1-1-2003

# DESIGN OF DIFFRACTIVE OPTICAL ELEMENTS: OPTIMALITY, SCALE AND NEAR-FIELD DIFFRACTION CONSIDERATIONS

Parker Kuhl

Okan K. Ersoy

# DESIGN OF DIFFRACTIVE OPTICAL ELEMENTS: OPTIMALITY, SCALE AND NEAR-FIELD DIFFRACTION CONSIDERATIONS

Parker Kuhl
Okan K. Ersoy

TR-ECE 03-09

School of Electrical and Computer Engineering
465 Northwestern Avenue
Purdue University
West Lafayette, IN 47907-2035

# TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

## ABSTRACT

In this research, two related research topics were investigated. The first one is the use of scalar diffraction theory for the purpose of simulating and implementing diffractive optical elements (DOEs). The main focus was on the optimal design of computer-generated holograms (CGHs). Two existing methods were combined to improve the reconstructed image of a CGH. The new method combines the Optimal Decimation-in-Frequency Iterative Interlacing Technique (ODIFIIT) with the Lohmann coding scheme. Simulations indicate that the reconstructed image produced with this method has less error than reconstructions with either method alone. Physical reconstructions were performed and compared to the simulated results.

Near field diffraction from DOEs with feature sizes on the order of a wavelength were also simulated and analyzed. Both the small feature size and the close observation distance make scalar diffraction theory inaccurate. For this situation, a complete electromagnetic theory is necessary to achieve accurate results. XFDTD software made by Remcom was used to simulate these elements and their diffracted fields. XFDTD calculates the diffracted fields by using a leapfrog finite difference time domain algorithm to solve Maxwell's curl equations directly. Fresnel zone plates and reflection and amplitude gratings were simulated. An amplitude grating was also combined with a Fresnel zone plate to imitate a Fourier lens system like that used for reconstructing the image from a computer-generated hologram. The angular spectrum of plane waves was also used to study these systems, and results were compared.

Results showed that XFDTD can accurately simulate small-scale DOEs. Precise knowledge of the diffracted field from small-scale DOEs will play an important and useful role in future optical systems as feature sizes reach the nano-scale level.

# 1 INTRODUCTION

Diffractive optics is used in many applications such as optical storage, processing, sensing, and communications. Diffraction is often described as the bending, or deviation of waves from a straight line due to interaction with an obstacle or aperture. To completely describe diffractive effects, electromagnetic theory is necessary. In many situations, an exact description of the diffracted field from an aperture cannot be obtained analytically, and requires numerical techniques. Under certain conditions, analytical results can be achieved using scalar diffraction theory. For it to be sufficiently accurate, two conditions must be met [1]: (1) the diffracting aperture must be large compared to the wavelength of the incident wave and (2) the diffracted field must not be observed too close to the diffracting aperture. The second condition often requires that the field be observed many meters behind the aperture called the far field, creating space issues in a confined lab. Fortunately, a lens can be used to overcome this problem.

Scalar diffraction theory is used in the first part of the thesis for the design of diffractive optical elements (DOEs). The general purpose of a DOE is waveshaping. While standard refractive optical elements such as mirrors and lenses are often bulky, expensive and limited to a specific use, DOEs are generally light-weight, compact, easily replicated, and can modulate complicated wavefronts. DOEs are also useful in manipulating multi-spectral signals.

A computer-generated hologram (CGH) is a specific type of DOE. CGHs have found applications in many areas involving waveshaping, laser machining [2], 3-D displays, and optical pattern recognition. CGHs have recently found uses in optical interconnects and security devices [3,4]. A particular setup for designing and testing a CGH is an optical Fourier transform system. In this system, the desired wavefront plane and its transform plane, called the input plane and observation plane, respectively, are placed in the front and back focal plane of a thin lens. The CGH is placed in the input plane and its far-field diffraction pattern, which will be seen to be the Fourier transform of the hologram's transmission function, occurs in the observation plane.

Advances in computing and integrated circuit production technology are making the application of DOEs more attractive. Designing a CGH consisting of a large number of points can be computationally demanding. However, many points are necessary to achieve the bandwidth required to reconstruct a complicated wavefront. As computers become faster, the ability to rapidly design CGHs and to adjust them in real time will increase, allowing them to be used in ways not previously possible. One example is in the use of spatial light modulators (SLM). In general, a CGH modulates the amplitude and/or phase of a wave to create a desired wavefront, just like a SLM. Rapid design of CGH can allow SLMs to implement the hologram in real time.

A limiting factor in the production of DOEs is the ability of the device to spatially modulate the amplitude and phase of a wavefront. Current technologies impose resolution limits, and require discretization and quantization of the amplitude and/or phase function of the DOE. The option is often limited to two levels of quantization. Two typical types of binary quantizations are to restrict the output function to 1 or 0 representing transmission and no transmission, respectively, and 1 or -1, where -1 represents a 180 degree phase shift.

Integrated circuit technology, specifically lithography and reactive ion etching (RIE), can be used to produce diffractive optical elements. Lithography and RIE are used to print patterns in materials like silicon using surface relief. Surface relief is used to implement phase modulation. IC technologies allow for better resolution and more levels of quantization, which increases efficiency. As these technologies mature, so will the output and efficiency of diffractive optical elements. This will result in DOEs being used in many more applications. Thus, it is important to study and optimize the design of diffractive optical elements.

This thesis discusses three methods of designing computer-generated holograms. The first method was introduced by Lohmann and Brown [5]. The hologram is divided into an array of rectangular cells each containing an aperture. The size of the aperture is adjusted to modulate amplitude, and the position of the aperture is shifted to control phase. This method is referred to throughout this thesis as the Lohmann method. The desired amplitude and phase at each point is obtained by computing the Fourier transform of the desired image. The second method is the Optimal Decimation-in-Frequency

Iterative Interlacing Technique (ODIFIIT) developed by Ersoy and Zhuang [6]. It is an iterative method which utilizes the decimation-in-frequency property of the Fourier transform to interlace several smaller holograms into one large hologram. Similar to Lohmann's method, the hologram plane is divided into an array of rectangles. However, the output function of each rectangle, called a cell, has the same value for all points inside that cell. The output of each cell is determined using the projection onto constraint sets (POCS) algorithm [7]. The third method is a new scheme developed in this thesis incorporating aspects of both Lohmann's method and the ODIFIIT. The new design procedure seeks to take advantage of the interlacing property of the ODIFIIT while using Lohmann's coding scheme to have more control over amplitude and phase when designing each cell. ODIFIIT is also applied to a quantized version of Lohmann's method to reduce the error in the reconstructed image.

All of the diffractive optical elements discussed so far are designed with the assumption that scalar diffraction theory applies. However, due to the advances in semiconductor technologies, there is an increasing demand for devices with feature sizes on the nanometer scale. On such a small scale, scalar diffraction theory breaks down. Refractive optical elements are not useful in these situations because their resolution is limited for small aperture sizes, which is due to diffractive effects. Therefore, a good understanding of diffractive effects on a small scale and the ability to use diffraction as an advantage is becoming more important. As devices reach the nano-scale level, efficient DOE design will be crucial.

Scalar diffraction theory becomes incomplete when the field is observed very close to an aperture, or when the size of the aperture is on the order of a wavelength. When this is the case, electromagnetic theory is needed to accurately predict diffractive effects. There are several numerical techniques used to simulate wave propagation using electromagnetic theory. Finite elements, beam propagation, and finite difference are examples of such methods [8,9]. In the second part of this thesis, the finite difference time domain (FDTD) method is used to simulate and analyze diffractive optical elements with small feature sizes and short characteristic lengths, like focal length. The FDTD method can be used to numerically solve partial differential equations, such as the wave equation and Maxwell's curl equations. FDTD methods use a grid representing the

region of interest, which is made up of many small cells. The quantities of interest are calculated at each cell. In the case of diffraction of electromagnetic waves, the quantities of importance are the electric and magnetic fields. For accurate results, each cell has a size on the order of 1/10 to 1/30 of a wavelength. In general, accuracy of results improves with decreasing cell size.

There are other factors besides the minimum resolution of the defined grid that can cause error in the results. The most notable is the application of absorbing boundary conditions (ABCs), which are not perfect. ABCs try to simulate a situation where waves encountering the boundary of the interest region are free to propagate as if there is no boundary. Effective FDTD methods reduce reflections at the boundary to a fraction of a percent.

In this research, XFDTD software by Remcom was used to simulate Fresnel zone plates, reflective gratings, and binary transmission gratings imitating computer-generated holograms. Feature sizes and observation lengths were made on the order of several wavelengths. The accurate simulation of these elements is important when considering if such diffractive devices could be used in micro and nano-scale applications.

The angular spectrum of plane waves [1] was also used to analyze the same elements as XFDTD. Theoretically, the angular spectrum method is valid for regions close to a diffracting aperture, but it is a Fourier transform method which does not directly solve Maxwell's equations. It also does not allow for specification of material properties and advanced geometries.

The results from XFDTD and the angular spectrum method were analyzed and compared. The angular spectrum gave good results in certain situations, but does not have the accuracy or versatility of XFDTD.

# 2 DIFFRACTION

## 2.1 Introduction

Diffraction is a phenomenon of considerable importance in the fields of physics and engineering whenever wave propagation is involved. Sommerfeld defined diffraction as "any deviation of light rays from rectilinear paths which cannot be interpreted as reflection or refraction" [1]. In 1665, the first account of diffractive phenomena was published by Grimaldi when he observed the shadow resulting from an aperture in an opaque screen illuminated by a light source. He observed that the transition from light to shadow was gradual rather than sharp, which we now know is due to diffraction. Sommerfeld's definition implies that diffraction only applies to light rays. In reality, diffraction occurs with all types of waves including electromagnetic, acoustic, and water waves, and is present at all frequencies. The content of this research deals exclusively with electromagnetic radiation at optical frequencies.

Diffraction was initially considered to be a nuisance when designing optical systems because diffraction at the apertures of an optical imaging system is often the limiting factor in the system's resolution. However, by the mid 1900's, methods and devices utilizing the effects of diffraction began to emerge. Examples include analog holography, synthetic aperture radar and computer-generated holograms and kinoforms, generally known as diffractive optical elements (DOE's). Computer-generated holograms will be one of the main topics of this thesis. Diffractive elements are also being combined with refractive elements to correct aberrations associated with refractive lenses.

The propagation of waves can often be described by rays which travel in straight lines. However, the behavior of wave fields encountering obstacles cannot be described by rays. Some of the wave encountering an obstacle will deviate from its original direction of propagation causing the resulting wave field to differ from the initial field at the obstacle. This is called diffraction. Classic examples include diffraction of light from a knife's edge and a wave field passing though an aperture in an opaque screen.

## 2.2 Rayleigh-Sommerfeld Diffraction

In scalar diffraction theory, predicting the diffracted wave field from an aperture is usually done by using Green's theorem to convert the Helmholtz equation into a meaningful integral equation. A discussion of this procedure and the various assumptions made along the way can be found in [1]. Scalar diffraction theory considers only the scalar amplitude of one transverse component of the wave field, and assumes that other components can be treated independently in a similar way. Scalar theory can lead to very accurate results if the diffracting aperture is large compared to the wavelength, and if the diffracted fields are observed at a reasonable distance from the aperture. What satisfies as a reasonable distance will be clarified later.

Fig. 2.2.1 illustrates diffraction from an aperture A illuminated by a wave field U, where the field at the output of the aperture is $U(x, y,0)$. Fig. 2.2.2 shows a cross section of the aperture and specifies the important geometries. On this plane, the field is assumed to be zero anywhere outside the aperture. Applying the Kirchhoff approximation [10], the diffracted field of a monochromatic wave at a point behind the aperture in vacuum is given by the expression [1]

$$U(x_0, y_0, z) = \frac{1}{j\,l}\iint_A U(x, y,0)\frac{z}{r_{01}}\frac{e^{jkr_{01}}}{r_{01}}dxdy\,. \qquad (2.2.1)$$

This is the Rayleigh-Sommerfeld diffraction formula, where $r_{01}$ is the distance from $P_0$ to $P_1$ given by $r_{01} = \left[(x_0 - x)^2 + (y_0 - y)^2 + z^2\right]^{1/2}$.

$l$ is wavelength, k is the wave-number defined as $k = \frac{2p}{l} = \frac{2pf}{c}$, and c is the velocity of light in vacuum.

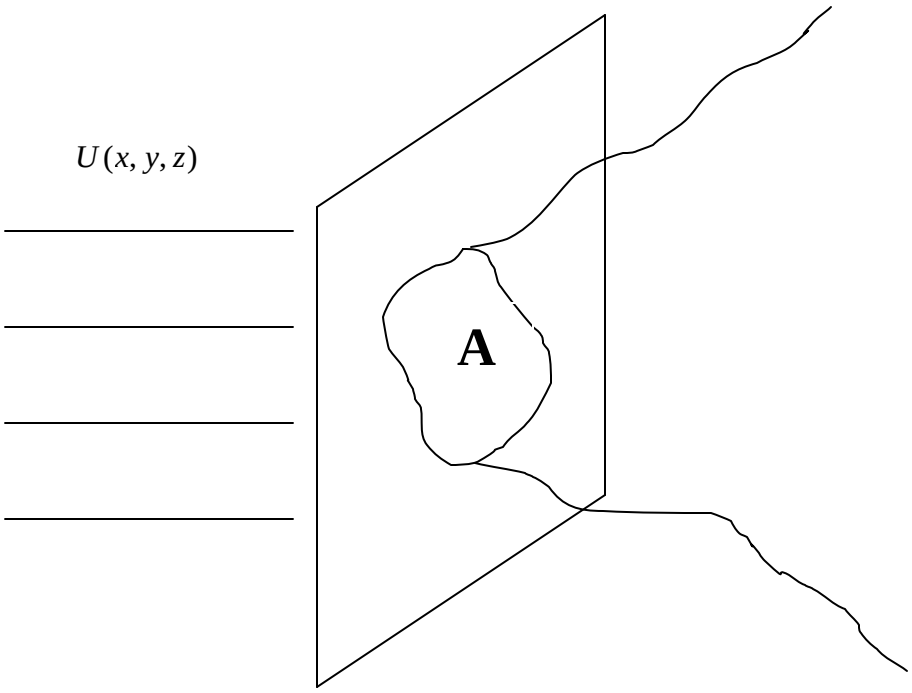



Fig. 2.2.1 Diffraction from an aperture illuminated by a wavefield

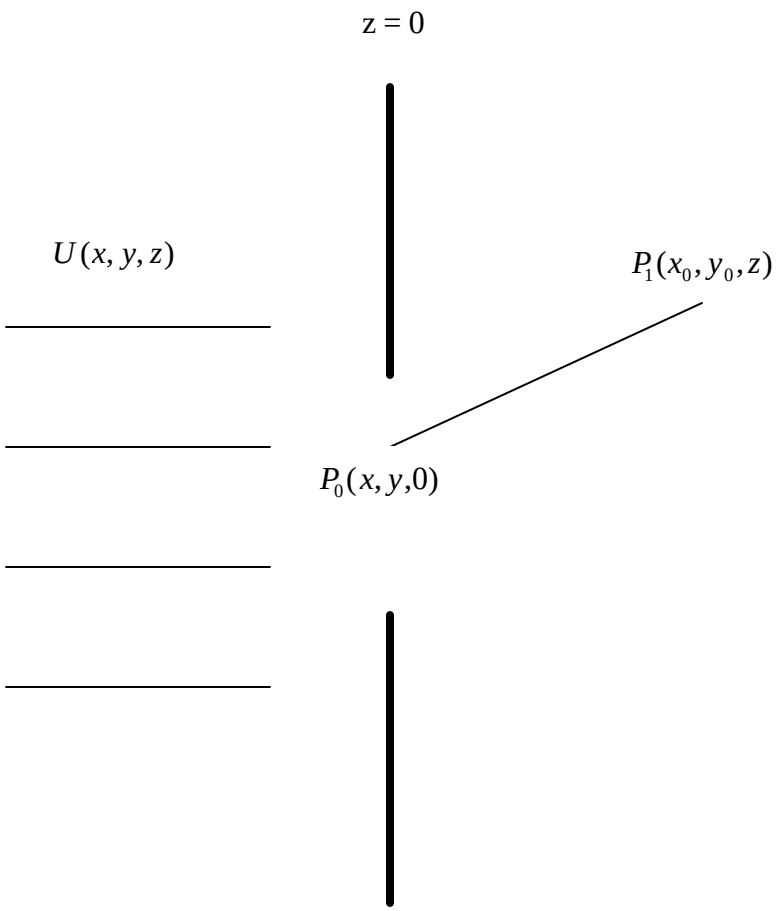



Fig. 2.2.2 Geometry of an aperture illuminated by a wavefield

## 2.3 Fresnel and Fraunhofer Diffraction

The Rayleigh-Sommerfeld formula can be made easier to compute by making approximations based on the size of the aperture, the distance z of the output plane from the aperture, and the observed region of the output plane. Let the aperture be contained inside a radius $L_1$ such that $U(x, y,0) = 0$ if $\sqrt{x^2 + y^2} > L_1$, and the observed wave field $U(x_0, y_0, z)$ at the output plane be restricted to a region $L_2$ so that $U(x_0, y_0, z) = 0$ if $\sqrt{x_0^2 + y_0^2} > L_2$. If

$$|z|^3 >> \frac{k(L_1 + L_2)^2}{8}, \tag{2.3.1}$$

then (2.2.1) can be approximated with good accuracy as [1]

$$U(x_0, y_0, z) = \frac{e^{jkz}}{j\lambda z} \iint_A U(x, y,0) e^{j\frac{\pi}{\lambda z}\left[(x_0 - x)^2 + (y_0 - y)^2\right]} dxdy. \tag{2.3.2}$$

This is known as the Fresnel diffraction formula, and any region satisfying (2.3.1) is called the Fresnel region.

For even greater z distances known as the Fraunhofer region, or far field,

$$z >> \frac{k}{2}(x^2 + y^2)_{\max} = \frac{kL_1^2}{2} \tag{2.3.3}$$

(2.2.1) is approximately given by [1]

$$U(x_0, y_0, z) = \frac{e^{jkz}}{j\lambda z} e^{j\frac{k}{2z}(x_0^2 + y_0^2)} \iint_A U(x, y,0) e^{-j\frac{2\pi}{\lambda z}(x_0 x + y_0 y)} dxdy. \tag{2.3.4}$$

This is the Fraunhofer diffraction formula. Notice that it is the Fourier transform of the field at the diffracting aperture $U(x, y,0)$ at the frequencies $f_x = \frac{x_0}{\lambda z}$ and $f_y = \frac{y_0}{\lambda z}$, multiplied by some factors.

It is important to notice that for a typical aperture size, the distance z required to meet the condition in (2.3.3) is relatively large for experiments in a lab or other confined space. Fortunately, a lens can be used to overcome this potential problem. Neglecting edge effects due to its finite size, the phase transformation generated by a lens produces a Fourier transform of the input wave at its focal plane. A lens with an illuminated

diffracting aperture placed at one focal plane will produce the Fraunhofer diffraction pattern of that aperture at the opposite focal plane. If the diffracting plane is moved away from the focal plane of the lens, the field in the opposite focal plane is just multiplied by a phase factor dependent on the distance moved. However, when viewing the intensity of light, which is the square of the field magnitude, the phase factor is irrelevant. Thus, the output is independent of the position of the diffracting aperture.

### 2.4 Conclusions

The Rayleigh-Sommerfeld (R-S) diffraction formula provides a starting point for finding the diffraction pattern from an aperture. For certain regions behind the aperture, approximations can be made to simplify the R-S formula. By considering the diffracted field in the Fresnel and Fraunhofer regions, diffraction patterns of simple apertures such as rectangles and circles can be determined analytically. Most importantly, the Fraunhofer diffraction pattern is simply the Fourier transform of the diffracting aperture. So, as long as the field at the output of a diffractive element can be represented mathematically, a fast Fourier transform algorithm can be used to determine its Fraunhofer diffraction pattern. This will be the basis for the design of the computer-generated holograms relevant to this thesis when scalar diffraction theory is used. The next chapter will present the basic theory behind computer-generated holograms, which fall under the general category of digital holograms.

# 3 DIGITAL HOLOGRAPHY

## 3.1 Introduction

Digital holography [11] is a branch of diffractive optics covering the theory, calculation and fabrication of thin diffractive elements.

Classically, to make a hologram, a wave reflected from an object is combined with a reference wave to produce an interference pattern which is recorded on a photosensitive material. This interference pattern contains amplitude and phase information which allows an image of the original object to be recovered. This process requires very stable conditions, and can take a considerable amount of time.

Computer generated holograms (CGHs), which are the main focus of this thesis, were introduced by Lohmann [5]. He introduced a purely digital implementation called the detour-phase method. The design of his CGH's is based on a desired Fraunhofer diffraction pattern. A more detailed description of Lohmann's method will be presented later. In general, CGHs are diffractive elements (DE's). DE's that modulate only the amplitude of an illuminating wave are called diffractive amplitude elements (DAE's). DE's which only affect the phase of an illuminating wave are known as diffractive phase elements (DPE's). Unlike DAE's which have a binary quantization (transmission or no transmission), DPE's may have several possible quantization levels. In general, the more possible levels of quantization a DPE has, the closer it can recreate the desired phase. This results in a more effective diffractive element in terms of reconstruction error and efficiency.

CGHs are usually designed based on scalar diffraction theory. This means that the typical feature size of the element must be large compared to the wavelength of the illuminating wave, and the element must be sufficiently thin [1]. Otherwise, rigorous diffraction theory is needed to describe diffraction from the CGH.

## 3.2 Methods of Coding

The optical function of a Fourier diffractive element with transmittance $H(\vec{u})$ is designed to generate the Fraunhofer diffraction pattern $f(\vec{x}-\vec{x}_0)$. Mathematically, we write [11],

$$F[H(\vec{u})] = h(\vec{x}) = af(\vec{x} - \vec{x}_0), \qquad \vec{x} \in R \tag{3.2.1}$$

The diffracted wave is proportional to the desired wave $f(\vec{x})$ inside a window $R$ in the observation plane. $\vec{x}_0$ represents an offset of the window $R$, and $F$ designates Fourier transform.

If only the intensity of the wave is considered, then (3.2.1) is changed to,

$$\left|h(\vec{x})\right|^2 = a^2 \left| f(\vec{x} - \vec{x}_0)\right|^2, \qquad \vec{x} \in R \tag{3.2.2}$$

This makes the phase of the desired wave a free parameter.

Three assumptions are made about the desired wave and its spectrum [11]:

1. The magnitude of the spectrum of $f(\vec{x} - \vec{x}_0)$ is normalized:

   $$\left|F^{-1}\left[f(\vec{x} - \vec{x}_0)\right]\right| = \left|F_T(\vec{u})\right| \le 1,$$

   where $F_T(\vec{u}) = F(\vec{u})e^{j2p\vec{u}\vec{x}_0}$

2. The desired wave $f(\vec{x})$ must be bandlimited due to the finite size of the diffractive element.

3. $f(\vec{x} - \vec{x}_0)$ is approximately restricted to the window $R$:

   $$f(\vec{x} - \vec{x}_0) \approx 0, \qquad \vec{x} \notin R.$$

Condition 3 implies the neglect of sinc-oscillations outside of $R$ due to the finite size of the element.

In order to make implementation easier, the transmittance $H(\vec{u})$ of the diffractive element has to satisfy certain digital constraints $C$. For a DAE, each point in $H(\vec{u})$ must allow either transmission or no transmission represented by 1 and 0 respectively. Therefore, $H(\vec{u}) \in \{0,1\}$ for every $\vec{u}$ representing a point in the hologram, and $C = \{0,1\}$. For a binary diffractive phase element $H(\vec{u}) \in \{-1,1\}$ or more generally, $\left|H(\vec{u})\right| = 1$. Coding of the input signal by the digital hologram requires finding a $H(\vec{u})$ that satisfies the transmittance conditions exactly and the conditions in (3.2.1) and (3.2.2) to a good approximation. In general, there are two different coding techniques [11]:

1. Iteratively search for the $H(\vec{u}) \in C$ that best approximates $f(\vec{x})$. Given an initial $H_1(\vec{u})$, successively generate $H_2(\vec{u})$, $H_3(\vec{u}), \ldots H_N(\vec{u})$ to satisfy the constraints in the image plane. This can be written as,

$$H_N(\vec{u}) = O_{IT} H_1(\vec{u}) \in C$$

where $h_N(\vec{x}) = F[H_N(\vec{u})]$ satisfies (3.2.1) or (3.2.2) to a good approximation. $O_{IT}$ represents the iterative algorithm. Or,

2. Starting with $F_L(\vec{u})$, find $H(\vec{u})$. The constraint in the image plane is already satisfied. A coding operator $O_C$ maps $F_L(\vec{u})$ onto $H(\vec{u})$ such that image plane constraints are satisfied, i.e.

$$H(\vec{u}) = O_C[F_L(\vec{u})] \in C$$

where $h(\vec{x})$ is given approximately by (3.2.1) or (3.2.2). The coding operator could be iterative or noniterative. Lohmann's method and ODIFIIT discussed later in this thesis both fall into this category.

### 3.3 Diffraction Efficiency

The diffraction efficiency of a DE with transmittance $H(\vec{u})$ is defined as [11]

$$h = \left\langle \left| h(\vec{x}) \right|^2 \right\rangle_R \tag{3.3.1}$$

where $\langle ... \rangle_R$ is an integration over $R$, the image region. This assumes the energy of the illuminating wave is normalized. Efficiency is a measure of how much of the energy in the initial illuminating wave is diffracted into the desired region $R$. In general, diffractive phase elements have greater efficiency than diffractive amplitude elements. The efficiency of a DPA increases with the number of phase levels.

### 3.4 Conclusions

The basics of digital holography provide a general foundation for the design of computer-generated holograms. All the CGHs designed in this research use the second general coding method mentioned. How the hologram plane is divided and the specific method used to encode amplitude and phase at each point in the hologram have a significant effect on the reconstructed image and the hologram's efficiency. The next three chapters discuss how the basic theory from this chapter is used to design computer-generated holograms using Lohmann's method, Optimal Decimation-in-Frequency Iterative Interlacing Technique, and an original method combining these two methods.

# 4 LOHMANN'S METHOD FOR DESIGNING COMPUTER-GENERATED HOLOGRAMS

## 4.1 Introduction

The goal in producing a synthetic hologram, or computer generated hologram (CGH), is to find a diffractive object with a predetermined Fraunhofer diffraction pattern. This is the inverse of the familiar problem where one wants to find the diffraction pattern resulting from an object. In generating a CGH, it is the diffraction pattern which is known, and the geometry of the object needs to be determined. The optical system assumed for generating synthetic holograms in the case of Fourier transform geometry consists of a hologram plane and an image plane, each a distance f from opposite sides of a lens (Fig. 4.1.1). f is the focal length of the lens. In general, the hologram plane is illuminated by a wave. Due to the Fourier transforming properties of a lens, the Fraunhoffer diffraction pattern, or Fourier transform, of the hologram plane occurs at the image plane. In other words, the wavefront at the hologram plane $F(n_x, n_y)$ has the Fraunhoffer diffraction pattern $f(x, y)$ at the image plane. By knowing the desired diffraction pattern, one can determine the wavefront needed at the hologram plane, namely,

$$F(n_x, n_y) = \int f(x, y) e^{-2pi(xn_x + yn_y)} dxdy. \qquad (4.1.1)$$

The coordinates $(n_x, n_y)$ are related to the physical coordinates $(x_f, y_f)$ by the following: $x_f = lfn_x$ and $y_f = lfn_y$. In actual implementation, one must find a way to implement the complex amplitude $F(n_x, n_y) = A(n_x, n_y) e^{if(n_x, n_y)}$.

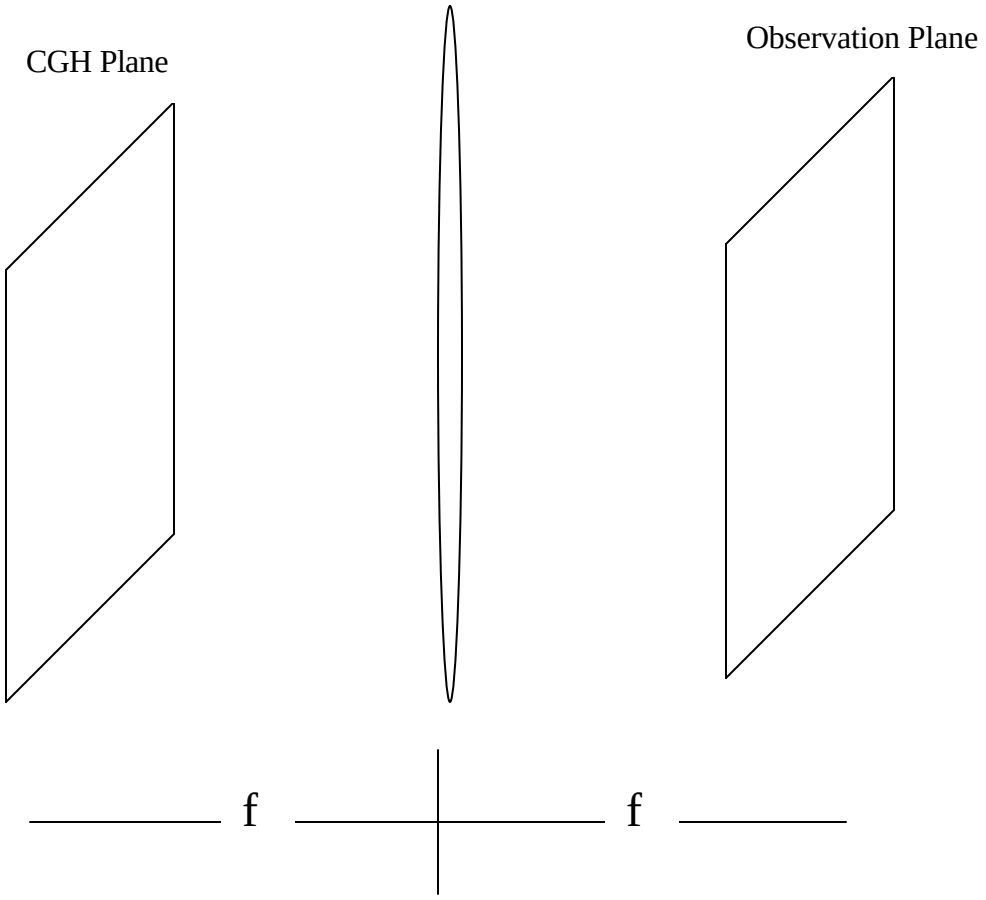



Fig. 4.1.1 A Fourier lens system for the design of CGHs.

