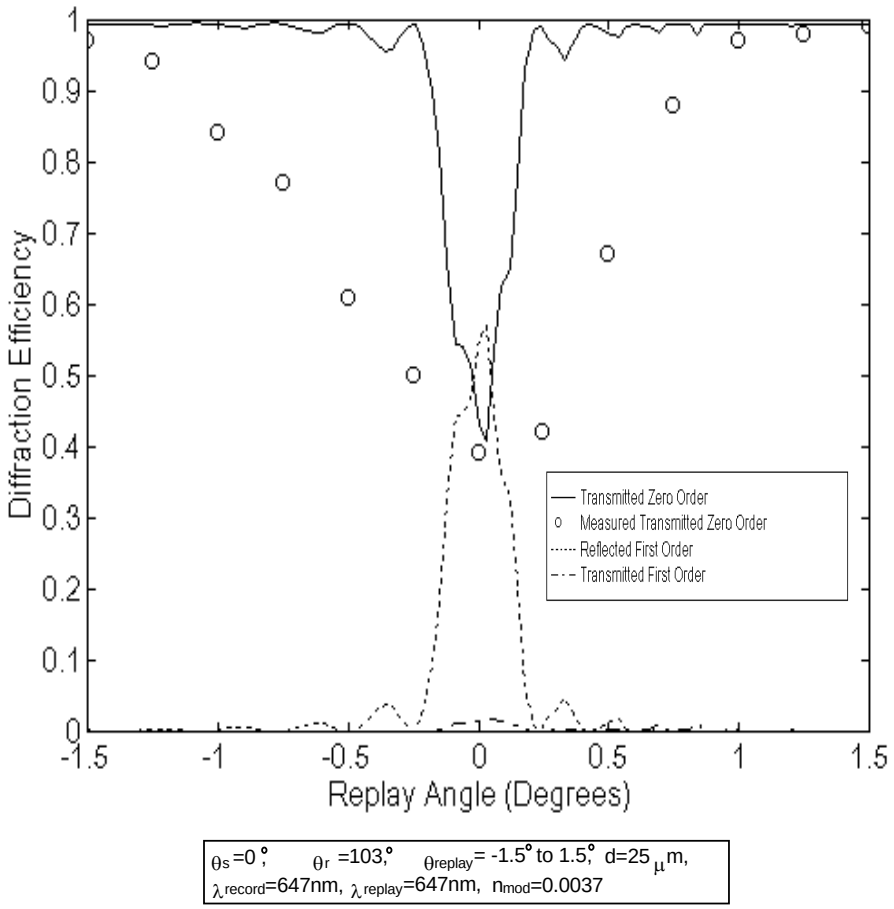


Figure 6.4 The measured zero–order transmitted diffraction efficiency versus deviation from the Bragg replay angle (0°) compared with the predicted diffraction efficiencies using the rigorous coupled wave theory.

From Fig. 6.4, the measured angular bandwidth is approximately 1° and the predicted angular bandwidth is 0.28°. The 0.5° resolution of the rotation stage introduces some error, although the bandwidth can clearly be seen to be slightly larger than predicted. Again, this slight increase in bandwidth is attributed to the tapering and non–perfect sinusoidal index profile within the film as described in the previous section.

6.4 Effects of Heating

As mentioned in §3.6.1, the normal procedure with Du Pont photopolymers is to heat the recorded film at about 100°C for 2 hours. This causes further monomer diffusion as well as diffusion of other components within the film, which can result in a higher index modulation depending on the initial recording conditions.

6.4.1 Heating Effects on Diffraction Characteristics

The effects of heating on Du Pont films has been reported by Masso [6.11] and Weber [6.12]. Weber reports that when recording non–slanted reflection holograms, the replay wavelength shifts toward the blue after heating, which would correspond to a shrinkage in the film. Masso reports similar results, but noticed that when the slant angle of a hologram is increased, the playback wavelength after heating red–shifts (replays at a longer wavelength), which would correspond to swelling. The highest slant angle he analyzes is 9°, and he makes the conclusion–"the effective shrinkage decreases with increasing slant angle and in some cases the optical thickness actually increases rather than decreases during processing for slanted fringe holograms."

Results obtained for heavily slanted holograms (40° fringe angles) show that the Bragg replay wavelength and replay angle are affected very little (if at all) from heat processing (i.e. no shrinkage) as shown in Fig. 6.5 and Fig. 6.6.

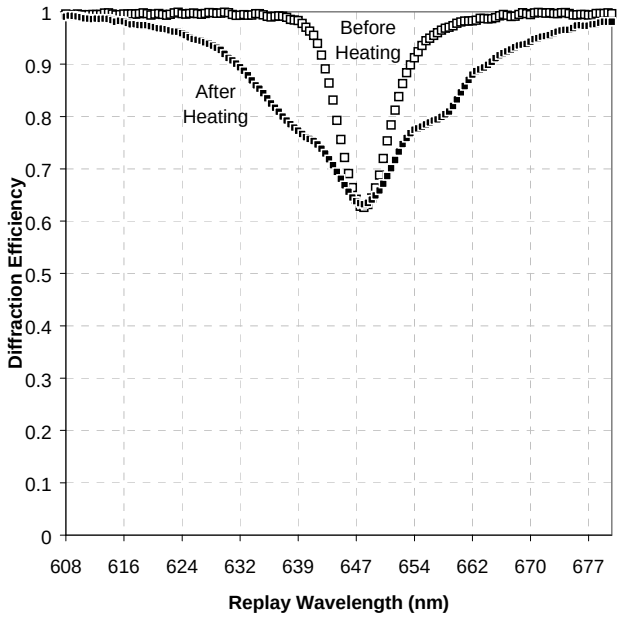


Figure 6.5 The effect of heat processing of a slanted fringe hologram on the replay wavelength. The wavelength bandwidth of the transmitted zero order increases from a FWHM of 8 nm to 25 nm, although the peak replay wavelength remains the same. The hologram was recorded with $\theta_s = 0°$, $\theta_r = 103°$, $\theta_{replay} = 0°$ and $\lambda_{record} = 647$ nm.

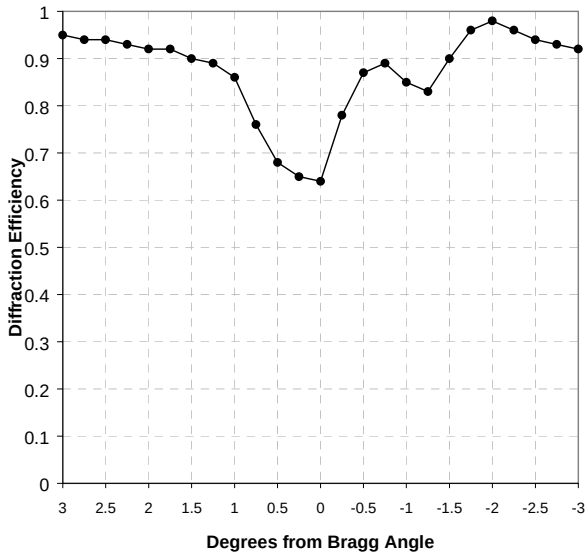


Figure 6.6 The effect of heat processing on the angular bandwidth for a slanted fringe hologram. The angular bandwidth of the transmitted zero order remains about the same (FWHM of 1°) as the similarly recorded (unheated) hologram in Fig. 6.4 except for the appearance of a side–lobe. The hologram was recorded with $\theta_s = 0°$, $\theta_r = 103°$, $\theta_{replay} = 0°$ and $\lambda_{record} = 647$ nm.

Initially, interpreting the effects of heating on edge–lit holograms was incorrectly attributed to swelling in Coleman [6.13] because of the increased bandwidth and the dramatically higher efficiency at another Bragg condition. When the heated holograms were viewed with white light, the apparent peak wavelength is shifted because of the increased bandwidth and the chromatic sensitivity of the eye. This effect showed up on display holograms as well as HOE's. In addition, the appearance of an increase in diffraction efficiency at another Bragg angle and wavelength, as presented in Fig. 6.7 [6.13], was interpreted to mean that the film had swelled.

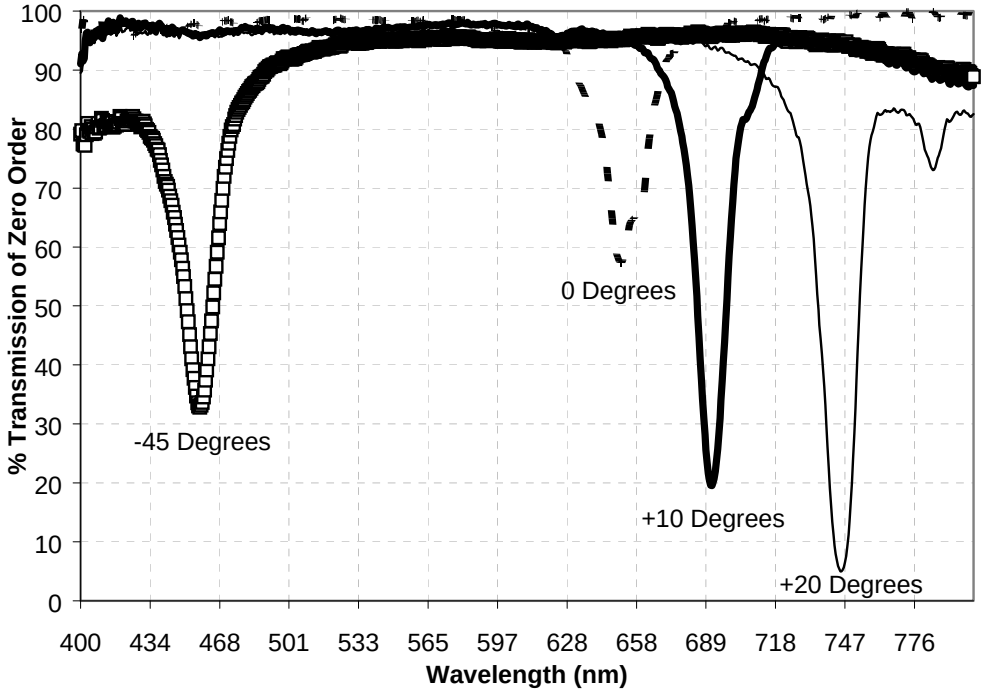


Figure 6.7 The decrease in transmission of the zero order (increase in diffraction efficiency of the first order) of a hologram as the Bragg replay angle is changed [6.13]. The hologram was recorded with $\theta_s = 0°$, $\theta_r = 103°$, $\theta_{replay} = 0°$ and $\lambda_{record} = 647$ nm.

If one were given the hologram of Fig. 6.7 without knowing how it was recorded, one might guess that the hologram was recorded at a wavelength of 746 nm since it has the highest diffraction efficiency. However, results and the theory predict that the diffraction efficiency can change for different Bragg conditions as shown in Fig. 6.8

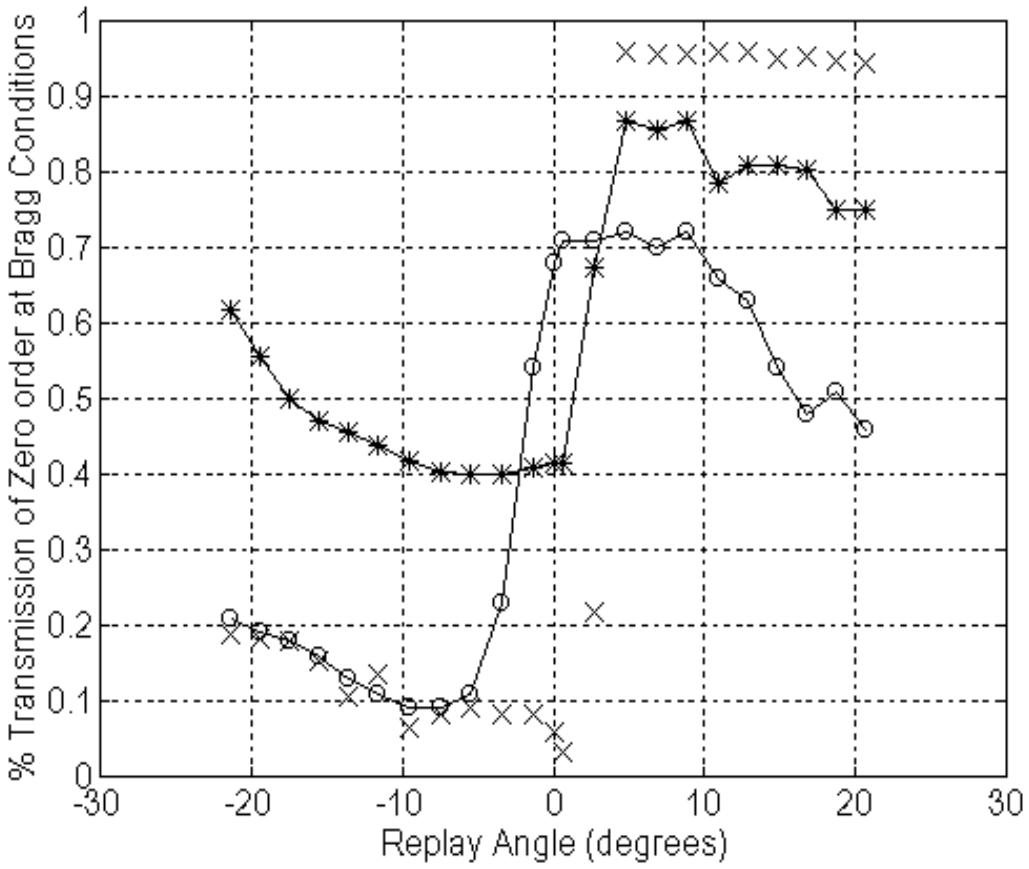


Figure 6.8 The percent transmission of the zero order illumination on a heated (O's) and unheated (*'s) steep referenced hologram for various Bragg conditions. The holograms were recorded as in Fig. 6.3 and the index modulation used to match the heated data was n_{mod} = 0.0105 for the RCW theory (X's). The hologram was recorded with $\theta_s = 0°$, $\theta_r = 103°$, $\theta_{replay} = 0°$ and λ_{record} = 647 nm.

The hologram in Fig. 6.8 was recorded with a 0° signal angle, and as can be seen, the zero order transmission for the heated hologram decreases significantly as the replay angle approaches -10°. The decrease in the zero order transmission corresponds to an increase in diffraction efficiency. This higher diffraction efficiency near the -10° Bragg condition is predicted by the rigorous coupled wave theory. However, the rigorous coupled wave theory predicts that the decrease in diffraction efficiency (increase in transmission) occurs just after the 0° Bragg condition and the slope is very steep. The recorded holograms do not demonstrate the steep rise in transmission probably due to a non–perfect sinusoidal refractive index profile. The curve is effectively smoothed by the deviation from a perfect sinusoidal refractive index modulation.

One might imagine the reason that the diffraction efficiency of a hologram could change for different Bragg conditions by looking at a reflection hologram in general. In a reflection hologram, for example, a steeper Bragg angle would reflect more light off of a fringe plane (a region of constant index of refraction) just as if it were a boundary to another medium as indicated by Fresnel's law. This high reflection off of the fringe plane would not allow as much coupling into the first diffracted orders as one with a shallower angle.

The effects of heating on the diffraction efficiency for all edge–referenced holograms is very dependent on the original recording conditions. In some samples, such as that in Fig. 6.5, the diffraction efficiency did not appear to change when the sample was heated. Other samples have shown a slight increase in diffraction efficiency. In some of the samples, the heated diffraction efficiency was still below other samples before heating at the recorded Bragg condition (Fig. 6.8). Normally, the samples were exposed for long times and large exposure values as this seemed to maximize the pre–heat diffraction efficiency. This may suggest that the maximum refractive index modulation can be obtained during exposure, and that when the sample is heated, the heat softens the film allowing further re–distribution of components (and the index of refraction) in–between the maximum and minimum values. This may or may not equate to a more sinusoidal index profile depending on the initial recording conditions. The increase in bandwidth after heating could be interpreted as a redistribution of index to a less sinusoidal profile.

As mentioned by Masso, the type of cover film used when heating can effect the shrinkage in a hologram. In the holograms examined here, the heating was done with only the Mylar as a cover material. Near the edges of the film and Mylar, non–uniform heating effects could be seen that were not visible in the middle. This suggests that the permeability and rigidity of the Mylar is unacceptable for heating and a cover film such as Microglass should be used.

6.4.2 Heating Effects on Physical Characteristics

As mentioned in §6.2.3, in order to view the fringe profile on the SEM, the hologram had to be heated for freeze–cracking. The hologram in Appendices D and E was heated before examination on the SEM and the fringe structure appears to be completely parallel confirming that there is no curve to the fringes (chirping) when heated. Unfortunately, using an SEM, one can not exactly determine the uniformity of the grating, nor compare a heated sample with an unheated sample.

6.4.3 Controlled Swelling/Shrinking

Du Pont photopolymers have the interesting ability to be controllably swelled as mentioned by Gambogi [6.15]. By laminating a tuning film onto a surface of the hologram before heating, the hologram can be swelled to replay at a longer wavelength. This process involves the further diffusion of monomers from the tuning film to the hologram during the heating process. Using different tuning films, Du Pont suggest that they can achieve a peak Bragg wavelength shift from 508 nm to 635 nm on a non–slanted reflection hologram recorded at 514 nm. Gambogi states that the "swelling is uniform and reproducible and maintains reflection efficiency and bandwidth." Lab results suggest that this may only be the case for precisely *uniform* exposures of *non–slanted* reflection holograms. Results of holograms with different pre–heat diffraction efficiencies seems to yield different replay wavelengths and bandwidths.

Du Pont later stated [6.16] that the bandwidth does increase moderately. Various other films were tried and results were obtained on slanted holograms which appear to have increased wavelength bandwidths after heating with a tuning film. Also, as discovered in the lab, the glue layer on a protective adhesive laminate seems to have the effect of shrinking the layer (blue-shifting) as shown in Fig. 6.9.

