

# **USC-SIPI REPORT #421**

**A Treatise on Cascaded Computer Generated Holograms**

**by**

**Ali Alkan Gulses**

**May 2014**

**Signal and Image Processing Institute**  
**UNIVERSITY OF SOUTHERN CALIFORNIA**  
Viterbi School of Engineering  
Department of Electrical Engineering-Systems  
3740 McClintock Avenue, Suite 400  
Los Angeles, CA 90089-2564 U.S.A.

# **A TREATISE ON CASCADED COMPUTER GENERATED HOLOGRAMS**

ALI ALKAN GÜLSES

---

A Dissertation Presented to the  
FACULTY OF THE USC GRADUATE SCHOOL  
UNIVERSITY OF SOUTHERN CALIFORNIA  
In Partial Fulfillment of the  
Requirements for the Degree  
DOCTOR OF PHILOSOPHY  
(ELECTRICAL ENGINEERING)

May 2014

## **ACKNOWLEDGEMENTS**

I would like to thank to my entire family first, especially Muge and Kemal Gulses for their support and encouragement, and all friends who were with me during my studies. I also wish to thank to USC SIPI faculty and staff for their help, and my supervisor Keith Jenkins for his guidance; in fact it will be really hard to find the taste of our discussions in elsewhere of the world.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
LIST OF FIGURES.....	vi
LIST OF TABLES.....	xii
ABSTRACT.....	xiii

## CHAPTER 1

### INTRODUCTION

1.1 Historical Overview.....	1
1.2 Thesis Overview.....	3

## CHAPTER 2

### DIFFRACTION THEORY

2.1 Light as an Electromagnetic Wave.....	6
2.2 Propagation in Spatial Domain.....	9
2.2.1 Diffraction from an arbitrary aperture.....	9
2.2.2 Effect of Lens.....	12
2.3 Propagation in Fourier Domain.....	13
2.4 Optics in Digital Domain.....	15

## CHAPTER 3

### IMAGING WITH PHASE ELEMENTS

3.1 Introduction.....	21
3.2 Definitions.....	24
3.3 Two Dimensional Images.....	27
3.3.1 One CGH case.....	27
3.3.2 Multi-CGH case.....	35
3.4 Three Dimensional Images.....	49

3.4.1 3D abstraction with two image planes.....	50
3.4.2 3D abstraction with $M$ image planes.....	64
3.5 Other Ways of Mitigating Noise.....	73
3.6 A Hybrid Method.....	74
<b>CHAPTER 4</b>	
<b>IMAGING WITH PHASE &amp; AMPLITUDE ELEMENTS</b>	
4.1 Introduction.....	83
4.2 Encoding Techniques.....	84
4.2.1 Overview.....	84
4.2.2 Simulations.....	85
4.3 Implementation with Cascaded CGHs.....	89
4.3.1 Two dimensional images.....	89
4.3.2 Three dimensional images.....	91
<b>CHAPTER 5</b>	
<b>EXPERIMENTAL</b>	
5.1 Introduction.....	95
5.2 Fabrication of Cascaded CGHs.....	95
5.2.1 General production methods.....	95
5.2.2 Production of proposed setup.....	100
5.3 Misalignment Errors.....	103
5.4 Experiments with Single Intensity-Modulated CGHs.....	106
5.4.1 The He-Ne laser.....	109
5.4.2 Production of holograms.....	110
<b>CHAPTER 6</b>	
<b>DISCUSSION &amp; CONCLUSION</b>	
6.1 Some Issues.....	113
6.1.1 Convergence.....	113
6.1.2 Sampling.....	114

6.1.3 Speckle formation.....	115
6.1.4 Other error metrics.....	116
6.1.5 Binary phase quantization.....	117
6.1.6 Lensless Functionality.....	117
6.2 Summary.....	118
6.2.1 Literature comparison.....	118
6.2.2 Thesis Summary.....	120
 <b>APPENDICES</b>	
APPENDIX A.....	123
APPENDIX B.....	129
APPENDIX C.....	130
APPENDIX D.....	132
APPENDIX E.....	133
 REFERENCES.....	 139

## LIST OF FIGURES

Fig. 2.1	In figure, arbitrarily shaped hole passes incident light through the surface $S$ . Planes are labeled with their own coordinate systems for convention. Physically, aperture plane can be anything. Output plane may be a screen, a CCD or an eye retina etc.....	10
Fig. 2.2	A conceptual grid represents a function on Cartesian coordinate system, where every point has a corresponding value. The function is divided to cells, called picture elements or pixels. Pixels are represented by the center dot in sampling process.....	16
Fig. 3.1	Basic image generation setup for a PE consisting of several elevation levels for each individual pixel (in figure they are shown by four levels for illustration purposes). Grey arrow on the optical axis shows the propagation direction of the flux. A lens with focal length $f$ , serves as a propagation tool as a Fourier transform operator. ‘Lena’ image is used for illustration.....	27
Fig. 3.2	Diagram of a simple bidirectional IFTA algorithm consisting of perpetual transforms to design and optimize PEs for imaging.....	29
Fig. 3.3	Diagram of flow chart for SA. ‘ $k$ ’ shows the loop number. $E_{new}$ means the error of the new trial and $E_{old}$ is the current error. The new attempt will be accepted if the conditional statement holds. The loop stops after certain number of iterations where the error doesn’t change anymore.....	31
Fig. 3.4	Hypothetical system configuration. $N$ thin PEs, each $L$ by $L$ , with a separation $d$ , are concatenated in order to create an output plane.....	34
Fig. 3.5	One sub-iteration cycle is shown to optimize one PE in Fig.3.4. Each element is calculated with an operation where the field coming from the image plane is divided with the field propagated up to that element. At the end, the element index is updated and another sub-iteration starts for the subsequent PE. When every component in the stack is finished, iteration is completed and the next one starts. This is controlled by mod ( $N$ ) factor.....	39
Fig. 3.6	Flow-chart for the SA algorithm in optimization of cascaded PEs for single-plane imaging. ‘ $k$ ’ shows the loop number. The pixel is randomly chosen from a random element for a perturbation.....	40
Fig. 3.7	Two different sample images have been used in experiments. (a) First image is a USC, Trojan mascot and (b) Second one is a Viterbi School of Engineering logo.....	41
Fig. 3.8	Results of computer experiments with 64 allowed phase quantization levels. Iterative algorithm is used. Error and diffraction efficiency are displayed versus number of phase elements for (a) first image and (b) second image.....	43

Fig. 3.9	Results of computer experiments with 4 allowed phase quantization levels. Iterative algorithm is used. (a) Error and diffraction efficiency are displayed versus number of phase elements for (a) first image and (b) second image.....	43
Fig. 3.10	Results of computer experiments with 4 allowed phase quantization levels. SA type algorithm is used. (a) Error and diffraction efficiency are displayed versus number of phase elements for (a) first image and (b) second image.....	44
Fig. 3.11	Resultant Images of computer experiments with 64 allowed phase quantization levels, referring to Fig.3.8. Iterative algorithm is used. (a) first image processed by 1 PE (b) second image processed by 1 PE (c) first image processed by 5 PEs (d) second image processed by 5 PEs...	46
Fig. 3.12	Resultant Images of computer experiments with 4 allowed phase quantization levels, referring to Fig.3.9. Iterative algorithm is used. (a) first image processed by 1 PE (b) second image processed by 1 PE (c) first image processed by 5 PEs (d) second image processed by 5 PEs...	47
Fig. 3.13	Resultant Images of computer experiments with 4 allowed phase quantization levels, referring to Fig.3.10. SA type algorithm is used. (a) first image processed by 1 PE (b) second image processed by 1 PE (c) first image processed by 5 PEs (d) second image processed by 5 PEs.....	48
Fig. 3.14	Hypothetical system configuration. $N$ thin PEs, each $L$ by $L$ , with a separation $d$ , are concatenated in order to create output, namely output (target) planes 1 and 2 with separation $\ell$ .....	50
Fig. 3.15	One iteration cycle of the extended IFTA which is used to optimize the system in Fig.3.14. This flow scheme is capable of handling successive 2D planes in continuum for both input and output in a general sense.....	51
Fig. 3.16	Flow-chart for the extended SA algorithm in optimization of cascaded PEs for multi-plane imaging. 'k' shows the loop number. In addition to the last flow chart, this time errors for the two output planes are averaged for comparison.....	53
Fig. 3.17	These gray scale images are used in the following experiments in place of the first and second output planes in order to demonstrate the effects of cascaded PEs in noise removal. The size of images used in experiments is 32 x 32 and 64 x 64 pixels.....	55
Fig. 3.18	Results of computer experiments with 64 phase quantization levels allowed. Extended IFTA is used. MSE is seen to decrease as more PEs are added. Although percentage decrease rate, $R$ and final values of the first output plane excel, the net amount of noise removed from the second output plane is more.....	56
Fig. 3.19	Results of computer experiments with 4 phase quantization levels allowed. Extended SA is used. MSE is seen to decrease as more PEs are added, similarly for both output planes.....	56
Fig. 3.20	Resultant Images (32 x 32 pixels) of computer experiments with 64 allowed phase quantization levels when Extended IFTA is used,	