**4.2 Lohmann's Method**

In the method developed by Lohmann [5,12], the hologram plane is divided into smaller rectangles each containing an aperture. An example of a Lohmann cell is shown in Fig.4.2.1. The size of the aperture is used to control amplitude, and its position is changed to adjust phase. This results in a binary transmission pattern. If $h(x, y)$ is taken as the diffracted amplitude from the hologram $H(n_x, n_y)$, then it should be proportional to the desired image $f(x, y)$. The binary amplitude of the hologram's transmission is described by

$$H(n_x,n_y)=\sum_n\sum_m rect\left[\frac{n_x-(n+P_{nm})dn}{cdn}\right]rect\left[\frac{n_y-mdn}{W_{nm}dn}\right] \tag{4.2.1}$$

When a tilted plane wave, $\exp(2pix_0n_x)$ falls on the binary hologram, the complex amplitude behind the hologram is $H(n_x,n_y)\exp(2pix_0n_x)$. A tilted plane wave is needed in order to manipulate phase in the hologram plane. Fourier transform of the hologram at the image plane results in the complex amplitude:

$$\int\int H(n_x,n_y)e^{2pi[(x+x_0)n_x+yn_y]}dn_xdn_y = \tag{4.2.2}$$

$$c(dn)^2\,\mathrm{sinc}\left[cdn(x+x_0)\right]\sum_n\sum_m W_{nm}\,\mathrm{sinc}(yW_{nm}dn)\exp\{2pi[dn((x+x_0)(n+P_{nm})+ym)]\}$$

Now, the parameters $W_{nm}$, $P_{nm}$, and the two constants $x_0$,c are chosen such that the complex amplitude in the image plane matches the desired image $f(x,y)$.

Eq.(4.2.2) can be compared to the desired image by writing $f(x,y)$ in the form

$$f(x,y)=\int\int F(n_x,n_y)e^{2pi(xn_x+yn_y)}dn_xdn_y=\sum_n\sum_m F(ndn,mdn)e^{2pi[dn(xn+ym)]} \tag{4.2.3}$$

The two sinc functions and the factor $\exp\left[2pi(xP_{nm}dn)\right]$ in (4.2.2) can all be assumed to be close to unity. The validity of this assumption will be discussed in Section 4.3. Then, by equating the Fourier coefficients, we get [12]

$$c(dn)^2W_{nm}\exp\{2pi[x_0dn(n+P_{nm})]\}\propto F(ndn,mdn);$$

$$F(ndn,mdn)\propto c(dn)^2A_{nm}\exp[2pi(f_{nm}/2p)];$$

$$W_{nm}\approx A_{nm};\qquad P_{nm}+n\approx f_{nm}/2px_0dn\ . \tag{4.2.4}$$

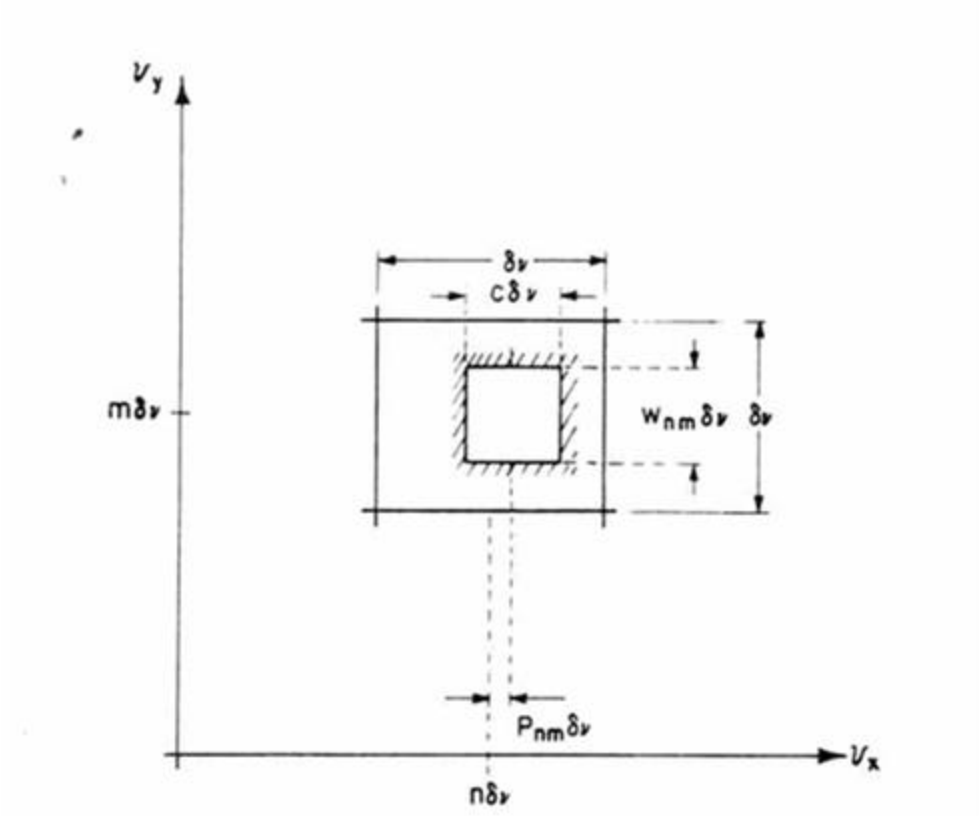



Fig. 4.2.1 The (n, m)th cell of a Lohmann-type binary hologram [12].

This shows that the height W and the position P of the aperture in each cell are responsible for generating the amplitude A and the phase $f$ of the complex amplitude $F$ at the cell. By choosing $x_0 dn$ equal to an integer M, we obtain

$$P_{nm} \approx f_{nm} / 2pM \ . \tag{4.2.5}$$

This completes the information needed to generate a Lohmann-type binary synthetic hologram.

### 4.3 Approximations in Lohmann's Method

The transmission pattern of a binary hologram was described in the previous section by a summation of rectangular functions that are shifted and scaled to represent the position and size of each aperture in the hologram plane:

$$H(n_x,n_y)=\sum_n\sum_m rect\left[\frac{n_x-(n+P_{nm})dn}{cdn}\right]rect\left[\frac{n_y-mdn}{W_{nm}dn}\right] \tag{4.3.1}$$

When the hologram is illuminated by a tilted plane wave $\exp(2pix_0n_x)$, the resulting Fraunhofer diffraction pattern is the Fourier transform of the product of the transmission pattern and the illuminating wave. The result is,

$$\int\int H(n_x,n_y)e^{2pi[(x+x_0)n_x+yn_y]}dn_x dn_y = \tag{4.3.2}$$

$$c(dn)^2\,\mathrm{sinc}\left[cdn(x+x_0)\right]\sum_n\sum_m W_{nm}\,\mathrm{sinc}(yW_{nm}dn)\exp\{2pi[dn((x+x_0)(n+P_{nm})+ym)]\}$$

In the previous section, the three approximations (a) $\mathrm{sinc}\left[cdn(x+x_0)\right]\approx const$, (b) $\mathrm{sinc}(yW_{nm}dn)\approx 1$, and (c) $\exp[2pi(xP_{nm}d_x)]\approx 1$ were made for simplicity. The effects of these approximations on the reconstructed image depend on several factors, and with proper design, can be minimized.

The sinc function $\mathrm{sinc}\left[cdn(x+x_0)\right]$ creates a drop-off in intensity in the x-direction proportional to the distance from the center of the image plane. Approximation (a) considers this sinc factor to be nearly constant inside the image region. If the size of the image region is $\Delta x\times\Delta y$, then at the edges $x=\pm\frac{\Delta x}{2}$ where the effects are most severe, we have $\mathrm{sinc}\left[cM\pm c/2\right]$. This implies that a small aperture size c results in less drop-off in intensity. However, this also reduces the brightness of the image. For $cM=1/2$, the brightness ratio between the center and the edge of the image region is 9:1. By reducing this product to $cM=1/3$, the ratio drops to 2:1 [12]. So, depending on the situation, or the preference of the designer, brightness can be sacrificed for a reduction in intensity drop-off at the edges of the reconstructed image.

The sinc function $\mathrm{sinc}(yW_{nm}dn)$ indicates a drop-off in intensity similar to (a), but in the y-direction. This approximation is a little less dangerous because $\Delta xdn=1$, which means $\left|yWdn\right|<\frac{1}{2}$. Therefore, the sinc factor in (b) has a minimum of 0.64 in the image region [12]. Basically, the sinc function acts like a slight decrease in amplitude transmission by a factor $\mathrm{sinc}(yW_{nm}dn)$. It is well known that amplitude errors in apertures of coherent imaging systems have little effect on the image because they do not

deviate rays like phase errors do [12]. To reduce the effects of this approximation, every W could be reduced by a constant factor. However, some brightness must be sacrificed.

A possible solution to the sinc roll-off in the x direction is to divide the desired image by $\text{sinc}\left[cdn(x+x_0)\right]$. The desired image $f(x,y)$ becomes $\frac{f(x,y)}{\text{sinc}\left[cdn(x+x_0)\right]}$. The same thing cannot be done for the y direction because $\text{sinc}(yW_{nm}dn)$ depends on the aperture parameter $W_{nm}$ which is yet to be determined. Luckily, this sinc factor is less influential than the x dependent sinc, and design of the hologram can be altered to reduce its effects.

Last, the phase shift $\exp[2pi(xP_{nm}d_x)]$ causes a phase error that varies with x location in the image plane. The range of this phase error depends on x and P. Since $|x| \le \frac{\Delta x}{2} = \frac{1}{2dn}$ and $|P| \le \frac{1}{2M}$, the phase error ranges from zero to $\frac{p}{2M}$. At its maximum, the phase error corresponds to an optical path length of $\frac{l}{4M}$. For M=1, this is within the $\frac{l}{4}$ Rayleigh criterion for wave aberrations [12].

The detrimental effects of these approximations are less when the size of the image region is restricted. Simply put, differences between approximated and actual results increase with distance from the center of the image plane. If the image region is smaller, its edges are closer to the center of the image plane, which lessens the difference. A rigorous solution to reduce the effects of these three approximations is worked out in the paper by Lohmann and Paris [12]. This procedure is not included in this thesis.

### 4.4 A Modification of Lohmann's Method: Constant Aperture Size

From the previous discussion on the approximations made in Lohmann's method, the sinc oscillation due to the variable size of each aperture is the most difficult approximation to deal with. Therefore, it would be desirable, also for the purpose of simpler implementation, to somehow have the height of each aperture constant. This would allow for the desired image to be divided in the y-direction by the y-dependent

sinc drop-off just as was done in the x-direction for the sinc factor associated with the constant width of the aperture.

Logically, if every aperture has the same size, then only the positioning of the apertures is affecting the output. This means that all the information is contained in the phase. The following method is used to "shift" information in the hologram plane from the amplitude to the phase.

If only the magnitude of the desired image is of concern, then the phase at each sampling point in the observation plane is a free parameter. For this situation, the range of height $W_{nm}$ values can be reduced by iterative methods [2]. Suppose the sampled desired image has amplitudes $a_{nm}$ each with an unspecified corresponding phase $q_{nm}$. The discrete Fourier transform of the image is $\{W_{nm}\exp(i2pP_{nm})\}_{nm}$, where $\{..\}_{nm}$ indicates the sequence for all points n and m. The first step in reducing the range of $W_{nm}$ values is to assign values of $q_{nm}$ to the initial desired image which are independent and identically distributed phase samples with a uniform distribution over $(-p, p)$ [2]. The resulting DFT of the image samples is denoted by $DFT[\{a_{nm}\exp(iq_{nm})\}_{nm}] = \{A_{nm}\exp(iy_{nm})\}_{nm}$. Then, the spectral amplitudes $A_{nm}$ are set equal to any positive constant A. The inverse DFT of the spectrum with adjusted amplitudes is $DFT[\{A\exp(iy_{nm})\}] = \{\tilde{a}_{nm}\exp(i\tilde{q}_{nm})\}$. The original image amplitudes $a_{nm}$ are now combined with the new phase values $\tilde{q}_{nm}$ to form the new desired image samples. This process is repeated for a prescribed number of iterations. The image phase obtained from the last iteration becomes the new image phase. The final image phase values are used with the original image amplitudes to generate $W_{nm}\exp(i2pP_{nm})$ used for designing the hologram function.

By constraining the amplitude in the hologram domain and performing iterations, information in the hologram plane is transferred from the amplitude to the phase. Therefore, this reduces the negative effects caused by making all the apertures the same height. If all the apertures have the same height, then approximation (b) in section 4.3 can be handled the same way as approximation (a).

**4.5 Experiments and Simulations with Lohmann's Method**

The first step in creating a computer-generated hologram was to pick an image and represent it mathematically. For example, a binary image can be represented by a two dimensional array of 0's and 1's. Next, the discrete Fourier transform of this image was computed. The Fourier transform becomes the desired hologram function. The binary transmission pattern was generated from the hologram function using Lohmann's method. Matlab code was written to implements Lohmann's coding algorithm. The binary transmission pattern obtained from Lohmann's method was displayed in one of two ways. To display the exact Lohmann hologram (i.e. aperture size and positions are exactly as specified), the pattern is drawn into a CAD layout. The code using Lohmann's method includes commands that generate a ~.dxf file, which draws the pattern into an AutoCAD layout. An alternative method involves dividing each hologram cell into an $NxN$ array. This restricts the possible center positions and heights of each aperture. Thus, the phase and amplitude of each cell is quantized. Specifically, there are $N$ possible positions for the center of the aperture (phase), and $N/2$ potential height values (amplitudes) since the cell is symmetric in the y direction. Obviously, a larger $N$ produces a pattern closer to that of the exact hologram. For convenience, when Lohmann cells are divided into $NxN$ pixels, this will be referred to as $N$ -level quantization.

All holograms were designed using parameter values of $c = 1/2$ and $M = 1$. Setting $c = 1/2$ means that each aperture has a width equal to half that of the cell. It can also be shown that this value of c maximizes the brightness of the image [2]. After choosing $c = 1/2$, the only allowable value is $M = 1$ [12]. Also, each approximation mentioned in Section 4.3 was assumed to be valid.

Fig. 4.5.1 shows a binary E image and its Fourier transform amplitude. The entire image is of size 64x64. Notice that the E image is placed completely on one side of the image plane. Because the field in the hologram plane is real, Fourier theory demands that the image plane have Hermitian symmetry. This means that for an image plane of size $NxM$, $h(n,m) = h(N-n, M-m)$, where h is the reconstructed image.

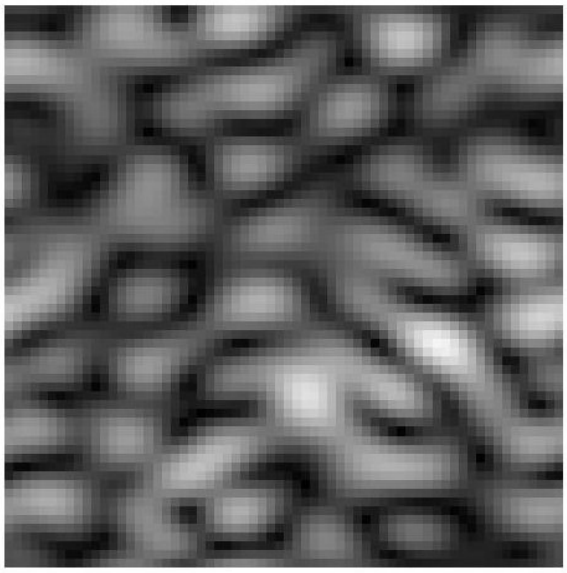
Fig. 4.5.2 displays the transmission pattern generated by Lohmann's method using the E image in Fig. 4.5.1. In this case, 16-level phase quantization, and 8-level amplitude quantization were used. Fig. 4.5.3 shows the simulated reconstruction. The

reconstructed image clearly shows the Hermitian symmetry due to the real-valued binary transmission pattern. In reality, the reconstruction seen here is repeated in both the x and y directions due to the finite size of the hologram.

Simulated reconstruction of a gray-scale image was also performed. Fig. 4.5.4 displays the image and its reconstruction. The entire image plane is 512x512.

One modification was made to Lohmann's method in the hope of improving the outcome of the reconstructed image. Since phase information in the Fourier, or hologram, plane is more important than amplitude in regards to the outcome in the image plane, Lohmann's method was modified to keep the amplitude of each cell constant and only manipulate phase. This was achieved by performing the iterative method discussed in Section 4.4 to move information from the amplitude to the phase in the hologram plane, thus allowing the size of each aperture to remain the same. Each aperture was still


(a)



(b)

Fig. 4.5.1 a) A binary E image b) Amplitude of the Fourier transform of the E image in (a).

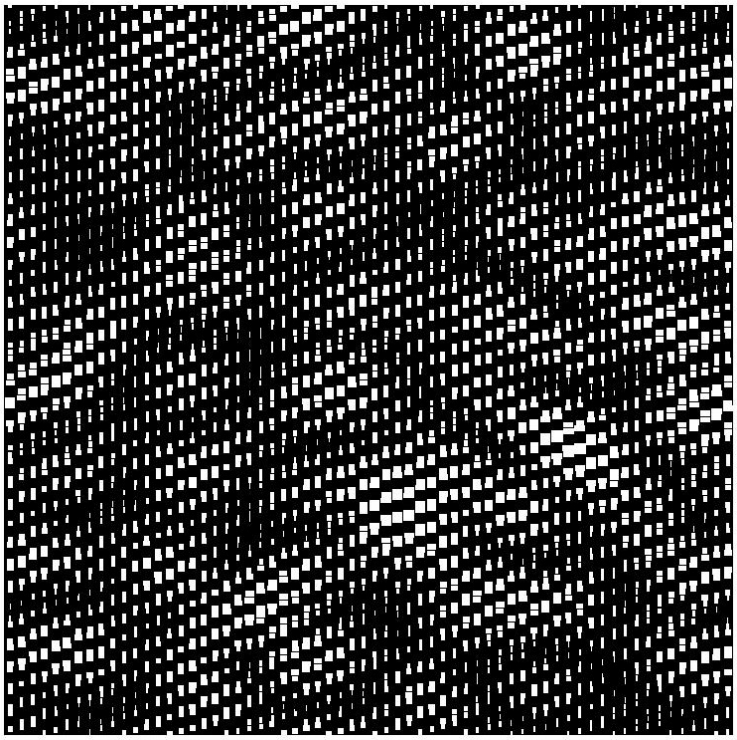


Fig. 4.5.2  The Lohmann hologram designed from the E image in Figure 4.5.1a.

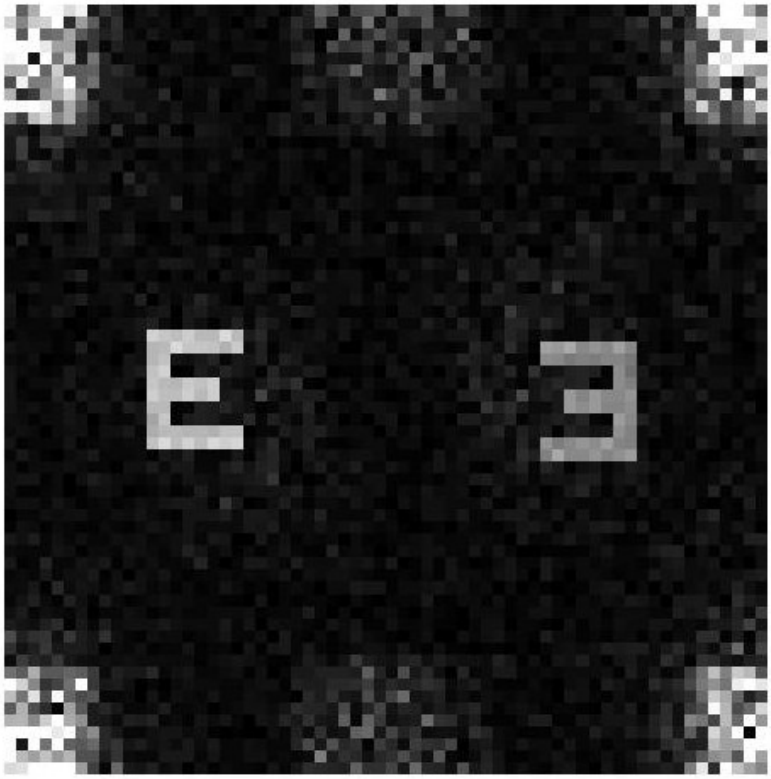
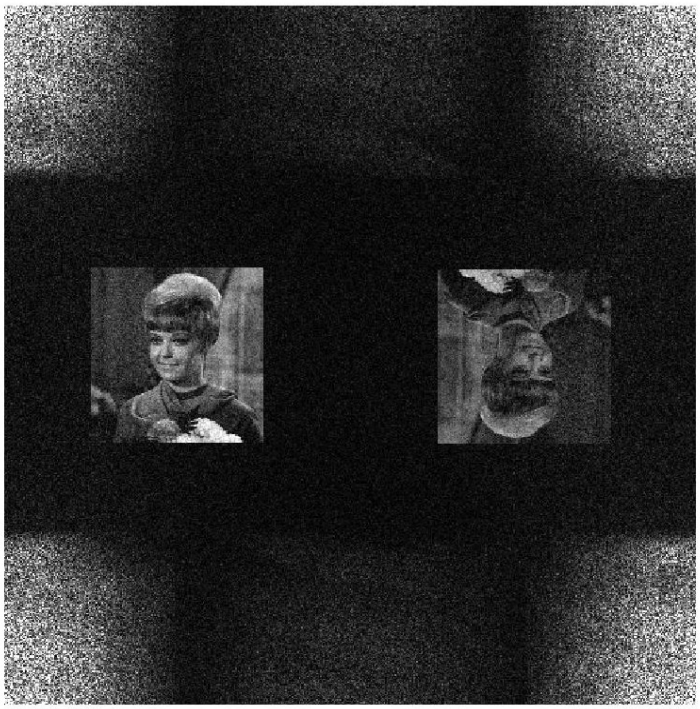


Fig. 4.5.3  The simulated reconstruction of the Lohmann hologram in Figure 4.5.2.

(a)



(b)

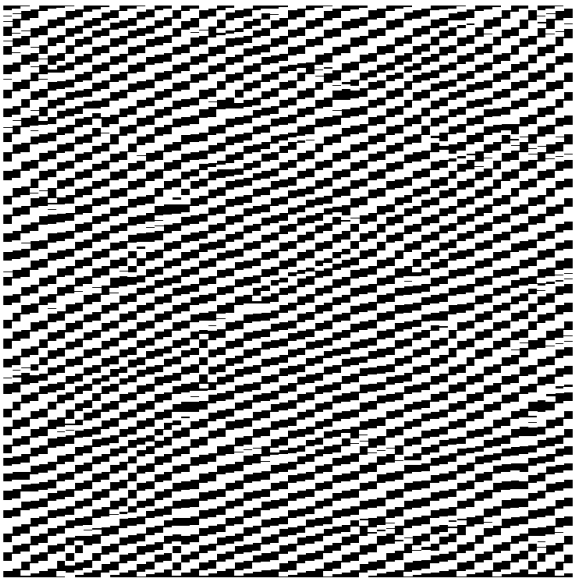
Fig. 4.5.4 a) A Girl image. b) Simulated reconstruction of the girl image by a Lohmann hologram.

shifted to produce the desired phase. This modification was made in the hope of increasing the brightness of the reconstructed image without significantly reducing its quality. As seen from Fig. 4.5.2, much of the hologram is black, which means that incoming light is being blocked, and thus energy is being lost. By keeping the amplitude of each block constant, and only adjusting phase, the average amplitude is greater, while hopefully maintaining reconstruction quality. An example of the transmission pattern generated using this modification is shown in Fig. 4.5.5a. The resulting simulated reconstruction is shown in Fig. 4.5.5b. Fig. 4.5.6 shows a reconstruction of the girl image from Fig. 4.5.4 using the constant amplitude method.

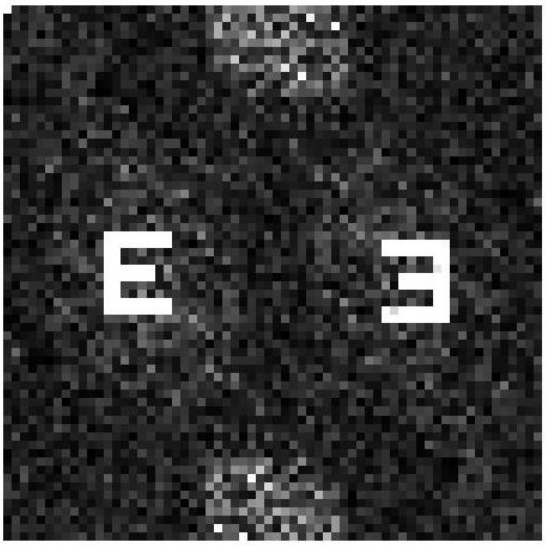
In order to make the transition between a computer generated hologram and a physical binary hologram, the designed binary transmission pattern must be realized. One way of achieving this is to print the designed hologram pattern onto a transparency. This was done for a Lohmann hologram, a modified Lohmann hologram with constant amplitude. Both hologram functions were represented in MATLAB using 16-level quantization, and saved as image files. The images were reduced in the computer and then printed at the desired size. The Lohmann holograms contained 64x64 cells, and were reduced to a size of 1.2x1.2cm. Therefore, the sampling period in the hologram plane was approximately 0.19mm. Each hologram was designed to reconstruct the E image seen in Fig. 4.5.1a, and tested using a HeNe laser. The reconstructions can be seen in Fig. 4.5.7.

**4.4 Discussion**

Probably the first thing noticed about a Lohmann hologram is that it closely resembles the amplitude of the actual Fourier transform of the desired image. This can be seen by comparing the hologram function in Fig. 4.5.1. Clearly, this should be the case if the hologram is representing the amplitude of the Fourier transform correctly. The greater the amplitude of the Fourier transform, the larger the aperture should be, which results in more transmittance. Depending on the preference of the designer, the transmission and no-transmission regions in the hologram may be interchanged. This can be done because the value of 1 for transmission and 0 for no-transmission was an



(a)



(b)

Fig. 4.5.5 a) Constant amplitude Lohmann hologram b) Simulated reconstruction by the const. amp. Lohmann method.

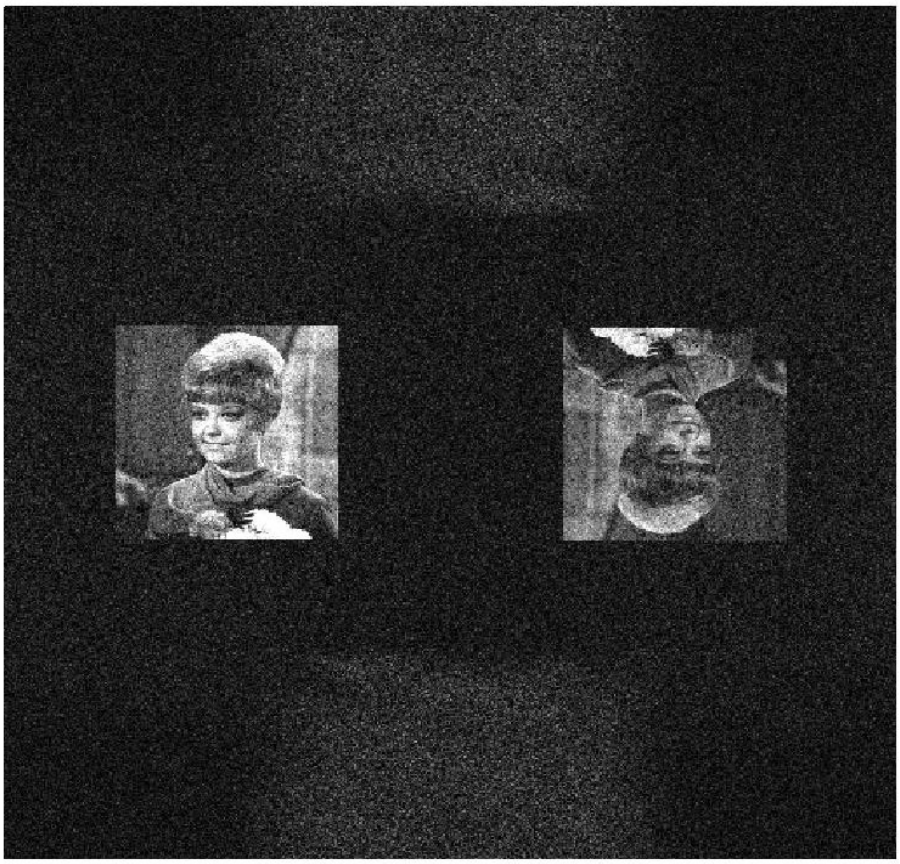
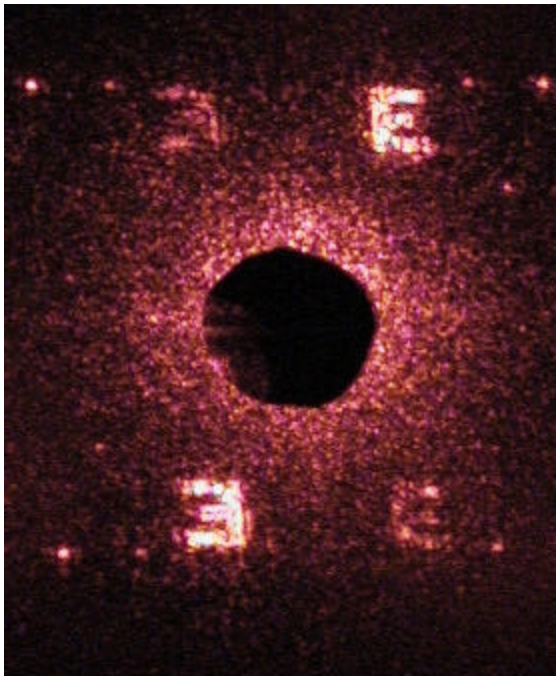
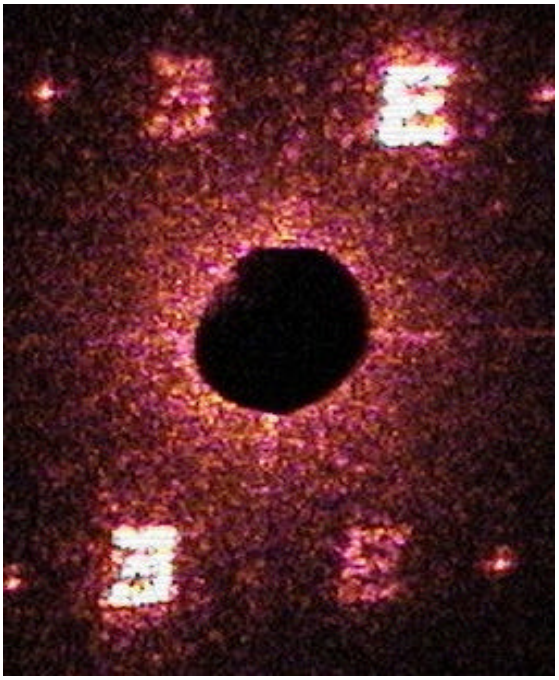


Fig. 4.5.6 Simulated reconstruction of girl image with the constant amplitude Lohmann method.