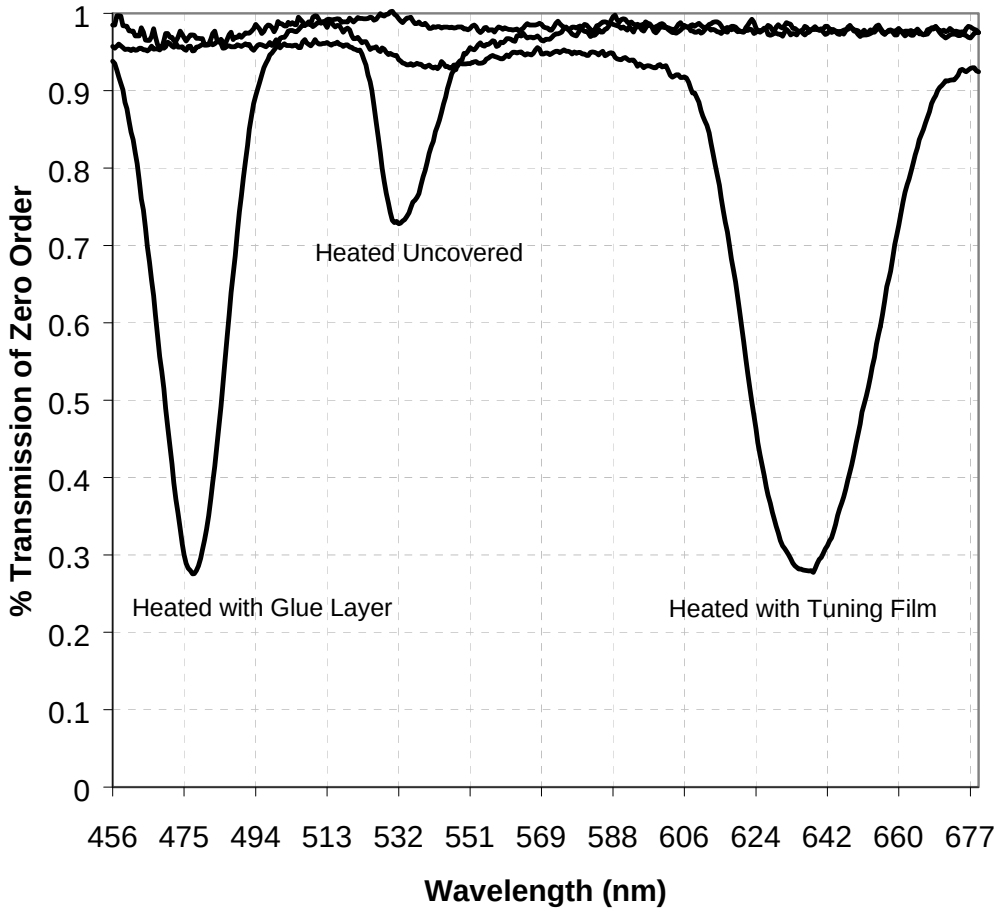


Figure 6.9 The replay wavelength of two different film laminates and no film laminate on a heated, slanted hologram recorded at 514 nm.

With slanted holograms such as the one in Fig. 6.9, the slanted hologram red–shifts when there is no cover layer (no Mylar) and it is heated. Unfortunately, it appears that for slanted holograms, this shift is a swelling effect, and the fringe angle also changes. This means that if one wanted to change the wavelength on a hologram, but have it replay at the same angle, one would have to pre–compensate the for the angle shift during recording. The shrinkage theory for holograms has been analyzed for

silver halide holograms by Beléndez et. al. [6.17, 6.18], photopolymer holograms by Masso [6.11] and general volume holograms by Gallo [6.19]. These models could be applied to the shrinkage effects of the Du Pont photopolymers to predict the resulting shrinkage or swelling.

The effects of the heating process appear to be diminished if their is a sufficient time between recording the hologram and the heating stage. This is believed to occur because the solvents evolve through the Mylar from the film. The solvent seems to help act as a medium for the components to diffuse, thus less of it would inhibit the diffusion. This effect is greatly increased if a portion of the film is left exposed to the air. Some samples heated with a Mylar cover exhibited non–uniform swelling effects near the edge which could be explained by solvent evolution from the edge.

By measuring the physical characteristics of edge-lit holograms using an SEM, one can obtain sufficient information (the fringe angle spacing) to verify the rigorous coupled wave diffraction modeling of the recorded samples. The results agree well with the theory except for the increase in angular and wavelength bandwidth which is probably due to the non-sinusoidal refractive index profile.

REFERENCES

6.1 Kogelnik, H., "Coupled wave theory for thick hologram gratings," Bell System Tech. Journal, Vol. 48, No. 9, pp 2909–2947, (1969)

6.2 Gambogi, W., Gerstadt, W., Mackara, S., and Weber, A., “Holographic transmission elements using improved photopolymer films,” SPIE, International Symposium on Optical Science and Engineering, San Deigo, CA, (1991)

6.3 Syms, R., *Practical Volume Holography*, Clarendon Press, London, 1990

6.4 Slayter, E., and Slayter, H., *Light and Electron Microscopy*, Cambridge University Press, 1992

6.5 Bjelkhagen, H.I., *Silver–Halide Recording Materials*, Springer–Verlag, Berlin, 1993

6.6 Kubota, T., “Characteristics of thick hologram grating recorded in absorptive medium,” Optica Acta, Vol. 25, No. 11, pp 1035–1053, (1978)

6.7 Uchida, N., “Calculation of diffraction efficiency in hologram gratings attenuated along the direction perpendicular to the grating vector,” JOSA, Vol. 63, No. 3, pp 280–287, March (1973)

6.8 Kermish, D., "Non–uniform sinusoidally modulated dielectric gratings," JOSA, vol. 59, No. 11, Nov. (1969)
6.9 Kowarschik, R., “Diffraction efficiency of attenuated sinusoidally modulated gratings in volume holograms,” Optica Acta, Vol. 23, No. 12, pp 1039–1051, (1976)
6.10 Morozumi, S., “Diffraction efficiency of hologram gratings with modulation changing through the thickness,” Jap. Journal of Applied Physics, Vol. 15, No. 10, pp 1929–1935, Oct. (1976)
6.11 Masso, J., and Ning, X., "Slanted fringe holograms in DuPont photopolymers," Holographic Materials, SPIE vol. 2405, pp 37–51
6.12 Weber, A., Smothers, W., Trout, T.J. and Mickish, D., “Hologram recording in Du Pont’s new photopolymer materials,” SPIE vol. 1212, Practical Holography IV, pp 30–39, (1990)
6.13 Coleman, Z.A., Metz, M., Phillips, N.J., “Holograms in the extreme edge illumination geometry”, Holographic Materials II, SPIE Proceedings vol. 2688, San Jose, CA (1996)
6.14 Metz, M., Phillips, N.J., Coleman, Z.A., Flatow, C., “Holographic optical element for compact fingerprint imaging system”, Optical Security and Counterfeit Deterrence Techniques, SPIE Proceedings vol. 2660, San Jose, CA (1996)
6.15 Gambogi, W., Weber, A. and Trout, T.J., “Advances and applications of DuPont Holographic Photopolymers,” SPIE vol. 2043, Quebec, Canada, Aug. (1993)
6.16 Zager, S. and Weber, A., “Display holograms in Du Pont’s OmniDex films,” SPIE vol. 1461, San Jose, CA, March (1991)
6.17 Belendez, A., Pascual, I., and Fimia, A., "Model for analyzing the effects of processing on recording material in thick holograms,” JOSA A, vol. 9, No. 7, pp 1214–1223 (1992)
6.18 Belendez, A., Pascual, I., and Fimia, A., "Efficiency of thick phase holograms in the presence of shear–type effects due to processing," Journal of Modern Optics, Vol. 39, No. 4, pp 889–899, (1992)
6.19 Gallo, J., and Verber, C., "Model for the effects of material shrinkage on volume holograms," Applied Optics, Vol. 33, No. 29, pp 6797–6803

Chapter 7

REPLAY CONSIDERATIONS

7.1 Introduction

Illuminating an edge–lit hologram has proven to be a complicated task. While one of the main aims of edge–lit holograms is to reduce the size of the final package (light source and hologram) this has to be done very carefully in order to maintain proper holographic replay. The replay considerations are actually more complicated than those for recording. The radiometry, directionality, and wavelength of the light source along with the replay substrate all affect the replay of the hologram. Some of the replay restrictions can be removed by the use of a coupling grating. Most of the comments presented here are applicable to replaying HOE's as well as display holograms.

7.2 Radiometry

In typical hologram illumination, the apparent brightness of the hologram depends on Fresnel reflections, diffraction efficiency, angular and wavelength bandwidths, and the radiometry of the light source. The Fresnel reflection from the front of a typical face–lit reflection hologram would be approximately 5% for a 20° illumination angle. In an edge–lit hologram, the Fresnel reflection before the hologram can easily be 24% (Table 3.2) plus an additional 4% reflection at the edge surface in most cases. The 4% reflections could be essentially eliminated with anti–reflection coatings for both cases.

The maximum diffraction efficiency for edge–illuminated holograms is theoretically 100%, although this has not been achieved in practice at the recorded wavelength with the work presented here. For face–lit diffuse object holograms, the maximum practically achievable diffraction efficiency would be closer to 64% [7.1] compared to 100% for face-lit HOE's. The edge–lit holograms recorded in this research have only achieved efficiencies near 62% for HOE's, and a similar or less value for a diffuse holograms. Thus, at least with HOE's and the current technology, non–edge–referenced holograms are more efficient.

The apparent brightness of a hologram is also dependent on the size and location of the illuminating light source. Light irradiating a hologram from a point source must obey the *cosine law of illumination*, and the *inverse square law* which are the basic equations of photometry. These equations state that the illumination, E, of a surface is proportional to the cosine of the angle, θ, between the normal to the surface and the line directly to the source, and is inversely proportional to the square of the distance, r, from the source with irradiance, I. Mathematically, this is expressed by [7.2]

$$E = \frac{I \cos\theta}{r^2} \tag{7.1}$$

With standard face–lit holograms, θ is approximately 20°, thus the cosine factor is approximately 0.94. With edge–lit holograms, θ is near 87° and the cosine factor is approximately 0.05. Therefore, based on the cosine law, the irradiance on a typical hologram is about 19 times brighter for face–lit holograms than with edge–lit holograms. This is one of the main difficulties with lighting a edge–lit hologram. The distance from the source to the hologram is usually about 2 meters for face–lit display holograms, while with an edge–lit hologram, it may be about 10 centimeters. Considering only the inverse square law, the edge–lit hologram is about 400 times brighter. Therefore, an edge-lit hologram has the potential to be much brighter at a given point on the hologram.

However, when displaying a typical face–lit hologram with something such as a halogen bulb, a reflector is usually used to gather a larger solid angle of light from the source. This increase in size of the source from a point source to an extended source dramatically increases the intensity reaching the hologram without degrading the image because the distance is large enough for it to be considered a point source. Due to the close proximity of the source with edge–lit holograms, a reflector can not be used without changing the appearance of the source from a point to an extended area. This creates spurious images in an edge–lit hologram which degrade the appearance of the hologram as described in the next section.

7.3 Directionality

The wavefront of the illuminating light is very crucial to the performance of a hologram. Ideally, a hologram should be replayed with the same wavefront as recording for virtual replay, or the conjugate wavefront for real replay. Light incident at angles far from the Bragg condition will not diffract efficiently due to the limits imposed by the angular bandwidth. Edge–lit reflection holograms can have angular bandwidths which are similar to non–edge–referenced holograms (excluding non–slanted holograms with high angular bandwidth), although the wavelength bandwidth

is usually higher. With transmission edge–lit holograms the angular and wavelength bandwidth are smaller than the typical transmission holograms.

In order for an image hologram to be clear (or an HOE to be selective), it should have a low angular bandwidth. Replay angles which deviate slightly from the Bragg replay angle diffract light at angles different from the desired angle of diffraction. The angular spread of light from undesired angular replay and the wavelength spread from different wavelengths diffracting at undesired replay angles can create significant dispersion.

In a practical holographic replay system, one must consider three dimensional diffraction, and therefore, the angle of polarization. The polarization can change when the angle of illumination is changed by such means as reflective optics. The polarization sensitivity of a hologram is a determining factor for the diffraction efficiency of a hologram. The polarization selectivity factors on replay are generally the same for all types of holograms, and the polarization requirements on edge–lit holograms are effectively the same as for traditional face–lit holograms.

If a hologram is recorded with collimated light, the resulting grating is composed of fringe planes that are parallel to each other. When the replay angle is changed, the apparent fringe spacing changes. If this hologram is then illuminated with diverging light of the same wavelength as recording, most of the light will not diffract because the apparent fringe spacing is different and the Bragg condition is not satisfied (as shown in Fig. 7.1). The fringe spacing defines which wavelength will diffract efficiently for a given angle.

Similarly, converging light will not efficiently diffract from a holographic grating recorded with collimated light. When collimated light is used to illuminate a grating recorded with converging or diverging waves, only the central, or axial, waves will diffract as they are the only rays at the correct replay angle.

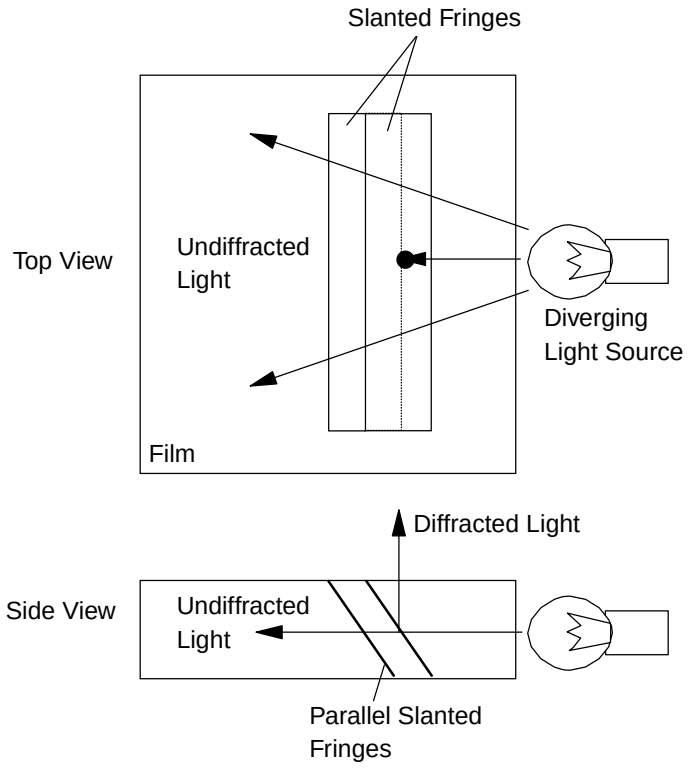


Figure 7.1 A model showing the inability of a hologram to efficiently diffract light when the apparent fringe spacing is not correct for the replay wavelength.

With edge–lit holograms, the recording could use collimated, converging or diverging beams. However, uniform intensities are easier to obtain for collimated than converging or diverging recording. The replay distance is critical in determining the type of replay source. For most face–lit display holograms, collimated or slowly diverging wavefronts can be used because the replay light sources are usually a meter or two away, thus the illumination is close to being collimated. The small distances from the light source to the hologram needed with edge–lit holograms implies that the replay optic must have a low f/# (strong divergence compared to its size) for diverging illumination. When recording a collimated edge–lit hologram, one must squeeze the replay collimating optics into a small area. Converging optics are not normally used because of the need for large, low f/# optics which are very expensive.

Linked with the directionality, the size of an illuminating source is also a crucial factor. Extended light sources can be thought of as an area of point light sources.

The angle of the light arriving in a specific area on a hologram is different from each point on an extended source. As shown in Fig. 7.1, the light at the incorrect angle will not diffract efficiently. In other words, extended sources have a very low spatial coherence and thus do not make good holographic illuminators. The effect of an extended source illumination on an HOE is chromatic and angular dispersion. For this reason, linear light sources do not replay holograms well as shown in Fig. 7.2.

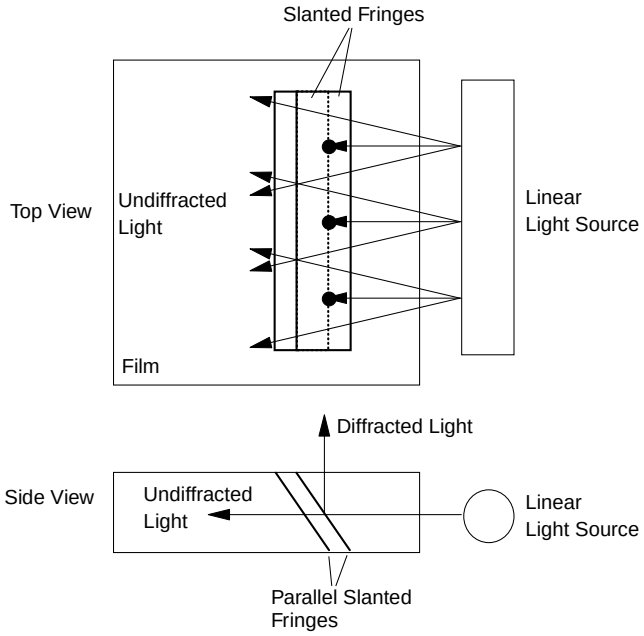


Figure 7.2 A model showing the undiffracted light from a linear light source when illuminating a grating formed from collimated beams.

Therefore, when using a linear white light source to illuminate a hologram recorded with collimated beams, the only high efficiency diffraction at the desired angle is from the collimated light near the recording wavelength. Uncollimated light at the recording wavelength will pass through the hologram (depending on the angular bandwidth). The light at angles and wavelengths far from those used in recording will create spurious images in a display hologram or increase the angular or chromatic dispersion of an HOE.

In a real system, some of the light reaching a hologram will be undiffracted. In display holograms, this is typically absorbed by a black light dump behind the

hologram (for reflection displays). For edge–lit reflection holograms the index of the black background is crucial in its ability to absorb, as the undiffracted light is incident at a steep angle. Also, in edge–lit holograms, all of the edges except the entrance must normally act as light dumps and absorb the light to prevent spurious images. However, a carefully designed edge–lit replay system could be designed so that the reflected light could be recycled to re–illuminate the hologram at the correct angle.

7.4 Replay Wavelength

The wavelength bandwidth associated with display holograms is usually a problem with regards to image brightness. Generally, the larger the wavelength and angular bandwidth, the brighter the image. The tradeoff with bandwidth for display holograms is image clarity as the more wavelength and angular selective a hologram is, the clearer the image. With HOE's, the low bandwidth is usually desired to achieve the selectivity.