	referring to Fig.3.14. (a) The first image set (for output plane 1 and output plane 2) is processed by 1 PE (b) the second image set (for output plane 1 and output plane 2) is processed by 5 PEs. The improvement is clear. Note that max scaling is applied here. In other words, in printing the results, all pixels are divided by the maximum value for normalization.....	57
Fig. 3.21	Resultant Images (32 x 32 pixels) of computer experiments with 4 allowed phase quantization levels when Extended SA is used, referring to Fig.3.14. (a) The first image set (for output plane 1 and output plane 2) is processed by 1 PE (b) the second image set (for output plane 1 and output plane 2) is processed by 5 PEs. Max scaling is applied here again.....	58
Fig. 3.22	MSE versus number of PEs used in two-output plane imaging, when extended IFTA is used with 64 allowed phase values.....	61
Fig. 3.23	Resultant Images (64 x 64 pixels) of computer experiments with 64 allowed phase quantization levels when Extended IFTA is used, referring to Fig.3.14. (a) The first image set (for output plane 1 and output plane 2) is processed by 1 PE. (b) The second image set (for output plane 1 and output plane 2) is processed by 5 PEs. Note that, no scaling is applied so reconstructed image intensities are unnormalized here and a little bit lower than the original due to other diffracted orders.....	62
Fig. 3.24	One iteration cycle of the extended IFTA in handling N PEs and M image planes.....	65
Fig. 3.25	Image set 1, used in experiments.....	67
Fig. 3.26	Image set 2, used in experiments.....	67
Fig. 3.27	Graphs, corresponding to image set in fig. 3.25. (a) Results for 2 image planes, '1' and '2'. Percentage decrease rates are 86% and 58% for output planes 1 and 2. (b) Results for 3 image planes, '1', '2' and '3'. Percentage decrease rates ( <i>R</i> ) are 84%, 49% and 34% for output planes 1, 2 and 3.....	68
Fig. 3.27	Graphs, corresponding to image set in Fig. 3.25. (a) Results for 2 image planes, 'U' and 'S'. Percentage decrease rates are 85% and 60% for output planes 1 and 2. (b) Results for 3 image planes, 'U', 'S' and 'C'. Percentage decrease rates ( <i>R</i> ) are 85%, 50% and 39% for output planes 1, 2 and 3.....	68
Fig. 3.28	Graphs, corresponding to image set in Fig. 3.26. (a) Results for 2 image planes, 'U' and 'S'. Percentage decrease rates are 85% and 60% for output planes 1 and 2. (b) Results for 3 image planes, 'U', 'S' and 'C'. Percentage decrease rates ( <i>R</i> ) are 85%, 50% and 39% for output planes 1, 2 and 3.....	68
Fig. 3.29	Resultant Images (64 x 64 pixels) of computer experiments coming from Fig. 3.25 when Extended IFTA is used. (a) The first image set is processed by 1 PE (b) the second image set is processed by 5 PEs.....	69
Fig. 3.30	Resultant Images (64 x 64 pixels) of computer experiments coming from Fig. 3.26 when Extended IFTA is used. (a) The first image set is	

	processed by 1 PE (b) the second image set is processed by 5 PEs.....	70
Fig. 3.31	Total MSE curves belonging to image sets 1 and 2. For the curves the decay constant is constrained in an interval. In fact, experiments with more than 3 image planes show also that the constant is still found in the same interval (0.6-0.7) for $\ell = 40$ mm. $R$ values are 68, 67, 51 and 48 from bottom to top.....	72
Fig. 3.32	Figure displays results for two different samples belonging to one phase element reconstruction using a noise window, with 64 quantization levels. The noise window used here is 4X the actual image.....	74
Fig. 3.33	Ultimate space usage concept in CGHs is displayed. PEs are placed as before from 1 to $N$ . Gray toned areas have the same dimensions with originally designed elements before. Pixels in the extra area are responsible for noise removal. Since this new configuration contains more pixels, its feature size is defined by $\kappa L$ and total area is $\kappa LN$ .....	75
Fig. 3.34	Result of cascaded PEs with also extended for dummy area. Two output planes are assumed. No phase quantization constraint is applied. $N$ shows the number of PEs as before and $\kappa$ denotes feature extension coefficient in x or y direction. The result of adding dummy area is compensated by changing the lens focal length in order to keep the same output dimensions. 3D graph reveals the combination of cascading and extending under these circumstances.....	78
Fig. 3.35	Result of cascaded PEs with also extended for dummy area. Two output planes are assumed. No phase quantization constraint is applied. $N$ shows the number of PEs as before and $\kappa$ denotes feature extension coefficient in x or y direction. The result of adding dummy area causes smaller output dimensions. 3D graph reveals the combination of cascading and extending under these circumstances.....	79
Fig. 3.36	Image outputs for the proposed system of ultimate expansion of a CGH when $N = 4$ and $\kappa = 2$ . (a) The actual image sizes are compensated by an imaging lens (b) Noise window cause decrease in output pixel size yielding a smaller image. For comparison reasons, both parts are shown in equal dimensions.....	80
Fig. 4.1	(a) According to above encoding scheme, cells represent pixel values and each cell is further divided to four sub-cells with a transparent rectangle inside. (b) Different gray tones may also be used instead of rectangles.....	85
Fig. 4.2	Sample 64 x 64 image for computer experiments.....	85
Fig. 4.3	Results from CGHs that are capable of modulating both phase and amplitude. Sinc modulation is seen at results and cannot be eliminated easily in this application since there is no pure phase modulating element. Lee's method is used. (a) CGH pattern (b) reconstruction (c) CGH pattern with 5 times more precise (d) reconstruction. ....	86
Fig. 4.4	Results from CGHs that are capable of modulating both phase and amplitude. Sinc modulation is seen at results and cannot be eliminated easily in this application since there is no pure phase modulating	

	element. Lee's method is used with grayscale levels instead of rectangular apertures. (a) CGH pattern (b) reconstruction (c) CGH pattern with 5 times more accurate (d) reconstruction.....	87
Fig. 4.5	Phase and amplitude modulating planes with separation $d$ . Phase element and amplitude element (AE) are placed successively.....	90
Fig. 4.6	Result of phase-amplitude modulation with cascaded CGHs. An almost perfect (noise-free) reconstruction is possible when proper adjustment of parameters and setup is implemented.....	91
Fig. 4.7	Phase and amplitude modulating planes with separation $d$ for a 3D reconstruction.....	92
Fig. 4.8	Result of phase-amplitude modulation with cascaded CGHs. An almost perfect 3D reconstruction is possible when proper adjustment of parameters and setup is implemented. Numerical MSE values are 0.03 for the first plane and 0.39 for the second. Image intensity has been adjusted for comparison with those in Fig.3.23.....	93
Fig. 5.1	A CGH is shown with magnified section module. One square-shaped pixel is displayed with feature size $\omega$ and etch-depth $s$ .....	96
Fig. 5.2	Production steps of a layer of phase relief structure in photolithography. (a) Substrate, usually glass. (b) Application of photoresist. (c) Masking application. (d) Result. (e) Etching. (f) Final form.....	99
Fig. 5.3	Final view of the compact device. After PEs are computed, they are replicated for speckle reduction (at least 3 or 5 times). And fabricated successively with separation distance $d$ . For the figure, there will be $N \times 5$ PEs contributing to image formation. Material 1 and material 2 are intended to be used for PEs and buffer layers, respectively.....	102
Fig. 5.4	Two PEs are shown. One of them is misaligned in three spatial coordinates, and this is denoted as $\Delta x$ , $\Delta y$ and $\Delta z$ .....	103
Fig. 5.5	This graph displays the change in total MSE versus the change in distance between two PEs in case of two image planes.....	105
Fig. 5.6	This graph displays the change in total MSE versus the change in lateral distance between two PEs in case of two image planes. This change can be in $x$ or in $y$ direction.....	105
Fig. 5.7	Optical table.....	107
Fig. 5.8	Conceptual drawing of the experimental setup.....	108
Fig. 5.9	(a) Desired image (b) Hologram (c) Numerical output (d) Experiment.	111
Fig. 5.10	(a) Desired image (b) Hologram (c) Numerical output.....	111
Fig. 5.11	Final device look fabricated and used in Optical Computing lab. Left hand side shows the device in Fig. 5.9 (b) and right hand side shows the device in Fig. 5.10 (b).....	112
Fig. 6.1	Number of PEs used versus (a) Co value (b) SNR, for the experiment results shown in Fig. 3.22.....	116
Fig. A1	A screen with tiny aperture is shown. Surface normal vectors are denoted by $\vec{n}$ . Hypothetical surfaces $S_R$ and $S_r$ are defined in the text and they serve the purpose of computing the field at $P_0$ , assumed to be the origin in spherical coordinates. $S'_R$ is a part of $S_R$ that intersects	

the screen.  $V$  defines corresponding volumes. Finally,  $\tilde{P}_0$  is the mirror image of  $P_0$  and it is symmetric with respect to screen..... 125

Fig.C1

Amplitude versus frequency (a) Spectrum of a function (b) Spectrum of a sampled function. Separation of two terms is shown by an arrow. Amplitude is shown with  $\chi$ , so that  $G = \chi e^{i\zeta}$  and  $G_{sampled} = \chi_{sampled} e^{i\zeta_{sampled}}$  respectively.  $f_{max}$  shows the bandwidth..... 131

## LIST OF TABLES

Table. 3.1	Output formulation regarding the system in Fig.3.4. In compact form, the transforms are represented by symbols. The transforms related to lens, are also denoted. In expressions, $\times$ shows multiplication.....	36
Table. 3.2	Output formulation regarding to system in Fig. 3.4. In explicit form, the transforms are expressed clearly (referring to Chapter 2). Mathematically, there is no difference between Fourier transforms coming from lens and free space propagation. Note that $\Theta_1$ , $\Theta_2$ and $\Theta_n$ are phase delays of the related elements.....	37
Table. 3.3	Performance merits for 1 and 5 PEs according to allowable phase levels on device.....	63
Table. 3.4	Performance merits for 1 and 5 PEs when allowed number of phase levels are 4, for extended SA case.....	64

## ABSTRACT

In diffractive optics and computer holography, generally, a common and well-studied scenario is using one hologram plane to generate a single image plane. As a continuation, in this work, cascaded phase only holograms are presented for single and multi-plane image formation. Traditional design methods, such as deterministic iterative and stochastic simulated annealing algorithms, for the one planar hologram case have been expanded and enriched to handle cascaded computer generated holograms. This special arrangement of holograms demonstrates interesting attributes in the reconstruction of classical 2D images. When it comes to 3D images, on the other hand, the performance of these cascaded-holograms is found to follow a trend and is even more remarkable: in order to facilitate the system, 3D images may be conceived to consist of successive 2D image planes; after application of the cascaded holograms for reconstruction, numerical results from computer experiments show that increasing the number of holograms for input, decreases the error in all output image planes for 3D setup. In addition, the cascaded holograms technique can be combined with the classical method of noise window which is achieved by planar extension of a single hologram by forfeiting some resolution. Thus when cascaded holograms are also expanded in the lateral domain, the results are even more promising. Specifically, it is observed that, while lateral extension of holograms as an application of the classical noise window concept is responsible for noise removal phenomena up to some degree in a general sense, longitudinal extension as a form of cascades may especially become a remedy for apparent incompatibility between image planes appearing in multi-focal systems, which is another remarkable outcome. As a continuation of the work, instead of cascaded phase only holograms, amplitude modulation is taken into account and phase plus amplitude modulation with a certain distance between them, is analyzed. This configuration shows some interesting properties and some significant superiority with respect to traditional techniques although some limitations, such as requirement of larger pixel sizes which is on the boundary of diffractive optics, exist.