(a)



(b)

Fig. 4.5.7 a) Optical reconstruction from a Lohmann hologram b) Optical reconstruction from a constant amplitude Lohmann hologram.

arbitrary choice. The choice of 0 for transmission and 1 for no-transmission could have been used. It does not match with the usual conception that zero means nothing, but as long as the binary coding is consistent throughout the hologram, either method will work.

Another option would be to use a (-1,1) binary system. This would be the desired choice if a phase shift is going to be used when physically implementing the hologram. Creating a phase hologram was not an option given the resources for this research, so only binary amplitude holograms where considered.

Next, the iterative method mentioned in Section 4.4 was used to design a Lohmann hologram where each aperture has constant amplitude. When compared to the other Lohmann holograms, the modified constant amplitude hologram clearly has a greater overall amplitude. In fact, each aperture was given the maximum relative height of 1 in order to maximize the brightness of the image.

Based on simulations, the image is indeed brighter in the constant amplitude case, but the sharpness of the E appears to have decreased slightly. The reduction in image sharpness might be attributed to ignoring any information that was still contained in the amplitude of the Fourier transform of the image. This amplitude information is of course lost when the aperture size is made constant. The reduction in image sharpness may also be due to noise around the edges of the E as a result of the all around increased intensity. If this is the case, this unwanted reconstruction at the edges simply may not have been visible in the original reconstruction due to a lack of brightness. Also, the constant amplitude hologram produced significantly less noise in the corners of the image plane, which is an advantage over the original method. The “noise” has moved from the corners of the image plane to the area around the image.

For the girl image reconstructed from a constant amplitude design, the image was again brighter and exhibited less noise in the corners of the image plane when compared to the original method. There is no noticeable error around the image for the constant amplitude case like there was for the binary image.

The two physical Lohmann holograms produced recognizable optical reconstructions of the letter E. A hole was cut in the observation screen to remove the bright, zero-order spot in the center of the image plane. As the simulations predicted, the original Lohmann method produced a more well-defined image than the constant

amplitude method. Also supporting the simulated results, the constant amplitude Lohmann hologram generated a brighter image and less reconstruction outside the image region. The amount of noise in the center of the image plane is noticeably greater using the original Lohmann method.

Holograms for larger, more complex images such as the girl seen in Fig. 4.5.4a were also made. The optical reconstructions from these holograms were not recognizable. This was due to resolution limits in printing. For the 512x512 hologram designed for the girl image, a hologram 1.2cm x 1.2cm has a sampling period of approximately 0.023mm. Furthermore, if each cell has only 8-level quantization, a resolution of 0.003mm is required. The accessible method was only capable of 0.02mm resolution.

### 4.7 Conclusions

Lohmann's method was used to design binary computer-generated holograms. Simulated reconstructions were performed for binary and gray-scale images. Holograms having constant aperture amplitude were designed in addition to standard Lohmann holograms. Holograms with constant aperture amplitude were designed following an iterative Fourier transform algorithm. Simulated reconstructions due to the constant amplitude Lohmann holograms had brighter images and less noise outside the image region, but displayed less image definition. Optical reconstructions from these holograms supported the simulation predictions.

# 5 INTERLACING AND INTERATIVE INTERLACING TECHNIQUES FOR DESIGNING COMPUTER-GENERATED HOLOGRAMS

## 5.1 Introduction

In the previous chapters we have seen that, in general, a hologram is an optical pattern that records the amplitude and phase information of a wavefront (or image) in order to reconstruct the original image. A computer-generated hologram can simulate the reconstruction of an image, as long as that image can be represented mathematically. Therefore, the original object is not needed in order to create the hologram.

The optical system assumed in the synthesis of a computer-generated hologram using the interlacing technique (IT) and the iterative interlacing technique (IIT) is the same Fourier lens system used for CGH design by Lohmann's method. Under the Fresnel approximation, the wavefronts at the two planes are related by a Fourier transform pair [1]. The image in the observation plane will be defined as the inverse Fourier transform of the wavefront at the CGH plane. Thus, the transmittance values of the CGH are designed based on the Fourier transform of the desired image. The objective, as before, is to design a transmittance pattern for the CGH, which results in a reconstructed image at the observation plane that resembles some desired image.

In IT and IIT, the CGH consists of an array of discrete points. The transmittance function of a hologram consisting of an $M \times N$ array of $\Delta n_x \times \Delta n_y$ sized pixels can be represented by the sum,

$$G(n_x, n_y) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} H(k,l)\, rect\left(\frac{n_x - k\Delta n_x}{\Delta n_x}\right) rect\left(\frac{n_y - l\Delta n_y}{\Delta n_y}\right) \tag{5.1.1}$$

where $H(k,l)$ is the binary transmittance of the $(k,l)th$ point. The reconstructed image in the observation plane is given by the Fourier transform of the transmittance:

$$g(x,y) = \int\int G(n_x, n_y) e^{2pi[xn_x + yn_y]} dn_x dn_y = \tag{5.1.2}$$

$$\Delta n_x \Delta n_y \text{sinc}\left[\Delta n_x x\right] \text{sinc}\left[\Delta n_y y\right] \sum\sum H(k,l) \exp\left[2pi(kx\Delta n_x + ly\Delta n_y)\right]$$

By ignoring the two constants and the two sinc factors outside the sums, the reconstructed image is approximated by the two-dimensional inverse discrete Fourier

transform (2D-IDFT) of the transmittance values $H(k,l)$. Ignoring the sinc factors is the same approximation that was made in the Lohmann procedure. For discussion on how to deal with this approximation, refer to section 4.3.

## 5.2 Projection-Onto-Constraint-Sets Method

As discussed in the basic theory of digital holography, there are several encoding methods available to generate binary transmittance values. The method used in IT and IIT is the Projection-Onto-Constraint-Sets (POCS) algorithm [7]. The POCS method uses the amplitude information of the desired image as the constraint that is imposed on the observation plane during transformation between the discrete-space domain (i.e. the observation plane) and the discrete-spatial-frequency domain (i.e. the CGH plane).

Letting the dimensions of both the observation and CGH planes be $MxN$, the relationship between the wavefronts at the observation plane $h(m,n)$ and the CGH plane $H(k,l)$ is given by the following discrete Fourier transform pair:

$$h(m,n) = \frac{1}{MN}\sum_{k=0}^{M-1}\sum_{l=0}^{N-1} H(k,l) W_M^{mk} W_N^{nl} \qquad (5.2.1)$$
$$\text{where} \quad 0 \le m \le M-1, \quad 0 \le n \le N-1$$

$$H(k,l) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} h(m,n) W_M^{-mk} W_N^{-nl} \qquad (5.2.2)$$
$$\text{where} \quad 0 \le k \le M-1, \quad 0 \le l \le N-1$$

and

$$W_u = \exp(i2p/u)$$

The goal of the POCS method is to generate the CGH whose reconstructed image most accurately resembles the desired image.

Given a desired image $f(m,n)$ in a region $R$ of the observation plane, the POCS method works as follows:

1) Using (5.2.2), compute $F(m,n)$ from $f(m,n)$.

2) Generate the binary transmittance values $H(k,l)$ from $F(k,l)$ based on the condition,

$$H(k,l) = \begin{cases} 1 & \text{if } \mathrm{Re}[F(k,l)] \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.2.3}$$

3) Using (5.2.1), find the reconstructed image $h(m,n)$. The accuracy of the reconstructed image is measured based on the Mean-Square-Error (MSE) between $f(m,n)$ and $h(m,n)$ within $R$, the region of the desired image. The MSE is defined as [13],

$$\mathrm{MSE} = \frac{1}{MN} \sum_{(m,n)\in R} \sum |f(m,n) - l\,h(m,n)|^2 \tag{5.2.4}$$

where $l$ is a scaling factor. The minimum MSE for $h(m,n)$ is achieved if,

$$l = \frac{\sum_{(m,n)\in R} \sum f(m,n) h^*(m,n)}{\sum_{(m,n)\in R} \sum |h(m,n)|^2} \tag{5.2.5}$$

4) Define a new input image $f'(m,n)$ such that

   a) Outside $R$, $f'(m,n)$ equals $h(m,n)$

   b) Inside $R$, $f'(m,n)$ has the amplitude of the original image $f(m,n)$ and the phase of $h(m,n)$.

5) Letting $f(m,n) = f'(m,n)$, go to step 1).

6) Repeat steps 1 through 5 until the MSE converges or specified conditions are met.

### 5.3 The Interlacing Technique (IT)

Another method of designing CGH's is the interlacing technique (IT) [6], which can be incorporated into any existing CGH synthesis method in order to improve its performance. The IT method divides the entire hologram plane into a set of subholograms. A subhologram consists of a set of cells, or points, referred to as a "block". All the subholograms are designed separately and then interlaced, or entangled, to create one hologram.

In the IT method, once the entire hologram is divided into smaller subholograms, the first subhologram is designed to reconstruct the desired image $f(m,n)$. The reconstructed image due to the first subhologram is $h_1(m,n)$. Because the subhologram cannot perfectly reconstruct the desired image, there is an error image $e_1(m,n)$ defined as,

$$e_1(m,n) = f(m,n) - l_1 h_1(m,n) \tag{5.3.1}$$

In order to eliminate this error, the second subhologram is designed with $\frac{e_1(m,n)}{l_1}$ as the desired image. Since the Fourier transform is a linear operation, the total reconstruction due to both subholograms is simply the sum of the two individual reconstructions. If the second subhologram were perfect and its scaling factor matched $l_1$, the sum of the two reconstructed images would produce $f(m,n)$. However, as with the first subhologram, there will be error. So, the third subhologram serves to reduce the left over error from the first two subholograms. Therefore, each subhologram is designed to reduce the error between the desired image and the sum of the reconstructed images of all the previous blocks. This procedure is repeated until each subhologram has been designed.

Each subhologram is generated sub-optimally by the POCS algorithm. However, the total CGH may not yet reach the optimal result even after all the subholograms are utilized once. To overcome this problem, there is the Iterative Interlacing Technique (IIT).

## 5.4 The Iterative Interlacing Technique (IIT)

IIT is simply an iterative version of the IT method, which is designed to achieve the minimum MSE [6]. After each subhologram has been designed using the IT method, the reconstruction due to the entire hologram $h_f(m,n)$ has a final error $e_f(m,n)$. To apply the iterative interlacing technique, start a new sweep through the subholograms, where the new desired image $f'(m,n)$ for the first subhologram is the original desired image $f(m,n)$ minus the reconstruction created by all of the subholograms except the first (divided by the scaling factor). So, the first subhologram attempts to reconstruct,

$$f'(m,n) = h_1(m,n) - \frac{e_f(m,n)}{l_f} \tag{5.4.1}$$

where $e_f(m,n) = f(m,n) - l_f h_f(m,n)$, and $l_f$ is the scaling factor after the last subhologram. Once the first subhologram is re-designed, the error image due to the entire hologram is calculated using the new reconstruction created by the first subhologram. Similarly, the second subhologram is designed to reconstruct $h_2(m,n) - \frac{e_f'(m,n)}{l_f'}$, which is the error image due to the reconstruction of all subholograms except the second, and $e_f'(m,n)$ is the updated final error. Convergence is achieved when the absolute different between successive reconstructed images

$$\Delta_i = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left|h_b^i(m,n) - h_b^{i-1}(m,n)\right| \tag{5.4.2}$$

reaches zero. By using the IIT method, the convergence tends to move away from the local-minimum MSE to reach the global-minimum MSE of the existing CGH [13].

### 5.5 Optimal Decimation-in-frequency Iterative Interlacing Technique (ODIFIIT)

Now that the IT and IIT methods have been presented, it is time to see how they can be applied most efficiently. The Optimal Decimation-in-Frequency Iterative Interlacing Technique (ODIFIIT) was designed to optimize the results of the IIT procedure [6]. ODIFIIT exploits the decimation-in-frequency property of the Fast Fourier Transform (FFT) when dividing the hologram into subholograms, and has two important advantages over IIT. It decreases computation time by reducing the dimensions of the Fourier transform and its inverse. Also, only the image inside the desired image region $R$ is involved in the transformation. This makes the design of each subhologram optimal because only contributions from the data of interest are taken into consideration.

The geometry of the image plane for the ODIFIIT method is shown in Fig 5.5.1. The desired amplitude $f_0(m,n)$ is a real-valued array of size $A$ x $B$. It will be the constraint throughout the design of the CGH. Therefore, the desired image $f(m,n)$ at any given time will be $f_0(m,n)$ times a floating phase, which is determined by the phase of the current reconstruction inside region $R$. $f(m,n)$ is placed within region $R$, which begins at the point $(M_1,N_1)$. The Hermitian conjugate of the reconstructed image exists in the region $R^+$ due to the real-valued CGH transmittance. Since the binary CGH has cell magnitude equal to unity, it is important that the desired image is scaled so that its DFT is normalized to allow a direct comparison between it and the reconstructed image $h(m,n)$.

The total CGH is divided into $\mu$ x $\nu$ subholograms, or blocks, where $m = \frac{M}{A}$ and $n = \frac{N}{B}$. $m$ and $n$ are guaranteed to be integers if $M$, $N$, $A$, and $B$ are all powers of two. Utilizing decimation-in-frequency, the blocks are interlaced such that the $(a,b)th$ block consists of the cells $(mk+a, nl+b)$, where $0 \le k \le A-1$, $0 \le l \le B-1$, $0 \le a \le m-1$, and $0 \le b \le n-1$. Fig. 5.5.2 shows an example with $m = n = 2$. Expressing $H(k,l)$ as the sum of all the blocks and using (5.4.1), the expression for the reconstructed image becomes

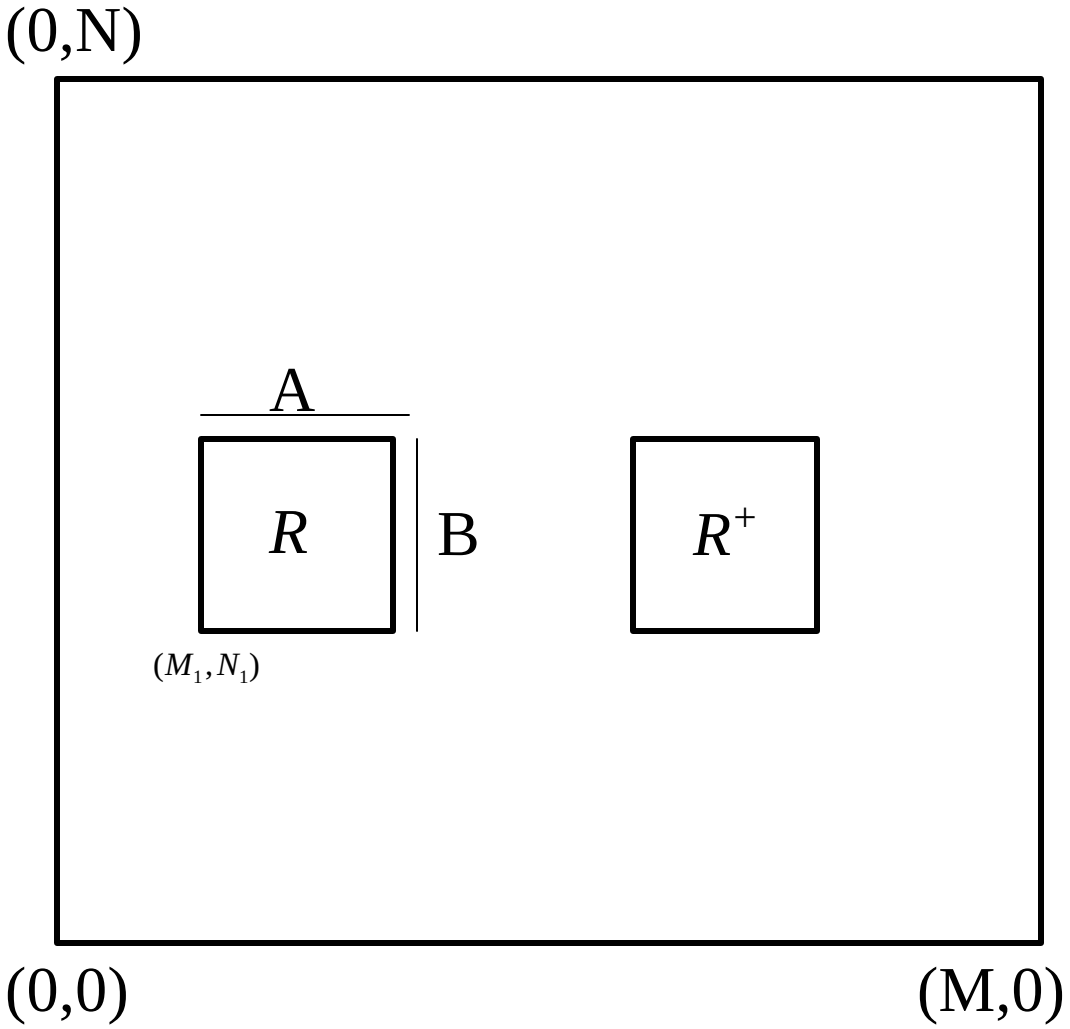



Fig. 5.5.1 Image plane geometry for ODIFIIT.

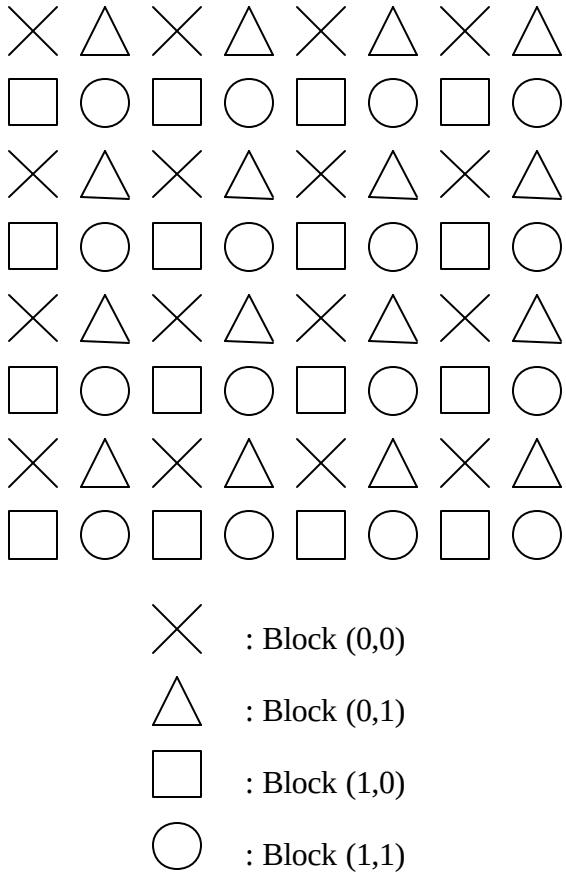



Fig. 5.5.2 Interlacing with $u=v=2$.

$$h(m,n) = \frac{1}{MN}\sum_{k=0}^{M-1}\sum_{l=0}^{N-1} H(k,l) W_M^{mk} W_N^{nl} \qquad (5.5.1)$$

$$= \frac{1}{mn}\sum_{a=0}^{m-1}\sum_{b=0}^{n-1}\left[\frac{1}{AB}\sum_{k=0}^{A-1}\sum_{l=0}^{B-1} H(mk+a, nl+b) W_A^{mk} W_B^{nl}\right] W_M^{ma} W_N^{nb}$$

$$\text{where} \quad 0 \le m \le M-1, \quad 0 \le n \le N-1$$

The reconstructed image in region $R$ is computed by replacing $m$ and $n$ by $m + M_1$ and $n + N_1$, respectively, and letting $m$ and $n$ span just the image region.

$$h(m+M_1, n+N_1) =$$
$$\frac{1}{mn}\sum_{a=0}^{m-1}\sum_{b=0}^{n-1}\left[\frac{1}{AB}\sum_{k=0}^{A-1}\sum_{l=0}^{B-1} H(mk+a, nl+b) W_A^{(m+M_1)k} W_B^{(n+N_1)l}\right] W_M^{(m+M_1)a} W_N^{(n+N_1)b}$$

$$\text{where} \quad 0 \le m \le A-1, \quad 0 \le n \le B-1 \qquad (5.5.2)$$

Let $h_{a,b}(m,n)$ be the size $A$ x $B$ inverse discrete Fourier transform of the $(\alpha,\beta)$th subhologram:

$$h_{a,b}(m,n) = IDFT_{AB}\left[H(mk+a, nl+b)\right]_{m,n}$$
$$= \frac{1}{AB}\sum_{k=0}^{A-1}\sum_{l=0}^{B-1} H(mk+a, nl+b) W_A^{mk} W_B^{nl} \qquad (5.5.3)$$

$$\text{where} \quad 0 \le a \le m-1, \quad 0 \le b \le n-1, \quad 0 \le m \le A-1, \quad 0 \le n \le B-1.$$

The IDFT above is of size $A$ x $B$. The reconstructed image inside region $R$ becomes

$$h(m+M_1, n+N_1) = \frac{1}{mn}\sum_{a=0}^{m-1}\sum_{b=0}^{n-1} h_{a,b}(m+M_1, n+N_1) W_M^{(m+M_1)a} W_N^{(n+N_1)b} \qquad (5.5.4)$$

$$\text{where} \quad 0 \le m \le A-1, \quad 0 \le n \le B-1$$

The indices $(m+M_1)$ and $(n+N_1)$ of $h_{a,b}(m+M_1, n+N_1)$ are implicitly assumed to be $(m+M_1)$ modulo A and $(n+N_1)$ modulo B, respectively. (5.5.4) gives the reconstructed image in $R$ in terms of the size $A$ x $B$ IDFT's of all the blocks. From this equation, it can be seen that the reconstructed image in $R$ due to the $(\alpha,\beta)$th block is

$$h'_{a,b}(m+M_1,n+N_1)=\frac{1}{mn}h_{a,b}(m+M_1,n+N_1)W_M^{(m+M_1)a}W_N^{(n+N_1)b} \tag{5.5.5}$$

which is the IDFT of the $(\alpha,\beta)$th block times the appropriate phase factor, divided by $mn$.

We now define an array, which will be useful later on:

$$\tilde{h}_{a,b}(m+M_1,n+N_1)=h(m+M_1,n+N_1)-h'_{a,b}(m+M_1,n+N_1). \tag{5.5.6}$$

This is the reconstruction in $R$ due to all except the $(\alpha,\beta)$th block.

Conversely, given the desired image in $R$, the transmittance values can be obtained. From (5.2.2)

$$H(k,l)=\sum_{m=0}^{A-1}\sum_{n=0}^{B-1}h(m+M_1,n+N_1)W_M^{-(m+M_1)k}W_N^{-(n+N_1)l} \tag{5.5.7}$$
$$\text{where}\quad 0\le k\le M-1,\quad 0\le l\le N-1$$

Dividing $H(k,l)$ into $u$ x $v$ blocks as before,

$$H(mk+a,nl+b)=\sum_{a=0}^{m-1}\sum_{b=0}^{n-1}\left[\sum_{m=0}^{A-1}\sum_{n=0}^{B-1}h(m+M_1,n+N_1)W_M^{-(m+M_1)a}W_N^{-(n+N_1)b}W_A^{-(m+M_1)k}W_B^{-(n+N_1)l}\right]$$
$$=W_A^{-M_1k}W_B^{-N_1l}\sum_{a=0}^{m-1}\sum_{b=0}^{n-1}DFT_{AB}\left[h(m+M_1,n+N_1)W_M^{-(m+M_1)a}W_N^{-(n+N_1)b}\right]_{k,l}$$

$$0\le k\le A-1,\quad 0\le l\le B-1,\quad 0\le a\le m-1,\quad 0\le b\le n-1. \tag{5.5.8}$$

Therefore, the transmittance values of block $(\alpha,\beta)$ that create the image $h(m+M_1,n+N_1)$ in region $R$ are given by

$$H(mk+a,nl+b)=W_A^{-M_1k}W_B^{-N_1l}DFT_{AB}\left[h(m+M_1,n+N_1)W_M^{-(m+M_1)a}W_N^{-(n+N_1)b}\right]_{k,l} \tag{5.5.9}$$
$$\text{where}\quad 0\le k\le A-1,\quad 0\le l\le B-1$$

Using (5.5.5) and (5.5.9), we can compute the reconstruction in $R$ due to each individual block, or, given a desired image in $R$, determine the transmittance values needed to reconstruct that desired image. Therefore, we can now utilize the IIT to design a CGH.

Letting $f_0(m+M_1,n+N_1),0\le m\le A-1,\ 0\le n\le B-1$, be the amplitudes of a desired image of size $A$ x $B$, the ODIFIIT algorithm procedure is as follows:

1) Define the parameters M, N, A, B, $M_1$, and $N_1$, and determine $\mu$ and $\nu$. Then, divide the total CGH into $\mu$ x $\nu$ interlaced subholograms, or blocks.
2) Create an initial M x N hologram with random transmittance values of 0 and 1.
3) Take the *M x N* IDFT of the total hologram. The reconstruction in the image region *R* is obtained simply by using only the points inside *R*. That is $h(m+M_1, n+N_1),\ 0 \le m \le A-1,\ 0 \le n \le B-1.$
4) The desired image $f(m+M_1, n+N_1)$ is obtained by applying the phase of each point $h(m+M_1, n+N_1)$ to the amplitude $f_0(m+M_1, n+N_1)$ as in the POCS method. So,

$$f(m+M_1, n+N_1) = f_0(m+M_1, n+N_1)\exp\left(i f_{m+M_1, n+N_1}\right) \qquad (5.5.10)$$

where $f_{m+M_1, n+N_1} = \arg\{h(m+M_1 n+N_1)\}$

5) Find the optimization parameter $l$ using (5.2.5).
6) Using (5.5.3), (5.5.5), and (5.5.6), find $\tilde{h}_{a,b}(m+M_1, n+N_1)$. This is the reconstruction in *R* due to all except the $(\alpha, \beta)$th block.
7) Determine the image that the $(\alpha, \beta)$th block needs to reconstruct (i.e. the error image) by subtracting $l\ \tilde{h}_{a,b}(m+M_1, n+N_1)$ from $f(m+M_1, n+N_1)$. Divide the difference by $l$. This yields,

$$e(m+M_1, n+N_1) = \frac{f(m+M_1, n+N_1)}{l} - \tilde{h}_{a,b}(m+M_1, n+N_1)\ ,$$

which is equivalent to the error image in the IIT method.
8) Using (5.5.9), find the transmittance values $E(mk+a, nl+b)$ for the current block that reconstructs the error image.
9) Design the binary transmittance values of the current block using the condition

$$H(mk+a, nl+b) = \begin{cases} 1 & \text{if } \mathrm{Re}[E(mk+a, nl+b)] \ge 0 \\ 0 & \text{otherwise} \end{cases}.$$

10) Find the new reconstruction in *R* due to the current block, $h'_{a,b}(m+M_1, n+N_1)$.
11) Determine the new total reconstructed image $h(m+M_1, n+N_1)$ by adding the new $h'_{a,b}(m+M_1, n+N_1)$ to $\tilde{h}_{a,b}(m+M_1, n+N_1)$.

12) With the new $h(m+M_1, n+N_1)$, use (5.5.10) to update $f(m+M_1, n+N_1)$.

13) Repeat steps 7) through 12) until the transmittance value at each point in the current block converges.

14) Update the total hologram with the newly designed transmittance values.

15) Keeping $\lambda$ the same, repeat steps 3) through 14) (except step 5) for all $mn$ blocks.

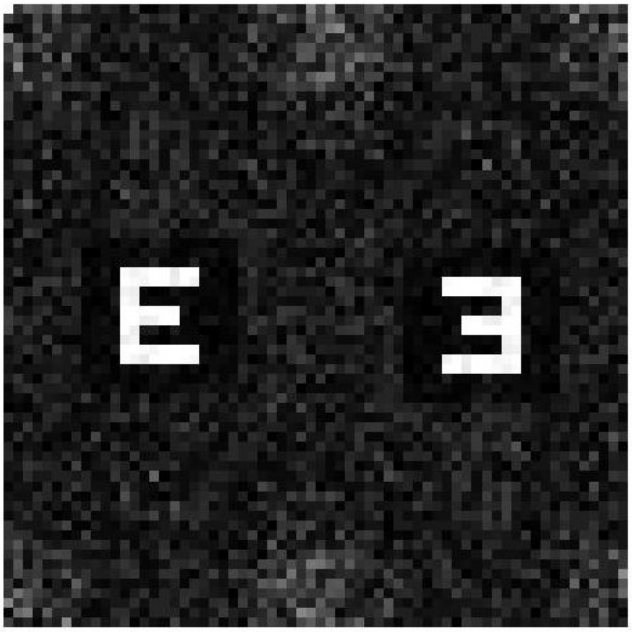
16) After all blocks are designed, compute the MSE from (5.2.4).

17) Repeat steps 3) through 16) until the MSE converges. Convergence indicates that the optimal CGH has been designed for the current $\lambda$.

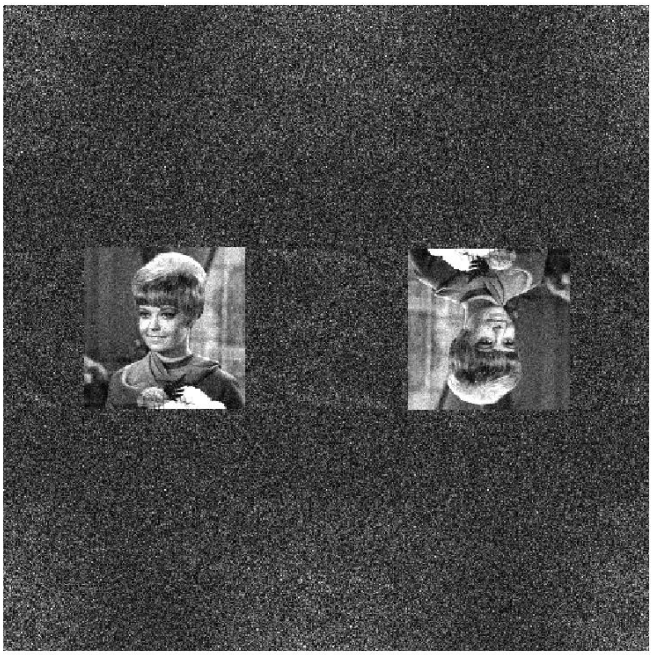
## 5.6 Experiments with ODIFIIT

The ODIFIIT method was used to design computer-generated holograms of the same binary E and girl images that were used in testing Lohmann's method. These two images are shown in Fig. 4.5.1a and Fig. 4.5.4a. Simulated reconstructions of both images are shown in Fig. 5.6.1. A higher resolution 256x256 grayscale image was also reconstructed using ODIFIIT. The desired image and simulated reconstruction are shown in Fig. 5.6.2.