In order for a hologram to have high resolution, the illuminating source should have a high degree of temporal coherence. The temporal coherence of a light source is usually measured in terms of coherence length which is related to wavelength bandwidth. Illumination sources with high wavelength bandwidths generally lower the resolution of the image for a display hologram. Two regimes around this problem which produce achromatic images are the dispersion compensation grating discussed generally in [7.3] and applied to edge–lit holograms in [7.4], or the achromatic angle as discussed in [7.3].

The diffraction efficiency of the hologram for light of a wavelength different that the Bragg wavelength (for a specific angle) is determined by the hologram's wavelength bandwidth. Generally, the wavelength bandwidth is low for reflection holograms and is high for transmission holograms. The wavelength selectivity of edge–lit transmission holograms is similar to typical face–lit transmission holograms. However, with regards to edge–lit reflection holograms, the wavelength bandwidth

for reference replay can be quite high while the wavelength bandwidth for signal replay for the same hologram is much lower. The appearance of a hologram with a high wavelength bandwidth illuminated by white light is quite bright relative to a low wavelength bandwidth hologram because of the combined intensity of all of the diffracted wavelengths. A relatively high wavelength bandwidth is sometimes very desirable for HOE's and edge–lit holograms as well.

Exactly how much the wavelength bandwidth of the hologram affects the replay depends on the wavelength bandwidth of the source. The ideal source for replay is usually the laser which recorded the hologram, although this depends on the application of the hologram. If a ion gas laser such as an argon or krypton is used, then the wavelength bandwidth for single frequency operation with an etalon is extremely low. For typical display holograms, the preferred light source is a halogen bulb with a FWHM bandwidth of around 325 nm. With edge–lit holograms, however, the heat, energy efficiency, and high bandwidth from a halogen bulb may inhibit it being used in a compact system. Laser diodes offer a solution with a wavelength bandwidth of around a FWHM of 3 nm. A cheaper, lighter, and even more compact source is a light emitting diode (LED). Bright red LED's usually have wavelength bandwidths around a FWHM of 25 nm while a superbright blue LED from Nichia has a wavelength bandwidth of FWHM of 70 nm.

7.5 Replay Substrate

Most commercially available silver halide films are coated onto glass and that is therefore the replay substrate. Du Pont photopolymer is a film laminate usually recorded on glass, which may be transferred to plastic for replay (§3.6.5). Normally, the preferred type of plastic for replay is acrylic because it is cheap and its index of refraction is close to that of the Du Pont photopolymers (they are both made from methyl methacrylate monomer). The major difference between glass and acrylic replay substrates are outlined in Table 7.1.

Glass	Acrylic
Heavy	Lightweight
Expensive if refractive index is matched to photopolymer	Cheap and no polishing necessary
No straie or blemishes	May have straie and blemishes
Can not be easily shaped (to fit in a device or to use the entrance edge as an optical element)	Can be molded to most any shape
Relatively heat resistant	Can melt or soften with heat
Relatively scratch resistant	Soft and susceptible to surface scratches
Can be left on recording substrate	Must be transferred

Table 7.1 The major advantages and disadvantages in using a glass or acrylic substrate for the replay of an edge–lit hologram.

The substrate preferred in this research was the glass substrate because of the additional variables introduced into the results when using acrylic. The effects of the epoxy on the photopolymer when transferring to acrylic, the difficulty recording on acrylic, and the acrylic edge polish contributed to this general decision to use the more expensive glass substrates.

The substrate thickness is also a crucial factor. Obviously, from a packaging point of view, the thinner the better. However, the thinner the substrate, the more difficult the edge–referenced recording and replay. During recording, the reference beam should underfill the edge to prevent diffraction lines created from the edges of the substrate, thus the thickness can not feasibly be very thin for direct edge reference recording (i.e. no waveguiding). A replay substrate should not normally need to be thicker than the recording substrate. However, replaying with a thick substrate can allow steep as well as shallow angles for recording and replay. Since the angular bandwidth for reference replay can be moderately high (Fig. 5.9e), this could create image blur. Therefore, in the results presented here, a compromise thickness of 3 mm was

normally used. Results with 0.8 mm thickness did look promising, although they were significantly more difficult to record.

7.6 Coupling Grating

Two of the difficulties encountered when replaying an edge–lit hologram include coupling the replay light through the edge of the substrate at the correct angle and the image degradation or dispersion introduced from the high wavelength bandwidth of the edge–lit hologram (§5.5.2). These difficulties can be overcome by the use of a coupling grating on the face of a recording substrate as illustrated in Fig. 7.3.

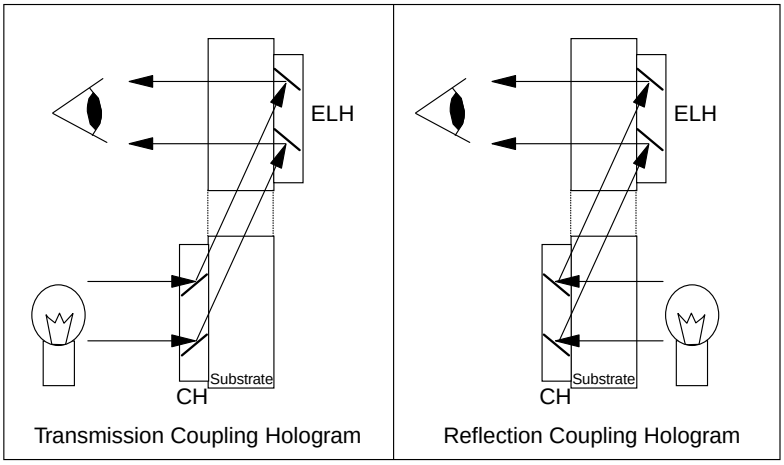


Figure 7.3 The two types of coupling holograms (CH) and their replay geometries for illuminating an edge–lit hologram (ELH). The ELH shown here is a reflection type, although it may also be a transmission or 'false transmission' hologram (§4.3.3)

The coupling hologram takes light incident normal to the face and diffracts it a steep angle towards the edge–lit hologram. The coupling and edge–lit holograms are made as typical edge–referenced transmission or reflection holograms. However, *in order for the coupling between the two holograms to be within same substrate, both the coupling and edge–lit holograms must be transferred to another substrate*.

The coupling hologram diffracts a small bandwidth of light from a broadband source at a the Bragg angle for the edge–lit hologram. The wavelength selectivity of the coupling hologram is very high for signal replay as can be seen in Fig. 5.6 (b). Also, since the light enters the face, less of the non–diffracted light is waveguided within the replay substrate than in the edge–lit case. This helps clarify the image and lower the dispersion. Thus, the non–selective edge–lit hologram is illuminated with the directed low bandwidth light diffracted from a 'face–lit' hologram. The result is a compact system which is clearer, yet just as bright as a conventional edge–lit hologram from the same source.

REFERENCES

7.1 Bjelkhagen, H.I., *Silver–Halide Recording Materials*, Springer–Verlag, Berlin, 1993
7.2 Born, M. and Wolf, E., "Principles of Optics," 4th Edition, Pergamon Press, London, 1970
7.3 Collier, R., Burckhardt, C., and Lin, L., *Optical Holography,* Academic Press, New York, 1971
7.4 Birner, S., "Steep reference angle holography: Analysis and Applications," Masters thesis, MIT, 1989

Chapter 8

APPLICATIONS AND MASS PRODUCTION

8.1 Introduction

Most of the markets for holography could benefit from the smaller space requirements that an edge–lit hologram can offer. Unfortunately, edge–lit holograms in some areas are virtually unfeasible, such as with very large display holograms (because of the disadvantages of the waveguiding light). Just as with traditional reflection or transmission holograms, the diffraction characteristics of edge–lit holograms often define the full range of applications in which the hologram can be used. Although, as mentioned in the previous chapter, the replay considerations can seriously limit the types of applications for edge–lit holograms.

First, two specific applications will be examined and then other general applications will be described. Since the production capability of a hologram is crucial to its general acceptance and use, various mass production techniques for edge–lit holograms will be reviewed.

8.2 Fingerprint Illuminator

Upon observation of one of the edge–illuminated holograms, Mike Metz of ImEdge Technology Inc., was able to see the image of a fingerprint when the finger was located in a certain position. This discovery led to the research into what is currently known as an Edge–Lit Fingerprint Illuminating Device (ELFID).

8.2.1 Background of Traditional Fingerprint Illuminators

With the growth of technology, the traditional means of ink pad and paper for obtaining a fingerprint is quickly becoming obsolete. Some systems have used a beamsplitter while most current fingerprinting devices use Frustrated Total Internal Reflection (FTIR) as a basis for imaging the fingerprint. These systems and the optics of general TIR systems are described by Harrick [8.1–8.7] and others [8.8–8.19]. Methods for obtaining a frustrated TIR image of a fingerprint usually require a camera with imaging lens, a light source, and the finger each at one face of a prism as shown in Fig. 8.1.

Unfortunately, the prism is very bulky, and the CCD camera has to be positioned laterally as well as a relatively large distance away due to the geometry and the imaging lens required. Other techniques have been recently developed which use a waveguide hologram to couple out the fingerprint image, such as those by Chennankara [8.20] and Igaki [8.21]. In these systems, the image of the fingerprint is actually waveguided back out the edge.

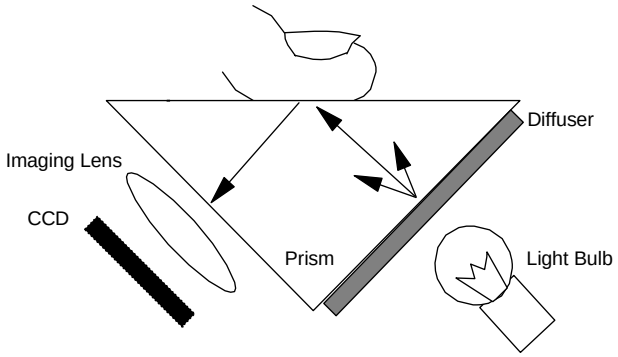


Figure 8.1 A traditional method for viewing a fingerprint using a prism and Total Internal Reflection.

8.2.2 Optics of the ELFID

The waveguide coupling method of Chennankara for viewing a fingerprint must use an imaging lens as the angles are not usually sufficiently collimated to maintain the relative locations of the ridges and valleys on the fingerprint image. The ELFID offers a clearer and less dispersive method for viewing the fingerprint with the CCD, hologram and finger co–axial.

The co–axial fingerprint imaging technology using the Fresnel reflection instead of a FTIR first developed by ImEdge for the ELFID can also be seen in the work of Drake [8.22]. In Drake's paper, he mentions looking at the 'grease print' which is left when an oily finger is removed from a substrate. This grease print was originally seen on ELFID's although the contrast is usually not as good as when the finger is in contact. The greaseprint is due to the surface profile left on a substrate when an oily finger in contact is removed. The uneven surface profile scatters light in some areas more than others, and in the far field view, light reflected from the substrate side appears brighter in the regions of the surface profile which are flatter or more planar (which usually correspond to the ridges of a fingerprint). The greaseprint image may have a higher contrast in some regimes where the substrate to CCD distance is small and the reflection from the finger is considerable (as explained later in §8.2.8).

Later in Drakes paper, he refers to the contrast in his waveguide holograms as being better in the "contact case" implying that the contrast is better when the finger is in contact (as his diagrams suggest) as opposed to in contact and removed. Essentially, there is very little difference in the optical systems of Drake than its predecessor the imaging ELFID (§8.2.3) except he uses a transmission waveguide hologram where the ELFID normally uses a reflection hologram. The basic viewing geometry of the reflection ELFID is illustrated in Fig. 8.2.

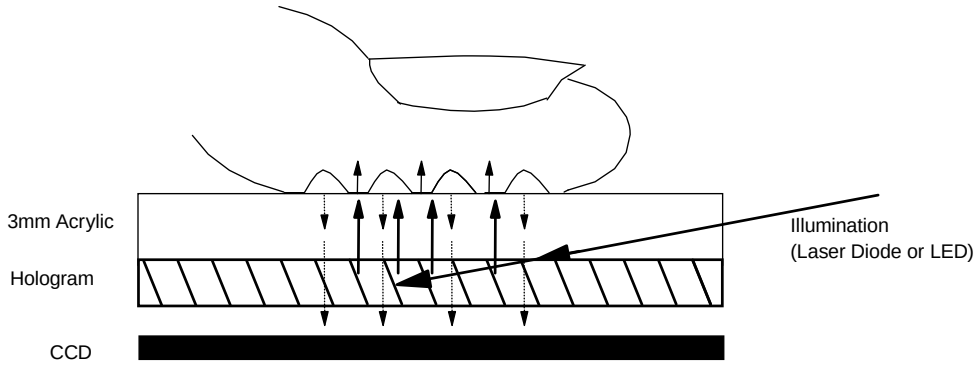


Figure 8.2 The basic optics involved with the ELFID. Where the finger is in contact with the substrate, the light is coupled into the finger. In the other areas, approximately 4% is directly reflected back toward the CCD.

Since the edge–lit hologram diffracts light in a single direction (collimated), the light reaching the CCD is a one to one correspondence to the exact size of the fingerprint area. In other words, an imaging lens is not necessary because there is just one direction of travel for all of the information bearing light reaching the CCD. This in its simplest form is much smaller than the prism methods and can be reduced to very small sizes. Since the hologram, finger, and CCD are in–line, the aberrations inherent with all TIR systems are not present. With the system setup of Fig. 8.2, much higher resolution can be obtained since the light carrying the fingerprint information is not imaged and reduced. With the ELFID system, remarkable information such as the detail of the pores on the ridges can be seen. Since the whole finger is illuminated directly, information of features within the valleys of the finger could also be obtained.

In an ELFID light enters the hologram from a steep angle through the edge of the substrate and then is directed normal to the finger. Where the finger is not in contact with the surface (no coupling), the light is reflected (normally about 4% from the Fresnel equations) back toward the CCD. Since the light is collimated from the hologram, the CCD needs to be the same size as the fingerprint area examined, but it does not require a lens to image onto the CCD.

An important factor regarding the replay of the hologram is matching the wavefront of the source for virtual replay (in the commonly used case of the non–transferred reflection edge–referenced hologram). Ideally, the light illuminating the hologram would be from a small halogen light bulb, an LED, or a laser diode very close to the edge. In order for these illuminators to work, the recording must have a diverging reference (the hologram is not being replayed in the normal phase conjugate mode). Thus, most of the research was done with diverging references during recording.

One difficulty with the replay configuration is spatial filtering of the reflected light. In other words, it is difficult to filter out the light which is coupled into the finger and remitted, and that which is scattered from the surface of the finger (both reach the CCD and degrade the image). Two possibilities of spatial filtering are shown in Fig. 8.3.

The two methods presented in Fig. 8.3 could spatially filter the light, thus removing a portion of the light reflected or scattered from the finger at angles which do not pass through the filtering elements. In (a) the light which is not within the numerical aperture of the microlens element or fiber optic faceplate will not pass through the lens. Unfortunately, a characteristic of the microlens array is that the image on each separate group of pixels of the CCD is inverted with respect to the group next to it.

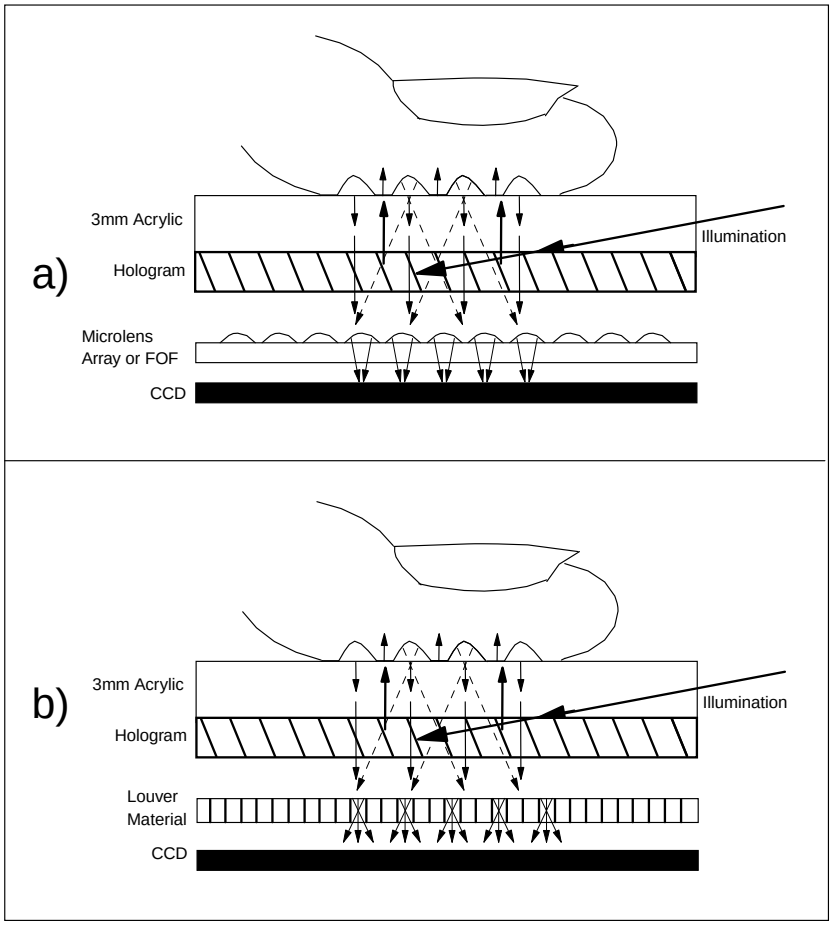


Figure 8.3 Two possible methods for spatially filtering the reflected light from the finger. (a) involves a microlens array or Fiber Optic Faceplate (FOF) while (b) uses a 'louver' type material.

In Fig. 8.3 (b) the light is filtered by being absorbed if it does not fall within the acceptable angle range for the vertical stripes of a clear material with black slats such as the louver material from 3M. The disadvantage with this method, however, is the linear shadows which are created on the pixels directly under the bands. While neither of these two methods are ideal, they do offer possibilities.

Optical clarity is also an important consideration. The reference and replay wavefronts must not be disturbed due to optical striae or poor polishing. Therefore, if one is considering recording or replaying on acrylic, it must have a very good edge polish. Even after careful hand polishing, a coherent, highly directional wavefront such as that from a laser may be disturbed through a poor surface finish on the edge of

the acrylic. Acrylic is not a good recording medium in general because of its optical properties when viewed at a steep angle. Also, its high elasticity means that it is more likely to bend or move than glass in recording. When recording an edge referenced hologram on acrylic, intensity variation bands are seen instead of a uniform reference. The inherently poor optical qualities of acrylic were one of the main reasons for the move to leaving the polymer on the recording substrate.