The proposed schemes in this work may be applicable to a formation of a simple and light-weight setup. Therefore, by using the techniques, results and algorithms here, especially on the cascaded phase only holograms, a functional static micro-holographic 3D display may be realizable. In addition, phase and amplitude modulating setup can be used on applications requiring large pixel size, such as on devices with large area electronics.

Additionally, methods to overcome possible hurdles in both computational design and physical application stages of the novel displays whose properties are mentioned above are briefly discussed. Although the work up to now is concentrated on algorithms and computation, experimental techniques for fabrication and error consideration coming from physical misalignment are also proposed. We believe that this work will be not only useful and functional in the analysis of cascaded phase elements but also a thorough examination on algorithms, noise removal or some other issues on computer holography and computational optics.

# CHAPTER 1

## INTRODUCTION

### 1.1 Historical Overview

In 1948, D. Gabor proposed a novel method which would then be called holography [1, 2]. According to this method, with a suitable reference wave in addition to the wave diffracted or scattered from an object, both phase and amplitude information of the object can be recorded. This would be one step beyond photography where only intensity might be recorded. As a result, holograms were useful as optical elements to correct aberrations, displays or data storage, as a few examples.

The meaning of the word hologram has expanded because holograms can also be generated by computers. Computer generated holograms (CGH) have many useful properties [3]. For example, an object which does not exist physically can be created computationally and an ideal wavefront can be computed on the basis of diffraction theory and encoded into a tangible hologram. There are four main steps for preparation. First, an object is defined computationally. It might be a digital image or a real object recorded with a camera as in digital holography. Next, wave propagation from the object to hologram surface is computed with theories which can be vector or scalar in format. Scalar theories are more common and appropriate for far-field or near-field diffraction.

Third step is to encode patterns with different representation schemes onto the hypothetical device. Finally, these patterns on the hologram surface are fabricated physically by using a printer, laser beam, or electron beam lithography on transparent materials, generally silica glass. At the end, the hologram can be used optically by illuminating with light.

There are diverse methods for the third step above in encoding information on the hologram surface. One way to encode is to use some specialized techniques so that both phase and amplitude modulation take place. One of them is so called detour phase hologram used by Brown and Lohmann in 1966 [4]. Another approach came from Lee [5]. These methods, creating both phase and amplitude modulation, would use cell approach, where every pixel is represented by cells with certain properties. One other method for encoding is to use some sophisticated algorithms and put the information directly on hologram surface only in terms of phase information without using cells. So, in this approach, the wave which would result in object reconstruction after hologram modulation, is calculated and written on the hologram. In this case generally phase modulation is used in a form of relief profiles. These sophisticated heuristic algorithms are supposed to compensate for the lack of amplitude and may be iterative in character or stochastic (such as an advanced brute force technique like simulated annealing or evolutionary like genetic algorithm).

In fact a CGH is a kind of diffractive optical element (DOE) which modulates a wave front in a desired manner. Another DOE is diffraction grating where again a controlled wavefront modulation is supplied but this time with periodic structures. Gratings attracted lots of attention around 80s and 90s.

The work on diffractive elements has evolved from using one grating plane, designed by a computer algorithm, to reconstruct one output plane, in which the goal is typically to achieve a specified intensity or phase pattern at the output. Recent efforts have studied the use of multiple DOE planes in a cascaded way to reconstruct patterns in

one output plane, which can provide additional features such as control over multiple colors, multiplexing capability, or high diffraction efficiency [6, 7].

Diffraction elements may also be used for 3D displays, since holography is a true-3D technique. This area of research has attracted attention starting from the invention of hologram which enables depth information due to phase recording. With the developments in computers and demand from entertainment market, 3D imaging is still a very hot field for new discoveries both in academic and industry environments.

Basic auto stereoscopic 3D displays are around for almost ten years; almost all currently available 3D displays are still stereoscopic, where different information comes to different eyes creating depth effect. This could be different polarization, different colors or combination of both. Due to the fundamental principles of stereoscopy, the discomfort of using an external device and distorted quality can not be eliminated. A “True-3D” system, in which none of the restrictions on the viewer exists due to physical duplication of light distribution, are more desirable and superior compared to stereoscopy; however such display systems are much more complicated [8], and require sophisticated setups.

Holographic reconstruction by liquid crystal devices is another common method for 3D outputs. Phase modulation is achieved by using a liquid crystal with its electrically modulated molecules’ ability to affect polarization state. The problem with those devices is mainly due to large pixels limiting the image size, undesired noise, image resolution and limited field of view.

## **1.2 Thesis Overview**

In this thesis, as a continuation of previous studies, an ultimate configuration will be proposed. This would be to combine diffractive elements as CGHs in cascaded manner to create 3D outputs. More descriptively, in this configuration both device and image may be allowed to be three dimensional represented by several planar elements and images in input and output domains, respectively. Our sophisticated micro-device is able to give the

image output with the best quality for static images. In addition, that concept could have a very easy and simple setup to operate and is in the limits of current microelectronics industry when it comes to physical realization.

Sequentially, we first define basics of electromagnetic theory related to diffractive optics in Chapter 2. Additionally, discretization of systems is analyzed since computational optics treatment requires discrete versions of equations due to digital environment. In Chapter 3 which is the core chapter in this thesis, computer designed diffractive phase elements in cascaded setup, are examined extensively. Miscellaneous results on both 2D and 3D image outputs exist in this chapter. Design algorithms are given and results of enormous computational experiments are discussed. Next part, Chapter 4, still examines cascaded diffractive elements but this time phase and amplitude modulation are applied on 2D and 3D image outputs. Methods, problems and solutions are discussed. Chapter 5 is devoted to some experimental work. In this section, production stages and experimental error associated with proposed system are briefly discussed. In addition, apart from the advanced topics treated herein, a low-cost fabrication method for diffractive elements is explained with experimental results. Finally, this work is concluded by Chapter 6 with a detailed discussion and summary.

In this thesis, we prefer to put some derivations in the appendices in order not to harm the flow of concepts. Some derivations, however, are placed in the context. These derivations and formulations can be considered to fit in the related context and of utmost importance for the sake of the thesis. The figures and derivations are all originally built to explain current concepts and to extend them for the creation of a display of the future.

Throughout our calculations, commonly a computer with AMD Opteron processor with 2.39 GHz speed is used with MATLAB® software under Windows 7 environment. In general, for computation, C language with some special commands in MATLAB® image processing toolbox is employed. Furthermore, reader may find some pseudo-code for the implementation of programs for certain simulations in Appendix E at the end.

Before developing and designing the concepts of novel configurations with diffractive elements, a brief overview of diffractive optics and its implementation in digital domain are reviewed now. Thus, most of the formulae that will be used later in this work are derived in this forthcoming chapter.

# CHAPTER 2

## DIFFRACTION THEORY

### 2.1 Light as an electromagnetic wave

Light is an electromagnetic (EM) wave. An EM wave is a kind of energy flow in space consisting of mutually interacting electric and magnetic fields [9]. If that electromagnetic disturbance propagating through space can be approximated by a single wavelength, it is called monochromatic wave; on the other hand if the electromagnetic disturbance has many wavelengths, then it is called polychromatic wave. The distribution of energy among the various constituent wavelengths is called the spectrum of the radiation and various regions of the spectrum are referred to by particular names, such as radio waves, ultraviolet radiation, visible radiation to which the human eye is sensitive, etc.

EM waves are described by four fundamental equations; namely, Maxwell's equations. In this work, we will accept Maxwell's equations as postulates and build everything upon them. Each of these equations represents a generalization of certain experimental observation and constitutes fundamentals of modern optics. In free space where there is no current or charge density, they can be written compactly as,

$$\vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t} \quad (2.1a) \quad \vec{\nabla} \cdot \vec{D} = 0 \quad (2.1c)$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (2.1b) \quad \vec{\nabla} \cdot \vec{B} = 0 \quad (2.1d)$$

Above,  $\vec{E}$  is the electric field vector,  $\vec{D}$  is the electric displacement vector,  $\vec{B}$  is the magnetic induction vector and  $\vec{H}$  represents the magnetic field vector. These vectors are all functions of position and time in free space and they are related linearly with simple equations;

$$\vec{D} = \epsilon_0 \vec{E} \quad (2.2a) \quad \vec{H} = \frac{1}{\mu_0} \vec{B} \quad (2.2b)$$

Where  $\epsilon_0$  is free space electrical permittivity and  $\mu_0$  is free space magnetic permeability. After defining major variables corresponding to physical quantities, we need a general equation to express a traveling wave in free space. In order to do that, first, curl operation is applied to both sides of Eq. (2.1b).