All holograms designed using ODIFIIT used the interlacing pattern shown in Fig. 5.5.2. There are many different ways in which the subholograms could be interlaced. However, it has been shown that this interlacing method minimizes the MSE and thus produces the best results [13]. Also, each hologram was divided into 4x4 blocks, or subholograms. So, the binary E image was contained in a 16x16 desired image region inside the entire 64x64 image plane. Similarly, the 128x128 desired image region of the girl image is within a total image plane of size 512x512. Various numbers of divisions have been tried [13]. Simulated experiments have shown that dividing the hologram into



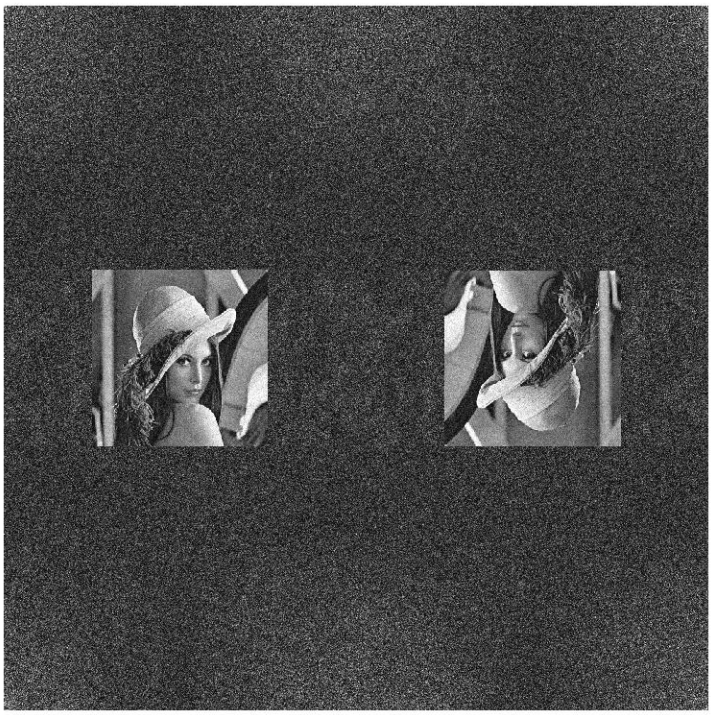
(a)



(b)

Fig. 5.6.1 Simulated reconstructed images from CGHs designed using ODIFIIT. a) Binary E image b) Gray-scale girl image.

(a)



(b)

Fig. 5.6.2  a) 256x256 gray

4x4 blocks leads to the lowest MSE, and that increasing the number of divisions does not improve outcome [13].

A physical CGH was designed with ODIFIIT to reconstruct the binary E image. The optical reconstruction is shown in Fig. 5.6.3.

### 5.7 Results and Discussion

The simulated reconstructions using ODIFIIT show very good results. The images inside the desired image regions are well defined, and little reconstruction is seen outside the desired image region for both the binary and gray-scale images.

By simple observation of the simulated reconstructions, there are several noticeable differences between the ODIFIIT method and Lohmann's method. In the reconstruction of the binary image, ODIFIIT produced a more uniform image. The intensity of the "E" reconstructed with ODIFIIT is almost constant for all points contained in the image. Conversely, the Lohmann reconstruction has light and dark patches inside the image. There is also no noticeable reconstruction inside the image region that is "misplaced." All extra reconstruction is moved out of the image region R. This is in contrast to the constant amplitude Lohmann's method where considerable noise is present around the E in the reconstruction.

Optical reconstructions from ODIFIIT also produced recognizable binary E images. The brightness of the image was comparable to that of the original Lohmann method, but not as bright as the constant amplitude method's image. The quality of the ODIFIIT and constant amplitude Lohmann images appear to be about the same. The ODIFIIT reconstruction is not as sharp as that produced with the original Lohmann method. However, the ODIFIIT images have less noise directly next to the E within the desired image region. This supports the simulated results. Also, reconstruction outside the image region seems to be relatively dispersed. This is in comparison to the concentrated noise in the center of the Lohmann reconstructions, and the excess noise around the E in the constant amplitude Lohmann image.

Like the CGHs designed with Lohmann's method, the transmittance function designed with ODIFIIT can also be coded as (-1,1) to make a binary phase hologram.

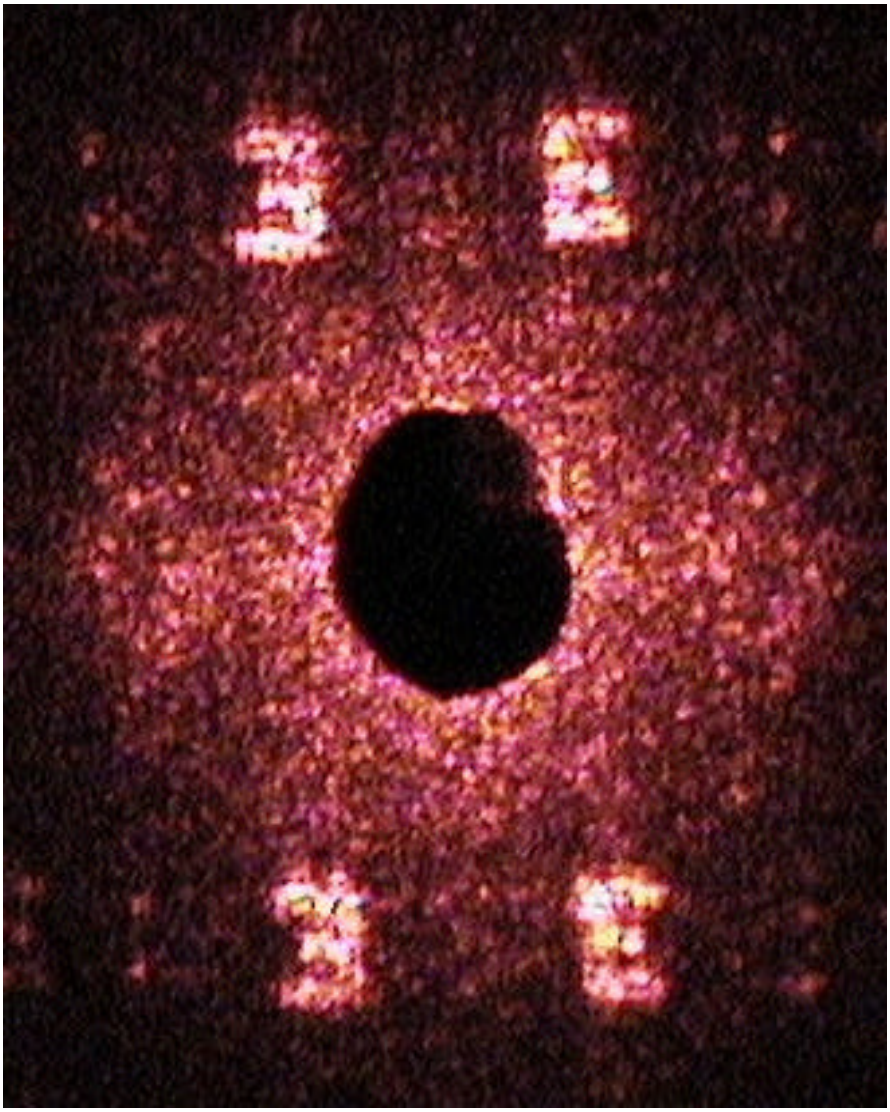


Fig. 5.6.3  Optical reconstruction of binary E image with an ODIFIIT hologram.

ODIFIIT also has the flexibility to code transmittance functions with multiple levels of phase quantization. This corresponds to a multi-level phase element. For example, four levels of phase quantization could be implemented when encoding the hologram. So at each point in the hologram there would be a phase shift of $0, p/2, p$, or $3p/2$. This generates a reconstruction with lower MSE because the hologram function is a closer approximation to the actual desired phase at each point. More levels of quantization also greatly increases the efficiency of the hologram. Since the transmittance can be imaginary at some points, the image plane does not have Hermitian symmetry, which also aids the increase in efficiency. Advances in semiconductor technology would make such high efficiency CGHs more practical. The current difficulty in making multi-level surface relief is that once the first round of surface relief is done, the element must be re-aligned very precisely. This is where much error can occur.

### 5.8 Conclusions

Computer-generated holograms were designed using ODIFIIT. Simulation results showed very accurate reconstructions of desired images. Compared to results from Lohmann's method, there was a significant increase in reconstruction quality for high resolution gray-scale images. Optical reconstruction was also performed, which produced a recognizable binary image. Optical reconstructions produced similar qualitative results to theoretical predictions. Quality of the optical reconstruction was similar to optical reconstruction using a Lohmann or constant amplitude Lohmann hologram. Optical reconstructions supported many of the expected differences between methods based on simulation.

Thus far, no quantitative descriptions of the reconstruction quality has been mentioned. In the next chapter, a new method is introduced. Quantitative analysis will be used to compare the new method with both Lohmann's method and ODIFIIT.

# 6 COMBINING LOHMANN'S METHOD WITH ODIFIIT FOR DESIGNING COMPUTER-GENERATED HOLOGRAMS

## 6.1 Introduction

Lohmann's coding scheme was implemented into the interlacing technique to create a new method for designing computer-generated holograms. This new method will be called the Lohmann-ODIFIIT method, or LM-ODIFIIT for short. The desired amplitude and phase of each subhologram point is encoded using a Lohmann cell, but the hologram is divided into interlaced subholograms like in ODIFIIT.

Recall that IT involves dividing the hologram plane into subholograms, where each subhologram is designed based on the difference between the desired image and the reconstruction due to all other subholograms. Since the aperture inside a Lohmann cell has an infinite number of possible positions, to determine the exact reconstruction of a Lohmann hologram, the hologram function must be made physical and the reconstruction performed optically. This can be a lengthy process. Therefore, a simulated reconstruction is needed to realistically use IT. The simulated reconstructions are done as they were in Chapter 4. The hologram plane is represented as the sum of the products of two rectangular functions and the reconstructed image is calculated by taking the Fourier transform of this function. This makes combining Lohmann's coding method with IT feasible.

## 6.2 Incorporating Lohmann's Coding Scheme into ODIFIIT

Based on the nature of Lohmann's coding technique, is there any advantage to using an iterative version of the Lohmann-IT method? The answer is "yes". Since the detour-phase method used in binary Lohmann holograms cannot code amplitude and phase exactly, there will always be some amount of inherent error in the reconstruction. Therefore, iterations should be effective in reducing error. What this ultimately means is that if Lohmann's method is incorporated into ODIFIIT, it will likely produce better results than Lohmann's method alone.

### 6.3 Quantized Lohmann's Method and ODIFIIT

Optimal decimation-in-frequency iterative interlacing technique (ODIFIIT) created by Ersoy and Zhuang [6] is an iterative interlacing technique, which exploits decimation-in-frequency to optimize the interlacing of subholograms. The ODIFIIT method uses a binary coding scheme that considers only the phase of each sampling point and incorporates projection onto constrained sets (POCS) which is an iterative Fourier transform algorithm.

The original goal in trying to combine Lohmann's coding technique with ODIFIIT was to supply more information at each sampling cell. But, Lohmann's method allows for an infinite number of aperture sizes and positions, which is not practical for many methods of implementation. To overcome this obstacle, a discrete method was used to quantize the size and position of the apertures in each cell. The general idea behind Lohmann's method is still used; size of each aperture controls amplitude, and phase is controlled by shifting the aperture. However, now the possible values of amplitude and phase are restricted (quantized). For example, imagine a Lohmann cell divided into 4x4 smaller squares. For a value of $c = 1/2$ in the Lohmann algorithm, which means that the width of the aperture is fixed at half the width of the entire cell, this cell permits three values of normalized amplitude (0,1/2, and 1) and three values of phase $(-p/2, 0, \text{and } p/2)$ for a total of seven possible combinations (an amplitude of zero has no phase). Values of phase greater than $p/2$, which create overlap between cells, were not allowed. Quantization means there will be error when coding amplitude and phase. Therefore, it makes sense to incorporate POCS and to design subholograms iteratively until a convergence condition is met. Simulations were performed for various cell quantizations.

Quantizing aperture positions makes sense when realization of the final CGH is considered. Current spatial light modulators (SLM) can be used in real time to control transmission or phase at each point. This is the same function that a binary amplitude hologram serves. Unfortunately, the resolution of a SLM cannot accurately re-create an exact Lohmann cell, thus quantization is needed to make realization practical. Similarly, technology used to make integrated circuits can be used for realizing CGH's as diffractive phase elements. Since precise, continuous surface relief is very difficult, this

technology requires quantization of phase. In general, when it comes to realization of a CGH, quantization is often preferred or even required. This is generally true for all advanced technologies of implementation.

### 6.4 Accounting for Approximations in Lohmann's Method with LM-ODIFIIT

It was mentioned in chapter 3 describing Lohmann's method that three approximations were made for simplicity: (a) $\text{sinc}\left[cdn(x+x_0)\right] \approx const$, (b) $\text{sinc}(yW_{nm}dn) \approx 1$, and (c) $\exp[2pi(xP_{nm}d_x)] \approx 1$. The affects of these approximations on the reconstructed image depend on several factors.

The sinc function $\text{sinc}\left[cdn(x+x_0)\right]$ creates a drop-off in intensity in the x-direction proportional to the distance from the center of the image plane. Approximation (a) considers this sinc factor to be nearly constant inside the image region. A small aperture size c results in less drop-off in intensity. However, this also reduces the brightness of the image.

The sinc function $\text{sinc}(yW_{nm}dn)$ indicates a drop-off in intensity similar to (a), but in the y-direction. This sinc function acts like a slight decrease in amplitude transmission by a factor $\text{sinc}(yW_{nm}dn)$. If one wants to reduce the effects of this approximation by sacrificing some brightness, every W could be reduced by a constant factor.

Last, the phase shift $\exp[2pi(xP_{nm}d_x)]$ causes a phase error that varies with x location in the image plane. This phase error depends on x and P, and ranges from zero to $\frac{p}{2M}$.

The solution used to account for the sinc roll-off in the x direction is to divide the desired image by $\text{sinc}\left[cdn(x+x_0)\right]$. The desired image f(x,y) becomes, $\frac{f(x,y)}{\text{sinc}\left[cdn(x+x_0)\right]}$. The same thing cannot be done for the y direction because $\text{sinc}(yW_{nm}dn)$ depends on the aperture parameters $W_{nm}$ which are determined based on the desired image. An iterative method in LM-ODIFIIT and quantized LM-ODIFIIT was

used to account for this sinc factor. The desired image is divided by the sinc factor affecting the x direction, and the hologram function is designed. Then, the y dependent sinc factors due to all apertures are calculated and summed to determine the effect on the output image. The desired image is now divided by the y dependent sinc factor just as it was for the x dependent factor. Next, the hologram is designed again, the effect on the output due to the new aperture heights is calculated, and the original desired image is divided by this new factor. This process is repeated until the reconstructed image does not change, or a convergence condition is met.

Lastly, by ignoring the phase factor $\exp[2pi(xP_{nm}d_x)]$, the output image is deteriorated as if there is a phase error due to an improper aperture shift [12]. In the quantized LM-ODIFIIT all of the apertures are shifted improperly due to quantization anyway. Neglecting this phase factor is accounted for by successive iterations while designing the hologram function.

### 6.5 Experiments and Simulations with LM-ODIFIIT Method

Table 1 shows mean square error (MSE) and efficiency data based on simulations. MSE represents the difference between the desired and reconstructed image inside the desired image region, and efficiency is a measure of how much of the incident wave is diffracted into the desired image region. The table includes data for ODIFIIT, Lohmann's method (LM), Lohmann's method using a constant amplitude for each cell (LMCA), ODIFIIT using Lohmann's coding method (LM-ODIFIIT), and LM-ODIFIIT with constant cell amplitude (LMCA-ODIFIIT). All results are for binary amplitude holograms. Values of $c = 1/2$ and $M = 1$ from Lohmann's method were used throughout. For comparison purposes, the MSE occurring from Lohmann's method was normalized to one, so all other MSE data is relative to Lohmann's method. Binary desired images

Table 6.5.1: Simulation data for binary amplitude computer-generated holograms

| **1.** | **Method** | **2.** | **MSE** | **3.** | **Efficiency** |
|---|---|---|---|---|---|
| LM | | 1 | | 1.2% | |
| ODIFIIT | | 0.33 | | 5.7% | |
| 4. | LMCA | 5. | 2.03 | 6. | 5.9% |
| 7. | LM-ODIFIIT | 8. | 5.9 • 10^-31 | 9. | 1.2% |
| 10. | LMCA-ODIFIIT | 11. | 1 | 12. | 2.1% |

Table 6.5.2: LM-ODIFIIT and LMCA-ODIFIIT data

| **13. Levels of Quantization (N)** | | **14.** | **MSE** | **15.** | **Efficiency** |
|---|---|---|---|---|---|
| 16. | 2 | 17. | 2.33 | 18. | 0.6% |
| 19. | 2-CA | 20. | 0.42 | 21. | 1.4% |
| 22. | 4 | 23. | 0.19 | 24. | 1.1% |
| 25. | 4-CA | 26. | 0.93 | 27. | 1.1% |

were used for the purposes of this experiment. Table 2 gives data from simulations where the amplitude and phase of each cell took on quantized values, and ODIFIIT was applied. Each Lohmann cell is divided into $NxN$ smaller squares. This is referred to as $N$-level quantization. Initiating quantization of amplitude and phase is done as mentioned in Section 6.3. MSE is still relative to the MSE from Lohmann's method alone.

Fig. 6.5.1 shows simulated LM-ODIFIIT reconstructions of the binary 'E' and gray-scale girl images. Fig. 6.5.2 displays optical results for LM-ODIFIIT. A comparison between the output of Lohmann's method, ODIFIIT, and LM-ODIFIIT for a 16x16 gray-scale image was also performed. The simulated and optical results are shown

in Fig. 6.5.3 and Fig. 6.5.4, respectively. The simulated and optical reconstructions from a CGH designed with LM-ODIFIIT and 4-level quantization are shown in Fig. 6.5.5.

## 6.6 Discussion

Looking at the simulation data of the two basic methods, ODIFIIT is better than Lohmann's method in terms of both MSE and efficiency. This supports the simulated results of these two methods. In Chapters 4 and 5, we saw that ODIFIIT produced a more uniform image with less "noise" inside the desired image region, and less reconstruction outside the desired image region when compared to Lohmann. However, the optical reconstructions did not support these predictions. Optically, Lohmann's method produced a sharper image than ODIFIIT, with about the same brightness. This indicates that the MSE for Lohmann should be lower than that of ODIFIIT, and that their efficiencies should be approximately the same. The difference in MSE is likely due to the fact that the MSE is only computed inside the desired image region. Although Lohmann produced a sharper image, the extra reconstruction near the image, which is considerably less for ODIFIIT, increases the MSE.

By altering Lohmann's method, applying the iterative Fourier transform algorithm, and using a constant amplitude in every cell, error in the image region increased, but efficiency was significantly improved. Basically, this means

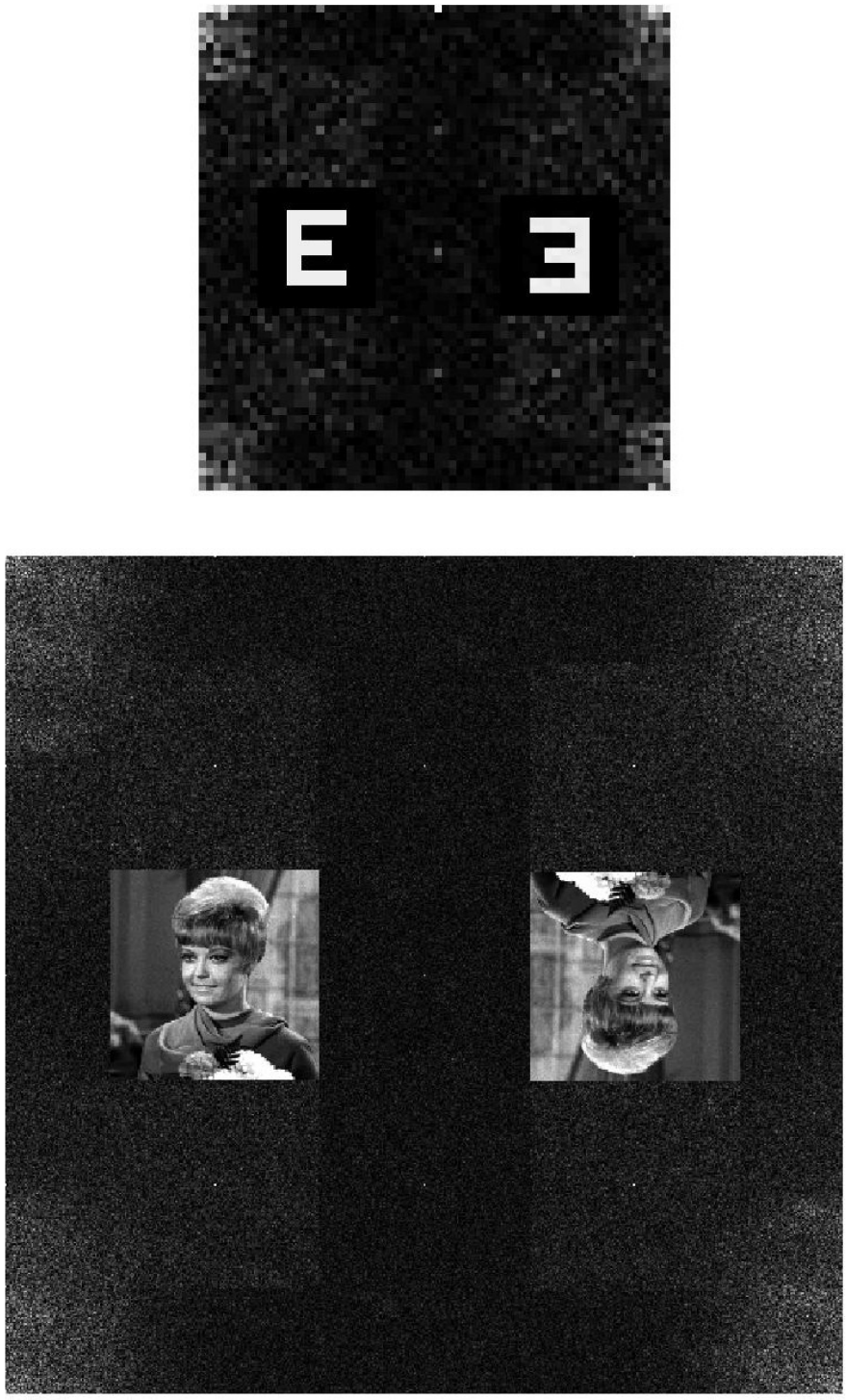


Fig. 6.5.1  Simulated reconstructions with LM-ODIFIIT.

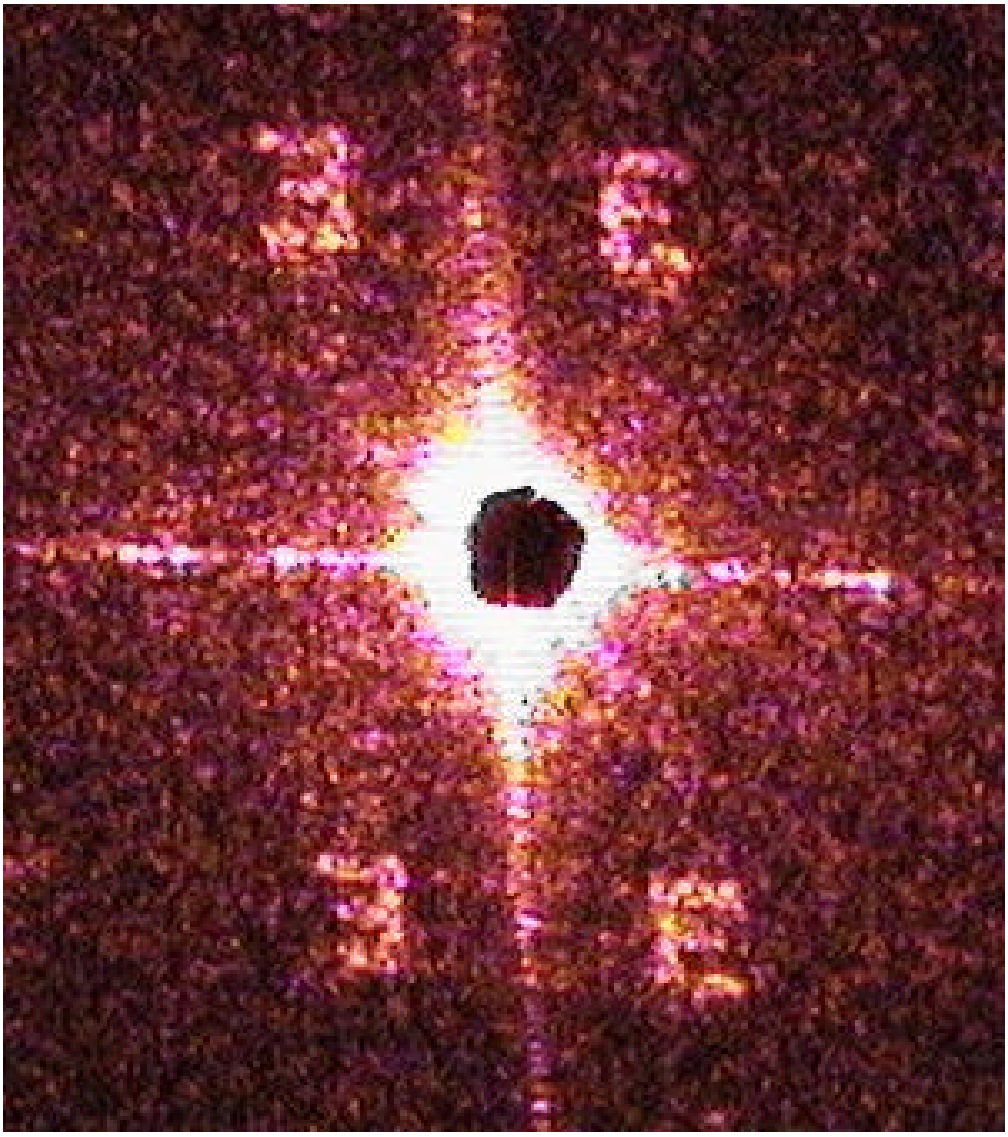
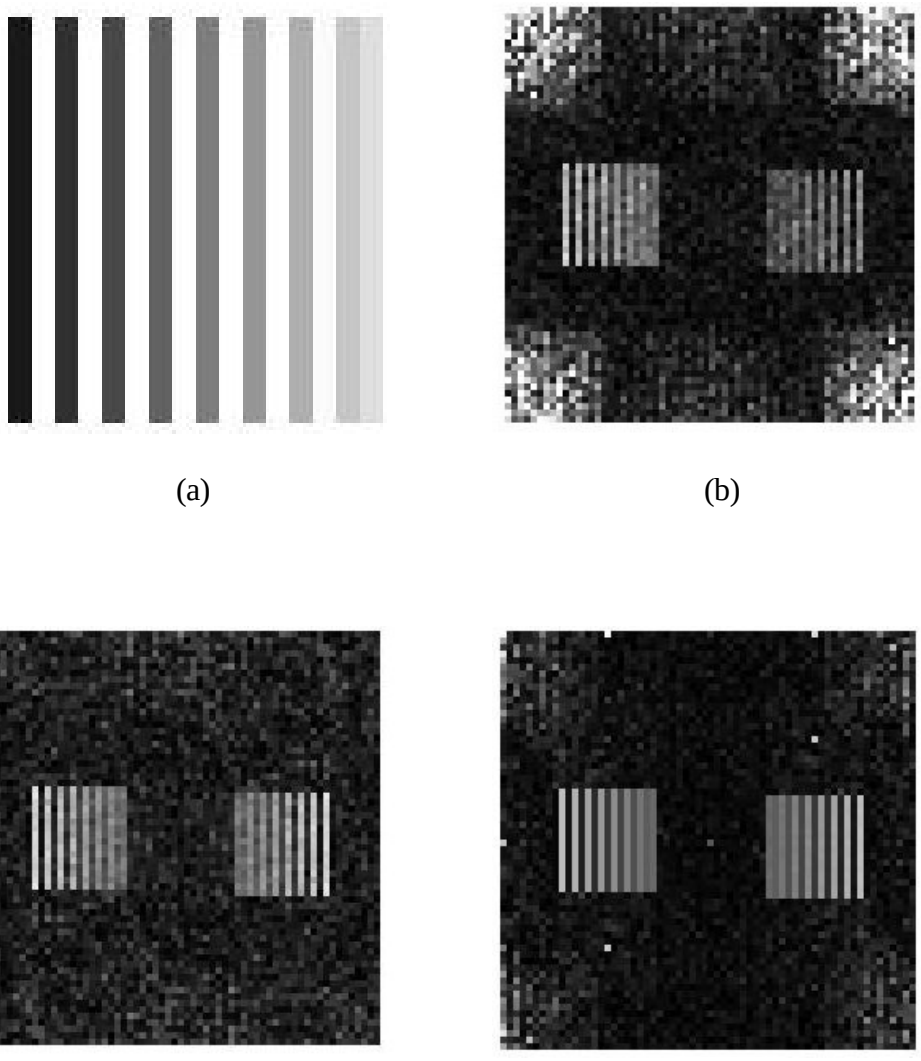


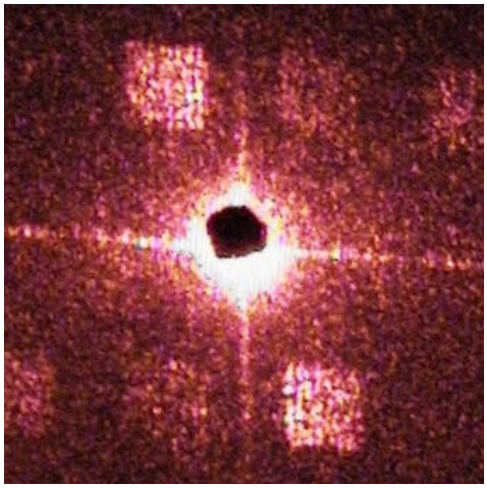
Fig. 6.5.2  Optical reconstruction from a LM



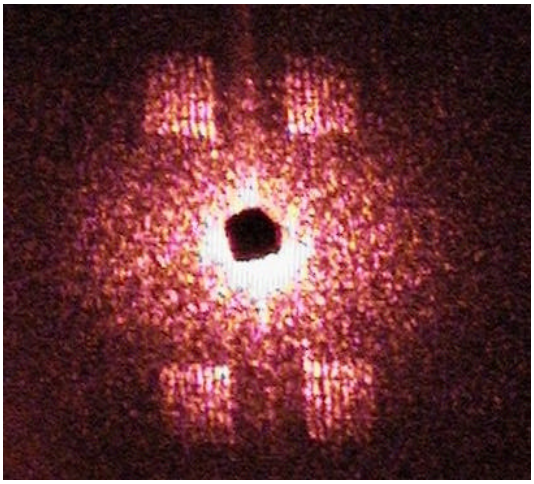
(a) (b)

(c) (d)

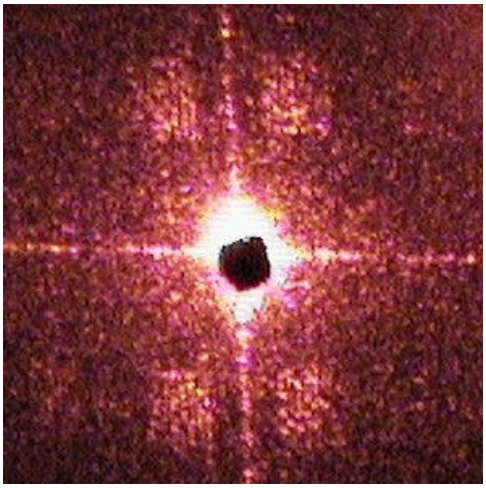
Fig. 6.5.3 Simulation results for a gray-scale images. a) Desired image b) Lohmann's method c) ODIFIIT d) LM-ODIFIIT.