8.2.3 The Imaging ELFID

While one of the unique properties of the ELFID is that it does not need a lens to capture the image on the CCD, one may use a lens in order to image the fingerprint onto a smaller CCD. The tradeoff here is resolution for cost in most cases as the smaller CCD's are cheaper, yet some of the information is lost when focusing the image down. The first ELFID's were created using the collimated signal wave reflected off of the substrate surface and focused onto a CCD with a lens. However, since the hologram is itself a recordable, directional optical element, the lens can be recorded within the hologram by putting a lens in the signal beam during recording. When the effect was first discovered, and in subsequent testing, the light had to appear from the whole finger, so it was necessary for the light to converge to the pupil of the eye as in Fig. 8.4.

In order for the light to converge on replay, the light had to be converging in the recording of the initial hologram since it was being used in virtual replay. This made a large area of the fingerprint visible with the naked eye or aided with a lens for normal inspection. One might imagine a system where the image of the fingerprint was focused onto a CCD using the hologram only. However, if one were to limit the distance from the CCD to the finger to 5 mm, with a CCD size of 1/3", and a maximum fingerprint length of 1.5 inches, the lens would need to be extremely fat (with an f–number of about *f*/0.169). This 5 mm distance from the finger to the CCD is an important distance and makes the ELFID unit very compact. Since the lens which would be required for such a thin device would have a very low f–number and

be very expensive, a more reasonable solution would be to use a collimated light–normal lens imaging system or the direct lensless system of Fig. 8.2.

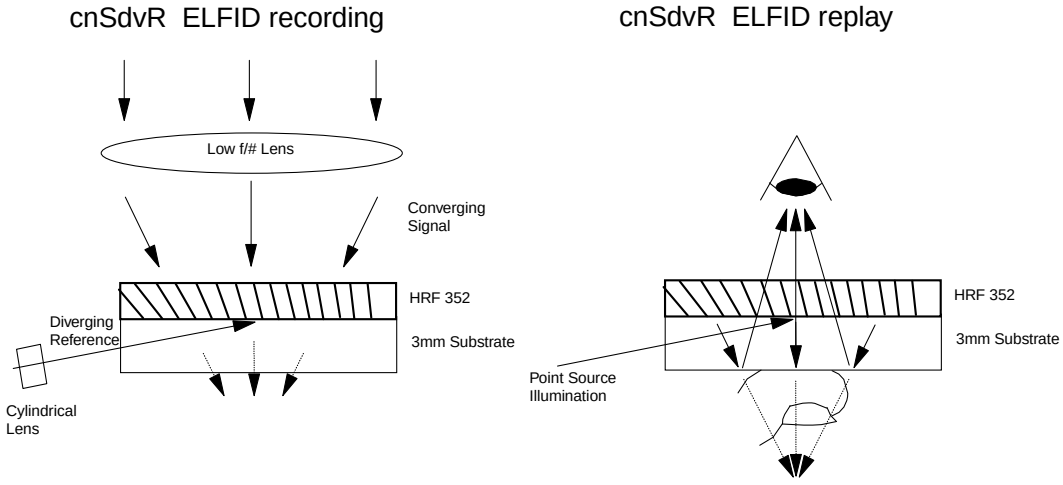


Figure 8.4 The recording and replay geometries of the cnSdvR ELFID (CoNverging Signal–DiVerging Reference). The fringes in this geometry are not parallel. The cnS series is good for viewing the entire fingerprint at a large distance with the eye or with the aid of a lens. The light from all of the fingerprint can converge to the pupil.

An important factor regarding the imaging is the distance of the CCD from the finger. When the image of the fingerprint is focused onto the CCD, this distance is increased (compared to the direct method of Fig. 8.2). The effect of this is to actually improve the image contrast because of the scattered light from the finger. The light remitted from the finger is scattered in many directions, and the amount reaching the CCD (through the solid angle of remittance) is proportional to the distance squared just as in the inverse square law of §7.2. This conclusion is dramatically realized when one tries to reduce the thickness of the device. Also, the numerical aperture of the lens can limit the angles 'seen' by the CCD which can also increase the contrast.

The detail available with the imaging lens ELFID system is very high. One can see the pore detail on the ridges with very high resolution. Depending on the imaging lens and CCD, one could see the large fingerprint pattern as in Appendix F or have high, detailed resolution of a small area as in Appendix G.

8.2.4 Angular Selectivity

When looking at Fig. 8.2, one would believe that the light from the surface of the substrate would bounce back to the hologram and then be diffracted out towards the edge instead of reaching the CCD. There are two main reasons why the system can still operate with light reaching the CCD. First, since the holograms recorded are not 100% efficient, some of the light reflected would pass through the hologram and reach the CCD. Since current CCD technology permits very low light level operation, this is not a problem.

Secondly, the inherent high angular selectivity of the edge–referenced hologram when illuminated from the face can help prevent the light from diffracting back out the edge. When replaying an ELFID, the reflected light from the substrate surface is incident to the hologram the second time near 0°. Therefore, some of the light at angles which deviate from 0° will pass through the hologram to the CCD.

If the incident light from the edge is directed and at angle which is slightly off–Bragg, it will still be diffracted towards the finger because of the reasonable angular bandwidth when illuminated through the edge. The diffracted angular deviation, θ, from the hologram normal will result in a reflection from the top surface at an angle 2θ. However, when reaching the hologram for the second time, the angular selectivity is much higher in this direction and the angle of incidence is 2θ after the reflection. Thus, a large percentage of this light passes through the hologram to the CCD. This can also be accomplished by recording the hologram at a very slight angle such as 2° which will be 4° off–Bragg after reflection and most of the light will pass through the hologram (although this may introduce unwanted reflections and aberrations).

These two reasons help explain why the light reflecting off of the substrate does not diffract back towards the edge. However, one can utilize the polarization selectivity

of the hologram to purposely minimize this diffraction back to the edge of the substrate.

8.2.5 Polarization Filtering

Since the hologram is polarization sensitive, one may rotate the polarization and the light will pass through the hologram undiffracted. This may be accomplished with the ELFID as in Fig. 8.5.

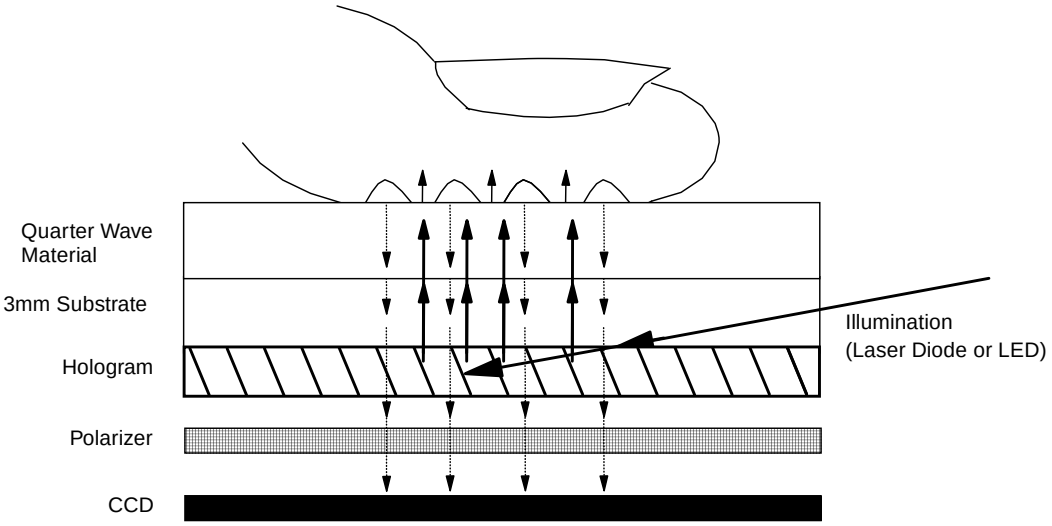


Figure 8.5 The ELFID replay geometry which uses a quarter–waveplate to rotate the reflection polarization by 90° relative to the beam incident on the hologram. The light passes through the hologram and the polarizer to the CCD.

The light diffracting from the hologram is still s–polarized. This light, after reflection at the substrate–air boundary, passes through the quarter wave material twice which rotates the polarization by 90°. This then passes through the hologram undiffracted. The polarizer in front of the CCD acts as a filter by absorbing any light which changed its polarization after scattering from the finger. To further increase the contrast, one can also polarize the illumination light (if it is an LED) or align the laser

diode polarization to s–polarization and possibly filter it as well. Results using the polarization filtering showed an increase in contrast.

8.2.6 Wavelength Considerations

Unfortunately, all of the wavelengths of *diffracted* light which do not carry information regarding the fingerprint may reflect back towards the CCD. Much of the light which is *not diffracted* may waveguide toward the finger and thus may also be scattered or reflected from the finger towards the CCD. These unwanted wavelengths fog the directed light which contains the fingerprint information. Originally, when gas ion lasers were used for illumination, wavelength filters were used before the CCD. Unfortunately, the laser line filters are very expensive and would not normally be used in a commercial device. Thus, not only is the spectrum of light entering the edge important, but just as important is the light which is reflected or scattered from the finger. The wavelengths of light reaching the hologram can be controlled by changing the light source to one with a small bandwidth as discussed in §7.4. The amount of light scattered from the finger can be altered by choosing a different recording wavelength as described in the next section.

8.2.7 Skin Optics

Ideally, in the configuration of Fig. 8.2, the image would look better if there was no reflectance from the finger. Unfortunately, there can be a significant reflectance from the finger depending on the incident wavelength. In order to determine the reflectance from the finger, the refractive index was needed, as well as information regarding the scattering and reflectance for various wavelengths. Many authors have examined the optics and reflectance of the skin [8.23–8.36] and the most thorough research which is applicable to the ELFID was conducted by Anderson and Parrish [8.37]. The three layers of the human skin–stratum corneum, epidermis, and the

dermis each have different optical qualities. The optical effects of each layer can be represented schematically as in Fig. 8.6 [8.37].

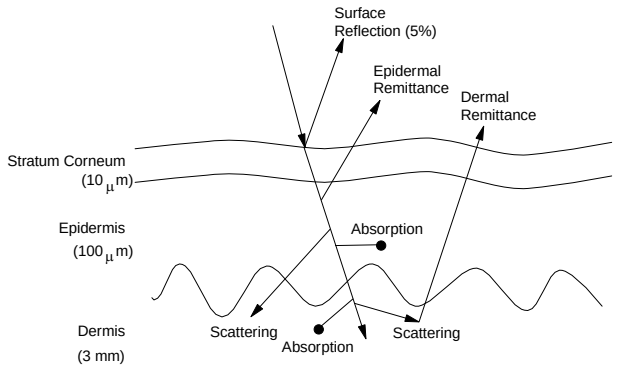


Figure 8.6 A schematic diagram of the optical properties of the three layers of skin [8.37].

The surface reflection of Fig. 8.6 is present in the valleys of the finger as on the ELFID of Fig. 8.2 where the finger is not in contact with the substrate. This surface reflection is not present in the ridges where the finger is in contact with the substrate surface. The stratum corneum and the epidermis contain melanin which absorbs shorter wavelengths more strongly than longer wavelengths. However, the melanin content can vary significantly between individuals. The thickness of the stratum corneum does not usually allow significant absorption to take place (although the thumb and fingers have a thicker stratum corneum than the one illustrated). The remittance of the incident light in the stratum corneum and the epidermis is minimal at visible and near IR wavelengths [8.37].

In the dermis, which does not contain any melanin, the scattering and reflecting is higher than in the other two layers. The transmittance through this layer is higher and forward directed for higher wavelengths within the range of 500 nm to 1.2 μm. Thus, the shorter wavelengths are absorbed and reflected within this layer.

Anderson presents data showing that the reflectance from a 200 μm layer of dermis is about 32% at 400 nm while the transmittance is near 10%. At 700 nm, the reflectance is about 28% and the transmittance is near 43%. Thus, if one were to look at only the

dermis it would appear blue. Without knowing exactly how much of the shorter or longer wavelengths is absorbed in the stratum corneum or the epidermis, one can not determine exactly which portion of the spectrum has a lower reflectance.

Since the reflectance is what needs to be minimized in the basic ELFID setup of Fig. 8.2, the total reflectance of the finger was measured in an edge–lit setup and is shown in Fig. 8.7

The reflectance of Fig. 8.7 was measured normal to the substrate with the light incident at a steep angle so that the spectral reflectance of stray light (or diffracted light) in the ELFID could be determined. The reflectance of light incident at a steep angle is essentially the same as many other measurements of skin reflectance using spectrophotometers and diffuse reflectance spheres [8.33–8.37].

As one can see from Fig. 8.7, the light reflecting off of the finger decreases with wavelength. This occurs despite the stronger tendency of the dermis to reflect the shorter wavelength light because of the stronger absorption of the shorter wavelength light in the epidermis and stratum corneum. Therefore, to reduce the reflectance which reaches the CCD, the recording wavelength was changed from 647 nm to 458 nm which should result in a reflection reduction of approximately 33% (from Fig. 8.7).

When recording with 647 nm, one is able to replay with laser diodes (thus narrow bandwidths near 3 nm). However, for replaying a blue recorded hologram, one must resort to blue LED's until the commercial availability of blue laser diodes. Unfortunately, the LED's reduce the contrast of the fingerprints because of the large bandwidth (FWHM of 70 nm).

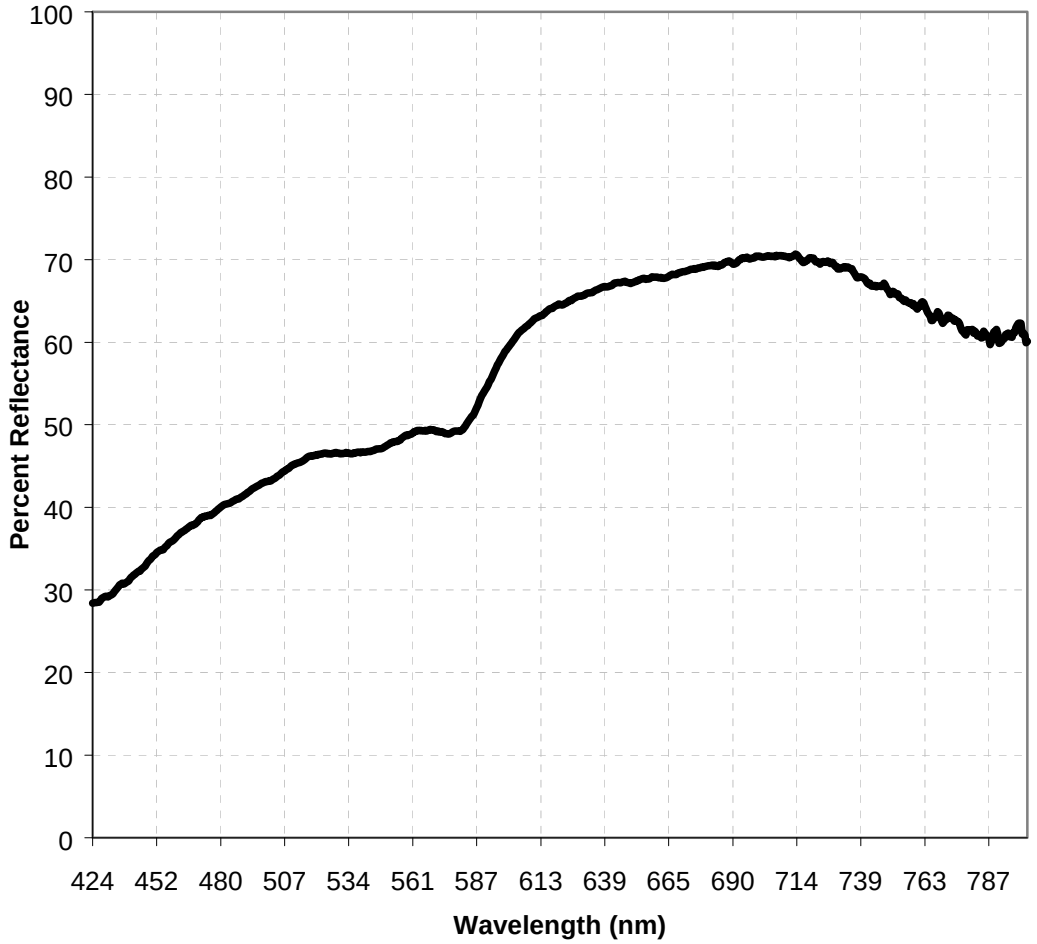


Figure 8.7 The normal reflectance of the author's finger when illuminated with white light at a steep angle.

8.2.8 Contrast Considerations

The final fingerprint image contrast on the CCD is based on the light arriving from all angles. Some incident light is remitted from the finger towards the CCD in all of the fingerprint illuminating methods. However, the effects of the remittance on the contrast are minimized in some illumination geometries as shown Fig. 8.8.