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{E}) = -\frac{\partial}{\partial t} (\vec{\nabla} \times \vec{B}) \quad (2.3)$$

This expression may be simplified by using Eqs. (2.1), (2.2) and a vector identity;  $\vec{\nabla} \times (\vec{\nabla} \times \vec{E}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \vec{\nabla}^2 \vec{E}$  for the left hand side. We now have,

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \vec{\nabla}^2 \vec{E} = -\mu_0 \frac{\partial}{\partial t} (\vec{\nabla} \times \vec{H}) \quad (2.4)$$

$$\vec{\nabla}^2 \vec{E} = \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} \quad (2.5)$$

Where  $\mu_0 \epsilon_0 = 1/c^2$  is also used. Eq. (2.5) is known to be ‘wave equation’ since it elegantly formulates time and space evolution of a wave. The solution of this equation gives ultimate form of the electric field propagating in space. Next, temporal part of the time harmonic electromagnetic fields with angular frequency  $\omega$  may be taken as  $\vec{E}(t) \propto e^{-i\omega t}$ . After substitution, the spatial wave equation gets,

$$\bar{\nabla}^2 \bar{E} = -k^2 \bar{E} \quad (2.6)$$

Where wave-number  $k$  is defined as  $\omega/c$ . Eq. (2.6) is called Helmholtz equation. From diffractive optics point of view, spherical coordinates are more appropriate in the solution since we are interested in a point source radiating equally to every direction, as a most general form. Since electric field is uniform along  $\theta$  and  $\phi$  directions, they are no longer variables and we only need to use  $r$  in spherical coordinates to re-express wave equation [10],

$$\frac{1}{r} \frac{\partial^2}{\partial r^2} (r\bar{E}) = -k^2 \bar{E} \quad (2.7)$$

$$\frac{\partial^2}{\partial r^2} (r\bar{E}) = -k^2 (r\bar{E}) \quad (2.8)$$

Since  $\bar{E}$  is only a function of  $r$ , Eq. (2.8) may easily be solved if a wise substitution is made, such that  $\bar{E}' = r\bar{E}$ . Thus, after this substitution,

$$\frac{\partial^2 \bar{E}'}{\partial r^2} = -k^2 \bar{E}' \quad (2.9)$$

After all those steps, wave equation has been put in a solvable form. Eq. (2.9) is an ordinary differential equation of second kind and has a well-known solution in an exponential form. Therefore,

$$\bar{E}' = \bar{E}_0 e^{\pm ikr} \quad (2.10)$$

And finally,

$$\bar{E} = \bar{E}_0 \frac{e^{\pm ikr}}{r} \quad (2.11)$$

Where  $\bar{E}_0$  shows the amplitude of the field pointing perpendicular to the propagation direction. Since spherically expanding waves are of interest, (+) sign will be chosen in Eq. (2.11). Thus, the complete expression for the electric field of a spherically expanding wave coming from a point source in free space is,

$$\vec{E}(r,t) = \vec{E}_0 \frac{e^{i(kr - \omega t)}}{r} \quad (2.12)$$

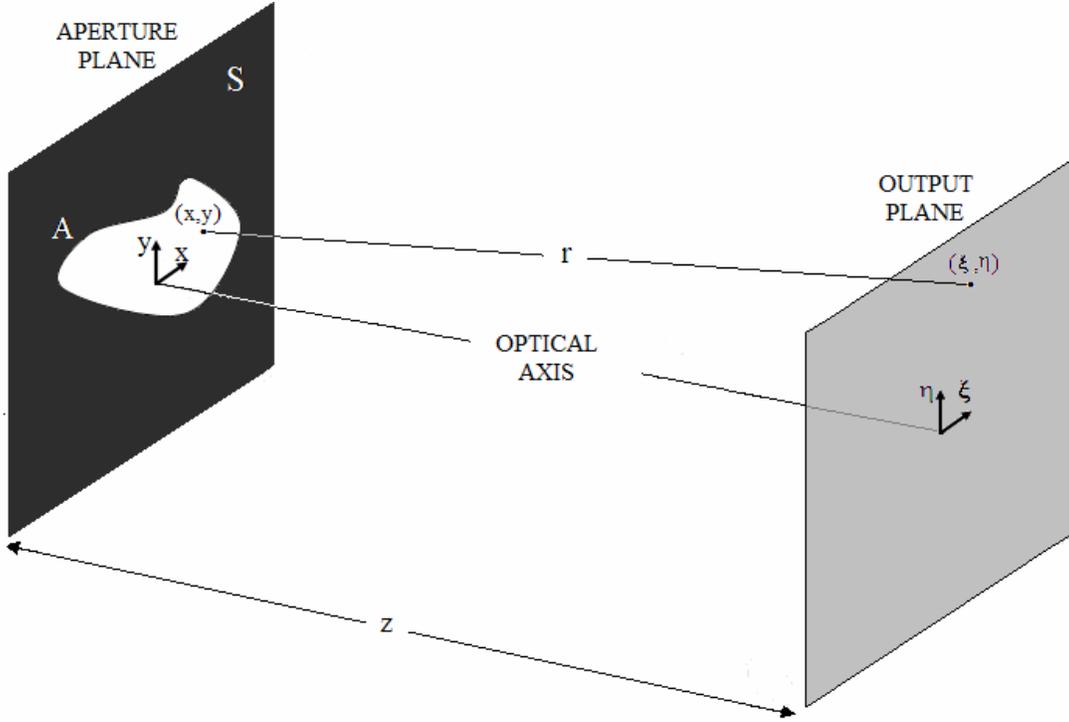
Eq. (2.12) is the final result for this derivation and of great interest for the rest of this work. Although it has been derived for a single point source, it can be used for multiple sources by using the superposition principle. We will just start making use of it as a tool in the next section while deriving the fundamental equations of diffraction. In addition, after now, the field amplitude will be denoted as  $\vec{\psi}$  instead of  $\vec{E}$ . Although, let's say  $\vec{\psi}(x, y)$ , generally means electric field as before, it is a more universal notation showing the wave disturbance at  $x$  and  $y$  coordinates.

## 2.2 Propagation in spatial domain

### 2.2.1 Diffraction from an arbitrary aperture

After reviewing the basics of EM wave theories and deriving the fundamental equations for light propagation in free space, the real problem should be defined briefly. It is desired to find the light intensity distribution at the output plane due to combination of point sources from an arbitrarily shaped aperture A on a surface S as shown in Fig. 2.1. It is assumed that there is no other source of illumination other than the A itself.

Diffraction can be considered as free space propagation after being affected by any kind of an obstacle or an arbitrary aperture. Before passing to fundamental discussion, it's necessary to review an important concept, namely Huygens principle. It basically states that each point on the wave-front of a disturbance can be considered as a new source of 'secondary' spherical disturbances and a new wave-front at a later time is found by adding these secondary wavelets cleverly. According to this principle, we can envision a wave-front hitting to the surface will make every point in the A, a self-spherical wave-



**Fig. 2.1** In figure, arbitrarily shaped hole passes incident light through the surface S. Planes are labeled with their own coordinate systems for convention. Physically, aperture plane can be anything. Output plane may be a screen, a CCD or an eye retina etc...

front emitter and thus the light intensity distribution on the output plane can be calculated by adding these ‘secondary wavelets’ one by one [11]. So for any point  $(\xi, \eta)$  at the output, the field can be written in terms of the source coordinates  $(x, y)$  as,

$$\psi(\xi, \eta) = \iint_A \psi(x, y) \frac{e^{ikr}}{r} \Omega \, dx dy \quad (2.13)$$

where,

$$r = \sqrt{(\xi - x)^2 + (\eta - y)^2 + z^2}$$

In above notation,  $r$  is the absolute value of the vector connecting points  $(\xi, \eta)$  and  $(x, y)$ . The term  $\psi(\xi, \eta)$  denotes initial wave amplitudes (in our case it corresponds to electric field  $E_0$ ) over the aperture and it is generally taken as unity for all points inside the aperture.  $\Omega$  is called inclination or obliquity factor, and related to different angular

spreads of source and target points. This factor has different values in different approaches such as, Kirchhoff or Rayleigh-Sommerfeld. According to Rayleigh-Sommerfeld solution, the factor  $\Omega$  is found to be  $\cos(\vec{n}, \vec{r})/i\lambda$  from Eq. (A12) (Appendix A), by using rigorous Green function formulation with  $\vec{n}$  being the normal of S. From Fig. 2.1, cosine factor can easily be seen to be  $z/r$ . So the final form of diffraction formula is given below.

$$\psi(\xi, \eta) = \frac{z}{i\lambda} \iint_A \psi(x, y) \frac{e^{ikr}}{r^2} dx dy \quad (2.14)$$

Eq. (2.14) is the final form for calculating any diffraction simulation. It is generally used in some kind of approximated forms for convenience. All assumptions described in this chapter are based on the fact that spreading angle is small with respect to separation  $z$  (paraxial approximation). When this assumption holds, we can write  $r \approx z$  for any multiplier at numerator or denominator. On the other hand, when dealing with the exponential term, binomial expansion (Appendix B) with two terms, is used for  $r$  in order to be more precise. In this case the magnitude for  $r$  becomes,

$$r \approx z \left[ \frac{1}{2} \left\{ \left( \frac{\xi - x}{z} \right)^2 + \left( \frac{\eta - y}{z} \right)^2 \right\} + 1 \right]$$

By putting this result into Eq. (2.14), we can reach the so-called Fresnel approximation,

$$\psi(\xi, \eta) = \frac{e^{ikz}}{i\lambda z} \iint_A \psi(x, y) e^{\frac{ik}{2z} [(\xi-x)^2 + (\eta-y)^2]} dx dy \quad (2.15)$$

Which is essentially a convolution. So, briefly, two important spatial domain propagation formulae are summarized here. The Rayleigh-Sommerfeld (Eq. (2.14)) which is useful for any propagation distance and Fresnel approximation (Eq. (2.15)) for relatively longer distances. Fresnel approximation is a handy formulation in various areas, such as understanding the lens effect. There is also another approximation which is called Fraunhofer approximation. This is generally used for almost infinite distance and because of that, the result is the same as using a lens which is the topic of next sub-section.

### 2.2.2 Effect of lens

A lens (with a focal length  $f$ ) is a device that converts a plane wavefront into a spherical wavefront of radius  $f$ . It's made of an optically transparent material, usually glass with refractive index around 1.5. By definition, a lens is said to be 'thin' if a ray exits from the same coordinates of where it enters [12]. An ideal convex (focusing) lens can be envisioned as a pure phase object with transformation function,

$$t(x, y) = e^{-\frac{ik}{2f}(x^2+y^2)} \quad (2.16)$$

By assuming a convergent lens ( $f > 0$ ).  $(x, y)$  coordinates denote to the center of the lens above which is supposed to be at the source location. Eq. (2.16) holds for paraxial approximation, where only wave-fronts that lie near the lens axis are considered. When the lens is placed at the input plane (right after the aperture plane S) which is also illuminated by a wave with amplitude of  $\psi(x, y)$ , the wave's amplitude right after the lens will become  $\psi(x, y) \Rightarrow t(x, y)\psi(x, y)$ , so the result at the focal plane can be found by using Eqs. (2.15) and (2.16).

$$\psi(\xi, \eta) = \frac{e^{ikz}}{i\lambda z} \iint_A \psi(x, y) e^{-\frac{ik}{2f}(x^2+y^2)} e^{\frac{ik}{2z}[(\xi-x)^2+(\eta-y)^2]} dx dy \quad (2.17)$$

After putting  $z = f$  and making necessary algebraic simplifications, a relatively simple form may be obtained, as written below,

$$\psi(\xi, \eta) = \frac{e^{ikz} e^{\frac{ik}{2f}(x^2+y^2)}}{i\lambda f} \iint_A \psi(x, y) e^{-i\frac{2\pi}{\lambda f}(\xi x + \eta y)} dx dy \quad (2.18)$$

Which can simply be recognized as Fourier transform (apart from the two phase terms and some constants in front of the integral) of the input. In fact, as will be developed in the next sections, Fourier transforms are commonly used in diffraction calculations and they are suitable for computer simulations with enormous theoretical research done for them for last two decades.