(a)

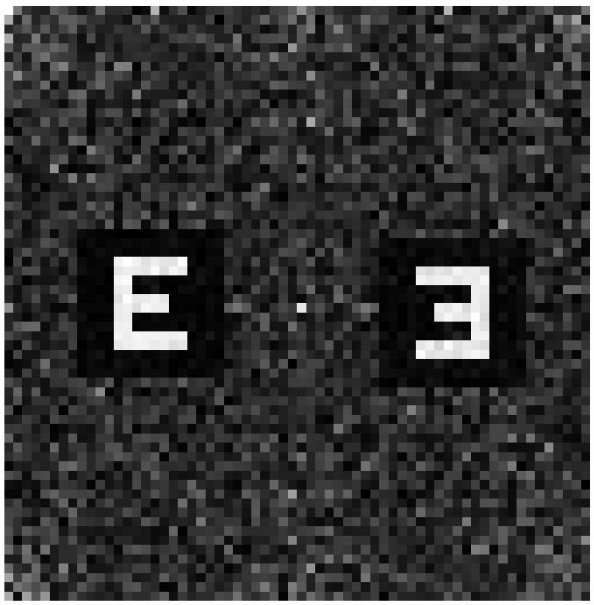


(b)

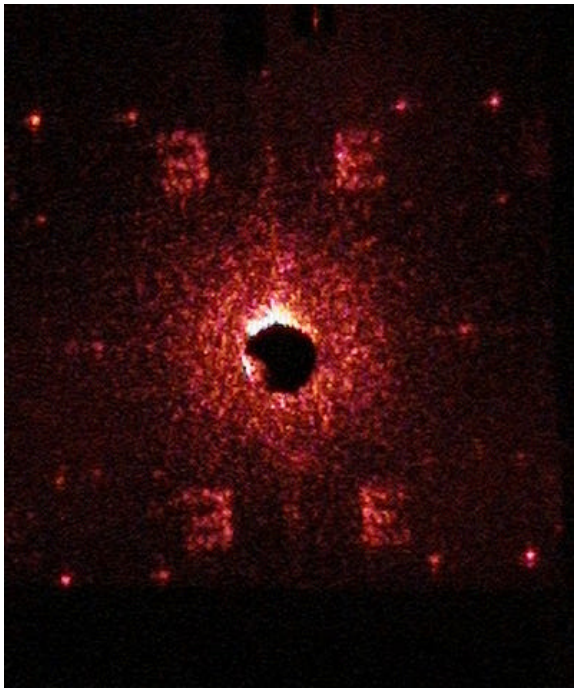


(c)

Fig. 6.5.4 Optical results for a gray



(a)



(b)

Fig. 6.5.5 Reconstructions using LM-ODIFIIT and 4-level quantization. a) Simulated b) Optical.

reconstruction of a brighter image with less definition. This result seems reasonable for the following reasons: less information is used in LMCA, which increases the error, but the amplitude values throughout the hologram are greater due to aperture size always being as large as possible, which increases the brightness. This quantitative analysis supports the visual results from Chapter 4. The output displays a reconstruction of the desired image which is brighter, but has "fuzzy" edges. Less reconstruction occurs outside the desired image region further confirming the efficiency increase. In fact, the efficiency of LMCA surpassed the efficiency of ODIFIIT, but the significant increase in MSE negates it as a superior method.

The data indicates that LM-ODIFIIT produces an output most resembling the desired image (lowest MSE), while its efficiency is the same as the original Lohmann method (LM). Simulations show that, qualitatively, LM-ODIFIIT generated an extremely sharp and uniform image inside the image region, while excess reconstruction outside the image region is limited.

Note how low the error from LM-ODIFIIT is when compared to all other methods. This indicates a reconstruction inside the image region which is far superior to any other method. This is supported by the visual results in Fig. 6.5.1. The simulated reconstruction of the girl image especially demonstrates the accuracy of LM-ODIFIIT, and should be compared to the LM and ODIFIIT results of the same image.

This dramatic increase in image quality for LM-ODIFIIT is not obvious in the optical reconstruction. The image in Fig. 6.5.2 shows a significant decrease in brightness from the LM and ODIFIIT results. Because of the large contrast in intensity, it is difficult to compare the quality of the image with the other two methods. At first, it appears that the LM-ODIFIIT image is the worst of the three. However, based on simulations, if the image were brighter, it is likely that the results would be similar to, or better than, the other methods.

To further test the legitimacy of the LM-ODIFIIT results, CGHs were designed to make the gray-scale image shown in Fig. 6.5.3a. LM-ODIFIIT shows a clear improvement over LM and ODIFIIT in the simulated results. Optical results followed the same trend as in the binary image reconstructions. The LM-ODIFIIT image is less

intense than the other two, and appears to be least accurate in representing the desired image.

The MSE for LMCA-ODIFIIT is equal to the MSE for LM, greater than that of LM-ODIFIIT, and less than the MSE of LMCA. We see that incorporating ODIFIIT into LMCA reduces the error, just as it did for LM. The increase from LM-ODIFIIT is expected because of the previously mentioned consequences of making the amplitude constant. However, due to the extremely small MSE of LM-ODIFIIT, it is somewhat surprising that the MSE of LMCA-ODIFIIT is not less. As expected, the efficiency of LMCA-ODIFIIT is greater than for LM-ODIFIIT.

Quantized LM-ODIFIIT with $N = 2$ had the largest MSE and the lowest efficiency of all the methods By using so few quantization levels, not enough information remains after encoding to allow for good reconstruction. Anytime the normalized amplitude is less than ½, the amplitude is quantized to zero and phase information is lost. This accounts for the high error and very poor efficiency. By implementing the constant amplitude technique, more phase information is retained. This results in better MSE and efficiency.

For $N = 4$, MSE dropped significantly, creating a very good reconstruction of the desired image. This method produced a low MSE, second only to the extremely low value of the original LM-ODIFIIT. Also, its efficiency is comparable to all other methods except ODIFIIT and LMCA. Making the amplitude a constant maintains the efficiency, but increases the error.

The optical reconstruction from LM-ODIFIIT with 4-level quantization is comparable to that of LM-ODIFIIT and has similar qualities to the simulated reconstruction.

All optical results so far have had 64x64 hologram and image planes. To improve reconstructions, the sampling rate was doubled in the image plane. Reconstructions resulting from the 128x128 holograms are shown in Fig. 6.6.1 and Fig. 6.6.2. The results indicate that reconstruction quality increases with the sampling rate as expected. The images are clearer and less reconstruction occurs in the center of the image plane.

The increase in image quality is most noticeable in the LM-ODIFIIT image. For this method, the noise in the center of the 128x128 images has dramatically decreased

from the 64x64 images, which helps to see the reconstructions better. But, the brightness of the other two methods is still greater. The 128x128 LM-ODIFIIT images appear to be as sharp as the other methods, but the lack of brightness still prevents a reliable conclusion. Based on improvement of the LM-ODIFIIT reconstruction with increased sampling, it is possible that with sufficient resolution and large enough sampling, the quality of its results could surpass those of the other methods.

Fig. 6.6.1 and Fig. 6.6.2 show that the 128x128 LM and LM-ODIFIIT images display periodic arrays of dots. The horizontal array in the LM image occurs because of the periodic nature of the hologram. This is due to all the cells being symmetric relative to the direction in which the aperture height is adjusted. The horizontal array in the LM-ODIFIIT image occurs for the same reason. However, in the LM-ODIFIIT hologram, it was observed that the phase at each point of the subholograms converged to the same value. This was observed for every subhologram except the first. This gives a periodic nature to the other direction of the hologram, which results in the vertical array of dots.

The experimental results presented in this thesis provide a starting point for more effective methods of CGH implementation. There are several aspects of the physical holograms used in this research which should be addressed when considering the accuracy of the results. The quantification of aperture positions due to finite resolution has already been mentioned. Two other topics are the resolution of the aperture edges and the actual transmission at each point in the hologram. Lohmann's method represents the edge of each aperture as a perfect step function. In reality, the physical holograms likely have a finite drop-off at the edge of the aperture due to finite resolution. Also, the print used was not perfectly opaque. Therefore, points where zero transmission is desired are actually slightly transparent. These are both reasons for error in the experimental reconstructions.

Up to now, all simulations neglected the sinc roll-off that actually occurs in the physical reconstruction from a binary CGH. The accommodating methods mentioned in Section 6.4 were used in simulations, which included this sinc roll-off behavior. This basically comes down to the ability to design a hologram of a gray-scale image. Simulated reconstructions showed excellent homogeneity in the x direction indicating that the x dependent sinc factor was accurately accounted for. Slight variations were

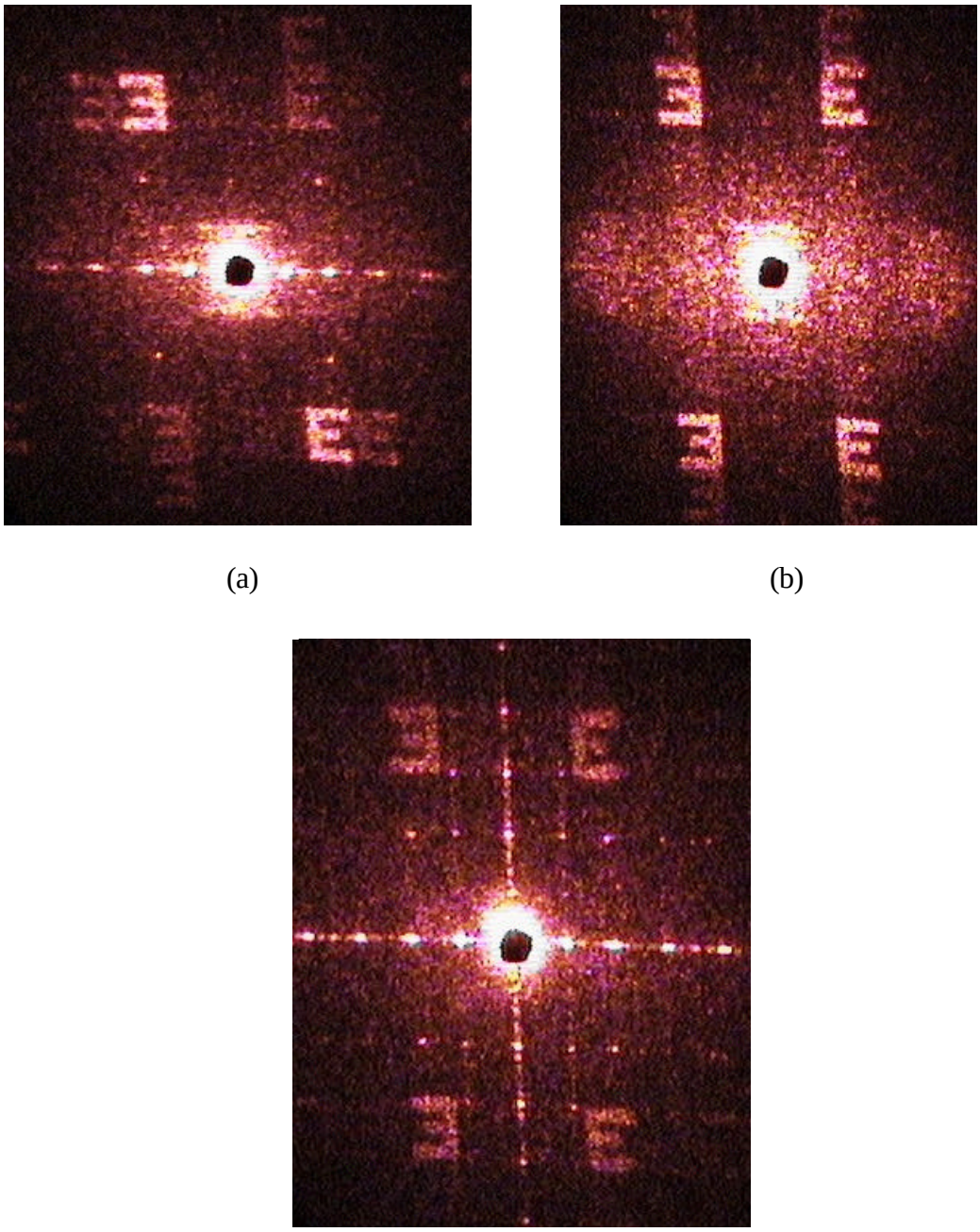
present in the y direction, but can be reduced to a desired level by continued iterations. Fig. 6.6.3 shows the simulated reconstructed gray-scale image and then the final image when the sinc roll-offs are factored into the output.

## 6.7 Conclusions

By incorporating ODIFIIT into the principles of the coding scheme used in Lohmann holograms, mean square error in the simulated reconstructed image was reduced. Interlacing and iterative interlacing techniques were effective in reducing reconstruction error. Overall, LM-ODIFIIT produced the best MSE, by far, with a MSE on the order of $10^{-31}$, relative to the MSE of Lohmann's method. LMCA had the highest efficiency.

For situations where quantization is required, a quantized Lohmann's method incorporating ODIFIIT was created. The best quantized technique was LM-ODIFIIT with $N = 4$, which decreased the MSE by 80% from LM while maintaining about the same efficiency.

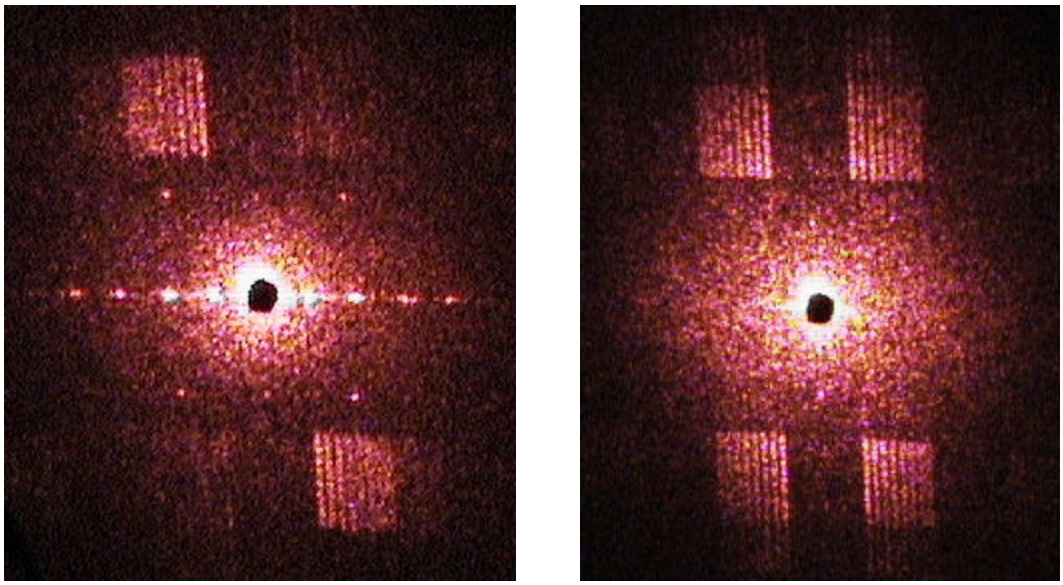
Optical reconstructions using LM-ODIFIIT did not support the simulated results. Images produced by LM-ODIFIIT were not as bright as the Lohmann and ODIFIIT images. The contrast in brightness made it difficult to compare the quality of the LM-ODIFIIT image with that of the other two methods.



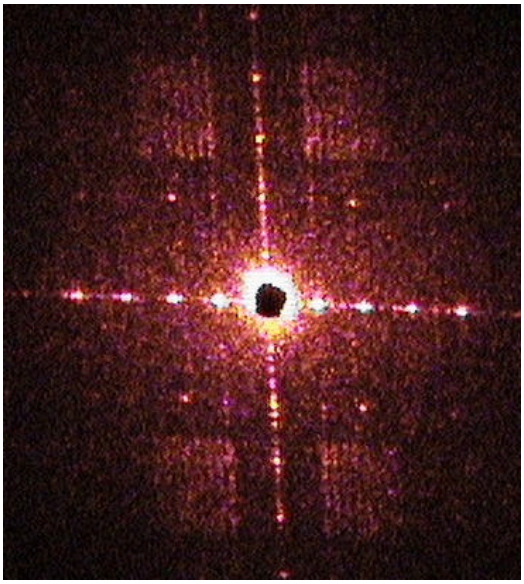
(a) (b)

(c)

Fig. 6.6.1 Optical reconstruction of 128x128 E image. a) Lohmann b) ODIFIIT c) LM-ODIFIIT.

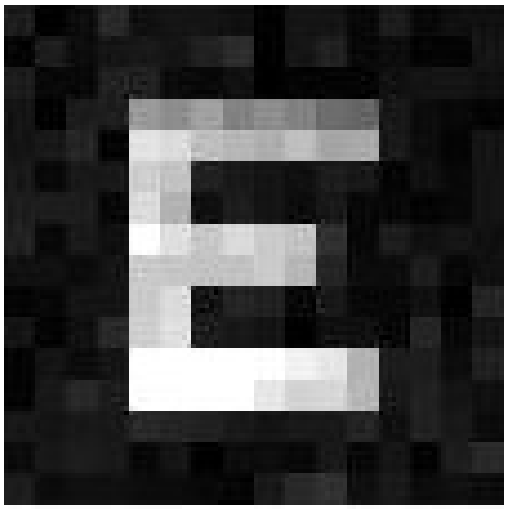


(a) (b)

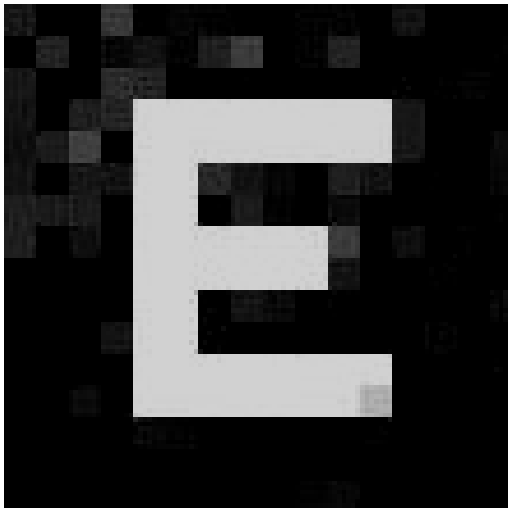


(c)

Fig. 6.6.2 Optical reconstruction of 128x128 gray image.
a) Lohmann b) ODIFIIT c) LM-ODIFIIT.



(a)



(b)

Fig. 6.6.3 a) Gray-scale image from LM-ODIFIIT hologram and b) Image after approximations are factored into the output.

# 7 IMPLEMENTATION OF DIFFRACTIVE OPTICAL ELEMENTS

## 7.1 Introduction

In general, lithography is the technology used to print patterns onto a substrate [14]. The most common use is printing the patterns of integrated circuits onto a semiconductor substrate. Advances in lithography techniques have allow for the production of precise, small-scale patterns. There exist various forms of lithography. These include optical lithography, e-beam lithography, EUVL (extreme ultraviolet lithography), and x-ray lithography [14]. There are two basic steps involved in lithography. The first step is to create a desired pattern. The second is to transfer this pattern into the substrate. The quality of a lithographic method is measured primarily by its resolution. Resolution is the minimum feature size that can be obtained. Resolution is not just determined by the lithography method, but also by the ability of the resist to reconstruct the pattern.

For producing unique DOEs, the popular lithography technique is e-beam lithography. E-beam lithography requires an intense, yet uniform, electron source with a small spot size, high stability, and long life. The main attributes of e-beam lithography technology are 1) it is capable of very high resolution; 2) it work with a variety of materials and patterns; 3) it is slow compared to optical lithography; and 4) it is expensive and complicated. Electron beam lithography tools can cost many millions of dollars and require frequent service to stay properly maintained.

Electron beam lithography is an effective and appropriate method for producing diffractive optical elements (DOEs). Combined with reactive ion etching (RIE), creation of DOEs with very fine resolution and high efficiency is possible. A binary phase hologram consists of a 2-dimensional array of points, and at each point a phase shift of either 0 or 180 degrees is induced. The phase shift is invoked by introducing a third dimension. At points where a 180 degree phase shift is desired, a hole is etched such that a wave reflected off a hole is 180 degrees out of phase with a wave of the same wavelength reflected from a point with no hole. E-beam lithography allows such holes to be made with great accuracy. The E-beam exposes a pattern onto an e-beam sensitive

resist indicating where holes are to be made, and RIE is used to "dig" the holes to the desired depth.

The JEOL JSM6400 scanning electron microscope (SEM) [15] at Purdue University has been modified to give it e-beam lithography capabilities. The JEOL machine is capable of generating a beam with a spot size around 10nm. Therefore, it is capable of exposing patterns with lines of widths less than $0.1\ mm$ or 100nm. For binary phase holograms, each square hole has a size of approximately $1\ mm$ x $1\ mm$. The JEOL is easily capable of exposing features with these dimensions. For feature sizes around one square micron, a total pattern size of 1mm x 1mm can be achieved, while maintaining high resolution. Patterns much greater than one square millimeter risk losing resolution around the edges of the pattern due to increased beam deflection at points along the edge.

The Semi Group 1000TP RIE System at Purdue University is a dry etch system that enables very fine surface relief. The 180 degree phase change needed for binary phase holograms occurs over a distance of half a wavelength. For visible light, that corresponds to a length on the order of half a micron. RIE makes it possible to induce surface relief of such small dimensions with very high accuracy.

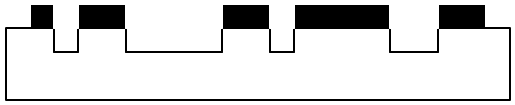
### 7.2 DOE Production Procedure

The basic procedure for producing a binary phase hologram using e-beam lithography is illustrated in Fig. 7.2.1. First, a resist is spin-coated onto a silicon substrate. Next, E-beam lithography is performed to write the desired pattern into the resist. The sample is then developed to remove the exposed resist. Once the exposed resist is removed, reactive ion etching is used to transfer the pattern into the substrate. Finally, the resist is removed. This process can be repeated to create multi-level DOEs.

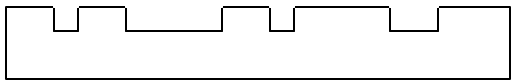
Since reactive ion etching will eventually be done on the sample, a high-percent resist spun to a relatively large thickness is used so that the RIE does not etch through the

1) Spin on e-beam resist

2) Expose resist with e-beam and develop



3) Reactive Ion Etching



4) Remove remaining resist

Fig. 7.2.1 Procedure to create a diffractive optical element.

resist. A realistic combination is a 9% 950K PMMA resist spun at 6000 rpm for 30 sec [16]. This produces a resist thickness of approximately 1micron. Next, e-beam lithography is used to write the desired pattern onto the resist. The first step in using the SEM for lithography purposes is to produce a good image using it as a microscope. This involves properly aligning and focusing the electron beam. Proper alignment of the SEM's electron gun is achieved when the beam of electrons is perpendicular to the surface of the sample. Optimum focusing of the beam occurs when the beam diameter is

minimized at the surface of the sample. The astigmatism of the beam and aperture centering are associated with the beam's focus. The aperture must not interfere with the e-beam. Since lithography requires the smallest aperture setting, positioning of the aperture is very important. Astigmatism of the beam should also be minimized, as it is the primary limiting factor in lithography performance. Proper adjustment of stigmation ensures that the beam is symmetrical at the sample surface. Thus, there is no variation in the resolution of vertical and horizontal lines. Both alignment and focusing are important in maximizing the machine's capabilities as a microscope and for lithography. In addition, lithography specifically requires an acceleration voltage between 35-40 kV and a working distance of 15mm between the electron gun and the sample surface [17]. Once a sample is prepared, mounted, and loaded into the SEM, and the beam is aligned and focused, lithography is done automatically by an NPGS system connected to the SEM.

Once the pattern is written, the exposed resist is removed by developing the sample with a combination of MIBK and IPA and then rinsing with just IPA. Next, reactive ion etching is performed to transfer the exposed pattern into the silicon substrate. Finally, the PMMA resist is striped off using acetone, leaving just the silicon substrate with the desired pattern etched into it. This process can be repeated to create multi-level diffractive elements.

### 7.2 Nanometer Pattern Generation System

The NPGS (Nanometer Pattern Generation System) package [17] and beam blanker incorporated into the JEOL system makes lithography possible. NPGS automatically controls the beam blanker and deflection of the E-beam during exposure. In order to run, NPGS requires an ~.RF2 file (known as a run file) of the desired exposure pattern. The run file is produced by first creating the desired exposure pattern with a CAD package called DesignCAD. Usually a pattern is simply drawn using DesignCAD just like one would create a pattern in any CAD software, such as AutoCAD. However, a binary hologram of size 512 x 512, for example, contains approximately 260,000 points. This is unreasonable to draw into DesignCAD by hand. Therefore, code was written to convert the binary hologram array into a DesignCAD file. DesignCAD generates a ~.DC2 file, which is just a text file representing a pattern drawn into

DesignCAD. Therefore, the code writes a ~.DC2 file, which represents each point in the hologram array as a 1 square micron box in DesignCAD.

Next, the ~.DC2 file is converted into a run file using the Run File Editor option in DesignCAD. The run file editor allows the user to adjust several exposure parameters. These include the center-to-center and line-to-line spacing of the e-beam. The center-to-center spacing determines the distance between consecutive exposure points, while line-to-line spacing indicates the e-beam step size as it moves from left to right. The exposure dose, which is the amount of electrons the sample gets exposed to, can also be varied. Dose can be specified in terms of current per length, current per area, or a point dose. The appropriate magnitude of the exposure dose depends on several factors, such as the type and thickness of the resist being used as well as the dimensions of the pattern being exposed. Experimenting with different values for these parameters allows the user to optimize the e-beam's capabilities for his or her particular needs. Once the JEOL machine is properly adjusted (aperture centered, astigmatism of the beam minimized, and magnification set appropriately) and the run file is processed, the pattern is exposed automatically by the NPGS system.

### 7.4 Conclusions

This chapter has discussed how semiconductor technology can be used to produce diffractive phase elements. As the minimum feature size capable of being produced by these methods decreases, the design of diffractive elements will have to change. Since scalar diffraction theory does not hold for aperture sizes less than a wavelength, and the near field problem is quite different from the far field problem, the Fourier transform based method of designing DOEs will have to be replaced. The next chapter discusses methods of analyzing diffraction in the near field and for small aperture sizes.

# 8 NEAR FIELD DIFFRACTION FROM SMALL-SCALE DIFFRACTIVE OPTICAL ELEMENTS

## 8.1 Introduction

Scalar diffraction theory relies on the assumptions that the size of the diffractive aperture is large relative to the wavelength of the incident wave. Characterization of diffraction is also quite different in the near field versus the far field. With the increasing trend to produce smaller and smaller devices where refractive optical components become impractical, there is a demand for diffractive elements with sizes and characteristics on the order of a wavelength or smaller. However, in this size range, scalar diffraction theory breaks down and analytical results are not possible. For near-field problems, one method is the angular spectrum of plane waves, which uses Fourier transform and propagation of plane waves. Another alternative is to solve Maxwell's equations directly using numerical methods. XFDTD software by Remcom uses a finite difference time domain method to solve Maxwell's equations. This allows simulation of diffractive elements with sizes and characteristics (such as focal length) that are too small for scalar diffraction theory to accurately predict. Simple elements like Fresnel zone plates, reflective gratings, and transmission gratings combined with zone plates were analyzed, as reported in this chapter.

## 8.2 The Angular Spectrum

One way to analyze how a wave behaves as it propagates is to look at its angular spectrum [1]. Consider a wave field $U(x, y, z)$ propagating in the z-direction. The wavelength is $l$ so that $k = 2p / l$. Letting z = 0 initially, the two dimensional Fourier representation of $U(x, y,0)$ is given by

$$U(x, y,0) = \iint_{-\infty}^{\infty} A(f_x, f_y,0) e^{j2p(f_x x + f_y y)} df_x df_y \tag{8.2.1}$$

where

$$A(f_x, f_y,0) = \iint_{-\infty}^{\infty} U(x, y,0) e^{-j2p(f_x x + f_y y)} dxdy \tag{8.2.2}$$

By including time variation, $A(f_x, f_y, 0)e^{j2p(f_x x + f_y y - ft)}$ represents a plane wave at z = 0. The direction of propagation is given by the direction cosines, $\cos a_x = 2pf_x$ and $\cos a_y = 2pf_y$. Therefore, $A(f_x, f_y, 0)$ is called the angular spectrum of $U(x, y, 0)$.