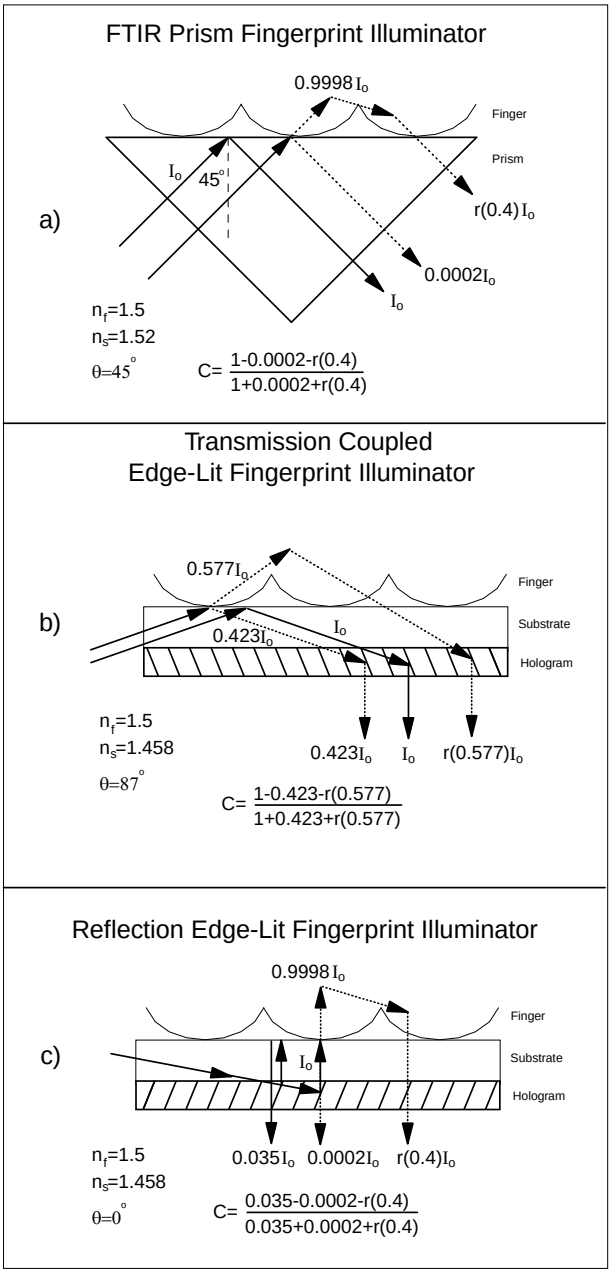


Figure 8.8 The direct and remitted light reaching the CCD in the (a) FTIR prism, (b) transmission coupled ELFID, and (c) reflection ELFID geometries for fingerprint illumination. The incident light, I_o, passes through the substrate with refractive index n_s at an angle, θ, and reaches the finger, n_f. The remittance from the finger which reaches a particular spot on the CCD is dependent on the angle, wavelength, individual skin characteristics and is represented here by the factor, *r*. The contrast, C, for the fingerprint image depends on Fresnel reflections, diffracted light from the hologram, and remitted light from the finger.

As can be seen from Fig. 8.8 in the FTIR and the reflection ELFID geometries, the light leaves the substrate to the CCD from Fresnel transmittance. In the transmission coupled ELFID, the light is coupled out of the substrate by means of a transmission hologram (this is similar to the waveguide hologram devices of Chennankara [8.20] and Igaki [8.21]).

The system with the best contrast is the FTIR prism geometry regardless of the remittance factor, *r*. The coefficient of 0.4 for the remittance in (a) and (b) is based on the finger reflectance at a wavelength of 458 nm from [8.34]. In the calculations, the holograms are assumed to have a diffraction efficiency of 100% for simplicity of comparison. The diagram does not illustrate the transmission ELFID geometry or the 'false' transmission ELFID geometry as these have similar contrast considerations as the reflection ELFID hologram. The transmission coupled and reflection ELFID cases (b) and (c) have lower contrasts which depend crucially on the remittance from the finger. In (b) the Fresnel reflection from the finger is a strong contributing factor which is believed to be larger than the remittance factor, thus playing a large role in decreasing the contrast. In (c), the Fresnel reflection from the finger is very small and virtually negligible. The remittance factor for all of these cases depends on the angle and the distance from the finger where it is measured. Thus, when the distance from the finger to the CCD is considerable (approximately 10 mm for example), the remittance factor *r*, is quite small. Therefore, between (b) or (c), the one with the higher contrast for small values of *r* is (c). This is believed to be the reason why the ELFID's with the best contrast have been produced by the reflection geometry of (c). In terms of the replay geometries (b) and (c), the only difference in the *r* values is essentially the difference in the light remitted normal to the finger surface when illuminated from a steep angle (b) and when illuminated normal to the surface (c). In both of these cases, the *r* value is believed to be similar.

In the prism geometry, the illumination light is at 45° and the examined light is at 45° from the plane of the fingerprint (which can vary depending on the FTIR geometry or prism chosen). However, in the prism geometry, since the large face must be approximately the size of the finger, this fixes the maximum distance from the fingerprint plane to the imaging lens. For example, in a 45° prism, if the large surface

of the prism is 2 cm x 2 cm, then the *minimum* distance to the lens for full field of view in the imaging system is 1.41 cm. This distance combined with the distance in the imaging optics dramatically reduces the *r* value, thus helping to explain the high contrast in the prism method. Also, this shows some of the size constraints associated with using the prism system.

One can determine the maximum contrast possible by letting *r* approach zero. While the maximum contrast in the prism system is approximately 0.99, the system occupies a large volume. The maximum contrast in the transmission coupled ELFID is 0.4 and the maximum in the reflection ELFID is 0.989. Thus, the reflection geometry (or a similarly equivalent transmission or false transmission geometry) has the greater potential for a high contrast compact fingerprint illumination system.

8.3 LCD Illuminator

One of the largest costs in full color liquid crystal displays is applying the red, green, and blue dyes on the individual pixels. Since holograms can inherently diffract selected wavelengths of light from a white light source, they seem a definite possibility for illuminating a LCD. An edge–illuminated LCD also seems feasible for lighting an LCD panel in a compact device. This opens the possibility for televisions hanging on the wall with no external illumination or projection.

The LCD illuminator termed Holographic Edge–Lit Panel, or HELP for short, consists of the basic edge–lit hologram, but includes a mask pattern and a special illuminating geometry. The HELP is created using the geometry of Fig. 8.9.

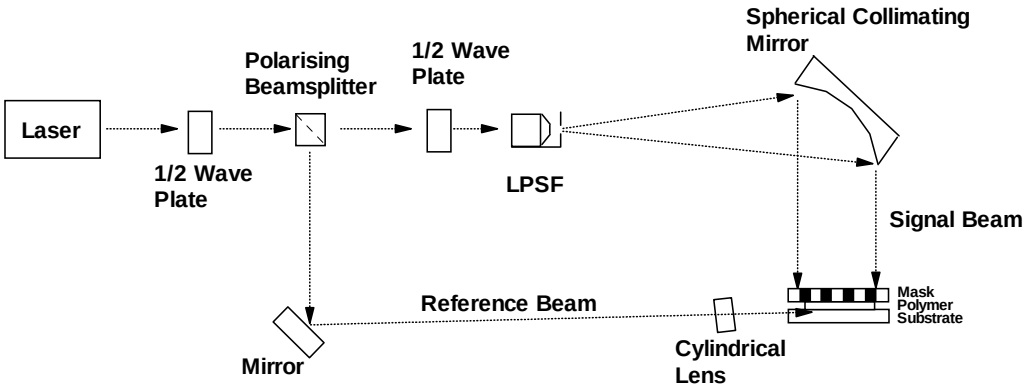


Figure 8.9 The basic recording geometry for a HELP

In Fig. 8.9 the cylindrical lens diverges the reference and the mirror collimates the light which then passes through the mask. The set–up can be modified for the below replay possibilities (The abbreviations cl, cn, and dv represent collimated, converging, and diverging beams respectively, and S, and R, represent the Signal and Reference. For example, clSdvR represents collimated signal and diverging reference):

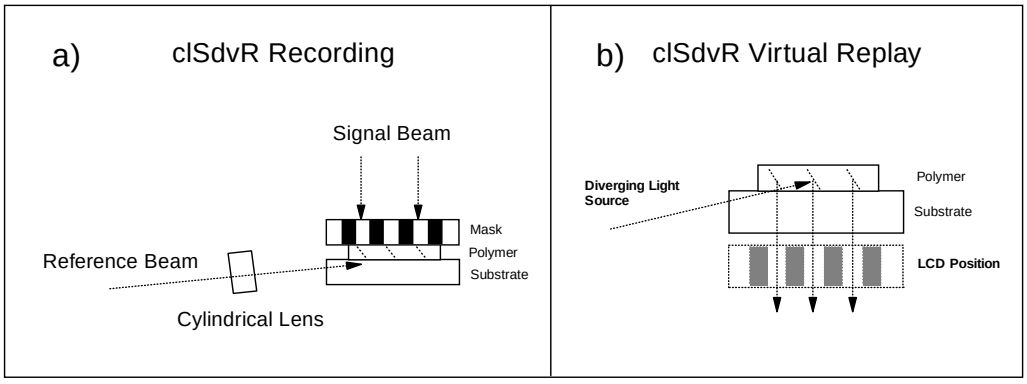


Figure 8.10 The clSdvR recording (a) and replay (b) geometry for the HELP.

With a clSdvR recording, a diverging point source for the virtual (non–conjugate) replay will reproduce the collimated signal beam which may pass through the individual pixels of an LCD. Note that there will be a minimum distance factor from the hologram to the LCD which is the thickness of the replay substrate. The real (conjugate) replay geometry of this recording would require a converging light source which is not practical in terms of size restraints.

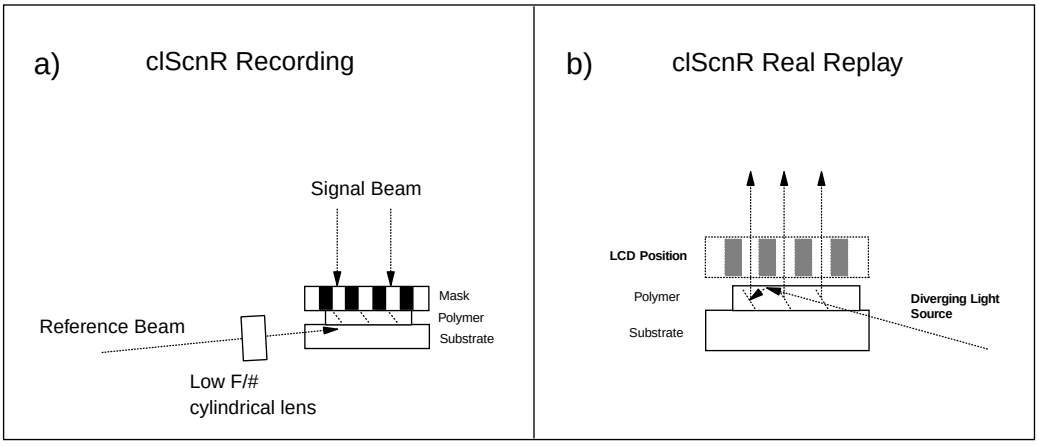


Figure 8.11 Recording a reflection clScnR hologram (a) and replaying it as a 'false transmission' hologram (b) as a result of the total internal reflection (TIR) inside the hologram.

In Fig. 8.11 the TIR results from the polymer–Mylar interface, or the polymer–air interface if the Mylar support layer is removed. The hologram diffracts the TIR light from apertures corresponding to the same positions in the mask. Note that in this geometry, the LCD can be essentially in contact with the polymer. Results seem to show that replaying in this geometry is not as bright as the normal reflection replay as in Fig. 8.10. The reason for this diminished brightness is uncertain, however it may partially be a result of absorption from residual dye as the light passes through the layer once before reaching the fringes at the Bragg angle.

The difficulty with a clSclR HELP (Fig. 8.12) is the reconstruction using collimated light. Tests have shown that a linear light source reconstructs several images because the light is still not collimated as described in §7.3. A collimated source would normally involve large space considerations relative to the display. A method using a film with black slats which absorbs light not within a defined angular range (such as 3M's louver material) in combination with a linear bulb may be possible to simulate linear collimated light (however, this may leave dark 'streaks' on the illumination area from the shadows of the slats or create diffraction effects if the slats are to small).

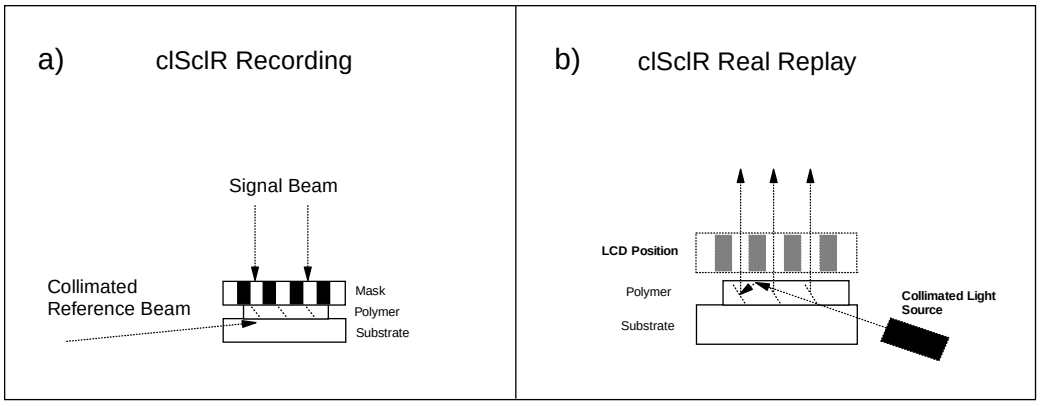


Figure 8.12 Recording (a) and replaying (b) an edge referenced hologram with a collimated reference and collimated signal.

In Figs. 8.10 through 8.12, the signal may be converging or diverging as well. One would hope that having diverging light from the pixels in replay would increase the angle of view, however, the light is specular and not diffuse. This would result in seeing only those pixels close to the viewer off to one side. For example, if the a viewer looked from the left side of the screen, only the pixels from the left side would be visible. The light from the pixels on the right side of the screen would be traveling in the opposite direction unless each individual pixel was recorded with a diverging beam.

The collimated light from the hologram on replay through the CCD essentially has no field of view. A simple method to increase the angle of view is to add a diffuser to the surface of the LCD (such as the Microsharp™ diffusion film developed by Nashua Corporation) so that the light diverges from the LCD as illustrated in Fig. 8.13.

Another method to increase the field of view of the light from the LCD would be to add a microlens structure on top of the LCD or introduce it into the recording mask. Thus, each individual pixel would have its own diverging wavefront.

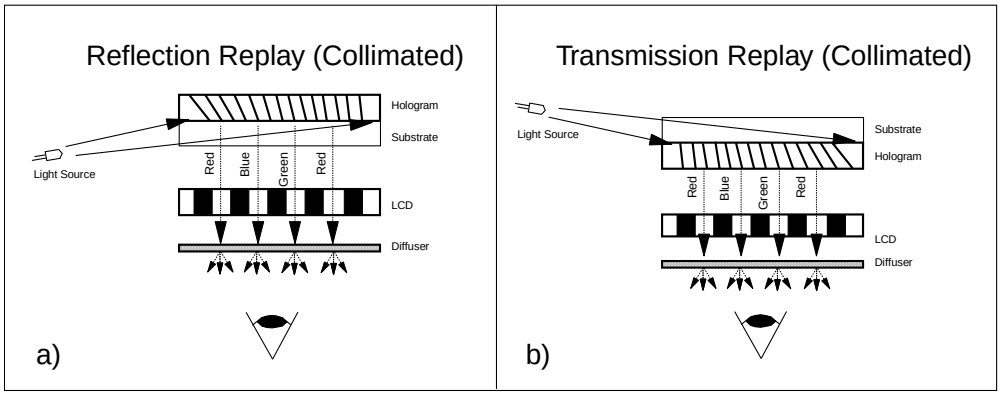


Figure 8.13 The reflection (a) and transmission (b) replay geometries for a collimated signal HELP with an LCD and diffuser.

One might also try to incorporate a diffuser in the hologram during recording. Tests were made using a diffuser in the recording by laminating it onto the mask so that the signal beam becomes diffuse just before entering the mask. In this geometry, the area of signal light that reaches the hologram is larger than the pixel size (due to the spread of the diffuser). Thus, the size of each holographic pixel (defined as the area of the grating which diffracts to one pixel in the mask) has a larger surface area which can use more of the incident light for one pixel.

Several advantages of the HELP over normal LCD backlights (where light directing sheets and crude diffusers are used) is the reduction in size, the correct replay of the corresponding color on the correct pixel (three holograms–red, green, and blue can be created to give the separate colors of the pixels), and no polarizer is needed because the hologram is polarization sensitive. Almost half of the cost of LCD's are involved in dying each of the pixels in the red/green/blue pattern. With the HELP, the color separation is in the holographic optics.

A HELP is in a sense, a series of minute holograms the same size as the pixels (unless a diffuser is used in recording). Thus, only registration is required for correct illumination of the LCD. To determine the ability for the hologram to illuminate the LCD at varying distances from the LCD with a white light source, a test hologram was illuminated with a halogen bulb. The light appeared to be emanating from the individual pixels out of the large face of the substrate. When the original mask is

place on the substrate as in recording, the alignment can be judged based on the Moire fringe pattern (the visible dark bands which appear when two amplitude masks are mis–aligned). The Moire pattern also proves that the recording was successful in generating the light in the proper mask pattern. The results appeared to be very successful for a single 514 nm recording with white light replay.

The wavelength bandwidth of edge–illuminated holograms is very crucial in creating a bright, colorful liquid crystal display system. Fortunately, the reflection edge–referenced holograms can have relatively high wavelength bandwidths when replayed through the edge of the substrate (Fig. 5.6e). The bandwidth is very dependent on the refractive index profile, but a FWHM of 60 nm should not be difficult to obtain and should have a considerable apparent brightness as an illuminator for an LCD.

A difficulty which may arise is the angular dispersion from the hologram. The off–Bragg wavelengths will diffract from the hologram (as defined by the wavelength bandwidth) and these wavelengths will diffract at different angles. If this dispersion is significant, it could result in a problem with image quality with the illuminated LCD. Some of the problems which arise from this dispersion may be alleviated by a using a diffuser on replay. By virtue of being an edge–lit hologram, the HELP suffers from illumination difficulties as mentioned in Chapter 7. Some of these, can be alleviated as described by the use of a coupling hologram (§7.6).

8.4 Other Applications

An edge–illuminated hologram can be used in place of a face–lit hologram in many different optical systems. The diffraction characteristics of the edge–lit hologram, are different than most typical holograms and may restrict its use in some applications. In addition, some of the special characteristics of the edge–lit hologram make it very useful in other applications. The different types of edge–lit hologram applications can be divided into HOE's and display holograms.

8.4.1 Holographic Optical Elements

The size reduction which an edge–lit hologram offers is very appealing in most applications and some systems may be easily altered to compensate for the diffraction characteristics distinctive to edge–lit holograms. In some cases, the diffraction characteristics may open up the edge–lit hologram to applications. The wavelength and angular bandwidths of an edge–lit hologram as the fringes approach 45° can be significantly small when replayed at the signal angle (Figs. 5.6b and 5.9b). Thus, the edge–lit hologram could be used as a wavelength or spatial filter which diffracts the unwanted light out the edge of the substrate.

Most HOE's can be converted to similar edge–lit hologram if desired. There is no difference with the signal beam since it can have any wavefront which a face–lit hologram can have. The only difference is the reference beam wavefront. The basic edge–lit geometry will limit the reference angles for recording and replay and the degree of divergence or convergence.