In Fraunhofer approximation case for relatively larger propagation distances, however, the transform in Eq. (2.18) is directly used without assuming any lens. This topic will be revisited from quantitative point of view later in Chapter 6.

In practical cases, some distance generally exists between the original source represented as an aperture  $A$  and a lens. In this case, field propagation through that distance should be considered as well.

### 2.3 Propagation in Fourier domain

When working with linear systems, it may sometimes be beneficial to decompose a signal to its constituent functions in order to evaluate the response of the system to each of these functions. A Fourier transform pair is commonly used for one of such decomposition methods. The transform and its inverse are defined as follows [12].

$$\mathfrak{S}[f(x, y)] = F(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi(f_x x + f_y y)} dx dy \quad (2.19a)$$

$$\mathfrak{S}^{-1}[F(f_x, f_y)] = f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(f_x, f_y) e^{i2\pi(f_x x + f_y y)} df_x df_y \quad (2.19b)$$

In the classical notation shown above,  $x$  and  $y$  correspond to the source coordinates again. Here,  $f$  and  $F$  are called Fourier transform pairs.  $f_x, f_y$  are spatial frequencies and they are related to angles of propagation scaled by wavelength for a propagating wave. Note that, as a special case, for lens Fourier transform, spatial frequencies are calculated at  $f_x = \xi / \lambda f$  and  $f_y = \eta / \lambda f$ , in Eq. (2.18) at the target location. For example  $f_{x,\max} = \xi_{\max} / \lambda f$ , so, highest spatial frequency is calculated by dividing the image size with wavelength and focal distance away from the lens.

By considering the complexity of integrals in Eqs. (2.14) and (2.15), propagation simulations are generally implemented in Fourier domain for convenience. Derivation follows the wave equation again. We want to examine a field on the  $x$ - $y$  plane and

propagating along  $z$ , by again, referring to Fig. 2.1. This time,  $\nabla^2$  operator in Eq. (2.5) will be written in Cartesian coordinates as,

$$\left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi = -k^2 \psi \quad (2.20)$$

Next step is to write the inverse Fourier decomposition of the wave function.

$$\left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi e^{i2\pi(f_x x + f_y y)} df_x df_y = -\left( \frac{2\pi}{\lambda} \right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi e^{i2\pi(f_x x + f_y y)} df_x df_y \quad (2.21)$$

Where  $\psi(x, y)$  and  $\Psi(f_x, f_y)$  have been used as a Fourier transform pair and wave number  $k = 2\pi / \lambda$  has been replaced. The derivative operation can be calculated directly for  $x$  ( $y$  component is likewise) and  $z$  components as shown below by noting that just  $\Psi$  is  $z$  dependent.

$$\frac{\partial^2}{\partial x^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi e^{i2\pi(f_x x + f_y y)} df_x df_y = -(2\pi f_x)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi e^{i2\pi(f_x x + f_y y)} df_x df_y \quad (2.22a)$$

$$\frac{\partial^2}{\partial z^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi e^{i2\pi(f_x x + f_y y)} df_x df_y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^2 \Psi}{\partial z^2} e^{i2\pi(f_x x + f_y y)} df_x df_y \quad (2.22b)$$

After putting Eq. (2.22) in Eq. (2.21) and equating the terms in integrals,

$$\frac{\partial^2 \Psi}{\partial z^2} = \left( \frac{1}{\lambda^2} - (f_x^2 + f_y^2) \right) \Psi \quad (2.23)$$

Finally, solution of Eq. (2.23) gives,

$$\Psi(f_x, f_y; z) = \Psi(f_x, f_y; 0) e^{-i2\pi z \sqrt{\frac{1}{\lambda^2} - (f_x^2 + f_y^2)}} \quad (2.24)$$

Where the exponential factor is the transfer function and is generally shown by  $H$ . Eq. (2.24) is the exact form of Fourier propagator and it explains the dynamic evolution of the wave in frequency domain. It is valid as long as the condition  $f_x^2 + f_y^2 \leq 1/\lambda^2$  is

satisfied. This equation is sometimes used in an approximated form. When propagation distance is relatively large, spatial frequencies are extremely small and binomial expansion may be used in the exponent, and approximated Fresnel propagator becomes

$$\Psi(f_x, f_y; z) = \Psi(f_x, f_y; 0) e^{-ikz} e^{-i\pi\lambda z(f_x^2 + f_y^2)} \quad (2.25)$$

Note that for calculations in this work, Fourier propagator in Eq. (2.24) will be used extensively but Fresnel propagator may appear in some special situations for convenience.

## 2.4 Optics in digital domain

The equations derived so far are for continuous domain in which there is no restriction on pixels or spatial frequencies. On the other hand, these integrals are generally impossible to solve and therefore numerical methods are used for computations. We need some other tools here, to convert the physical information about diffractive optics to such a state so that a computer is able to process it for numerical manipulations. Therefore, when implementing diffractive optics simulations on the computer, it is necessary to represent functions by uniform discrete arrays of sampled values and apply processing methods designed for these discrete signals. When sampling a signal (in our case as a form of image), care should be taken for Nyquist-Shannon sampling theorem which states that for a perfect reconstruction of a band-limited signal, the signal must be sampled at a sampling rate more than or equal to two times of its bandwidth (Appendix C). So from Eq. C4,

$$\frac{1}{\Delta x} \geq 2f_{\max} \quad (2.26)$$

Where  $\Delta x$  is sampling interval,  $1/\Delta x$  is sampling rate and  $f_{\max}$  is the bandwidth. The starting point of discussion is a sampled version of a two dimensional function  $g(x, y)$  in spatial domain, which may be written as,

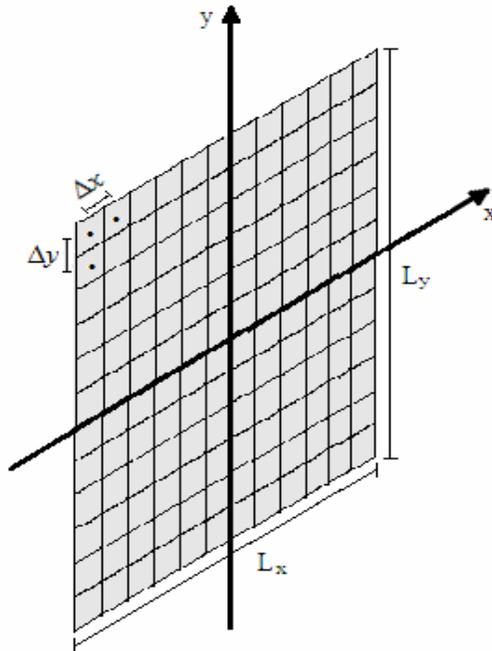
$$g(x, y) \Rightarrow g(m\Delta x, n\Delta y) \text{ or } \text{comb}\left(\frac{x}{\Delta x}\right) \text{comb}\left(\frac{y}{\Delta y}\right) g(x, y) \quad (2.27a)$$

where,

$$m = -\frac{M}{2}, \dots, \frac{M}{2} - 1, \quad n = -\frac{N}{2}, \dots, \frac{N}{2} - 1 \quad (2.27b)$$

$$\text{comb}\left(\frac{x}{\Delta x}\right) \text{comb}\left(\frac{y}{\Delta y}\right) = \Delta x \Delta y \sum_{m=-M/2}^{M/2-1} \sum_{n=-N/2}^{N/2-1} \delta(x - \Delta x m) \delta(y - \Delta y n) \quad (2.27c)$$

In above definitions, sample intervals are denoted by  $\Delta x$  and  $\Delta y$  along axes, and  $m$  and  $n$  are numbers corresponding to indices of samples. Number of samples will be shown by  $M$  and  $N$  (or number of pixels in some applications where every pixel is represented by one sample. This is not always the case; sometimes more than one sample is required to represent a pixel), for  $x$  and  $y$  coordinates respectively, so that  $L_x = M\Delta x$  and  $L_y = N\Delta y$ . These variables are shown on a conceptual grid in Fig. 2.2. According to the figure and Eq. (2.27), a function, say  $g(x,y)$  again, is sampled at points  $x = \Delta x m$  and  $y = \Delta y n$ . Evidently the values of the function at these points are registered for processing.



**Fig. 2.2** A conceptual grid represents a function on Cartesian coordinate system, where every point has a corresponding value. The function is divided to cells, called picture elements or pixels. Pixels are represented by the center dot in sampling process.

Next step is to write the transforms in discrete domain. In fact, the Discrete Fourier Transform (DFT) is one of the most useful tools in computer simulations, where it's implemented by a special algorithm called Fast Fourier Transform (FFT). Before deriving the DFT, we should take a look at frequency domain where discretization is not as obvious as in spatial domain. The convention for spatial frequencies is,

$$f_x = \frac{p}{M\Delta x}, \quad f_y = \frac{q}{N\Delta y} \quad (2.28a)$$

Where

$$p = -\frac{M}{2}, \dots, \frac{M}{2} - 1, \quad q = -\frac{N}{2}, \dots, \frac{N}{2} - 1 \quad (2.28b)$$

Maximum spatial frequency for  $x$  coordinate is  $1/2\Delta x$ . This is in agreement with Nyquist-Shannon sampling theorem stated briefly before. Same argument is valid for  $y$  coordinate too. After these fundamental definitions, a DFT can be written by using continuous Fourier transforms in Eq. (2.18) and discrete Riemann sums. After making necessary changes for position and frequency, below definitions are obtained.

$$\mathfrak{S}(g(m, n)) = G(p, q) = \sum_{m=-M/2}^{M/2-1} \sum_{n=-N/2}^{N/2-1} g(m, n) e^{-i2\pi\left(\frac{pm}{M} + \frac{qn}{N}\right)} \quad (2.29a)$$

$$\mathfrak{S}^{-1}(G(p, q)) = g(m, n) = \frac{1}{MN} \sum_{p=-M/2}^{M/2-1} \sum_{q=-N/2}^{N/2-1} G(p, q) e^{i2\pi\left(\frac{mp}{M} + \frac{nq}{N}\right)} \quad (2.29b)$$

The factor in front of inverse DFT in Eq. (2.29b) follows from the spectral theorem and essential for energy conservation. By using the same logic, propagation transfer function in Eq. (2.24) can also be sampled. For every allowed  $p$  and  $q$  values, the below given function is calculated which results in a numerical grid.

$$\Psi(f_x, f_y; z) = \Psi(f_x, f_y; 0) e^{-i2\pi\sqrt{\frac{1}{\lambda^2} - \left(\left(\frac{p}{M\Delta x}\right)^2 + \left(\frac{q}{N\Delta y}\right)^2\right)}} \quad (2.30)$$

Thus, the transfer function is now used for propagation simulations for an image on a computer. Calculation and application of this function will be discussed further in Chapters 3 and 4 and in Appendix E.