Similarly, the angular spectrum at any z position of a wave field $U(x, y, z)$ is

$$A(f_x, f_y, z) = \iint_{-\infty}^{\infty} U(x, y, z)e^{-j2p(f_x x + f_y y)}dxdy. \tag{8.2.3}$$

So, in terms of its angular spectrum,

$$U(x, y, z) = \iint_{-\infty}^{\infty} A(f_x, f_y, z)e^{j2p(f_x x + f_y y)}df_x df_y \tag{8.2.4}$$

Now, $U(x, y, z)$ must satisfy the Helmholtz equation at all points in a source-free region.

$$\nabla^2 U(x, y, z) + k^2 U(x, y, z) = 0 \tag{8.2.5}$$

Substituting (8.2.4) into (8.2.5),

$$\iint_{-\infty}^{\infty} \left\{ \frac{d^2}{dz^2} A(f_x, f_y, z) + \left[k^2 - 4p^2(f_x^2 + f_y^2)\right] A(f_x, f_y, z) \right\} df_x df_y = 0 \tag{8.2.6}$$

If this is to be true for all waves, the integrand must be zero.

$$\frac{d^2}{dz^2} A(f_x, f_y, z) + \left[k^2 - 4p^2(f_x^2 + f_y^2)\right] A(f_x, f_y, z) = 0 \tag{8.2.7}$$

The solution to the differential equation in (8.2.7) is

$$A(f_x, f_y, z) = A(f_x, f_y, 0)e^{jmz} \tag{8.2.8}$$

where

$$m = \sqrt{k^2 - 4p^2(f_x^2 + f_y^2)} \tag{8.2.9}$$

So, each angular spectrum component at $z = 0$ is multiplied by the phase factor $e^{jmz}$ to give the new angular spectrum.

When $k^2 > 4p^2(f_x^2 + f_y^2)$, $m$ is real, and the wave propagates freely. Plane waves with this property are known as homogeneous waves. When $k^2 < 4p^2(f_x^2 + f_y^2)$, $m$ is imaginary, and

$$A(f_x, f_y, z) = A(f_x, f_y, 0)e^{-kz} \tag{8.2.10}$$

where $m = jk$ . Therefore, the amplitude of the plane wave component is attenuated in the z-direction. Waves attenuated in the z-direction are called evanescent waves.

Given $A(f_x, f_y, z)$ in terms of $A(f_x, f_y, 0)$, substituting (8.2.8) into (8.2.4) yields

$$U(x, y, z) = \iint_{-\infty}^{\infty} A(f_x, f_y, 0)e^{jz\sqrt{k^2 - 4p^2(f_x^2 + f_y^2)}} e^{j2p(f_x x + f_y y)} df_x df_y \tag{8.2.11}$$

So, if $U(x, y, 0)$ is known, $U(x, y, z)$ can be determined by the following procedure:

1) Take the Fourier transform of $U(x, y, 0)$ to determine $A(f_x, f_y, 0)$.
2) Multiply $A(f_x, f_y, 0)$ by $e^{jmz}$ where $m$ is given in (8.2.9) to get $A(f_x, f_y, z)$.
3) Take the inverse Fourier transform of $A(f_x, f_y, z)$ to determine $U(x, y, z)$.

The angular spectrum method can be implemented in the computer using FFT to compute the diffracted field from a DOE. First, the field at the DOE $U(x, y, 0)$ is sampled in increments of $\Delta x$ and $\Delta y$ so that $x = m\Delta x$ and $y = n\Delta y$. The initial field is represented by $U(m, n, 0)$. The angular spectrum at z=0 is computed by taking the FFT of the sampled field at the diffractive element. So, $A(k, l, 0) = FFT[U(m, n, 0)]$, where $f_x = k\Delta f_x$ and $f_y = l\Delta f_y$. The angular spectrum at z=0 is then multiplied by the appropriate phase factor as in (8.2.8) to get the angular spectrum at the new z position. Therefore, $A(k, l, z) = A(k, l, 0)e^{jmz}$, where $m = \sqrt{(2p/l)^2 - 4p^2[(k\Delta f_x)^2 + (l\Delta f_y)^2]}$. Finally, the inverse Fast-Fourier transform of the result is computed to determine the new sampled field at z; $U(m, n, z) = IFFT[A(k, l, z)]$.

### 8.3 Maxwell's Equations

Maxwell's equations connect the electric and magnetic fields in a set of partial differential equations which successfully predict electromagnetic phenomena. Maxwell's equations need to be solved to completely describe diffractive effects when scalar diffraction theory fails.

For a source-free region (no electric or magnetic current sources) which may contain material capable of absorbing electric or magnetic field energy, Maxwell's equations are given by [9]:

$$\frac{\partial \vec{B}}{\partial t} = -\nabla \times \vec{E} - \vec{J}_m \tag{8.3.1}$$

$$\frac{\partial \vec{D}}{\partial t} = \nabla \times \vec{H} - \vec{J}_e \tag{8.3.2}$$

$$\nabla \cdot \vec{D} = 0 \tag{8.3.3}$$

$$\nabla \cdot \vec{B} = 0 \tag{8.3.4}$$

$\vec{E}$ is electric field, $\vec{D}$ is the electric flux density, $\vec{H}$ is magnetic field, $\vec{B}$ is magnetic flux density, $\vec{J}_e$ is the electric conduction current density, and $\vec{J}_m$ is the equivalent magnetic conduction current density used to account for magnetic loss. In linear, isotropic nondispersive media, $\vec{B}$ is related to $\vec{H}$ and $\vec{D}$ to $\vec{E}$ by a scalar:

$$\vec{B} = m\vec{H} \tag{8.3.5}$$

$$\vec{D} = e\vec{E} \tag{8.3.6}$$

where $m$ is magnetic permeability and $e$ is electric permittivity. Similarly,

$$\vec{J}_e = s\vec{E} \tag{8.3.7}$$

$$\vec{J}_m = r'\vec{H} \tag{8.3.8}$$

where $s$ is electric conductivity and $r'$ is an equivalent magnetic resistivity. Substituting (8.3.5)-(8.3.8) into (8.3.1) and (8.3.2) yields

$$\frac{\partial \vec{H}}{\partial t} = -\frac{1}{m}\nabla \times \vec{E} - \frac{r'}{m}\vec{H} \tag{8.3.9}$$

$$\frac{\partial \vec{E}}{\partial t} = \frac{1}{e}\nabla \times \vec{H} - \frac{s}{e}\vec{E} \tag{8.3.10}$$

The vector components of the curl operators in (8.3.9) and (8.3.10) result in six coupled scalar equations equivalent to Maxwell's curl equations in three dimensions [9]:

$$\frac{\partial H_x}{\partial t} = \frac{1}{m}\left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y} - r'H_x\right) \tag{8.3.11a}$$

$$\frac{\partial H_y}{\partial t} = \frac{1}{m}\left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} - r' H_y\right) \tag{8.3.11b}$$

$$\frac{\partial H_z}{\partial t} = \frac{1}{m}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x} - r' H_z\right) \tag{8.3.11c}$$

$$\frac{\partial E_x}{\partial t} = \frac{1}{e}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - s E_x\right) \tag{8.3.12a}$$

$$\frac{\partial E_y}{\partial t} = \frac{1}{e}\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - s E_y\right) \tag{8.3.12b}$$

$$\frac{\partial Ez}{\partial t} = \frac{1}{e}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - s E_z\right) \tag{8.3.12c}$$

These coupled partial differential equations are the foundation of FDTD algorithms for describing the interaction of electromagnetic waves with arbitrary three dimensional objects. Theoretically, (8.3.3) and (8.3.4) for source free regions are a direct consequence of the curl equations. Therefore, they do not need to be explicitly enforced. However, the space grid used in the FDTD must be formed to uphold the relations.

## 8.4 Finite Differences

The Taylor's series expansion of $u(x,t_n)$ about the point $x_i$ to $x_i + \Delta x$ for a fixed time $t_n$ is given by [9]

$$\begin{aligned} u(x_i + \Delta x, t_n)\Big|_{t_n} = u\Big|_{x_i,t_n} &+ \Delta x \frac{\partial u}{\partial x}\Big|_{x_i,t_n} + \frac{\Delta x^2}{2}\frac{\partial^2 u}{\partial x^2}\Big|_{x_i,t_n} \\ &+ \frac{\Delta x^3}{6}\frac{\partial^3 u}{\partial x^3}\Big|_{x_i,t_n} + \frac{\Delta x^4}{24}\frac{\partial^4 u}{\partial x^4}\Big|_{x_i,t_n} \end{aligned} \tag{8.4.1}$$

The last term is an error term, where $x_1$ is a point in the interval $(x_i, x_i + \Delta x)$. Similarly, an expansion to the point $x_i - \Delta x$ for fixed time $t_n$ is given by

$$\begin{aligned} u(x_i - \Delta x, t_n)\Big|_{t_n} = u\Big|_{x_i,t_n} &- \Delta x \frac{\partial u}{\partial x}\Big|_{x_i,t_n} + \frac{\Delta x^2}{2}\frac{\partial^2 u}{\partial x^2}\Big|_{x_i,t_n} \\ &- \frac{\Delta x^3}{6}\frac{\partial^3 u}{\partial x^3}\Big|_{x_i,t_n} + \frac{\Delta x^4}{24}\frac{\partial^4 u}{\partial x^4}\Big|_{x_2,t_n} \end{aligned} \tag{8.4.2}$$

where $x_2$ is in the interval $(x_i - \Delta x, x_i)$. Adding the two expansions gives

$$u(x_i + \Delta x) + u(x_i - \Delta x) = 2u\Big|_{x_i,t_n} + \Delta x^2 \frac{\partial^2 u}{\partial x^2}\Big|_{x_i,t_n} + \frac{\Delta x^4}{12}\frac{\partial^4 u}{\partial x^4}\Big|_{x_3,t_n} \tag{8.4.3}$$

where $x_3$ lies in the interval $(x_i - \Delta x, x_i + \Delta x)$. Rearranging the above expression, we get

$$\frac{\partial^2 u}{\partial x^2}\Big|_{x_i,t_n} = \left[\frac{u(x_i + \Delta x) - 2u(x_i) + u(x_i - \Delta x)}{(\Delta x)^2}\right]_{t_n} + O\left[(\Delta x)^2\right] \tag{8.4.4}$$

This is a second-order accurate, central-difference approximation to the second partial derivative of u. The remainder term $O\left[(\Delta x)^2\right]$ goes to zero as the square of the space increment. Using shorthand notation, (8.4.4) takes the form

$$\frac{\partial^2 u}{\partial x^2}\Big|_{x_i,t_n} = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2} + O\left[(\Delta x)^2\right] \tag{8.4.5}$$

Following a similar procedure, the second partial derivative of u with respect to time is given by

$$\frac{\partial^2 u}{\partial t^2}\Big|_{x_i,t_n} = \frac{u_i^{n+1} - 2u_i^n + u_i^{n-1}}{(\Delta t)^2} + O\left[(\Delta t)^2\right] \tag{8.4.6}$$

## 8.5 The Yee Algorithm

Yee's algorithm [18] solves Maxwell's curl equations from a set of finite-difference equations. The original set of equations introduced by Yee in 1966 assumed lossless mateials $r' = 0$ and $s = 0$, but show the fundamentals of FDTD algorithms. Each electric field component in three dimensional space is surrounded by four circulating magnetic field components, and every magnetic field component is surrounded by four circulating electric field components (Fig. 8.5.1). The finite difference expressions for the space derivatives are central in nature and second order accurate [9]. The field component locations in the Yee grid and the central difference operations on the field components ensure the two Gauss' Law relations are upheld.

Electric and magnetic field components are also centered in time in a leapfrog arrangement. This means all the electric components are computed and stored for a

particular time using previously stored magnetic component data. Then, all magnetic data is determined using the electric field data just computed.

For a uniform, rectangular lattice, a point in space is represented by three components $(i, j, k) = (i\Delta x, j\Delta y, k\Delta z)$ where $\Delta x, \Delta y,$ and $\Delta z$ are the space increments in the x, y, and z directions respectively. i, j, and k are integers. So, a function $u(\vec{r}, t)$ evaluated at a discrete space point at a discrete point in time is $u(i\Delta x, j\Delta y, k\Delta z, n\Delta t) = u^n_{i,j,k}$ where $\Delta t$ is the time increment and n is an integer. Now, Yee's centered finite-difference expression for the first partial space derivative of u in the x-direction, evaluated at time $t_n = n\Delta t$ is given by [9]:

$$\frac{\partial u}{\partial x}(i\Delta x, j\Delta y, k\Delta z, n\Delta t) = \frac{u^n_{i+1/2,j,k} - u^n_{i-1/2,j,k}}{\Delta x} + O\left[(\Delta x)^2\right] \tag{8.5.1}$$

Notice that Yee uses data only a distance $\Delta x/2$ away from the point in question as opposed to $\Delta x$. Similarly, the first partial derivative of u with respect to time for a particular space point is given by

$$\frac{\partial u}{\partial t}(i\Delta x, j\Delta y, k\Delta z, n\Delta t) = \frac{u^{n+1/2}_{i,j,k} - u^{n-1/2}_{i,j,k}}{\Delta t} + O\left[(\Delta t)^2\right] \tag{8.5.2}$$

These difference equations are now applied to Maxwell's curl equations. Consider (8.3.12a)

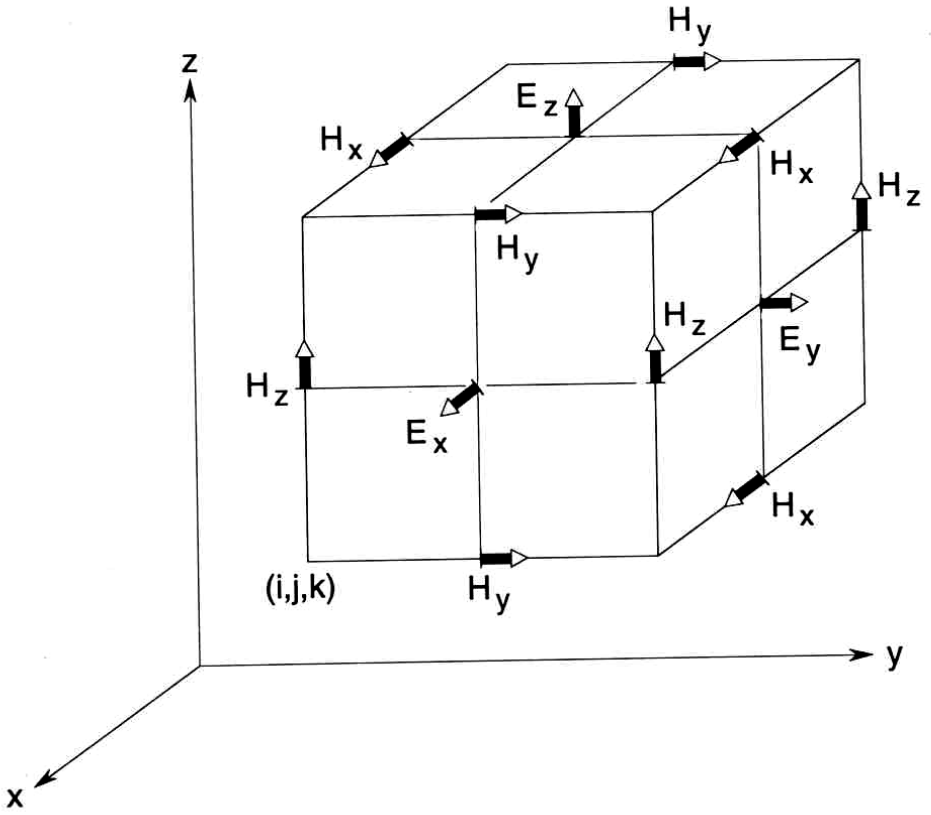



Fig. 8.5.1 Electric and magnetic field vectors in a cell of the Yee mesh [9].

$$\frac{\partial E_x}{\partial t} = \frac{1}{e}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - s E_x\right) \tag{8.3.12a}$$

Substituting in the space and time derivatives for point $(i, j, k)$ at time $n$ yields

$$\frac{E_x\Big|_{i,j,k}^{n+1/2} + E_x\Big|_{i,j,k}^{n-1/2}}{\Delta t} = \frac{1}{e_{i,j,k}}\left(\begin{array}{l}\dfrac{H_z\Big|_{i,j+1/2,k}^{n} - H_z\Big|_{i,j-1/2,k}^{n}}{\Delta y} - \\ \dfrac{H_y\Big|_{i,j,k+1/2}^{n} - H_y\Big|_{i,j,k-1/2}^{n}}{\Delta z} - s_{i,j,k} E_x\Big|_{i,j,k}^{n}\end{array}\right) \tag{8.5.3a}$$

Now, all terms on the right side of the equation are evaluated at time step n. Remember, all the magnetic field components at time n have already been determined and stored, but

$E_x$ at time n has not. Only values of $E_x$ up to time $n-1/2$ are stored. For a region without loss, this term is zero. If $s_{i,j,k}$ is non-zero, then $E_x$ can be estimated by a semi-implicit approximation [9]:

$$E_x\Big|_{i,j,k}^{n} = \frac{E_x\Big|_{i,j,k}^{n+1/2} + E_x\Big|_{i,j,k}^{n-1/2}}{2} \tag{8.5.3b}$$

which is the average of the known value of $E_x$ at time $n-1/2$ and the unknown value at time $n+1/2$. Using (8.3.21) and multiplying both sides by $\Delta t$, (8.3.20) becomes

$$E_x\Big|_{i,j,k}^{n+1/2} + E_x\Big|_{i,j,k}^{n-1/2} = \frac{\Delta t}{e_{i,j,k}}\left(\begin{array}{l}\dfrac{H_z\Big|_{i,j+1/2,k}^{n} - H_z\Big|_{i,j-1/2,k}^{n}}{\Delta y} - \dfrac{H_y\Big|_{i,j,k+1/2}^{n} - H_y\Big|_{i,j,k-1/2}^{n}}{\Delta z} \\ - s_{i,j,k}\dfrac{E_x\Big|_{i,j,k}^{n+1/2} + E_x\Big|_{i,j,k}^{n-1/2}}{2}\end{array}\right) \tag{8.5.3c}$$

Collecting like terms and solving for $E_x\Big|_{i,j,k}^{n+1/2}$ yields the final expression:

$$E_x\Big|_{i,j,k}^{n+1/2} = \left(\frac{1-\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}{1+\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}\right)E_x\Big|_{i,j,k}^{n-1/2} + \left(\frac{\dfrac{\Delta t}{e_{i,j,k}}}{1+\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}\right)\left(\begin{array}{l}\dfrac{H_z\Big|_{i,j+1/2,k}^{n} - H_z\Big|_{i,j-1/2,k}^{n}}{\Delta y} \\ - \dfrac{H_y\Big|_{i,j,k+1/2}^{n} - H_y\Big|_{i,j,k-1/2}^{n}}{\Delta z}\end{array}\right) \tag{8.5.4a}$$

It follows that [9]

$$E_y\Big|_{i,j,k}^{n+1/2} = \left(\frac{1-\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}{1+\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}\right)E_y\Big|_{i,j,k}^{n-1/2} + \left(\frac{\dfrac{\Delta t}{e_{i,j,k}}}{1+\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}\right)\left(\begin{array}{l}\dfrac{H_x\Big|_{i,j,k+1/2}^{n} - H_x\Big|_{i,j,k-1/2}^{n}}{\Delta z} \\ - \dfrac{H_z\Big|_{i+1/2,j,k}^{n} - H_z\Big|_{i-1/2,j,k}^{n}}{\Delta x}\end{array}\right) \tag{8.5.4b}$$

$$E_z\Big|_{i,j,k}^{n+1/2} = \left(\frac{1-\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}{1+\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}\right)E_z\Big|_{i,j,k}^{n-1/2} + \left(\frac{\dfrac{\Delta t}{e_{i,j,k}}}{1+\dfrac{s_{i,j,k}\Delta t}{2e_{i,j,k}}}\right)\left(\begin{array}{l}\dfrac{H_y\Big|_{i+1/2,j,k}^{n} - H_y\Big|_{i-1/2,j,k}^{n}}{\Delta x} \\ - \dfrac{H_x\Big|_{i,j+1/2,k}^{n} - H_x\Big|_{i,j-1/2,k}^{n}}{\Delta y}\end{array}\right) \tag{8.5.4c}$$

$$H_x\Big|_{i,j,k}^{n+1} = \left(\frac{1-\dfrac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}{1+\dfrac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}\right)H_x\Big|_{i,j,k}^{n} + \left(\frac{\dfrac{\Delta t}{m_{i,j,k}}}{1+\dfrac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}\right)\left(\begin{array}{l}\dfrac{E_y\Big|_{i,j,k+1/2}^{n+1/2} - E_y\Big|_{i,j,k-1/2}^{n+1/2}}{\Delta z} \\ - \dfrac{E_z\Big|_{i,j+1/2,k}^{n+1/2} - E_z\Big|_{i,j-1/2,k}^{n+1/2}}{\Delta y}\end{array}\right) \tag{8.5.5a}$$

$$H_y\Big|_{i,j,k}^{n+1} = \left(\frac{1-\frac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}{1+\frac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}\right) H_y\Big|_{i,j,k}^{n} + \left(\frac{\frac{\Delta t}{m_{i,j,k}}}{1+\frac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}\right)\left(\frac{\frac{E_z\Big|_{i+1/2,j,k}^{n+1/2} - E_z\Big|_{i-1/2,j,k}^{n+1/2}}{\Delta x}}{-\frac{E_x\Big|_{i,j,k+1/2}^{n+1/2} - E_x\Big|_{i,j,k-1/2}^{n+1/2}}{\Delta z}}\right) \quad (8.5.5b)$$

$$H_z\Big|_{i,j,k}^{n+1} = \left(\frac{1-\frac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}{1+\frac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}\right) H_z\Big|_{i,j,k}^{n} + \left(\frac{\frac{\Delta t}{m_{i,j,k}}}{1+\frac{r'_{i,j,k}\Delta t}{2m_{i,j,k}}}\right)\left(\frac{\frac{E_x\Big|_{i,j+1/2,k}^{n+1/2} - E_x\Big|_{i,j-1/2,k}^{n+1/2}}{\Delta y}}{-\frac{E_y\Big|_{i+1/2,j,k}^{n+1/2} - E_y\Big|_{i-1/2,j,k}^{n+1/2}}{\Delta x}}\right) \quad (8.5.5c)$$

## 8.6 XFDTD

The software package XFDTD made by Remcom uses a leapfrog, finite difference, time domain algorithm to solve Maxwell's curl equations. XFDTD allows the user to design a structure and visualize results in a graphical interface. Geometry design, project setup, and result viewing can all be done utilizing pull-down menus.

The region of interest is a cubical mesh, where each mesh edge is given different material properties in order to simulate a specified geometry. For each cell, the user may choose the material to be a perfect conductor or free space, or may define the exact physical properties of the material. The sampling in space is sub-wavelength, typically in the range of 1/10 to 1/30 of a wavelength. The region of interest is excited by either a plane wave or multiple voltage sources. The excitation may be pulsed or sinusoidal.

After the geometry and excitation are specified, the duration of the simulation is set by specifying the number of desired time steps. Time sampling is chosen to ensure numerical stability of the algorithm [9].

When the modeled region extends to infinity, absorbing boundary conditions (ABCs) are used at the boundary of the grid. This allows all outgoing waves to leave the region with negligible reflection. The region of interest can also be enclosed by a perfect electrical conductor such as in the case of a waveguide.

Once all these parameters are set, XFDTD calculates the magnetic field at the surface of every cell inside the three dimensional region of interest for the first time step. From these magnetic field components, all electric field components can be calculated and saved. Time is stepped forward by one increment and magnetic field is computed

again for the new time. The electric field is then determined from the newly calculated magnetic field. This procedure continues until the final time step is reached.

Once the fields are calculated for the specified number of time steps, near zone transient and steady state fields can be seen as color intensity images, or a field component at a specific point can be plotted versus time. When performing simulations where the steady-state output is desired, observing a specific point over time helps to indicate whether a steady-state had been reached.

## 8.7 More Diffractive Optical Elements

### 8.7.1 Fresnel zone plates

Fresnel zone plates are diffractive elements, which can often serve the same purpose as a lens. Fig. 8.7.1 shows the geometry of a circular aperture from the side. The path difference between a ray traveling along SOP and SAP is

$$\Delta(r) = (r_0 + r) - (z_0 + z) = \sqrt{r^2 + z_0^2} + \sqrt{r^2 + z^2} - (z_0 + z) \qquad (8.7.1.1)$$

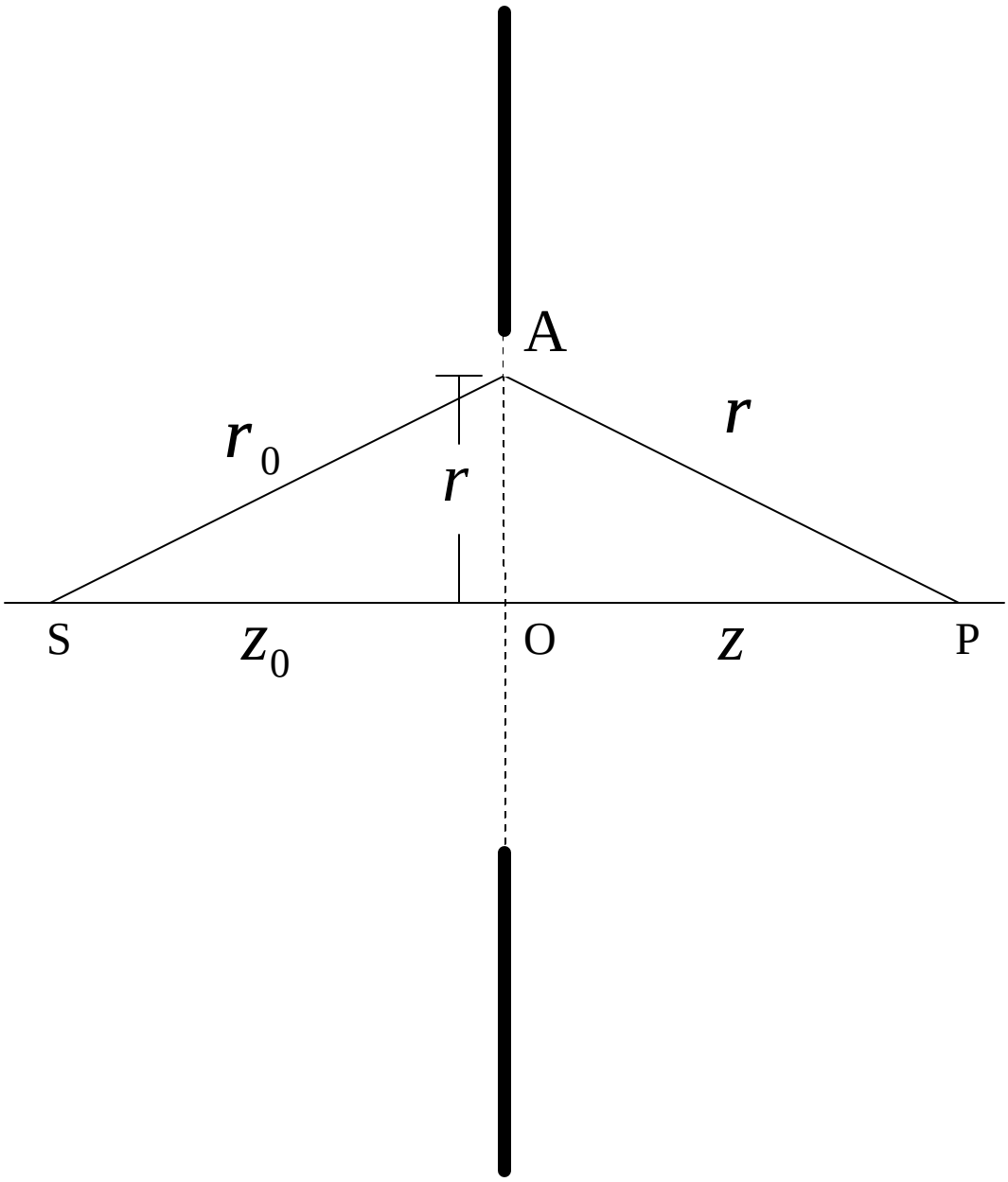



Fig. 8.5.1 Electric and magnetic field vectors in a cell of the Yee mesh [9].

The Fresnel-zone parameter, $n$, is defined such that the path difference is an integer multiple of half wavelengths [19]:

$$n\frac{l}{2} = \Delta(r) \tag{8.7.1.2}$$

The area between $r_n$ and $r_{n-1}$ is called the nth Fresnel zone. Since the field at P coming from the radius $r_n$ is half a wavelength out of phase with the field from radius $r_{n\pm1}$, if the field coming from each zone has equal amplitude, then adjacent zones serve to cancel each other out. Therefore, the total field at P will increase if either all even or all odd zones are blocked out, creating a focal point at distance z. When this is done, the resulting structure is called a Fresnel zone plate (FZP).

If the zone plate is illuminated by a plane wave, $z_0 \approx \infty$, then solving (8.7.1.1) for r yields,

$$\frac{r_n}{l} = \sqrt{\frac{f_0}{l}\left(n - \frac{1}{2}\right) + \frac{(2n-1)^2}{16}} \tag{8.7.1.3}$$

where z is a constant $f_0$ called the focal point. In this case, each zone is "centered" around the radii found if (8.7.1.2) is substituted into (8.7.1.1). There is still a $\frac{l}{2}$ path length difference across each zone, but the edges occur where there is a $\pm\frac{l}{4}$ path length difference from the $\Delta = n\frac{l}{2}$ points. (8.7.1.3) is found by using the following relation in place of (8.7.1.2)

$$\frac{2n-1}{4}l = \Delta(r) \tag{8.7.1.4}$$

Similarly, the focal length can be expressed as a function of $n$, implying that there are several focal points. Indeed, there exist points of increased intensity at $f_0/3$, $f_0/5$, $f_0/7$, and so on [19]. This is one main difference between a FZP and a conventional refractive lens.

### 8.7.2 Diffraction gratings

A diffraction grating is a repetitive array of diffracting elements which creates periodic modulations in the phase, amplitude, or both of an incoming wave [19]. These diffracting elements may be apertures or obstacles. An array of alternating opaque and transparent regions is called a transmission amplitude grating. If the entire grating is transparent, but varies periodically in optical thickness, it is called a transmission phase grating. A reflective material with periodic surface relief also produces distinct phase relationships upon reflection of a wave. These are known as reflection phase gratings.