One HOE mentioned early in the history of edge–illuminated holograms is the Head–Up Display, or HUD. This is a very useful application of an edge–lit hologram as first described by Upatnieks [8.38]. In display systems such as those in a cockpit or even in cars of the future, the illumination must not occupy a large volume and the transmitted zero–order light must not be seen.

A potentially large market for edge–lit holograms is using them as HOE's for illuminating signs. Illuminating signs usually wastes a tremendous amount of light and electricity. Since the edge–lit hologram can recycle the replay light which is not diffracted, it potentially has a much better optical efficiency. In addition to the obvious size benefits, the diffraction characteristics of an edge–lit hologram such as bandwidth or field of view may be beneficial for some types of signs such as those which should not be viewed at some angles.

8.4.2 Display Holograms

As one of the first applications of holography, display holograms have always had the potential for creating a beautiful three–dimensional image which could hang on anyone's wall. Unfortunately, one of the main drawbacks in display holography is the lighting requirements. One of the main driving forces in edge–lit holography research has been the potential for removing the external lighting difficulties of holograms and simply bring the light source into the frame.

Most of the researchers in the field of edge–illuminated or waveguide holography have demonstrated display holograms because it is usually a simple method for demonstrating the specific type of hologram. Special techniques which apply to traditional face–lit display holograms can also apply to edge–lit holograms. Benton *et al.* [8.39] successfully demonstrated that a transmission rainbow hologram could be made as an edge–lit hologram. Birner [8.40] showed that a dispersion compensation grating could be used with an edge–lit hologram to create an achromatic hologram.

An interesting phenomenon discovered in this research is the ability to display a reflection hologram in a 'false transmission' mode as illustrated in Fig. 8.14.

The false transmission mode of Fig. 8.14(c) offers and unique method for viewing the orthoscopic image of a reflection H2 hologram by using the reflected TIR light within the hologram as the illumination. This bounce can occur at the Mylar–air boundary or at the hologram–air boundary if the Mylar is removed. This replay geometry is unique to the edge–lit hologram and offers the possibility to replay a reflection hologram in a transmission geometry while maintaining the diffraction characteristics of a an edge–referenced reflection hologram.

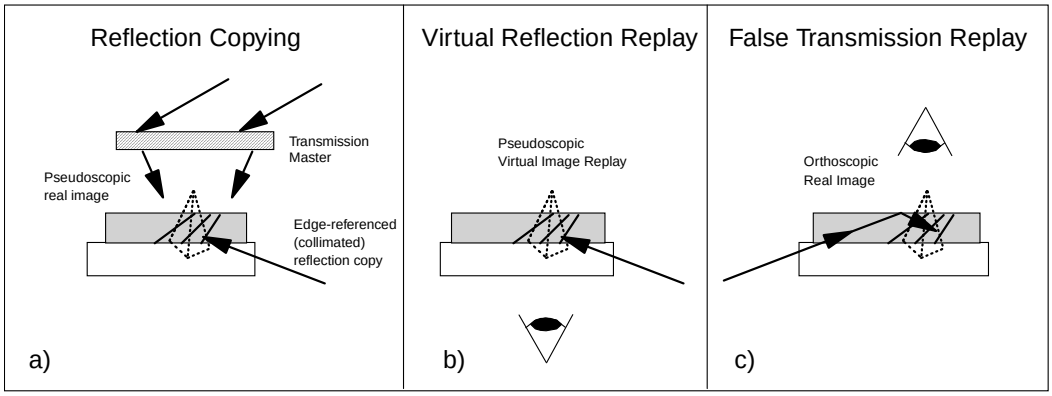


Figure 8.14 Two replay possibilities of an edge–referenced reflection hologram recorded and replayed with collimated light.

8.5 Mass Production

The stability and coherence requirements for direct recording of volume holograms often rules out the quick and typical mass production techniques employed with other optical elements. However, three different techniques have emerged which can possibly allow mass reproduction of edge–lit volume holograms. These techniques are *reflection contact copying*, *alternative Bragg recording*, and *direct fringe copying*.

8.5.1 Reflection Contact Copying

Reflection holograms have the ability to be copied by illuminating the hologram through a recording medium in contact with the hologram. The first order reflected (diffracted) light then interferes with the incident light inside the unexposed film. This creates an exact copy of the original hologram in a single beam geometry. Contact copying was attempted with an ELFID and yielded reasonable efficiency (slightly less than the original). The set–up for copying was arranged as in Fig. 8.15.

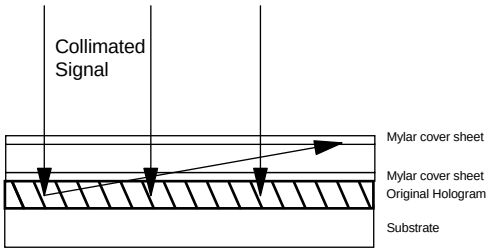


Figure 8.15 The contact copying geometry for ELFID's with a collimated laser for the replayed signal beam. The light diffracts in the direction of the original reference beam and interferes with the replay light within the film.

The efficiency of the contact copy holograms could have been less than the original because the effective beam ratio is much less than unity (due to the absorption) as is the case with most single beam recording geometries. Contact copying using the reference beam would be extremely difficult to set–up on a mass production scale due to the tight angle constraints. Therefore, copying using the signal beam seems to be the best method for copying using a laser. Unfortunately, this is virtually limited to collimated signal holograms. When contact copying display holograms with the signal beam (i.e. the projected image from a transmission master) it is very difficult to align the image at exactly the same angle and focus depth as the copy hologram. The same difficulties applies for HOE's when matching a non-collimated wavefront for signal beam contact copying. The high angular selectivity of the edge–referenced hologram when replayed through the signal makes alignment for non–collimated wavefronts very difficult, thus yielding inefficient copies.

In order to copy more than one hologram at a time, one must record the holograms using a collimated signal beam. With a collimated signal one could copy several holograms at a time by putting them together and using a large collimated beam, or alternatively, one could use a laser line scanner to scan across a sample such as in the Du Pont copying machine [8.40]. This continuous copying roll–fed format can offer excellent production times and capacities.

One of the benefits of using the contact copying method is the lower spatial coherence required on the copying laser. Since the copy and the original hologram are in contact with each other, movement relative to each other is much less likely. Also, the distance between the original and the copy is usually just the thickness of the Mylar cover layer (50 μm). This thickness could be reduced in the packaging which could lower the coherence requirements of the laser (or other light source). There must, however, be a separation layer between the original and copy if they both are photopolymer because the solvents in the fresh, unexposed photopolymer will react with the original recorded hologram. If the original is silver–halide, then solvent damage is much less likely, although a protective film is advised.

8.5.2 Alternative Bragg Condition Recording

Using the alternative Bragg recording method described in §4.5, a hologram with parallel fringes can be directly made with relaxed angles at a different Bragg condition. This technique is limited to collimated wavefronts for the reference and signal beams (unless a complex wavefront conversion technique is applied). This could be accomplished on a roller for continuous or on a pair of prisms for non–continuous recording as illustrated in Fig. 8.16.

The reflective component in Fig. 8.16 could be a mirror, retroreflector, blazed grating or another device depending on the desired fringe angle within the photopolymer. Large format holograms could be made using this technique. If glass rollers are used instead of prisms, then the recording angles must be adjusted for the angles of refraction in the rollers. One must note that if the fringe angle is not at 45°, then the angle introduced by the reflective component means that the area of interference is smaller than the total area exposed. One must realize this and be careful not to fog the area before the subsequent exposure.

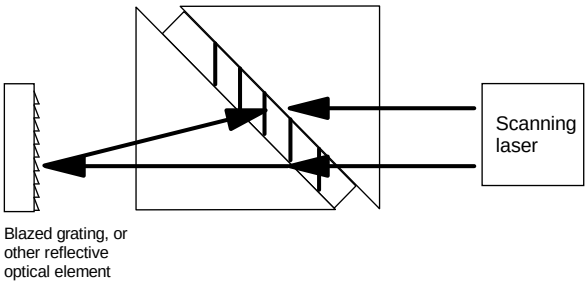


Figure 8.16 One possibility for recording a large hologram using the alternative Bragg recording geometry.

8.5.3 Direct Fringe Copying

The non–coherent conversion of amplitude fringes to volume phase fringes has been described in §4.7 and is referred to here as direct fringe copying. Some of the benefits of this method are very useful when considering mass manufacture techniques. The light source required for copying does not need to be a laser as a broadband UV source will work very well. The cost of a high power lamp is very cheap compared to that of a laser and the maintenance is significantly easier. However, the directionality of the light from the lamp must be controlled to match the amplitude holograms fringe angles for proper copying.

The size of the copy hologram in the direct fringe copying method is only limited to the size of the transmission master or similar amplitude grating. Also, the system can easily be set up on a roll–feed continuous copy arrangement. This method is limited to slanted holograms, and the coupling into the copy film becomes more difficult as the angle is increased due to reflections. A prism allowing UV penetration would then be required. Also, as the fringe angle increases towards a pure reflection hologram, the fringe spacing will decrease and diffraction effects from the edges of the amplitude grating may become a problem. Testing the diffraction effects on small

fringe spacing holograms in the direct fringe copying method is left for further research.

REFERENCES

8.1 Harrick, N.J, *Internal Reflection Spectroscopy*, Harrick Scientific Corporation, New York, 1987

8.2 Harrrick, N.J., "Use of frustrated total internal reflection to measure film thicknesses and surface reliefs," Journal of Applied Physics, Vol. 13, No. 9, pp 2774–2775

8.3 Harrick, N. J., "A continuously variable optical beam splitter and intensity controller," Applied Optics, Vol. 2, No. 11, pp 1203–1204

8.4 Harrick, N. J.,"Electric Field Strengths at Totally Reflecting Interfaces," JOSA, Vol. 55, No. 7, pp 851–857, (1965)

8.5 Harrick, N. J., "Determination of refractive index and film thickness from interference fringes," Applied Optics, Vol. 10, No. 10, pp 2344–2349,(1971)

8.6 Harrick, N. J., and Carlson, A. I., "Internal reflection spectroscopy: Validity of effective thickness equations," Applied Optics, Vol. 10, No. 1, pp 19–23, (1971)

8.7 Harrick, N.J. and du Pre, F.K., "Effective thickness of bulk materials and of think films for internal reflection spectroscopy," Applied Optics, Vol. 5, No. 11,pp, 1739–1743 (1966),

8.8 Chakravarty, B. N., "On the diffraction of light incident at nearly the critical angle on the boundary between two media," pp 503–511??(1921)

8.9 Bryngdahl, O. “Evanescent Waves in Optical Imaging,” *Progress in Optics* XI, E. Wolf, North–Holland, pp 167–221, (1973)

8.10 Burghardt, T. and Thompson, N., “Evanescent intensity of a focused Gaussian light beam undergoing total internal reflection in a prism,” Optical Engineering, Vol. 23, No. 1, pp 62–67, (1984)

8.11 Edwards, J., Ausserre, D., Hervet, H. and Rondelez, F., “Quantitative studies of evanescent wave intensity profiles using optical fluorescence,” Applied Optics, Vol. 28, No. 10, pp 1881–1884, May (1989)

8.12 Renard, R., “Total Reflection: A new evaluation of the Goos–Hanchen shift,” JOSA Vol. 54, No. 10, pp 1190–1197, Oct. (1964)

8.13 Lee, E., Benner, R., Fenn, J. and Chang, R., “Angular distribution of fluorescence from liquids and monodispersed spheres by evanescent wave excitation,” Applied Optics, Vol. 18, No. 6, pp 862–868, (March 1979)

8.14 Allain, C., Ausserre, D., Rondelez, “Direct optical observation of interfacial depletion layers in polymer solutions,” Physical Review Letters, Vol. 49, no. 23, pp 1694–1697, Dec. (1982)

8.15 Hall, E., “The penetration of totally reflected light into the rarer medium,” Phys. Rev. , Vol. 15, pp 73–106, (1902)

8.16 Nassenstein, H, “Superresolution by diffraction of subwaves,” Optics Communications, Vol. 2, No. 5, pp 231–234, Oct. (1970)

8.17 Carniglia, C. and Mandel, L., “Absorption and emission of evanescent photons,” JOSA, vol. 62, No. 4, pp 479–486, April (1972)

8.18 Zhu, S., Yu, A., Hawley, D. and Roy, R., “Frustrated total internal reflection:a demonstration and review,” Am. J. Phys., vol. 54, No. 7, pp 601–607, July, (1986)
8.19 Jong, A. E., "A novel prism for total reflection," Oct. (1962)
8.20 Chennankara, B., et al.,"Optical fingerprint recognition using a waveguide hologram," Applied Optics, Vol. 34, No. 20, pp 4079–4082 (1995)
8.21 Igaki, S., et. al.,"Real–time fingerprint sensor using a hologram," Applied Optics, Vol. 31, no. 11, pp 1794–1802, (1992)
8.22 Drake, M., Fiddy, M.,"Waveguide hologram fingerprint entry device," Optical Engineering, Vol. 35, pp 2499–2505, (1996)
8.23 Penrose, L, "Fingerprints, palms, and chromosomes," Nature, Vol. 197,No. 4871, pp 933–938
8.24 Harrick, N.J., "Fingerprinting via total internal reflection," Philips Technical Review, Vol. 24, No. 9, pp 271–274, (1962)
8.25 Li, Hui, and Xie, Shusen,"Measurement method of the refractive index of biotissue by total internal reflection" Applied Optics, Vol. 35, No. 10, pp 1793–1795, April (1996)
8.26 Tuchin, V. V., "Laser light scattering in biomedical diagnostics and therapy," Journal of laser Applications, Vol. 5(2,3), pp 43–60 (1993)
8.27 Gandjbakhche, A. H., Schmitt, R., Bonner, R., and Nossal, R., "Random walk theory applied to non–invasive in vivo optical measurements of human tissue," Annual Conference on Engineering in Medicine and Biology, pp 332–333, (1992)
8.28 Cheong, W, Prahl, S., Welch, A., "A review of the Optical properties of Biological Tissues," IEEE Journal of Quantum Electronics, Vol. 26(12), pp 2166–2185, (1990)
8.29 Kashima, S., “Noncontact laser–tissue blood–flow measurement using polarization to reduce the specular reflection artifact,” Optics and Laser Tech., Vol. 26, No. 3, pp 169–175, (1994)
8.30 Bolin, F. and Preuss, L., “Refractive index of some mammalian tissues using a fiber optic cladding method,” Applied Optics, vol 28, pp 2297, (1989)
8.31 Vangemert, M., Jacques, S., Sterenborg, H. and Star, W., “Skin Optics,” IEEE Trans. on Biomedical Eng., Vol. 36, No. 12, pp 1146–1154, (1989)
8.32 Anderson, R.R..,"Polarized Light Examination and Photography of the Skin," Archives of Dermatology, Vol. 127, pp1000–1005, (1991)
8.33 Kopola, H., et. al.,"Two–channel fiber optic skin erythema meter," Optical Engineering, Vol. 32, No. 2, pp. 222–225, (1993)
8.34 Jacquez, J., et. al.,"Spectral reflectance of human skin in the region 235–700mμ," Journal of Applied Physiology, Vol. 8, pp 212–214, (1956)
8.35 Everett, M., et. al.,"Penetration of epidermis by ultraviolet rays," Photochemistry and Photobiology, Vol. 5, pp. 533–542, (1966)
8.36 Buettner, K.,"The effects of natural sunlight on human skin,"The Biologic effects of ultraviolet radiation, Proceedings of the first conference, Permagon Press, London, pp 237–249 (1969)
8.37 Anderson, R. and Parrish, J., “The optics of human skin,” Journal of Investigative Dermatology, Vol. 77, No. 1, pp 13–19
8.38 Upatnieks, J., "Compact holographic sight," SPIE Vol. 883, Holographic Optics:Design and Applications, pp 171–176 (1988)

8.39 Benton, S.A., Birner, S., and Shirakura, A., "Edge–lit rainbow holograms," Practical Holography IV, SPIE vol. 1212, pp 149–157
8.40 Armstrong, M. and Tipton, D., “Machine for continuous hologram production,” SPIE vol. 1914, pp 134–154

Chapter 9

CONCLUDING REMARKS AND DIRECTIONS FOR FURTHER RESEARCH

The large potential for edge–illuminated holograms still remains largely untapped. As a result of the difficulties involved with recording and replaying edge–lit holograms, they have not been the center of widespread research. The recording and illumination difficulties are carefully examined here and the limits imposed by current technology, such as replay light sources, crucially affects edge–lit hologram performance.

When recording an edge–referenced hologram, the absorption of the dye in the film can play a significant role in reducing the recorded fringe contrast. Knowing the thickness, absorption of the dye, and the angles involved in recording, one can maximize the average contrast and improve the diffraction efficiency of the hologram.

The recording material used for recording an edge–lit hologram is very crucial to the performance of the hologram. Results shown here illustrate the extreme importance

of closely matching the refractive index of the substrate and the film. With silver halide holograms, the indices are not matched close enough for efficient recording and replay. The coupling of light from the substrate to the film is much higher when using a recording film such as Du Pont photopolymer which can be index matched with a recording substrate. Excellent results have been obtained from holograms recorded and replayed at steep angles using special glass substrates with the index of refraction very near (less than) that of the Du Pont photopolymers.

Other photopolymers or recording films may be used, but the refractive indices, packaging, and possible real–time effects must be considered. Fortunately, the real–time effects of the Du Pont photopolymers can offer diagnostic information and the fogging and real–time diffraction do not seem to significantly degrade the performance in experimental samples.

Both transmission and reflection edge–lit holograms have traditional as well as unique replay geometries. The false transmission geometry enables a reflection recorded hologram to be replayed as a transmission hologram from a total internal reflection within the film layer (and similarly with a false reflection hologram). Edge–lit holograms can also be evanescent holograms. These holograms are usually very thin and are therefore not usually very efficient. However, edge–referenced evanescent holograms lead to a new type of hologram where the reference beam coupling into the film is gradual as the index of the film increases. These self–induced index matching holograms can offer unique possibilities for further research into precise index matching and diffusion characteristics of photopolymers.

The complications and non–uniformities involved with recording and replaying edge–lit waveguide holograms prohibit them from many applications where uniformity and efficiency are major requirements.