One important concern here is again sampling of this transfer function and the allowable values of variables. These are vital for the validity of simulation results, because even a correct simulation may cause incorrect experimental results. Due to the complex exponential, Eq. (2.24) can be defined as Chirp-like function [13]. The phase term can be written for  $x$ -component as,

$$\varphi = 2\pi z \sqrt{\frac{1}{\lambda^2} - f_x^2} \quad (2.31)$$

Sampling a chirp function, as required for a propagation simulation, can be tricky due to the increasing slope of the phase with frequency. Two definitions are made to find sampling condition; A dual of sampling theorem and local position. With comparison to Eq. (2.26), sampling condition in frequency space is,

$$\frac{1}{\Delta f_x} \geq 2|x_{\max}| \quad (2.32)$$

And, we are defining position as,

$$|x| = \frac{1}{2\pi} \frac{\partial \varphi}{\partial f_x} \Rightarrow |x|_{\max} = \frac{1}{2\pi} \left| \frac{\partial \varphi}{\partial f_x} \right|_{\max} \quad (2.33)$$

Now, solving Eqs. (2.32) and (2.33) together gives,

$$\Delta f_x \left| \frac{\partial \varphi}{\partial f_x} \right|_{\max} \leq \pi \quad (2.34)$$

This expression states that the maximum change in the absolute phase must be no more than  $\pi$  between any two adjacent samples. Violation of this constraint causes aliased phase values [14]. From Eq. (2.31), the slope of the phase function is,

$$\left| \frac{\partial \phi}{\partial f_x} \right| = \frac{2\pi z f_x}{\sqrt{\lambda^2 - f_x^2}} \quad (2.35)$$

Maximum slope occurs when  $f_x$  is maximum. Eq. (2.35) is used in Eq. (2.34), yielding,

$$\Delta f_x \leq \frac{\sqrt{\lambda^2 - f_{x,\max}^2}}{2 z f_{x,\max}} \quad (2.36)$$

Assuming frequency and spatial domain sampling are related through the scaling properties of the FFT, i.e.,  $\Delta f_x = 1/L_x$ ,  $f_{x,\max} = 1/2\Delta x$  from Eq. (2.28), in which  $L_x$  is the side length and  $\Delta x$  is the sampling interval. Finally the required sampling condition is,

$$\Delta x \geq \frac{\lambda}{2} \sqrt{1 + \left(\frac{2z}{L}\right)^2} \quad (2.37)$$

when  $z \gg L$ , Fresnel sampling condition is obtained by ignoring '1' in the square root.

$$\Delta x \geq \frac{\lambda z}{L} \quad (2.38)$$

In this chapter, some important parts in diffraction theory formulation have been summarized with special emphasis on digital domain applications. Sampling conditions have been shown both for spatial and frequency domains.

In summary, two kinds of propagation transfer function exist. Fourier transfer function simulates propagation of a wave for any distance and will be used throughout calculations in the next chapters. On the other hand Fresnel transfer function is for relatively larger distances and may not be applicable for our purposes except for special cases introduced and discussed later. From sampling point of view, some restrictions may be applied for propagation in Fourier domain. Propagation in spatial domain, however, is not easy due to tough formulation procedure.

The information and knowledge treated in this chapter is going to be used now in the demonstration and creation of some novel ideas about CGHs in Chapter 3 which is the main part of the entire work.

# CHAPTER 3

## IMAGING WITH PHASE ELEMENTS

### 3.1 Introduction

DOEs or CGHs are mainly used to modulate wavefronts in a most general sense [15]. They find applications in numerous areas such as aspherical surface testing, optical interconnections, beam shaping, head-up displays, 3D image reconstruction and security based optical image encryption systems. From an imaging point of view, the importance of holography comes from the fact that, its reconstructions include all key visual information used by the human visual system. This is especially valuable in the case of 3D outputs, where the viewer needs to perceive depth information. When it becomes ‘computer-generated’, a digital environment is used for design and even a physically non-existing object is possible to reconstruct. CGHs can be classified as phase only, amplitude only or both phase and amplitude, based on the device’s function. In order to create phase modulation, different relief profiles can be used in pixel locations on a silica glass substrate for instance, which also allows almost complete transmittance of incident power thus maximizing the diffraction efficiency. On the other hand, for amplitude modulation, some black pixels (or gray-scale) might be used. In some cases, halftoning can also be employed to emulate different gray tones. Phase and amplitude modulation together may be possible by dividing each pixel into sub cells and encoding the

information in a sophisticated way as will be shown in the next chapter briefly. However, this decreases resolution limit of the device and severely reduces diffraction efficiency; the encoding scheme brings some extra noise too. Thus, phase only CGHs are generally preferred, considering diffraction efficiency and pixel resolution of the device.

Work on diffractive elements has evolved from using a single element, to reconstruct one target plane, in which the goal is typically to achieve a specified intensity or phase pattern at the output [16]. In addition, using stratified structures has been an issue of interest for almost 20 years. Some significant examples in literature might be appropriate to summarize as follows. A system of two phase elements has been studied in some  $4f$  and similar configurations in order to compensate for the lack of amplitude modulation by adding a second phase element for efficiency [17]. Diffraction properties of stratified volume holographic optical elements have been researched and, for example, angular selectivity of such devices has been demonstrated as a function of buffer layer parameters [6]. Several works have examined angular and wavelength selectivity with possible practical applications for iteratively designed multiple CGH planes in a cascaded setup [7, 18 and 19]. In addition, a general design procedure has been proposed, and some multiplexing characteristics have been shown, for classes of volumetric optical devices [20, 21]. For applications to optical interconnections, several operations can be done effectively such as simultaneous wavelength-division demultiplexing by using cascaded relief profiles [22].

On the other hand, some other efforts have studied the possibility of 3D reconstruction by specifying the intensity pattern in multiple target planes, using one diffractive element [23]. In this case, typically two planes have been studied as a minimal set to evaluate feasibility. Because the output is specified as the intensity in these multiple planes, instead of as the amplitude and phase in one target plane (which would determine the intensity in the other target planes due to propagation), there is no guarantee of optical compatibility of the specified intensities in the multiple target planes. Thus these efforts include techniques like the ‘Ping-pong algorithm’ that iteratively find phases at the target planes in order to minimize the possible optical incompatibility of

specified intensities [24]. Some applications and extensions can also be found in literature [25-27]. The concept of creating multiple images has found numerous applications in various areas in addition to displays such as multiple-image hiding in optical security systems as an example [28].

There are several methods used in the design of CGHs. These methods might broadly be classified within two general approaches as being deterministic or stochastic [29]. Iterative Fourier Transform (IFTA) and Simulated Annealing (SA) algorithms are common examples of the former and the latter, respectively. The SA algorithm tends to provide higher accuracy in exchange for longer computation time. But as today's computers become more powerful day by day, time consuming calculations for complicated structures can be implemented in return of moderate computation times. Although, some other algorithms or approaches exist in literature, there is no perfect reconstruction of the target whatever method is used, due to several reasons such as destruction of amplitude information for phase only elements, quantization noise coming from relief profiles or cross-talking between different target image planes in case of 3D imaging. In addition to aforementioned optical incompatibility, there are also manufacturing defects causing degradations in the final reconstruction from the physical device.

In this chapter, by assuming a coherent, monochromatic light with zero angle of incidence, we first start with the simplest case where only one CGH plane and target plane exist, and then cascaded CGHs will be introduced as form of phase elements (PEs) to create 2D images computationally. Finally, we present different approaches for multi-plane imaging to simulate 3D views by again using cascaded CGHs. Multi-layer device and multi-plane image concepts, which are summarized in the previous two paragraphs, are combined and cascaded PEs are used to improve overall output quality. In this process every plane is treated separately for both device and image domains by extending the current algorithms. The final result is an ultimate version of an imaging system: 3D device and 3D image. This can be thought as a continuation of the design algorithms and some device configurations explained above; our first algorithm combines and extends

aspects of both IFTA and Ping-pong algorithm. We will call it extended IFTA. Our second algorithm is a derivative of the SA algorithm and will be called extended SA. By using appropriate design approach, algorithm and parameters, it is shown that reconstruction noise and crosstalk between multiple target images can be reduced. One other relevant approach would be to use dummy areas which can be put around target images to improve performance by either increasing the number of pixels at the hologram plane or shrinking the target image dimensions [23]. With clever usage of this extra area in amplitude imposition stages, it is possible to remove some noise from the desired image area. However, one of the obvious consequences is reduced resolution at the image area, or higher resolution requirements at the hologram plane. Another factor is the diffraction efficiency for the target image, which is substantially lower for image reconstructions with dummy areas. Finally, the noise that is distributed to the dummy area around the image region might expose an undesired effect for specific display applications, especially for the multi-plane imaging case. We present that extending the hologram in the ‘z-dimension’ (as a cascade of CGHs) instead, leads to error reduction without loss of resolution or significant amount of diffraction efficiency. In addition, this enhanced approach can be combined with a dummy area and is open to further optimization.

### 3.2 Definitions

While working with digital images, we will use two main types of parameters, namely, image quality metric and diffraction efficiency in assessing image reconstruction performance. For the image quality metric, there are mainly three approaches: mean squared error (MSE), correlation coefficient (Co) and signal to noise ratio (SNR).