Fig. 8.7.2 illustrates a transmission amplitude grating. For a wave normally incident upon the grating, the path difference between parallel, transmitted waves is $d \sin q_m$. Now, if this path difference is an integer multiple of wavelengths, then the transmitted waves interfere constructively. Otherwise, there is destructive interference, and the transmitted waves are canceled. Therefore, the grating equation for normal incidence is

$$d \sin q_m = m l \tag{8.7.2.1}$$

This equation holds for all the types of gratings previously mentioned. For a given grating period, $d$, equation (8.7.2.1) shows that constructive interference may happen at several angles corresponding to different values of $m$, referred to as modes.

### 8.7.3 Fourier transform of a periodic grating

Consider a periodic amplitude grating with period $d$ and opening size $x_0$. For an incident wave of constant amplitude A across the grating, the field at the output of the grating can be represented by the Fourier series [20]:

$$U_1(x_1) = \sum_{n=-\infty}^{\infty} c_n e^{j 2pnx_1 / d} \tag{8.7.3.1}$$

where

$$c_n = \frac{A}{pn} \sin\left( \frac{npx_0}{d} \right) \tag{8.7.3.2}$$

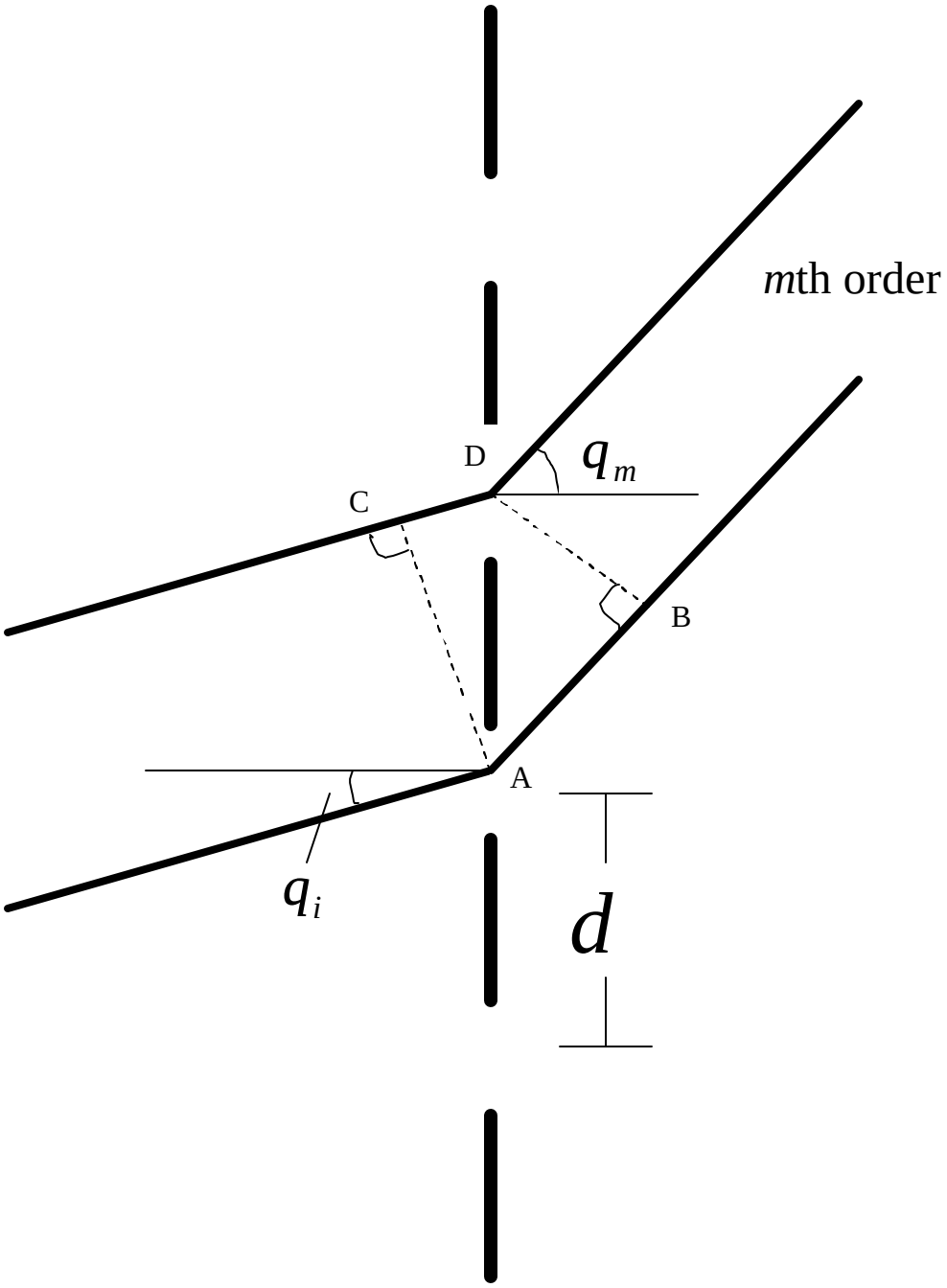



Fig. 8.7.2 Periodic amplitude grating.

Now, suppose a lens of focal length $f_1$ is used to produce the Fourier transform of the output field at the grating. The Fraunhofer diffraction pattern is given by

$$U_2(x_2) = C\int_{-\infty}^{\infty}\sum_n c_n e^{jx_1(2pn/d + kx_2/f_1)}dx_1 \qquad (8.7.3.3)$$

where C is a complex factor. Using the fact that

$$d(w) = \frac{1}{2p}\int_{-\infty}^{\infty} e^{jwx}dx \qquad (8.7.3.4)$$

The Fourier transform of the grating's output becomes [20]

$$U_2(x_2) = 2pC\sum_n c_n d\left(\frac{kx_2}{f_1} + \frac{2pn}{d}\right) \qquad (8.7.3.5)$$

Analyzing this last expression, the intensity pattern at the focal plane of a lens resulting from a grating of period $d$ is a series of evenly spaced lines. The spacing is $\frac{lf_1}{d}$. Thus the separation is proportional to the focal length and inversely proportional to the grating period relative to wavelength. It should also be noted that $c_n$ decreases with increasing $n$, which means the intensity of the lines gets smaller the farther they are from the axis.

### 8.8 Angular Spectrum and XFDTD Results

All XFDTD simulations used a cell size of $l/20$. The excitation was a y-polarized sinusoidal plane wave propagating in the z-direction. All diffracting structures are made of perfect electrical conductors. All edges of the diffracting structures are parallel to the x-axis to avoid canceling the y-polarized electric field.

Using XFDTD, a one dimensional FZP (Fig. 8.8.1b) with a thickness of $0.1l$ and focal length $3l$ was simulated. Its output was analyzed for the first three modes. The mode corresponds to the number of even or odd zones that are blocked. The intensity along the axis passing through the center of the plate is plotted as a function of distance from the plate. This is seen in Fig. 8.8.2. The plot shows that the intensity peak near $3l$ behind the plate gets higher and narrower as the mode increases. The peak also gets closer to the desired focal length of $3l$ for higher modes. We can also look at the cross

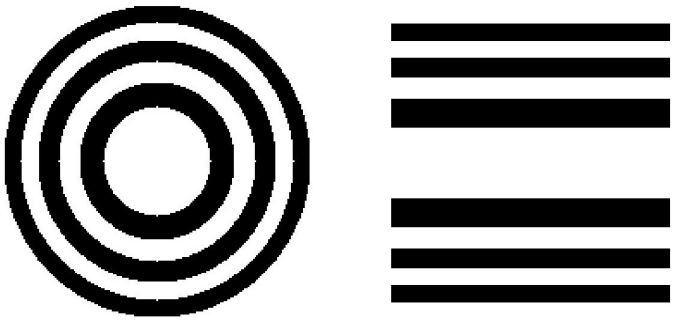


Fig. 8.8.1 a) 2

section of the intensity along the plane of maximum intensity. The intensity in the focal plane is plotted in Fig. 8.8.3. The plot suggests that the spot size decreases with increasing mode and that side lobe intensity is reduced for higher modes. It should be noted that intensity at $f_0/3$, where an additional focal point is expected, increases with the mode. However, the increase could not be described as a peak.

The same FZP was simulated using the angular spectrum method. A plane wave incident on the zone plate is represented by a sampled version of the field at the output of the plate. The field is zero inside every blocked zone, and given a value of 1 between blocked zones. The polarization of the wave is not specified in angular spectrum computations because it is based on scalar diffraction theory. All diffractive elements investigated with the angular spectrum method were represented in this way. The main difference from the XFDTD simulation is that the angular spectrum method does not allow for the width of the zone plate to be specified. The intensities along the axis of propagation and in the focal plane are plotted in Fig. 8.8.4 and Fig. 8.8.5, respectively.

The angular spectrum results again show a higher and narrower peak for higher order modes. They also predict lower side lobes in the focal plane as the mode increases.

Two characteristics stand out between the different methods. The angular spectrum method shows the build up of a second focal point at $f_0/3$ as the mode increases. This peak is not as pronounced in the XFDTD results. Secondly, there is a distinctly different intensity drop-off past the main focal point.

Next, using XFDTD FZPs of order three were made with thicknesses of $0.2l$ and $0.4l$. The intensity in along the axis and in the focal plane is plotted in Fig. 8.8.6 and Fig. 8.8.7. Now, there is a slight buildup of intensity at $f_0/3$ as the thickness of the plate increases. Another notable feature is that the focal point gets closer to $3l$ as the thickness increases. There was no significant change in the focal plane as the thickness was varied.

Fig. 8.8.8 shows the XFDTD output for a FZP of mode three and thickness $0.4l$. Angular spectrum results could not be performed to simulate varying plate thickness.

Reflective gratings with periods of $l$ and $1.5l$ were designed in XFDTD. For a period of $l$, only the zero order mode exists. Therefore, the reflected wave should form a standing wave pattern with the normally incident incoming wave. Fig. 8.8.9 shows the

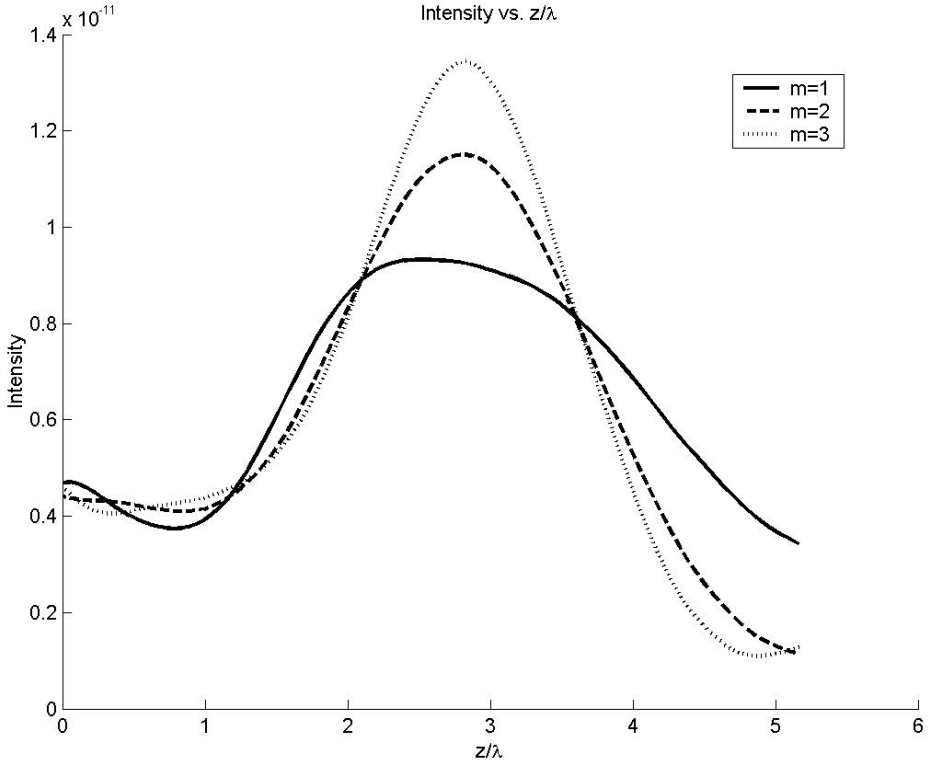



Fig. 8.8.2 XFDTD results for the intensity along the axis passing through the center of a FZP for modes 1

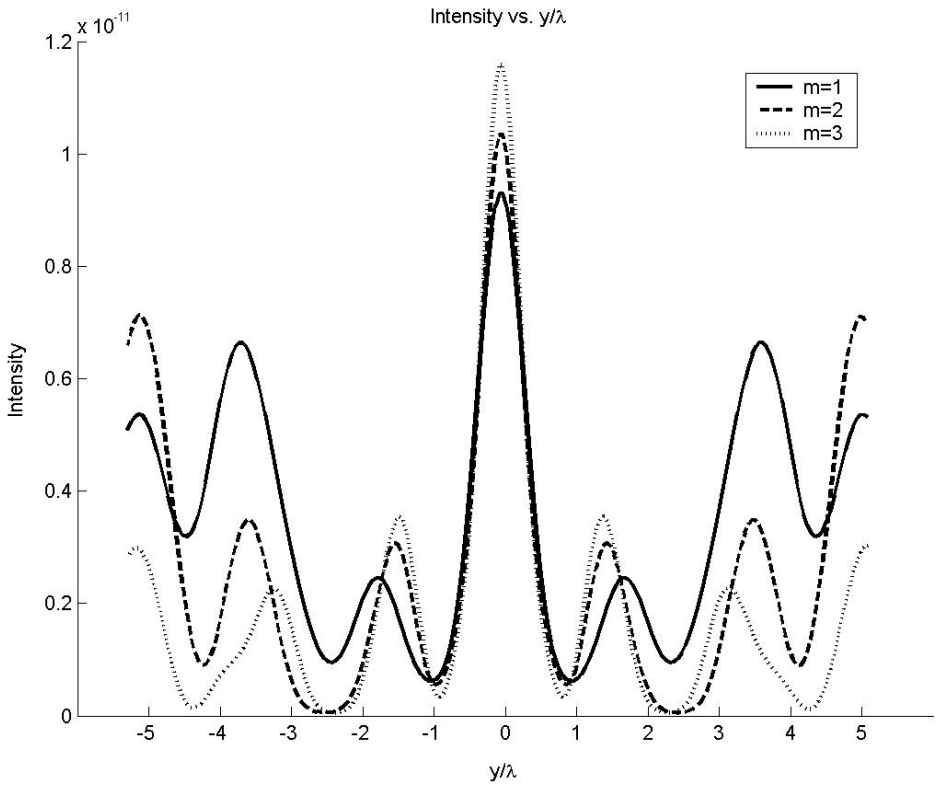



Fig. 8.8.3 XFDTD results for the intensity in the focal plane of a FZP for modes 1

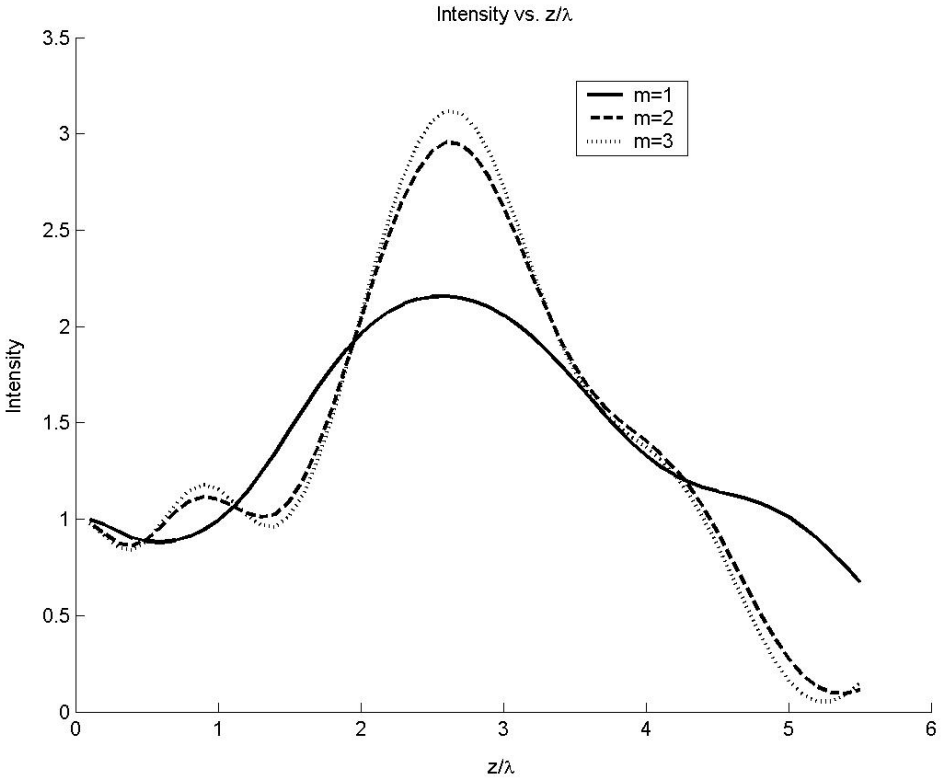



Fig. 8.8.4 Angular spectrum results for the intensity along the axis of propagation for FZPs of modes 1

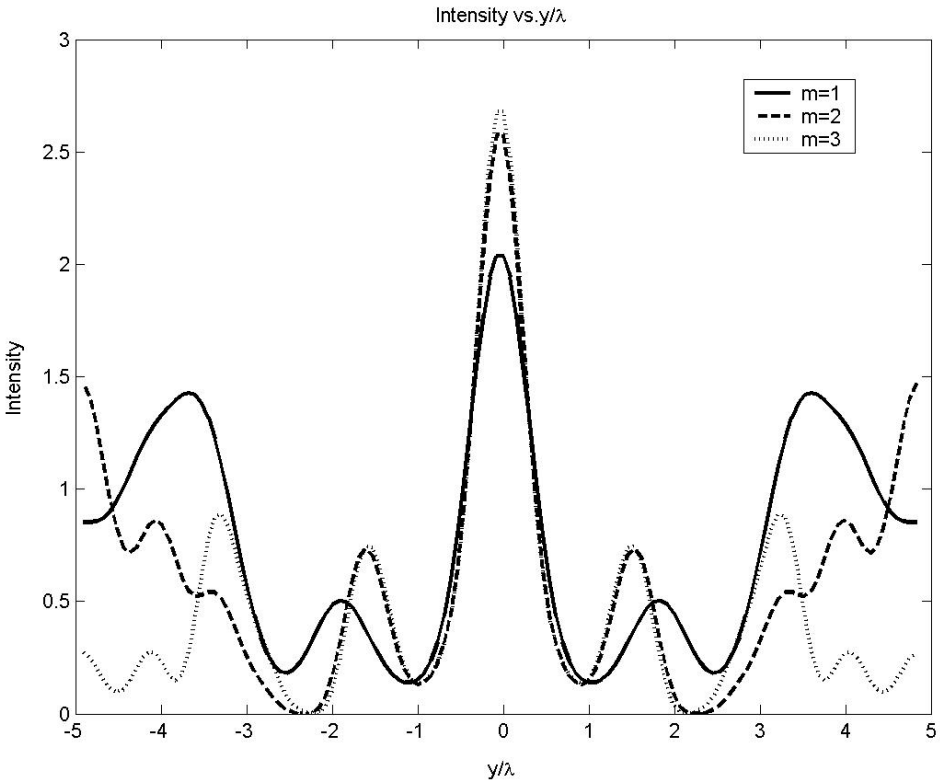



Fig. 8.8.5 Angular spectrum results for the intensity in the focal plane of a FZP for modes 1

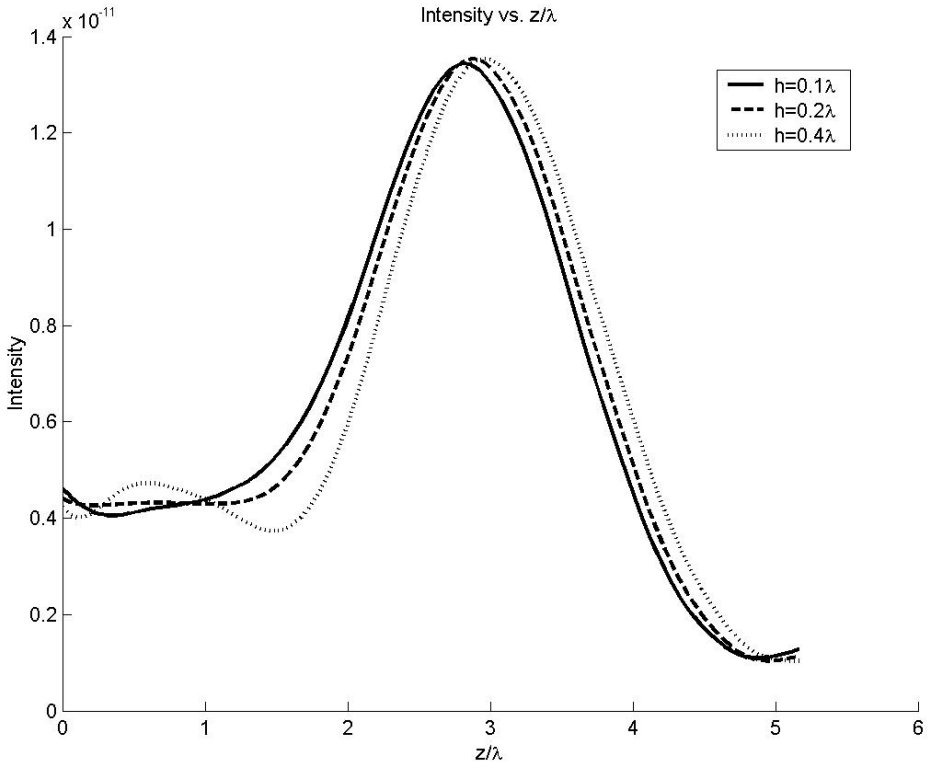



Fig. 8.8.6 Intensity plot from XFDTD results for third order FZPs of varying thickness. Focal length is 3 $l$ .

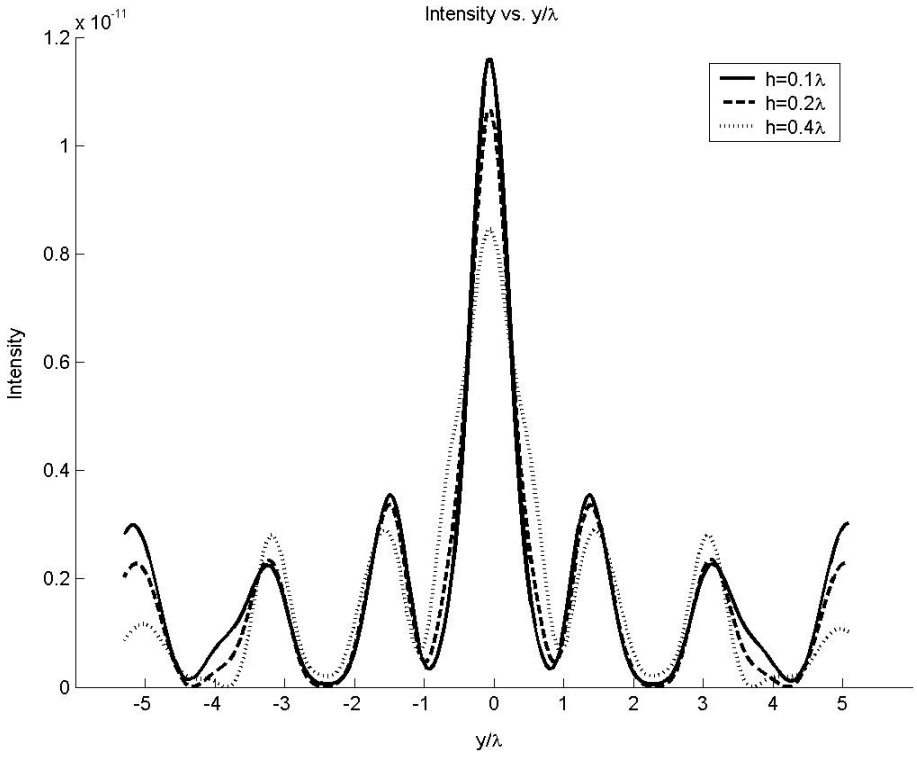



Fig. 8.8.7 Intensity in the focal plane for various FZP thicknesses ($m$ =3).

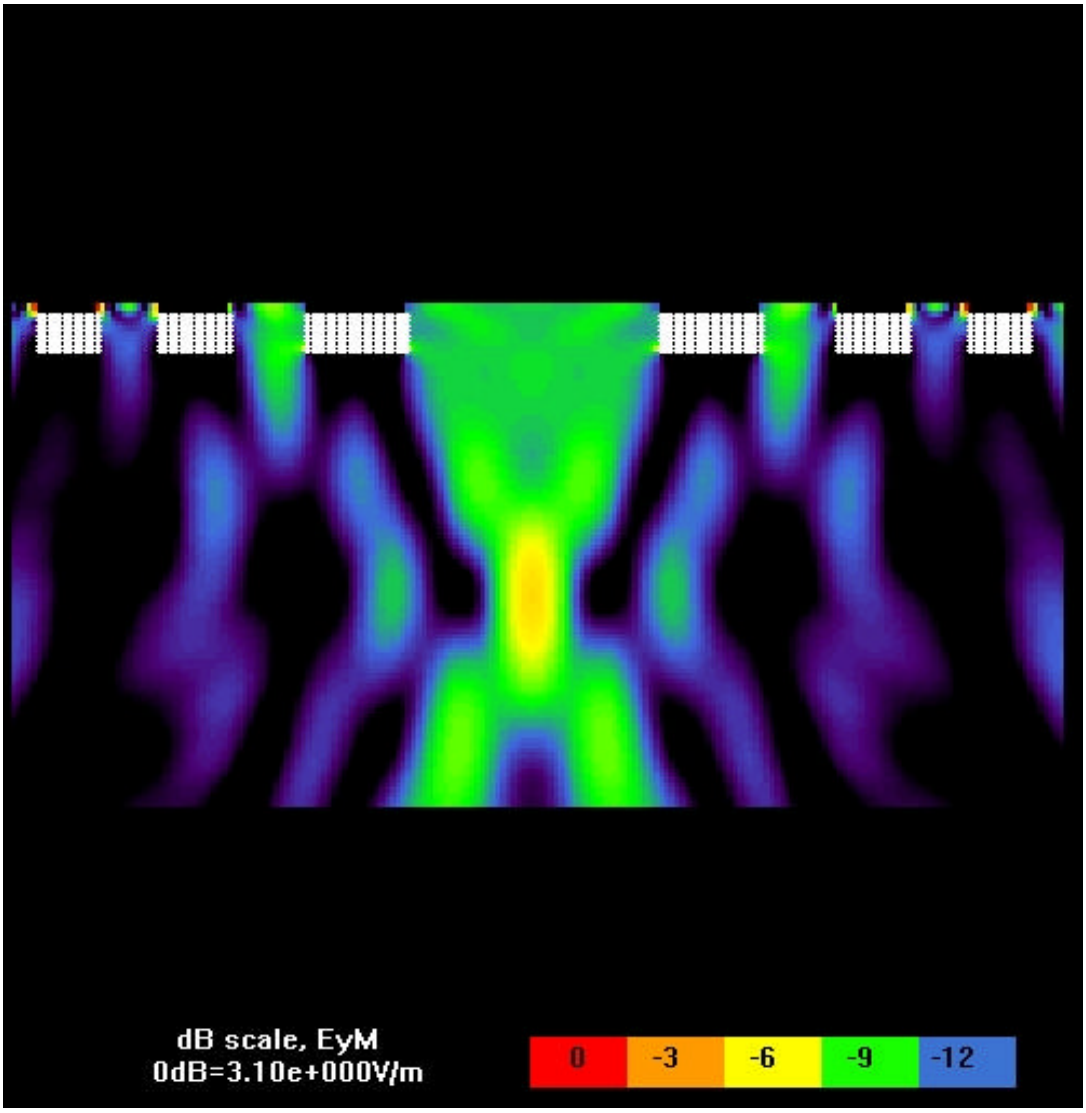



Fig. 8.8.8 XFDTD image for Fresnel zone plane of mode 3 and thickness $0.4l$ .

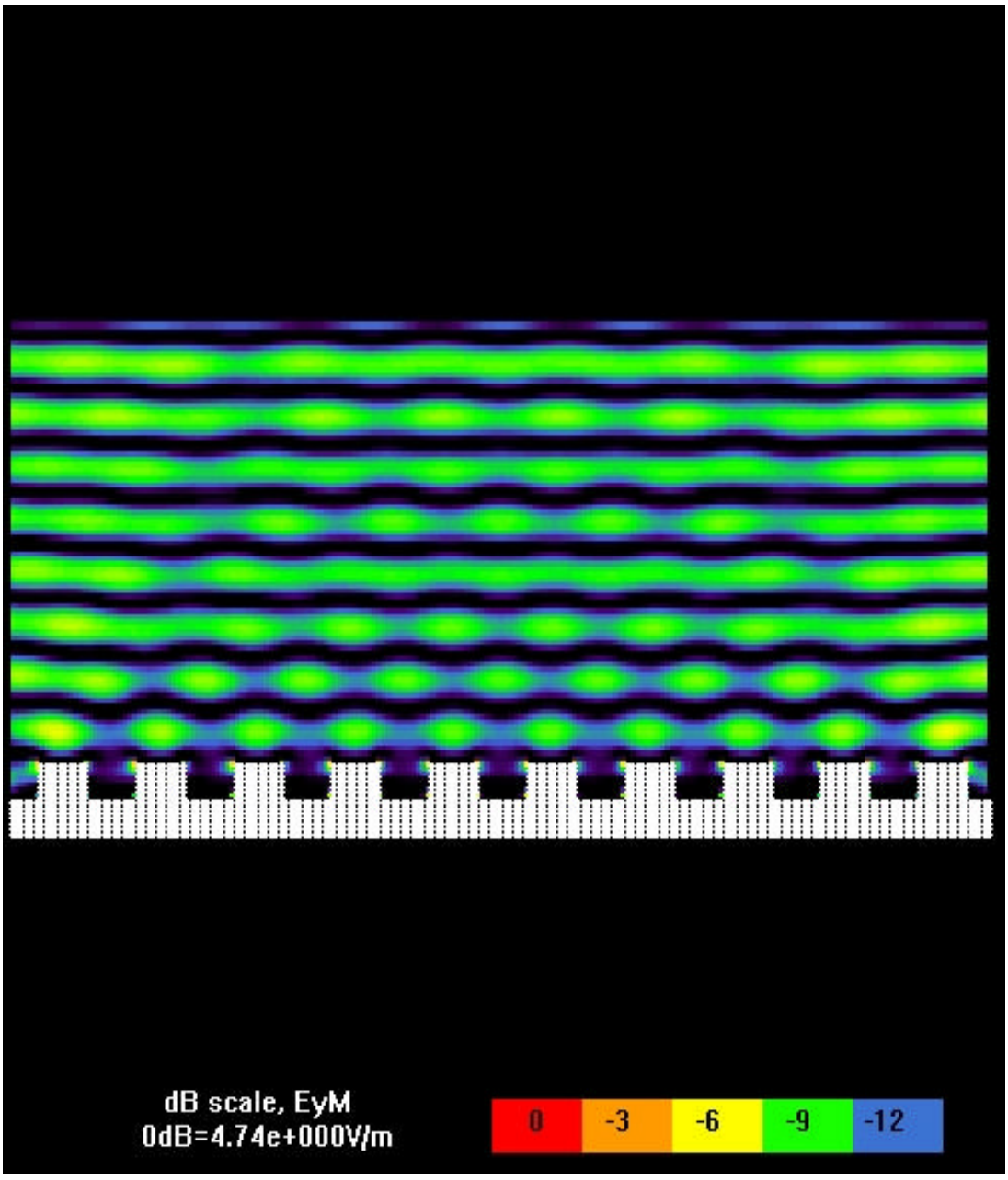



Fig. 8.8.9 Reflective grating with period $l$ .

simulation results of this grating. The results support the expected outcome. It is seen that a standing wave pattern is produced. For a grating with period $1.5l$, the grating equation says that there will be a zero order mode reflected back onto the normally incident wave as well as a first order mode propagating at an angle of 41.8 degrees. The simulation results of the $1.5l$ period grating are shown in Fig. 8.8.10. The results show that a standing wave pattern is created at an angle of zero degrees, but the pattern is not uniform across the image. At the edges, the field is propagating out to the sides. This represents the expected first-order mode.