Two new methods for making slanted fringe holograms offer great potential for simple recording and mass manufacturing. The alternative Bragg condition recording can bypass many of the difficulties associated with steep–referenced recording. Holograms recorded at other Bragg conditions have proven to be very

useful and efficient as well as being much simpler to record. An apparent limitation is that one can only record parallel fringes from collimated wavefronts. An even simpler method involving copying amplitude gratings to volume holograms in photopolymer has been proposed and tested for amplitude transmission holograms. This direct fringe copying method has a very large potential for mass production as a laser is not required for copying.

A new adaptation of the rigorous coupled wave theory was developed to analyze the diffraction characteristics of edge–lit holograms. With the theory, the regions of validity of two approximate models were illustrated for the particular case of slant angles near 45°. While some further modifications are needed on the rigorous theory presented here to perfectly model experimental edge–referenced holograms, the general diffraction characteristics of measured samples were similar to those predicted by the rigorous model. In general, edge–referenced holograms near 45° fringe angles have high wavelength and moderate angular bandwidth when replayed as edge–lit holograms. When they are replayed with the signal, the holograms have very low angular and wavelength bandwidths. Heat processing the samples generally increases the bandwidth and the index modulation.

Replaying an edge–lit hologram is more complicated than recording. The radiometry, directionality, and wavelength of the illuminating source are all important considerations complicated by the compact optical system of the edge–lit hologram. While the radiometry can improve the brightness of the edge–lit hologram, the directionality restrictions severely limit the efficiency and size of the system. Also, the wavelength bandwidth of the source can seriously limit the use of the edge–lit hologram in some systems such as fingerprint illuminators. Coupling gratings can offer some relief to the stringent illumination restrictions imposed on the edge–lit hologram while still maintaining a compact system size.

Many optical systems incorporating a holographic optical element would benefit tremendously if the HOE was a compact edge–lit hologram. The fingerprint and LCD illuminators are excellent examples of optical systems which can use an edge–lit hologram to reduce their size.

The edge–lit fingerprint illuminating device offers many advantages over the traditional fingerprint illuminators. It is inherently smaller, can resolve higher detail, and is free of optical aberrations. There are many considerations regarding the replay of a fingerprint illuminator. The most important considerations are the bandwidth and reflectance from the finger which will lower the contrast on a detector very close to the finger. Edge–lit display holograms or HOE's, such as an LCD illuminator, have many important limitations and restrictions on the replay. High wavelength bandwidth and off–Bragg diffraction are two important considerations which can degrade the image or increase the dispersion in an HOE.

The traditional technique of reflection contact copying, along with the new techniques of alternative Bragg condition recording and direct fringe copying are possible avenues for mass production. The reflection contact copying and the direct fringe copying offer lower coherence requirements for copying and would normally be preferred to the alternative Bragg condition recording. When using the direct fringe copying method, the ability to copy features with a very small fringe spacing is undetermined.

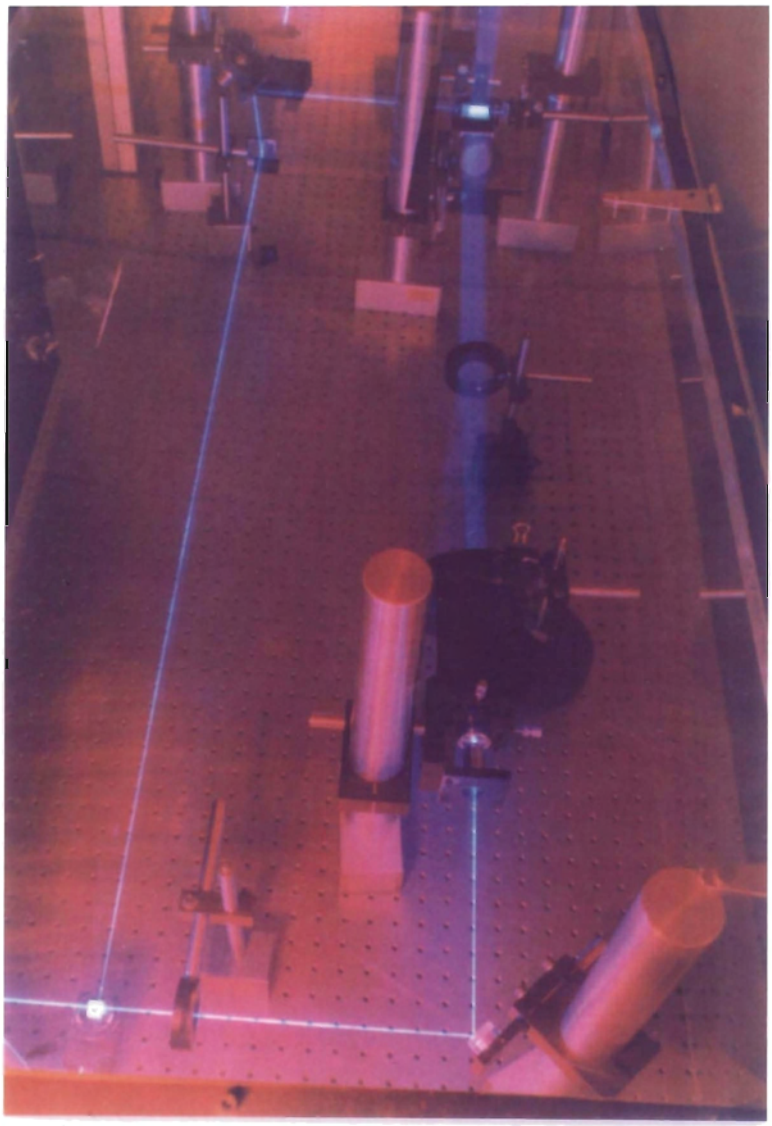
Further areas of research on edge–lit holograms include diffraction modeling, further heating experiments, measurement techniques, recording techniques, and testing various photopolymers. The rigorous coupled wave theory presented here could be modified to include 90° diffraction and incidence. The dynamic effects of the photopolymer could also be introduced to the theory along with non–perfect cosinusoidal modulation. The polarization selectivity can be modeled using the rigorous theory to model three–dimensional conical diffraction. Since in real holograms the absorption inherent in recording adds to tapering, this and other non–uniformities could be incorporated into the theory. Evanescent holograms could be incorporated into the interference contrast calculations for recording and the rigorous theory could be altered to examine evanescent replay.

The effects of heating on the index profile of photopolymers is not fully understood. A spectrophotometer setup such as the one illustrated here could be used to analyze

the diffraction while the sample is heated under various conditions and cover layers (such as tuning film or Microglass). The spectrophotometer could also be transferred to an angular stage to look at the diffraction at steep angles of incidence. Other recording techniques such as the direct fringe copying need to be evaluated at high spatial frequencies and for various thicknesses.

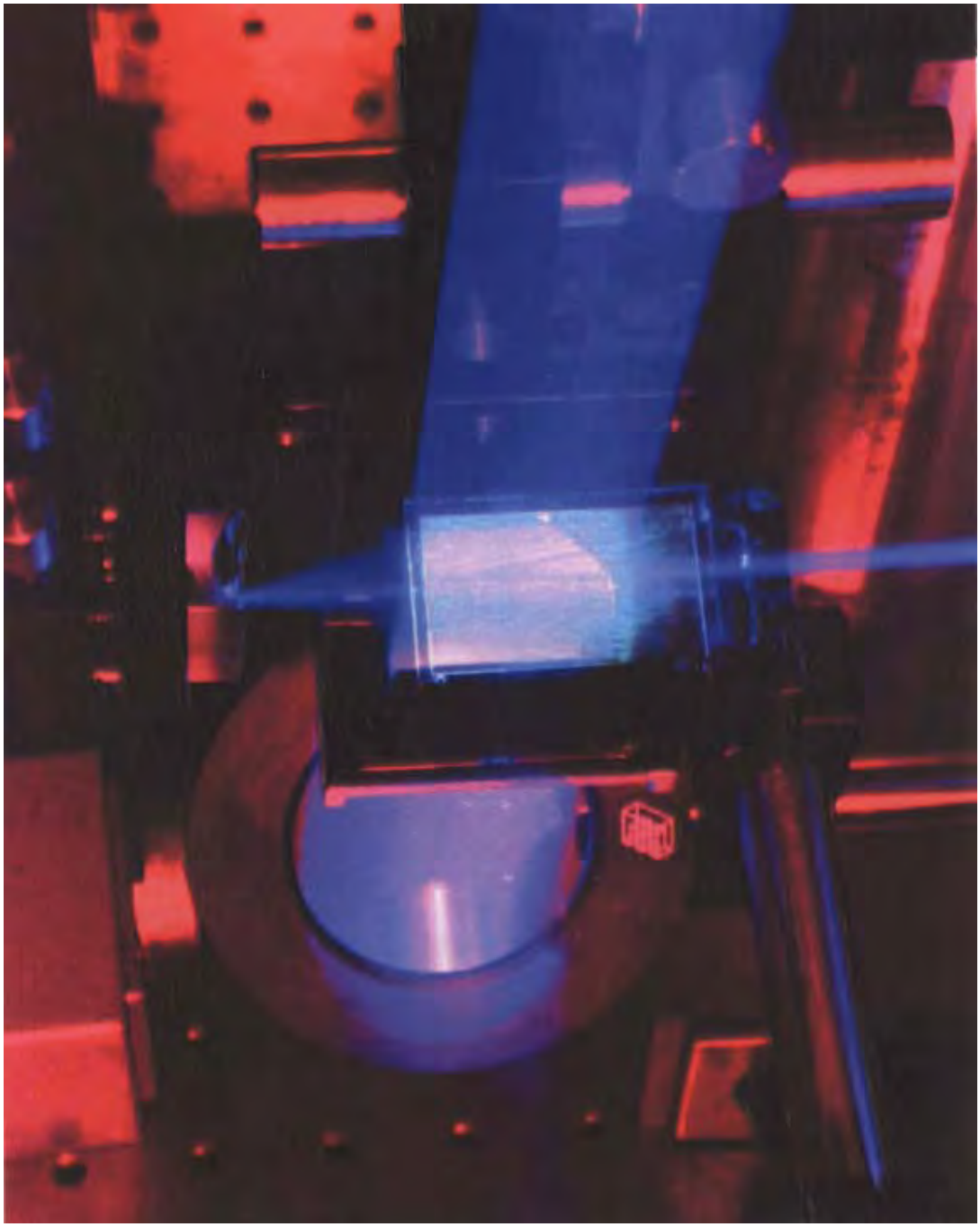
New types of photopolymer are currently being developed in Du Pont and other companies which may result in higher refractive index modulation, index profiles closer to a cosinusoidal response, or other benefits. Continuously testing and experimenting with these films may soon enable edge–lit holograms to be used in many optical systems, helping to fulfill the great promise of holography.

Appendix A



A photograph showing the setup for recording edge-referenced holograms as in Fig. 3.1. The blue area is 458nm light from an argon ion laser. The red light is from a safelight and is used to see the components of the setup.

Appendix B



A photograph showing the real-time interferometric fringes generated from the fluorescence of the dye when a recorded edge-referenced hologram is moved on the setup .

Appendix C

The following program determines the mathematical solution of the rigorous coupled wave theory of Moharam (JOSA 1981, p814) for any number of desired space harmonics using the Coleman adaptation for thick gratings and a large number of space harmonics. The program was written for the MATLAB™ software version 4.0.

Notational remarks are made using a "%" for the introduction and variable descriptions and references to the equations in the text of the thesis.

```
% This MATLAB program uses the rigorous coupled wave theory of Moharam
% (JOSA '81 p814) with the Coleman adaptation for thick gratings with
% many possible space harmonics. This will work for cosinusoidal
% modulation with slant and many diffraction orders. The boundary
% condition matrix is well conditioned. The fringe angle and phi vary between
% -90:0:90 degrees. This allows for input angles from -90 to 90 degrees.

% This particular version will show the higher Diffraction Efficiency
% as a darker area in a 3-Dimensional graph versus the Bragg
% wavelength deviation and versus the reference beam angle where the
% replay angle is at the signal angle.  This program will generate
% the graph in Figure 5.6(d) and will take appx. 40 minutes
% to compute.  Where possible, the notation for section 5.2 will
% be used.  This program is NOT valid for 0 degree slant, 90 degree diffraction
% for any order, or 90 degree incidence.

%%% INITIALIZATION

clear;
count=0;
count1=0;
num_points=90;          % This determines the number of points to graph in a
% square plot (i.e. 90 x 90 points). Decrease this to decrease the calculation time


%%% HARMONIC VARIABLES
```

```
MM=3;          % which means that the total number of retained orders is 2*MM+1
               % Thus in this case, the orders -3,-2,-1,0,1,2,3 are included
nn=2*MM+1;
NN=2*MM+2;

%%% R E C O R D I N G   V A R I A B L E S

% This section is used to compute the slant and period of a grating
% recorded and to be analyzed.

theta_s=0;             % This is the angle of incidence for the signal IN THE FILM
theta_r_min=.001;      % This is the minimum reference angle
theta_r_max=180;       % This is the maximum reference angle
for theta_r=theta_r_min:(theta_r_max-theta_r_min)/num_points:theta_r_max;
% This is the reference angle loop (for recording in this case)

fringe_angle=(theta_r+theta_s)/2;     % This is the fringe angle
rlambda=.647;                         % This is the replay wavelength in microns
rt=1;    % rt is a variable in case one has defined a negative fringe angle which
         % allows the correct computation of the slant phi.
if fringe_angle<0
  rt=-1;
end
phi=fringe_angle-90*rt;

nholo=1.5;    %This is the average refractive index of the hologram
ninc=nholo;npost=nholo;e_1=ninc^2;e_3=npost^2; e_avg=nholo^2;
% These define the other permittivities and refractive indices.  In this example, they
% are equal across the boundaries.

period=rlambda/(2*nholo*abs(sin(pi/180*0.5*(theta_s-theta_r))));
% This is the absolute period (or fringe spacing) of the hologram grating.

%%% R E P L A Y   V A R I A B L E S

%%%% REPLAY ANGLE LOOP
theta=theta_s;         % This is the replay angle in first medium (degrees). In this
                       % case it is the signal angle.

if theta_r==90         % This is because the program can not handle 90 degree
  theta=89.9999;       % incidence.
end

if theta>90            % This is necessary to change the direction for the same grating
  theta=theta-180;     % when replaying a reflection hologram. (i.e. so that the zero
end                    % order transmitted light always travels in the same direction).
```

```
if theta<-90              % This is in case one wants to consider a large negative replay
  theta=theta+180;     % angle as above for a reflection hologram
end
% The previous section essentially ensures that -90<theta replay<90

%tmin=85;
%tmax=89.99;
%for theta=tmin:(tmax-tmin)/num_points:tmax  % This is the replay angle loop
thetaprime=180/pi*asin(nholo/ninc*sin(pi/180*theta));
% This is the replay angle in the hologram

%%%% REPLAY WAVELENGTH LOOP

lmax=.647+.03;          %This shows the Bragg deviation in microns
lmin=.647-.03;
for lambda=lmin:(lmax-lmin)/num_points:lmax
% This is the replay wavelength loop
%lambda=.647;           % This is needed if one is not varying the replay wavelength


%%%% THICKNESS LOOP

%dmin=0;
%dmax=50;
depth=25;                % This is the thickness of the film in microns
%for depth=0:(dmax-dmin)/num_points:dmax         % This is the thickness loop

%%%% MODULATION LOOP

%nmodmin=.000000001;
%nmodmax=.02;
%for nmod=nmodmin:(nmodmax-nmodmin)/num_points:nmodmax
nmod=.02;      % This is the amplitude of the ref. index modulation
%emod=.2526;
emod=e_avg*nmod*2/nholo; % This is the amplitude of the permittivity modulation
%delta_n=2*nmod;    % This is the relationship between delta n and the refractive
                    % index modulation amplitude
%nmod=emod/(2*nholo);      % This relates the permittivity modulation amplitude
                           % with the ref. index modulation amplitude



%%%% CALCULATIONS

k=pi*emod/(2*lambda*nholo);  %This is the coupling constant, defined in Eq. 5.14
mu=lambda/(period*nholo);       % Defined as in Eq. 5.14
p=2*lambda^2/(period^2*emod); % Defined as in Eq. 5.14
```