For the MSE, although there are slightly different versions, the version below will be in use throughout this work, so that,

$$MSE(\%) = \frac{\sum_{All\ pixels} (|\psi_{target} - C|\psi_{reconstructed}|)^2}{Number\ of\ pixels} \times 100 \quad (3.1)$$

According to the above definition, target image's absolute field values (square root of intensity) are subtracted from reconstructed image field values pixel by pixel. The denominator functions as a normalization factor and ensures results to be independent from size. The coefficient  $C$  is chosen such that the MSE becomes independent of scaling factor of reconstruction (Appendix D).

$$C = \frac{\sum_{\text{all pixels}} |\psi_{\text{target}} - \psi_{\text{reconstructed}}|}{\sum_{\text{all pixels}} |\psi_{\text{reconstructed}}|^2} \quad (3.2)$$

If a curve containing several points of MSE values is considered, a rate definition might also be useful in order to characterize the curve quantitatively. At that point, we define MSE percentage decrease (error reduction) rate by comparing the first and the last data on the corresponding decrease curve. This parameter is a part of image quality metric and tells what percentage of the noise can be removed from an image under consideration.

$$R(\%) = \frac{|MSE_{\text{last data}} - MSE_{\text{first data}}|}{MSE_{\text{first data}}} \times 100 \quad (3.3)$$

Apart from the MSE, another image quality metric is  $Co$  which is used extensively in literature. It basically compares similarity in intensities which is absolute square of fields of the target and reconstructed images ( $I = |\psi|^2$ ).

$$Co(I_{\text{target}}, I_{\text{reconstructed}}) = \frac{\text{cov}(I_{\text{target}}, I_{\text{reconstructed}})}{\sigma_{I_{\text{target}}} \sigma_{I_{\text{reconstructed}}}} \quad (3.4)$$

Where cov denotes covariance between target and reconstructed images.

$$\text{cov}(I_{\text{target}}, I_{\text{reconstructed}}) = \sum_{\text{all pixels}} (I_{\text{target}} - \langle I_{\text{target}} \rangle) (I_{\text{reconstructed}} - \langle I_{\text{reconstructed}} \rangle) \quad (3.5)$$

So, the similarity is computed by first subtracting the averages from individual pixel values for target and reconstructed images then multiplying and adding together. In the denominator of the Co expression, standard deviations work as normalization factors. According to the definition, the standard deviations of the target and reconstructed intensity patterns are found by,

$$\sigma_{I_{target}} = \sqrt{\sum_{all\ pixels} (I_{target} - \langle I_{target} \rangle)^2} \quad (3.6a)$$

$$\sigma_{I_{reconstructed}} = \sqrt{\sum_{all\ pixels} (I_{reconstructed} - \langle I_{reconstructed} \rangle)^2} \quad (3.6b)$$

Note that unlike the MSE, Co increases as the reconstruction gets improved and becomes unity for the perfect case, corresponding to zero error.

The next one, signal to noise ratio (SNR), is a simple metric. Calculation involves division of target field to existing noise which can be considered to be target minus reconstructed fields.

$$SNR = \frac{\sqrt{\sum_{All\ pixels} |\psi_{target}|^2}}{\sqrt{\sum_{All\ pixels} (|\psi_{target}| - C|\psi_{reconstructed}|)^2}} \quad (3.7)$$

Thus, root mean squared (RMS) value of the target is basically divided by RMS value of the noise. The coefficient  $C$  was defined before.

Another parameter is the diffraction efficiency ( $\eta$ ). It is defined as the zeroth diffracted order divided by the entire optical energy on the image plane including all diffracted orders.

$$\eta = \frac{I_{zeroth}}{I_{total}} \quad (3.8)$$

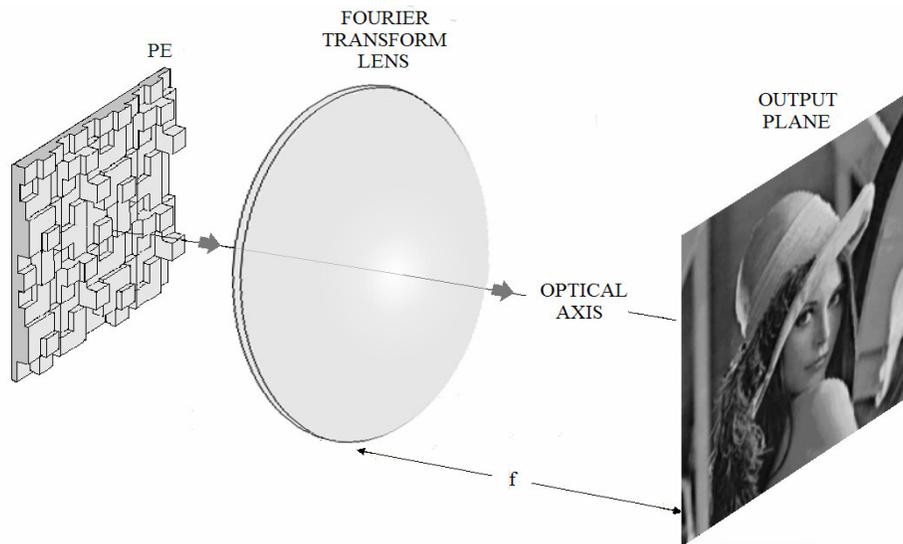
So, these parameters will become essential measure to evaluate performance and certain characteristics of images coming from digital simulations. For image quality metric, MSE

will play the main role in the majority of results given in Chapters 3-5 but Co and SNR will be exemplified too for ease of comparison to some partially-related works in the literature. For quantitative evaluations about these parameters, Chapter 6 can be referred. As a footnote, MSE will sometimes be denoted as ' $E$ ' for error briefly in the text or in the diagrams.

### 3.3 Two Dimensional Images

#### 3.3.1 One CGH case

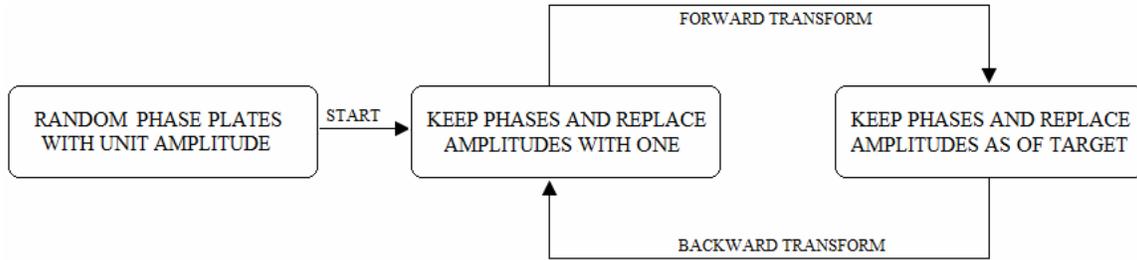
Forming two dimensional images by using one CGH as a form of PE is a well known area of diffractive optics. One is usually interested in determining phase from two intensity measurements and this process is called phase retrieval. The phase retrieval technique is also applied commonly in electron microscopy, astronomy or in crystallography in addition to diffractive element design. For the purposes of this work, the steps of phase retrieval of a computer generated diffractive phase element, which will be denoted as PE from now, is demonstrated here. A general setup is shown in Fig. 3.1 below.



**Fig. 3.1** Basic image generation setup for a PE consisting of several possible elevation levels for each individual pixel (in figure they are shown by four levels for illustration purposes). Grey arrow on the optical axis shows the propagation direction of the light. A lens with focal length  $f$ , serves as a propagation tool as a Fourier transform operator. 'Lena' image is used for illustration.

In the figure the PE is depicted with four elevation levels conceptually. In reality, the number of elevation levels which denotes number of phase quantization levels can be anything as powers of 2 in binary optics fabrication process (see, Chapter 5 for more information). The target image resides on the right side of the scheme. The two intensity measurements here are the target image and the PE; the latter is all 1s at a 0-1 intensity scale. Being a pure phase object, it is ideally supposed to transmit all the flux. In the middle, Fourier transform lens helps imaging by making propagation distance shorter and indirectly having every pixel contribute to the image formation. Principally, since a thin lens takes Fourier transform of the device in the front, from Chapter 2 (Eq. (2.17)), taking the inverse Fourier transform of the desired image should give the modulation required by the device. On the other hand, since amplitude information is to be destroyed for pure phase element device, resulting performance will be degraded. Then, the main problem is to find the phase distribution that gives the best quality image reconstruction. More technically, given a set of constraints placed on an object and another set of constraints placed on its Fourier transform, which is the target image intensity, the aim is to find an object (and its transform) that satisfies all constraints and conditions.

One way of solving the problem and increasing the quality is to use iterative methods such as IFTA. And one basic way of implementing it, is called Gerchberg-Saxton or error-reduction algorithm for phase element design [30]. Due to the constraints on the object side, which are mainly phase quantization and amplitude extraction, a perfect, error-free solution is not possible, theoretically. Once the best solution is found, the uniqueness of the solution is often relatively unimportant for synthesis problems. The flow chart is shown in Fig. 3.2. In its simplest form, a loop starts with a random numerical grid of phase values. Forward Fourier transform is taken and desired amplitude distribution is imposed at the output plane while keeping the phases unchanged. An inverse Fourier transform is then applied and all intensity values are reset to one for a pure phase device at the input plane. So, phases are constantly iterated while desired amplitudes are imposed at the input and output. The mathematical Fourier transforms are representing lens transforms referring to Fig. 3.1.



**Fig. 3.2** Diagram of a simple bidirectional IFTA algorithm consisting of perpetual transforms to design and optimize PEs for imaging.