The third set of simulations explored the Fraunhoffer diffraction pattern of a diffractive element. The general purpose of computer-generated holograms is to change the shape of a beam into a desired pattern. This is usually done by creating a binary filter that approximates the amplitude and phase of the Fourier transform of the desired beam shape.

When the binary filter is placed in the path of the beam, a lens is used to produce the Fourier transform of the modulated beam, which produces the desired shape at the opposite focal point of the lens. With conventional optics, good results require a lens with a large aperture and focal length. XFDTD simulations were done to see if this might be possible at short distances, where a refractive lens is replaced by a Fresnel zone plate.

A system where a transmission grating with period $d = 2l$ is placed a distance $2.5l$ in front of a Fresnel zone plate with focal length $3l$ was simulated in XFDTD. From the preceding analysis, several evenly spaced lines decreasing in intensity should be observed in the focal plane of the FZP. The lines should be separated by $1.5l$. The XFDTD simulation results are shown in Fig. 8.8.11 and Fig. 8.8.12. They show the field along the direction of propagation and the field in the focal plane respectively. Fig. 8.8.13 plots the intensity across the focal plane of the lens, which is where the desired pattern of lines is to be observed. The plot displays several peaks in intensity. From (8.7.3.5), we expect to see peaks at 0, $1.5l$, and $3l$ away from the center of the focal plane. The results show a series of three evenly spaced peaks at 0, $2l$, and $4l$ that decrease in intensity. The spacing is not precisely what was expected, but the general pattern is there. There is also a peak at about $l$ away from the center of the plane. This peak is likely “left over” from the peak that is seen at this same place in the focal plane

when no grating is present. But, the intensity of this peak has decreased due to the presence of the grating and the field it is trying to impose there.

A primary reason for the discrepancy in the spacing of the lines is probably the reflection at the incident surface of the zone plate, which interferes with the diffracted wave from the grating. Even so, the qualitative results compare well with theory.
The same Grating/FZP system was simulated using the angular spectrum method. The intensity in the focal plane of the FZP is shown in Fig. 8.8.14. In this case, the angular spectrum results are considerably different from XFDTD. The intensity in the focal plane shows a series of peaks, but their positions do not correspond to those predicted in (8.7.3.5). Most notably, there is no peak in the middle of the plane. This suggests that the XFDTD results are more accurate in this case.

## 8.9 Conclusions

The angular spectrum method and XFDTD software were used to simulate small-scale diffractive elements in the near field. Simulation results of Fresnel zone plate and diffraction gratings compared well with theoretical predictions. XFDTD provided greater freedom in generating the desired geometry and specifying material parameters than the angular spectrum. Also, the angular spectrum method is still based on scalar diffraction theory. XFDTD is more useful, and yields more accurate results, when scalar diffraction theory cannot be used with sufficient accuracy (i.e. when aperture sizes are on the order of a wavelength).

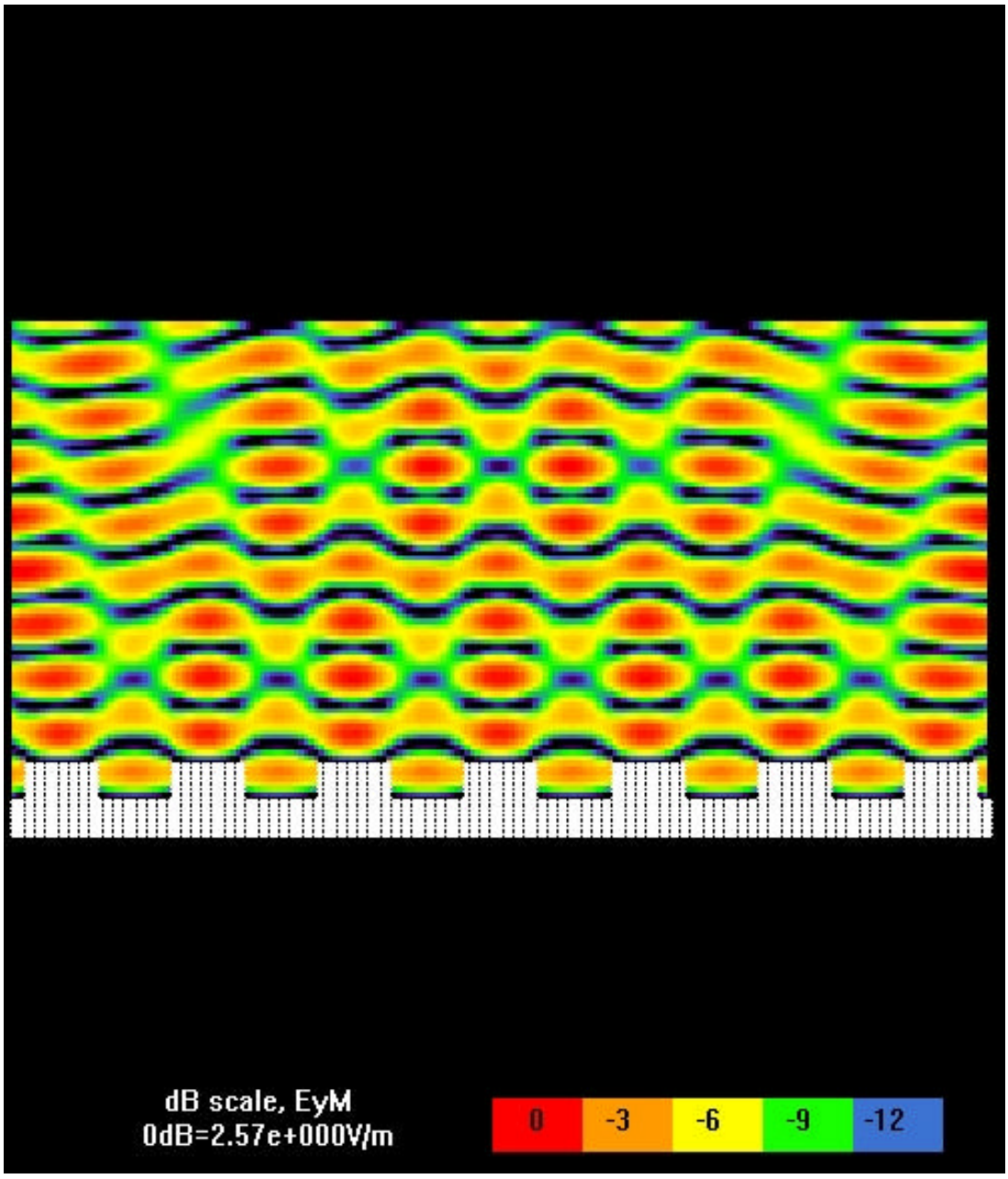



Fig. 8.8.10 Reflective grating with period $1.5l$ .

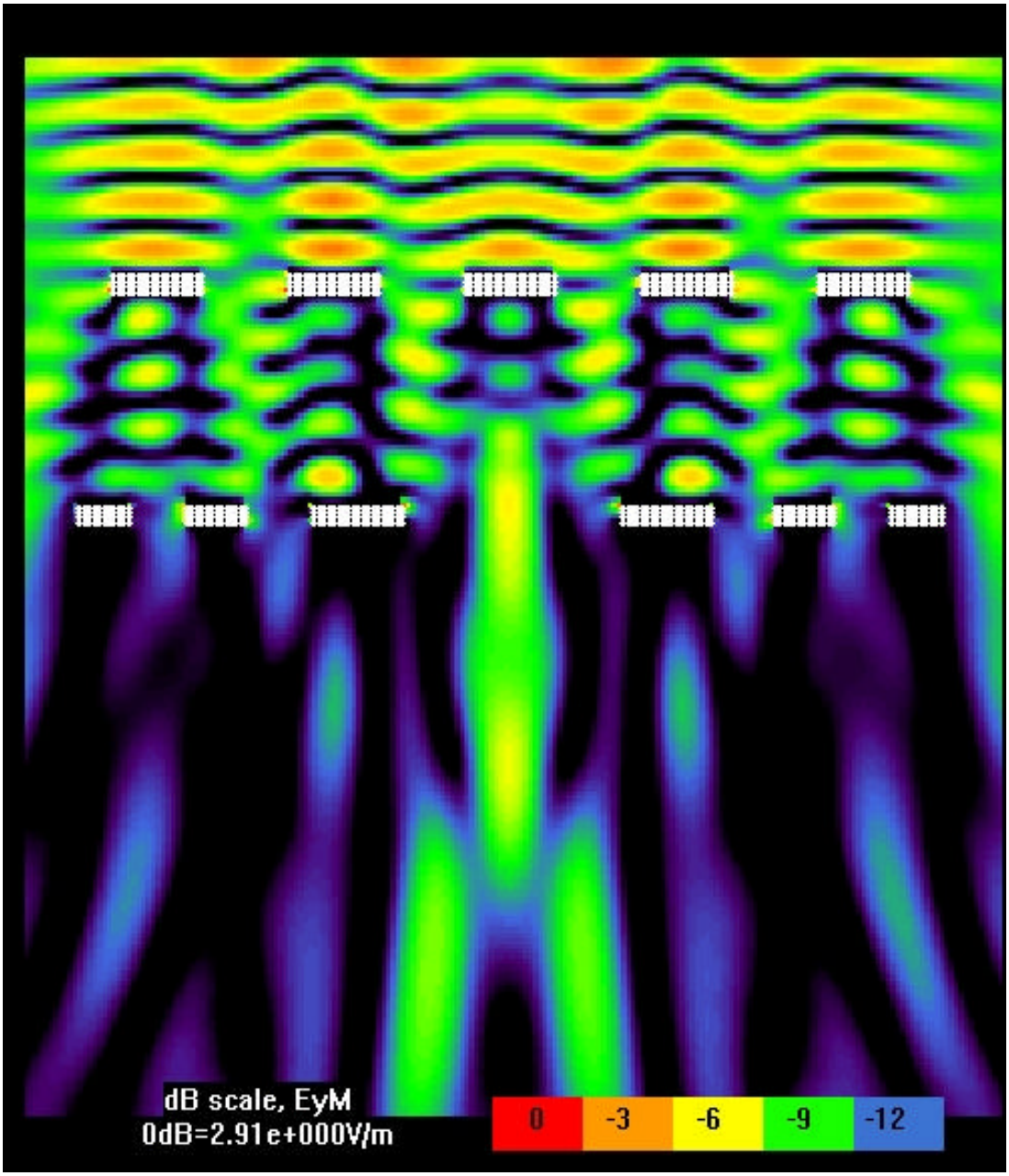



Fig. 8.8.11 XFDTD image of grating ( $d = 2\,l$ ) and FZP ( $f = 3l$ ) results.

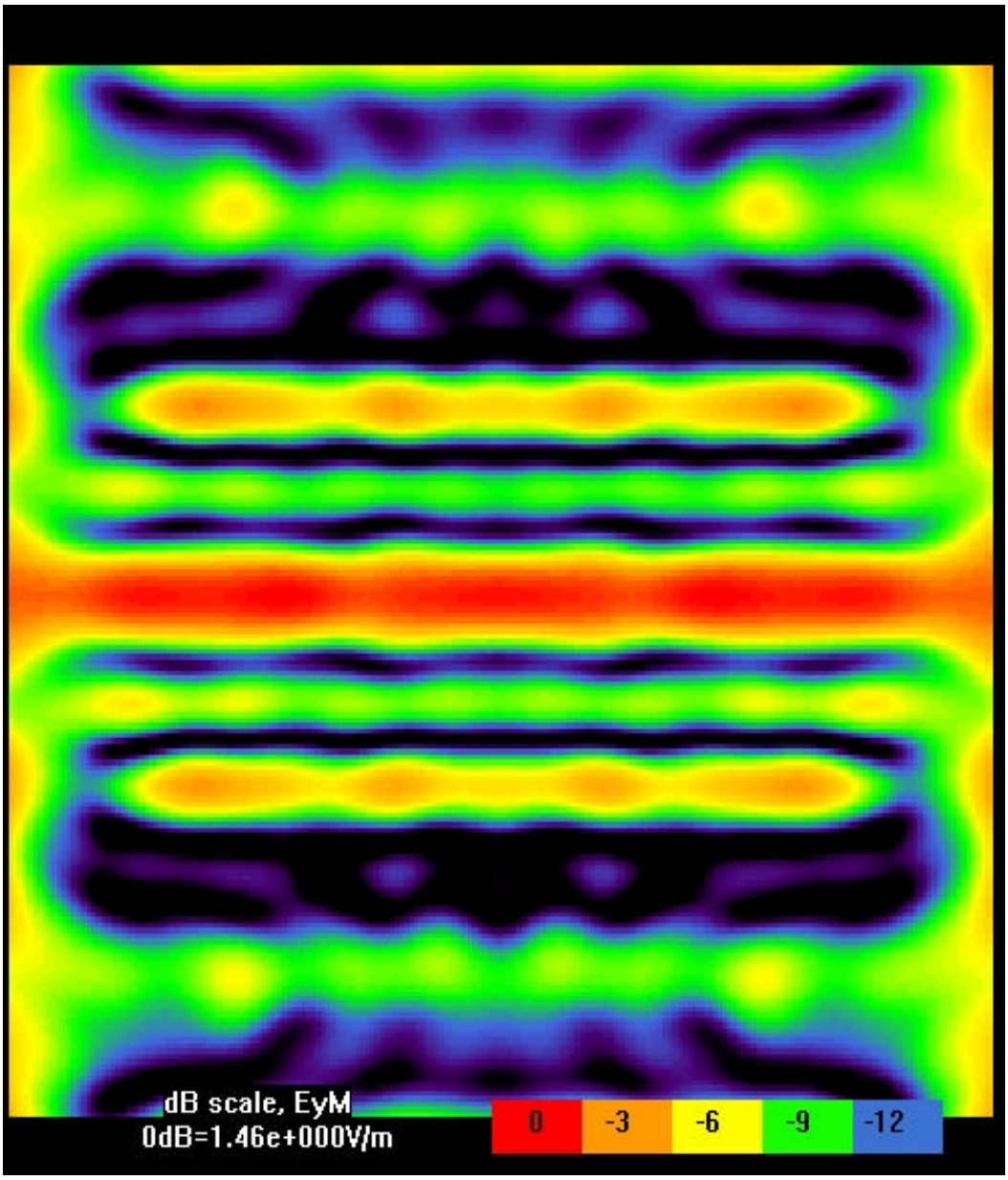



Fig. 8.8.12 XFDTD image of grating ( $d = 2l$ ) and FZP ( $f = 3l$ ) results in the focal plane of the FZP.

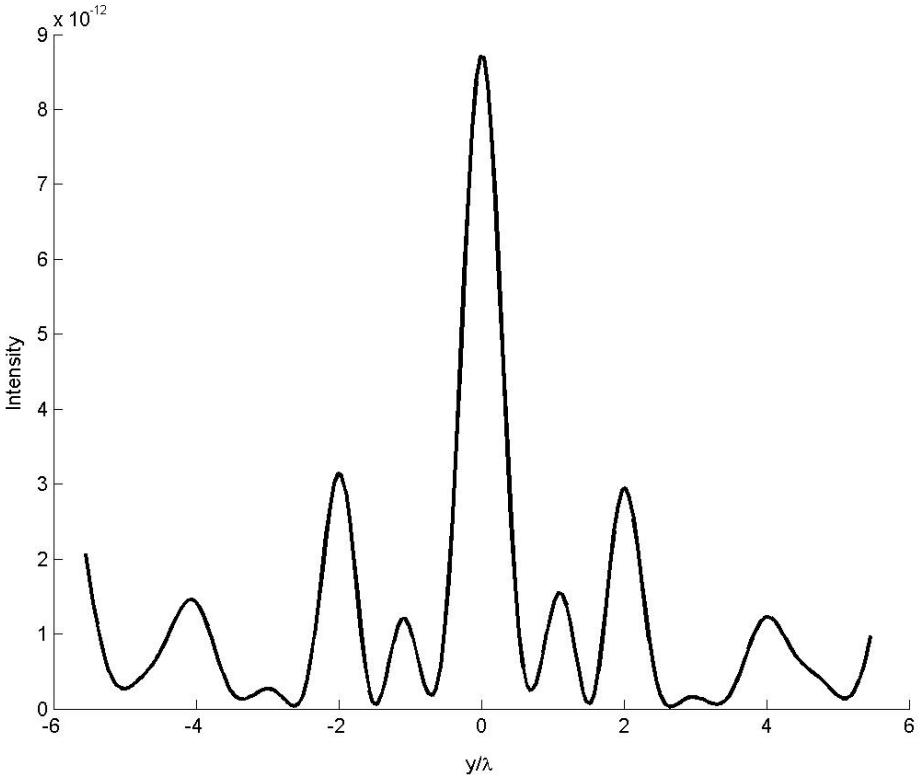



Fig. 8.8.13 XFDTD results for the intensity in the focal

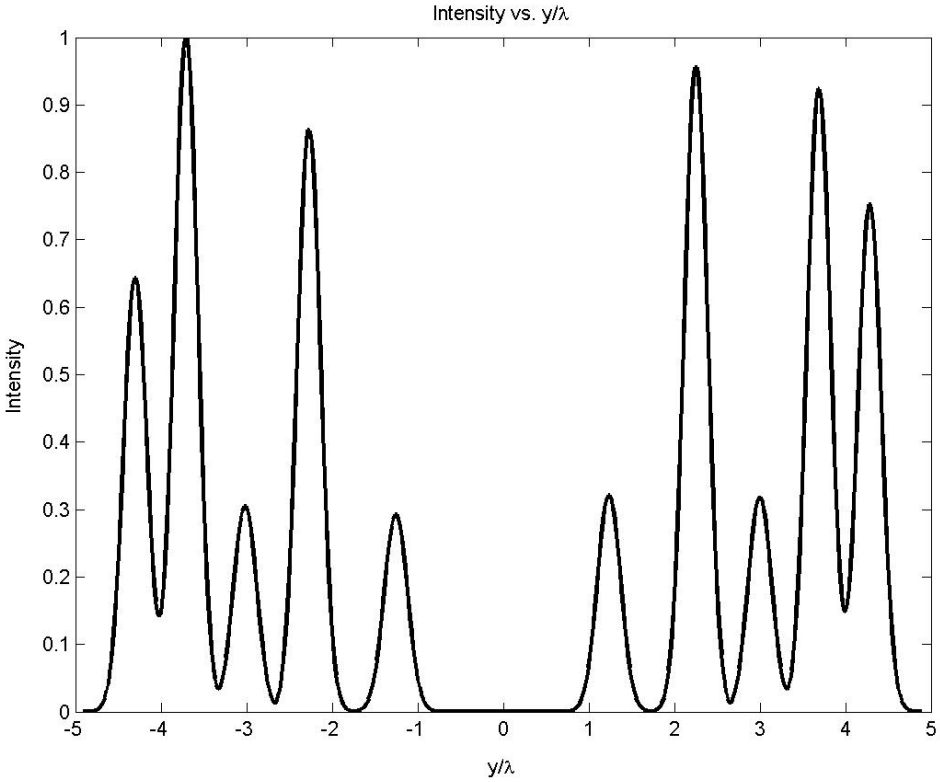



Fig. 8.8.14 Angular spectrum results for the intensity in the focal plane of a grating/FZP system. Grating period is $2l$ .

# 9 CONCLUSIONS

This research discussed scalar diffraction theory and its use in designing diffractive optical elements. Specifically, computer-generated holograms were designed using several methods. Simulated and optical reconstructions were performed for holograms designed using Lohmann's method (LM). LM utilizes a two dimensional array of apertures with varying size and position to modulate the amplitude and phase of a wavefront. Simulated reconstructions from Lohmann holograms showed similar characteristics to actual optical results. Brightness of the reconstructed image was increased by making the amplitude of each cell the same maximum value. The increased brightness came at the expense of a decrease in image quality in both simulated and optical outputs.

The interlacing technique (IT) and iterative interlacing technique (IIT) were studied and explored in order to implement the optimal decimation-in-frequency iterative interlacing technique (ODIFIIT). The IT works by dividing the hologram plane into smaller subholograms. Each subhologram is designed to reduce the error of the image created from all the other subholograms. The subholograms are then interlaced together to create one larger hologram. When this procedure is performed iteratively, it is called the IIT. ODIFIIT exploits the decimation-in-frequency characteristic of the FFT when dividing the hologram plane to optimize the IIT. Simulated and optical reconstructions from holograms designed using ODIFIIT produced accurate results, and shared many similar qualities. Brightness in the optical results of ODIFIIT was similar to the brightness of the optical Lohmann results, but LM produced a distinctly sharper image. The image quality was similar to that of the constant amplitude Lohmann's method (LMCA), but with a decrease in brightness.

Next, ODIFIIT was incorporated into Lohmann's coding scheme in the hope of improving performance. Simulations indicate a significant decrease in the error of the reconstructed image using Lohmann's method with ODIFIIT (LM-ODIFIIT). The MSE, which measures the difference between the desired image and the actual reconstructed image inside the desired image region, was $10^{31}$ times less than the MSE using LM alone. This means that the reconstructed image inside the image region matches the desired

image almost exactly, within a constant scalar factor. ODIFIIT was also used with the constant amplitude Lohmann's method (LMCA-ODIFIIT). Reconstructed error decreased from LMCA alone, but MSE and efficiency were almost the same as the original Lohmann's method.

A quantized Lohmann's method was created and used with ODIFIIT. The iterative nature of ODIFIIT makes accurate reconstruction possible with the quantized LM. The significant error introduced by course quantization of amplitude and phase in LM is effectively overcome by successive iterations. Quantization is also practical when considering implementation of the CGH. In general, current methods of implementation have limited resolution, which makes quantization necessary. Quantized LM-ODIFIIT effectively reduced MSE below the value set by LM, while maintaining approximately the same efficiency.

The second part of this thesis explored the near-field diffraction patterns from small-scale diffractive optical elements. Scalar diffraction theory breaks down under these conditions. The angular spectrum method, which is theoretically valid in the near-field, was one of the methods used. The angular spectrum of plane waves is a Fourier-transform-based method which still depends on scalar diffraction theory. It becomes inaccurate when aperture sizes get too small. The angular spectrum method was used to analyze the effects that small aperture size has on its results.

XFDTD software was also used to determine the near-field diffraction patterns from small-scale DOEs. XFDTD uses a finite-difference time-domain algorithm to solve Maxwell's curl equations directly. Diffractive elements are represented by specifying the physical properties at each point in a cubical mesh. All elements were considered to be perfect electrical conductors excited by a polarized plane wave.

Diffraction from Fresnel zone plates (FZPs) of various modes and thicknesses was simulated. Angular spectrum and XFDTD simulations both showed a build-up of intensity at the focal point of the zone plate with increasing mode. Both methods also showed a reduction in spot-size and a decrease in side-lobe intensity as the mode increased. These results support theoretical predictions. Only the angular spectrum method indicated a build-up of the expected second focal point at 1/3 the distance of the primary focal point. XFDTD simulations indicate that the point of maximum intensity of

a third-order FZP moved closer to the expected focal point as the thickness of the zone plate increased. As the plate was widened, the second focal point began to grow.

Reflection gratings with periods of $l$ and $1.5l$ were explored with XFDTD. The standing wave pattern resulting from reflection off the first grating ($d = l$) supports the theoretical prediction that all modes except the zero-order mode is suppressed. The grating equation predicts that a grating with period $1.5l$ should have a zero and a first-order mode. XFDTD results supported this claim by illustrating a standing wave pattern at an angle other than zero relative to the axis normal to the grating surface.

Finally, a transmission grating/FZP pair was analyzed to simulate a system similar to those used for computer-generated holograms. XFDTD results produced a diffraction pattern at the focal point of the FZP resembling the Fourier transform of the grating. In this case, the angular spectrum method results did not make any particular sense.

As the desire to produce devices on the micro and nano-scale becomes reality, the ability to implement effective diffractive elements will be very important. This thesis has focused on only a small portion of the possible research in the area of computer-generated holograms and small-scale diffractive elements.

There are several possibilities for future research in the areas of CGHs and small-scale DOEs. If high resolution implementation techniques are available, high quality physical CGHs could be made using all the coding methods discussed in this thesis. Optical reconstructions could be performed to better verify the simulated results that were presented. Of particular interest is a high resolution CGH designed using LM-ODIFIIT, which produced results far superior to all other methods in terms of the reconstruction error. Also, multi-level DOEs are of considerable interest. In general, more levels of quantization allows for significantly greater efficiency. This potential increase in efficiency should be taken seriously. The information on E-beam lithography and reactive ion etching could be used as a starting point for producing such high resolution diffractive phase elements.

The advantage of using CGHs in optical communications is another area for future research. CGHs could be tested at optical communications wavelengths to verify their effectiveness at such wavelengths. Then one could explore techniques to optimize

CGHs for a particular wavelength. This would allow for specialized elements, which perform optimally at the desired frequency, or range of frequencies.

Realization and testing of small-scale elements like the Fresnel zone plates and gratings in Chapter 8 to verify simulated and theoretical predictions would expand upon the work in this thesis. Once it is determined how well actual results compare with theory, research could be performed to optimize the design of small-scale elements. Or, a new theory for determining the diffracted field from small-scale elements, which is less computationally demanding, might possibly be developed.

For the previously mentioned future research ideas, once experimental results are compared with theoretical predictions and optimization techniques are developed, ideas for revolutionary products should be addressed. Efficient diffractive optical elements could offer a promising solution to many current and future engineering problems. With dedicated research and some educated intuition, diffractive elements have the potential to be applied in situations not previously considered.

## REFERENCES


[1] J.W. Goodman, *Introduction to Fourier Optics*, San Francisco: McGraw Hill, 1968.

[2] N.C. Gallagher and D.W. Sweeney, "Infrared Holographic Optical Elements with Applications to Laser Material Processing," *IEEE J. of Quantum Elec*. vol. QE-15, Dec. 1979, pp. 1369-1380.

[3] M.R. Feldman and J.E. Morris, "Holographic Optical Interconnects for VLSI Multichip Modules," *IEEE Transactions on Components, Packaging, and Manufacturing Technology – Part B*, vol. 17, May 1994, pp. 223-227.

[4] J. Dittman, L.C. Ferri, C. Vielhauer, "Hologram watermarks for document authentications," in *International Conference on Information Technology: Coding and Computing*, Apr. 2001, pp. 60-64.

[5] B.R. Brown and A.W. Lohmann, "Complex Spatial Filtering with Binary Masks," *Applied Optics*, vol 5., June 1966, pp. 967-969.

[6] J.Y. Zhuang, "Optimal Decimation-in-frequency Iterative Interlacing Technique for Synthesis of Computer-generated Holograms," *J. Opt. Society of America*, vol. 12, pp. 1460-1468, 1995.

[7] H. Stark, ed., *Image Recovery: Theory and Applications*, Academic Press 1986.

[8] J.M. Jin, *The Finite Element Method in Electromagnetics*, New York: Wiley, 2002.

[9] Allen Taflov, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Boston: Artech House, 1995.

[10] M. Born and E. Wolf, *Principles of Optics*, 7th ed., New York: Cambridge, 1999.

[11] F. Wyrowski, "Theory of Digital Holography," *Holographic Systems, Components, and Applications*, 1991, Third International Conference on, 16-18 Sep. 1991, pp. 93-98.

[12] A.W. Lohmann and D.P. Paris, "Binary Fraunhofer Holograms, Generated by Computer," *Applied Optics*, vol. 6, Oct, 1967, pp. 1739-1748.

[13] J.Y. Zhuang, *Iterative Nonlinear Optimization Techniques for Diffractive Optical Applications and Their Implementations*, PhD Thesis, School of Electrical and Computer Engineering, Purdue University, 1994.

[14] M. Bowden, L. Thompson, C. Wilson, eds., *Introduction to Microlithography*, American Chemical Soc., 1983.

[15] JSM 6400 Manual by JEOL – USA, 111 Dearborn Rd., Peabody, MA 01960. In Japan, JEOL Ltd., 1-2 Musashino 3-Chome, Akishima Tokyo 196.

[16] A. Weill, "The spin coating process mechanism," in *The Physics and Fabrication of Microstructures and Microdevices*, M. J. Kelly, C. Weisbuch, eds; (Springer-Verlag, erlin, 1986) p. 51.

[17] NPGS Manual by J. C. Nabity Lithography Systems, P.O.Box 5354, Bozeman, MT 59717, USA, jcnabity@aol.com.

[18] K.S. Yee, "Numerical solutions of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas and Propagation*, vol. 14, 1966, pp. 302-307.

[19] E.Hecht, *Optics*, Mass.: Addison-Wesley, 2002.

[20] S. Ramo, J.R. Whinnery, T. VanDuzer, *Fields and Waves in Communication Electronics*, 3rd ed., New York: Wiley, 1994.