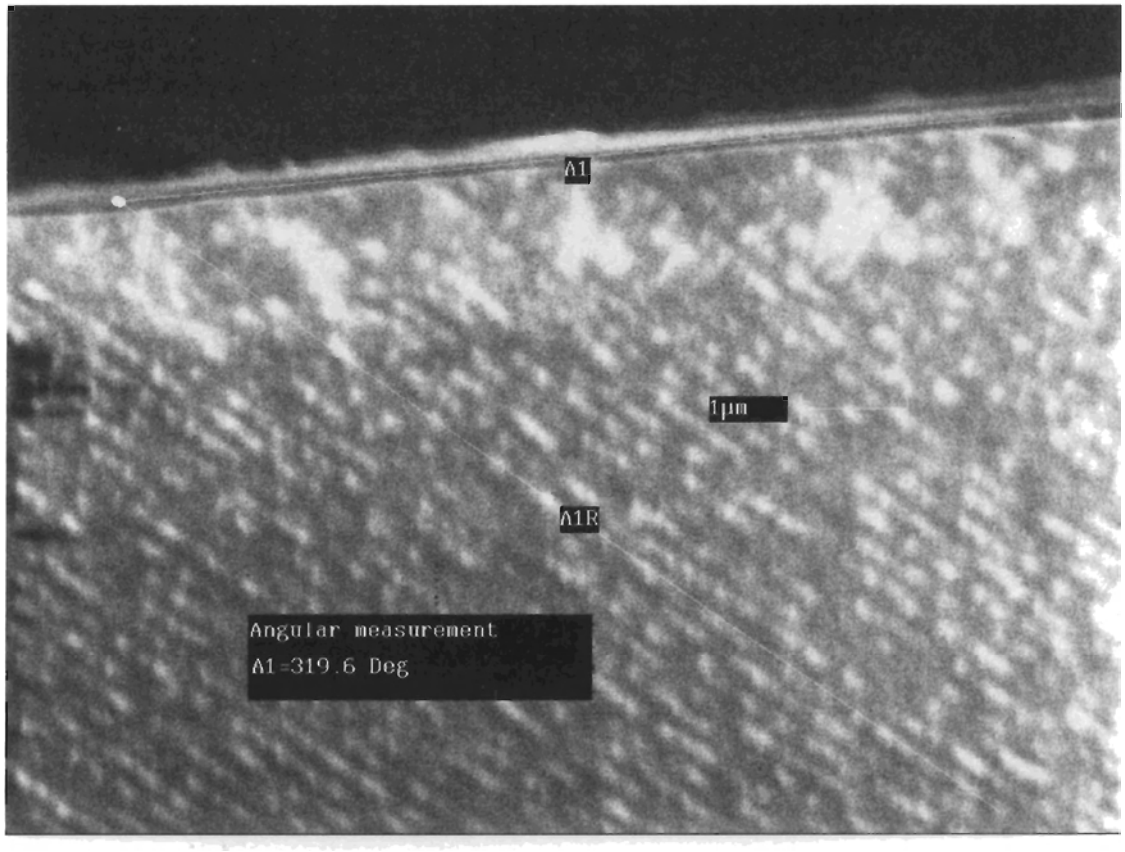
```
B=2*period*nholo*cos(pi/180*(phi-thetaprime))/lambda;
% This is the Bragg indicator (B=1 is at first Bragg condition) defined in Eq. 5.15
k0=2*pi/lambda;

%%% CALCULATE e1,e2,e3 Matrix NOTE positive order on top (1 0 -1)

e1=zeros(nn);e2=zeros(nn);e3=zeros(nn);kx=zeros(nn);
for m=1:nn
kx(m,m)=sin(pi/180*thetaprime)-(MM+1-m)*mu*sin(pi/180*phi);
% This is Beta of Eq. 5.8 divided by k1 of 5.9

e1(m,m)=k0*nholo*sqrt(e_1/e_avg-(kx(m,m)^2))*sign(e_1/e_avg-(kx(m,m))^2);
e3(m,m)=k0*nholo*sqrt(e_3/e_avg-(kx(m,m)^2))*sign(e_3/e_avg-(kx(m,m))^2);
% These are the matrices for e1 and e3 as in Eq. 5.10
e2(m,m)=k0*nholo*(cos(thetaprime*pi/180)-(MM+1-m)*mu*cos(pi/180*phi));
end
%This is the matrix for e2 as in Eq. 5.11

%diff_angles=180/pi*atan(diag(kx)*k0*nholo./diag(e3));
% This can calculate the angles of diffraction

%%%% Calculate the variables for the matrix brs of Eq. 5.16
% These use Eqs. 5.21 NOTE positive orders on top
b=eye(nn);cc=eye(nn);g=zeros(nn);
a=8*e_avg/emod;
g(NN-1)=a;
g(NN+1)=a;
for m=1:nn
   for n=1:nn
     b(m,n)=g(m-n+NN);
   end
end
for m=1:nn
   cc(m,m)=a*(cos(pi/180*thetaprime)-(MM+1-m)*mu*cos(pi/180*phi));
   b(m,m)=-a*p*(MM+1-m)*(MM+1-m-B);
end

A=[zeros(nn) eye(nn) % This is the matrix brs represented in Eq. 5.22
  b        cc];


[W,eigenvalue]=eig(A);  % W is a full matrix of eigenvectors
                        % eigenvalue is a diagonal matrix of corr. eigenvalues

%This section re-orders the eigenvectors and eigenvalues so that the matrix
% X has the higher values in the bottom half
test=diag(i*eigenvalue);
old=W;
```

```
temp=imag(test);
[yy,pp]=sort(real(test));
for n=1:2*nn
  ok(n,1)=yy(n)+i*(temp(pp(n)));
  testW(:,n)=old(:,pp(n));
end
eigenvalue=diag(ok)/i;
W=testW;

%%%%%%%%% BOUNDARY CONDITIONS

%%% Calculate the total amplitude matrix (tam) at the 2 boundary conditions
tam=zeros(2*nn,1); % Initialize
tam(3*MM+2)=2*e1(MM+1,MM+1); % This corresponds to D and Z in Eq. 5.35
% by being 3*MM+2, the transmission orders are on top of the bcm matrix

elements=diag(eigenvalue); % This is a matrix of all the eigenvalues on the diagonal
EV=k*eigenvalue;
pEV=k*diag(elements(1:nn)); % This is matrix V+ of Eq. 5.35
qEV=k*diag(elements(nn+1:2*nn)); %This is matrix V- of Eq. 5.35
X=expm(i*EV*depth);% This is full matrix of X
xp=expm(i*pEV*depth);% This is the top half of X and is X+ of Eq. 5.35
xq=expm(-i*qEV*depth); % This is the bottom half of X and is X- of Eq. 5.35

Y=expm(-i*e2*depth);% This is Y of Eq. 5.31
E1=(e1+e2);% This is G1 of Eq. 5.35
E3=(e2-e3);% This is G3 of Eq. 5.35


% bcm will use the eigenvectors in the ROWS!
ww=W(1:nn,1:2*nn); %This is the full matrix of W
wwp=W(1:nn,1:nn);% This is W+ of Eq. 5.35
wwq=W(1:nn,nn+1:2*nn);% This is W- of Eq. 5.35

bcm=[E3*wwp*xp-wwp*pEV*xp    E3*wwq-wwq*qEV
    E1*wwp-wwp*pEV    E1*wwq*xq-wwq*qEV*xq];
% This is the large matrix of Eq. 5.35

c=bcm\tam;     %This solves Eq. 5.35
P=c(1:nn);      % This is the solution P of Eq. 5.35
Q=c(nn+1:2*nn);% This is the solution Q of Eq. 5.35

%%%%%%%%%CALCULATE THE COEFFICIENTS OF REFLECTION

R_amp=wwp*P+wwq*xq*Q;
%This is the solution for the amplitude of the reflection coefficients as in Eq. 5.28
R_amp(MM+1)=R_amp(MM+1)-1;% This is the amplitude of the zero order
% reflection coefficient which needs the kronecker delta in Eq. 4.28
```

```
R_int=real(e1/(e1(MM+1,MM+1)))*(abs(R_amp).^2);
% This is the intensity of the reflected diffracted orders as in Eq. 5.36

%%%%%%%%%% CALCULATE THE COEFFICIENTS OF TRANSMISSION

T_amp=Y*wwp*xp*P+Y*wwq*Q;
%This is the solution for the amplitude of the transmission coefficients as in Eq. 5.30
T_int=real(e3/(e1(MM+1,MM+1)))*(abs(T_amp).^2);
% This is the intensity of the transmitted diffracted orders as in Eq. 5.37


% The following tests that the total diffraction efficiency is not greater than 1.
DEtest=sum(T_int)+sum(R_int)-1;
        if abs(DEtest) > 0.0000001
        disp('error of the energy conservation')
        disp(DEtest)
        end

% The following stores the values for the various diffraction orders
count=count+1;
DE2(count)=T_int(MM-1); % Diff. Eff. of the +2 Transmitted diffracted order
DE1(count)=T_int(MM);  % Diff. Eff. of the +1 Transmitted diffracted order
DE0(count)=T_int(MM+1); % Diff. Eff. of the 0 Transmitted diffracted order
DE_1(count)=T_int(MM+2); % Diff. Eff. of -1Transmitted diffracted order
DE_2(count)=T_int(MM+3); % Diff. Eff. of -2 Transmitted diffracted order
RDE3(count)=R_int(MM-2); % Diff. Eff. of +3 Reflected diffracted order
RDE2(count)=R_int(MM-1); % Diff. Eff. of +2 Reflected diffracted order
RDE1(count)=R_int(MM); % Diff. Eff. of +1 Reflected diffracted order
RDE0(count)=R_int(MM+1); % Diff. Eff. of 0 Reflected diffracted order
RDE_1(count)=R_int(MM+2); % Diff. Eff. of -1 Reflected diffracted order
RDE_2(count)=R_int(MM+3); % Diff. Eff. of -2 Reflected diffracted order
lam(count)=(lambda-.647)*1000;      %This stores the wavelength from the loop
%aDE1(count)=-diff_angles(MM); % This can store the diffracted angles
%aDE_1(count)=diff_angles(MM+2);
end
count1=count1+1;
THETA(count1)=theta_r; % This stores the replay angle
Diff(1:num_points+1,count1)=DE1';
% This stores the diff. eff. of the +1 transmitted order
if theta_r>90
  Diff(1:num_points+1,count1)=RDE1';
%This stores the diff. eff. of the reflected order if it is a reflection grating
end
count=0;
end

% The following draws the graph
figure
```

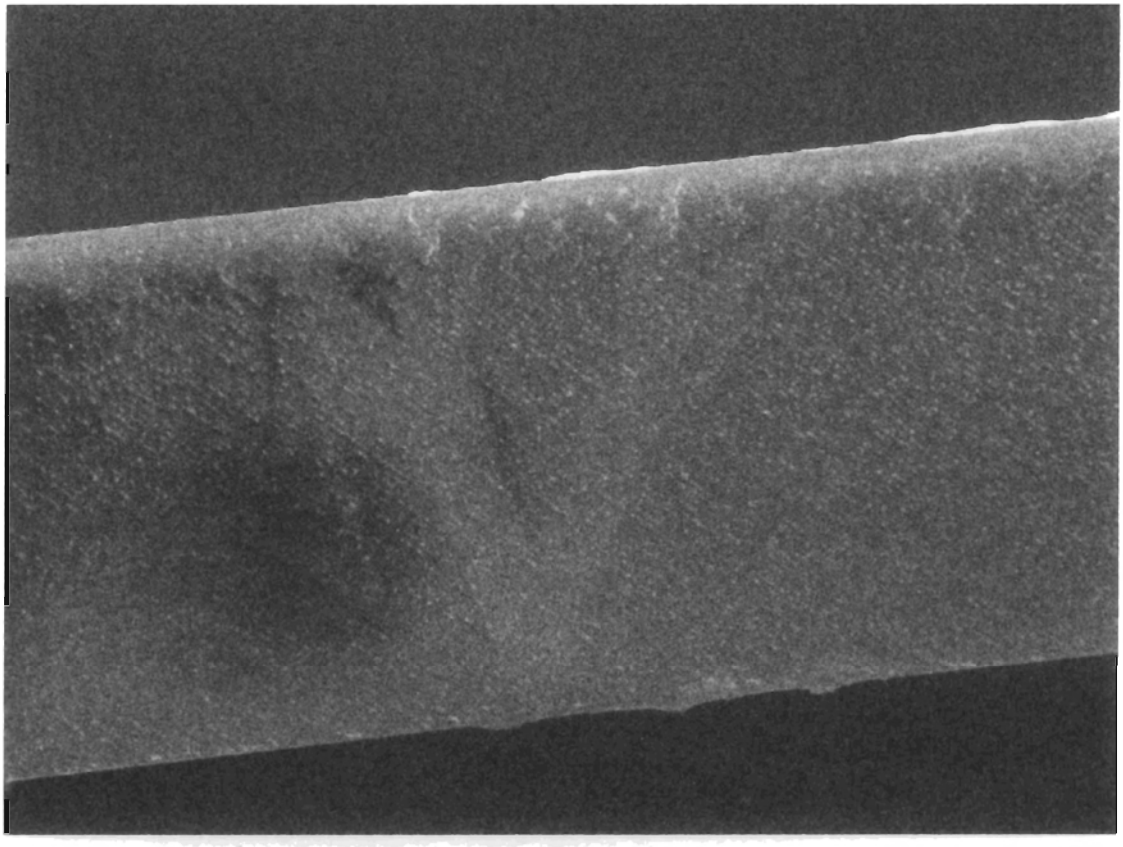
```
[replay,bandwidth]=meshgrid(THETA,lam);
pcolor(replay,bandwidth,Diff);
xlabel('Reference Angle');
ylabel('Bragg Wavelength Deviation (nm)');
mymap=1-gray;
colormap(mymap);
shading interp;
title('1st Order Diffraction Efficiency, Replay at Signal (0 Degrees)');
```

Appendix D



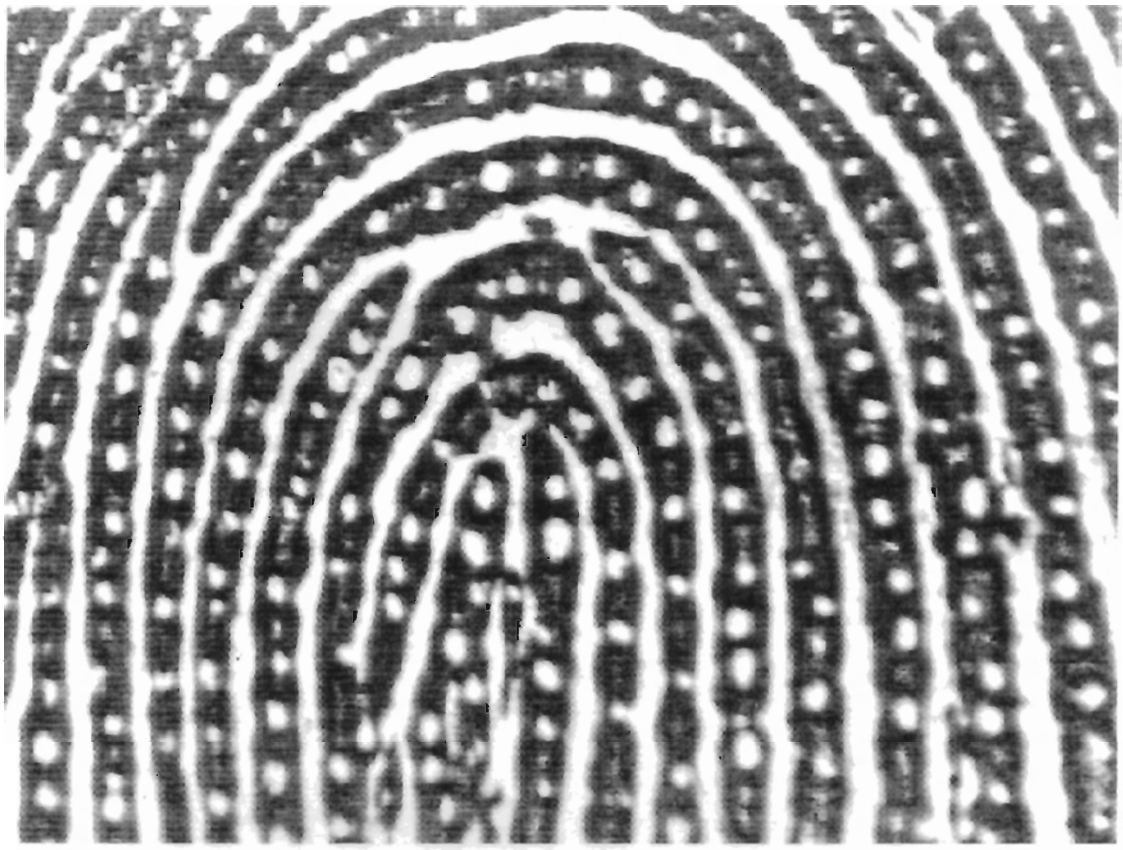
An SEM photograph of a steep reference reflection hologram showing the fringe angle measured as approximately 40° (360°-319.6°). The non-uniformities on the top surface of the picture could be from multiple sets of fringes recorded from a reference beam reflection.

Appendix E



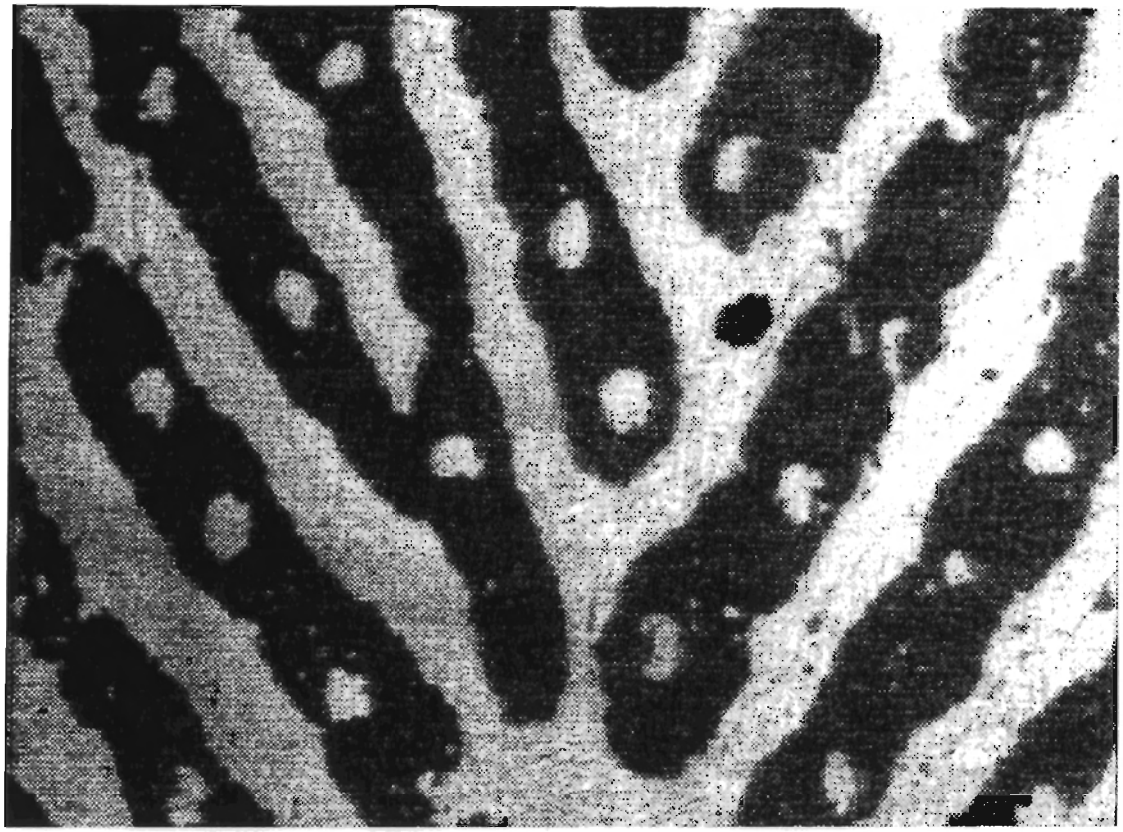
An SEM photograph of a steep reference reflection hologram with the full 20μm film thickness. The fringes of the hologram are slightly visible.

Appendix F



An example of a fingerprint taken from an imaging ELFID. The detail of the pores and the absence of aberrations are clearly visible. The resolution is near 2500 dots per inch. (Photo courtesy of U.S. Dept. of Defense).

Appendix G



An high resolution view of a small portion of a fingerprint taken from an imaging ELFID. The resolution is near 8000 dots per inch. (Photo courtesy of U.S. Dept. of Defense).