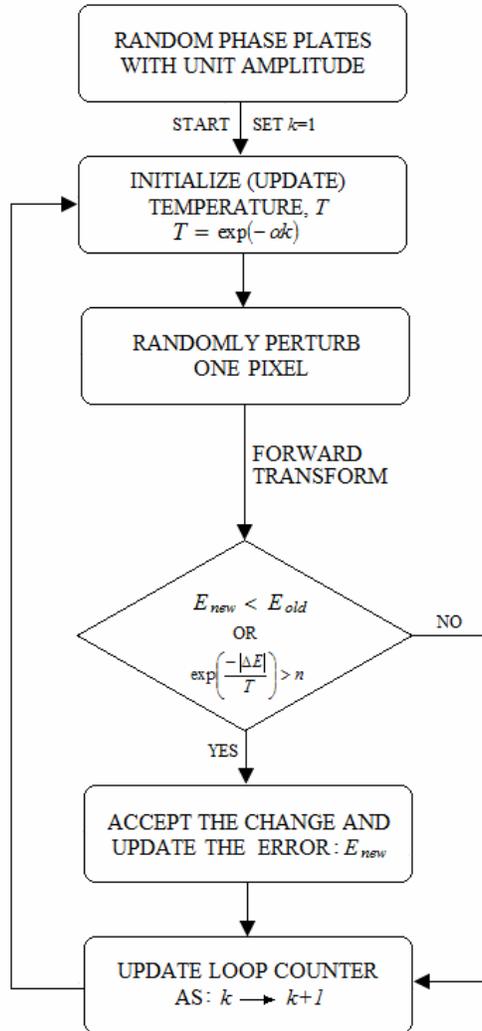
IFTA is a member of bidirectional algorithms set since it consists of forward and backward transforms. Fienup showed that MSE converges and is a monotonically decreasing function of iteration number [31]. The loop halts when error reduction stops which happens after enough iterations. In some cases, the process stops when a pre-defined error value is reached or in some other cases, the loop is ended when the decrease of error between several successive iterations approaches a limiting value.

Since IFTA is a deterministic algorithm, the main issue with it is the stagnation problem where solution falls into a local minimum and misses the global one in all solution space. This fact might be important in some situations. In the next paragraph we will examine briefly the basics of one of the stochastic approaches which can partially solve the stagnation problem. After all, IFTA type methods are easy to apply and fast to compute in spite of stagnation. Therefore, it has found and is still finding extensive usage in the literature. There are some other kinds or variations of the IFTA available also. Input-output approach is one of them [32]. It is basically the same with the previously defined error-reduction approach, except that the field at the input plane is further modified before a subsequent forward transform in the loop, so that a faster convergence is obtained. Apart from those, an alternative approach in solving the phase-retrieval problems is to employ one of the gradient search methods. Being a non-iterative algorithm, the steepest-descent method, seeks minimum error by computing partial derivatives forming the gradient of error function through solution space [31]. Afterwards, assumed solution proceeds in a direction opposite to that of the gradient, minimizing the error. Steepest descent method is likely to fall into local minima too and physically

similar to error-reduction method. Since it is deterministic in nature, stagnation problem persists.

Algorithms that are deterministic in nature tend to get stuck in local minima in solution space, providing limited performance. In order to fix this problem partially, some other design techniques might be used. SA is one of them and unlike deterministic approaches; it is stochastic in nature [33]. Another difference is the flowing directions of the two algorithms; the previously explained IFTA above, is classified as bidirectional in character since it goes forward and backward, whereas SA flows only forward, making it a unidirectional algorithm. SA is a general optimization technique which is used broadly in science and engineering. Its approach evolved from crystallization of metals. Molten metal crystallizes as it cools down. Slower cooling increases the order of molecules making the material more stable, whereas rapid cooling results in discontinuous crystal structures causing brittle materials [29]. So, at the beginning while in liquid phase, the orientation of each molecule is an independent variable, as they freeze under control, the molecules orient themselves into larger and stronger crystals. In CGH design, molecules correspond to pixels and stability is represented by image quality reconstructed by those pixels. In the optimization process of the classic system in Fig. 3.1, the steps are shown in flow chart below in Fig. 3.3. So, the implementation of SA is initiated by a random phase plate. One pixel is randomly changed and the performance of this new configuration is tested; the new configuration is accepted to be permanent if its error is lower and conditionally accepted if it is not. Then another loop starts and another pixel is perturbed for trial, etc., until the error converges (stops decreasing and becomes almost constant according to certain metrics). The fact that even the higher error configurations may be accepted, allows the system to escape from local minima. The probability of acceptance of a worse candidate is an exponentially decreasing function of  $T$ . In the literature,  $T$  denotes temperature due to physical roots of annealing process pertaining to cooling process of metals. The condition for acceptance in our case is [34]

$$\exp\left(\frac{-|\Delta E|}{T}\right) > n \quad (3.9)$$



**Fig. 3.3** Diagram of flow chart for SA. ‘ $k$ ’ shows the loop number.  $E_{new}$  means the error of the new trial and  $E_{old}$  is the current error. The new attempt will be accepted if the conditional statement holds. The loop stops after certain number of iterations where the error doesn’t change anymore.

Where  $n$  is a random number between 0 and 1,  $\Delta E$  is the difference between new attempt and previous error values ( $E_{new} - E_{old}$ ). Numerical value of  $T$  should be decreasing as iterations take place and in our case it is an exponentially decreasing function of loop counter;  $T = \exp(-\alpha k)$  with  $\alpha$  being a case-specific constant and  $k$  is the iteration number.

Note that, Eq.3.9 represents the ease of acceptance of higher error configurations and is more probable to be satisfied in the beginning and becomes harder as iterations keep going and temperature decreases. Acceptance of higher error trials makes the system avoid from any possible stagnation partially. The methodology summarized here is typical although some slightly different versions exist especially in the stages of pixel(s) selection for perturbation, temperature adjustment or conditional statement. In addition, there are other variants one of which is called quantum annealing (QA) [35, 36]. Being inspired by tunneling of electrons in quantum mechanics, instead of changing a pixel at a time, a region of pixels is modified this time. The size of the perturbed-region is determined by tunneling field strength which itself is a parameter, and can be treated similarly like the others. It may start high, with a large neighborhood and is slowly reduced through the computation or it may directly depend on the result which makes it adaptive in nature.

One of the important parameters in CGH design is the number of phase quantization levels of the element, where number of represented phases is allowed to be powers of 2, as mentioned before. Specifically, for IFTA, quantization constraint is either applied during iterations which makes it a direct design, or it can be applied after all iterations are finished, in which case it has been named as indirect design [37]. On the other hand, in SA case, quantization is inherent from initial conditions since each pixel is treated individually one by one and quantization level is chosen randomly. So, nothing can be said about direct or indirect design. In this work, indirect design is employed for IFTA and frequently 64 and 4 phase quantization levels are used in order to exemplify the two different situations where the effects of phase quantization is negligible or dominant, respectively. However, there will be cases that other numbers (between 64 and 4) are used sometimes. In addition, an ideal situation without any phase quantization

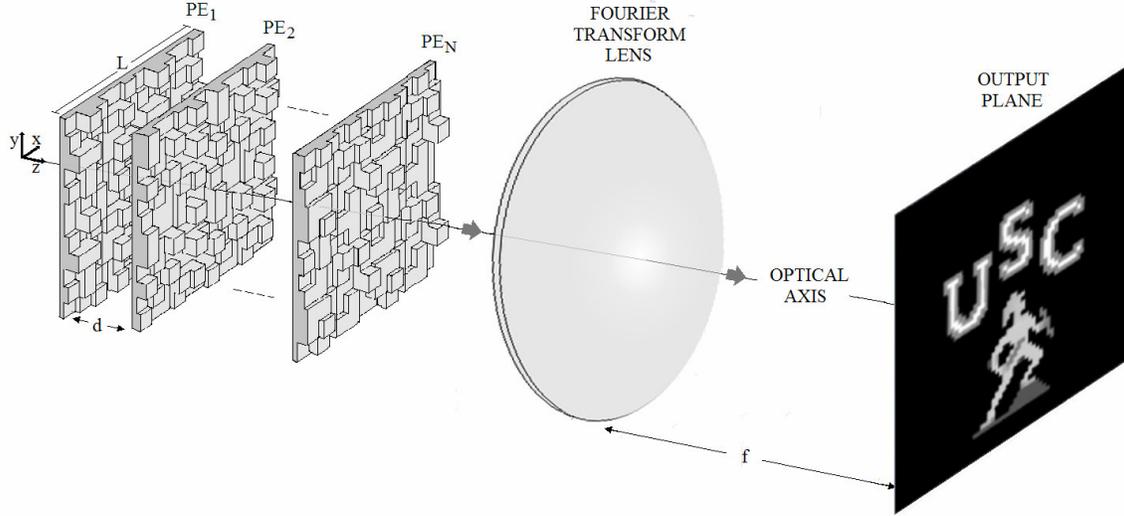
constraint will also be examined in some special situations where mathematical abstraction is needed. These points will be emphasized again when required.

The above outlined algorithms cause error reduction up to some point. Unfortunately, when the reconstructed images are observed, too much remaining noise may be seen. The most important part of this noise comes from the constraints on object as mentioned. For further improvement, as mentioned before, number of pixels may be increased or target image dimensions may be decreased, which consequently enables designer to use a dummy area in addition to image itself. This increases the degrees of freedom by extra pixels contributing to imaging. By using that extra area in amplitude imposition stage properly, it is possible to remove some noise from the desired image area. But there are some issues arising when a dummy window is intended to use. One of the issues is the reduced resolution at the target since some of the area would be used as a noise region. Instead, one might use extra pixels by keeping the image as it is. In that case, higher resolution is required for the device and since there is a limit for pixel dimension for scalar diffraction theory, that resolution limit may not be exceeded. Another point is the diffraction efficiency for the target, which is lower for image reconstructions with dummy areas. In addition, the removed noise is collected and distributed to dummy area region around image, which might expose an undesirable effect for specific display applications. Especially in 3D imaging cases, this approach may not function as well. As will be shown, noise can also be eliminated by using cascaded phase elements, which are placed successively before a Fourier transform lens.

### 3.3.2 Multi-CGH case

Now the idea of imaging with CGHs as a form of PEs can be extended to cascaded CGHs. New configuration is shown in Fig. 3.4.  $N$  thin\* CGHs with distance  $d$  between each other, are placed successively. These elements are conceived to be replicated in the  $x$ - $y$  plane to handle periodic boundary conditions which is not shown in the figure. An image with 'USC' letters with Trojan statue is intended to be reconstructed. Note that for  $N$  equals to 1, the situation is exactly like in Fig. 3.1, so that we will have chance to compare performances.

\* By 'thin', we mean that the propagation within the structure is unimportant and thus can be neglected.



**Fig. 3.4** Hypothetical system configuration.  $N$  thin PEs, each  $L$  by  $L$ , with a separation  $d$ , are concatenated in order to create an output plane.

First of all, we will consider the formation and formulation of output of such a system. Then some design techniques will be mentioned. From left to right a unit plane wave is first processed by  $PE_1$  then  $PE_2$  until  $PE_N$ . Between elements, free space propagation is used. By assuming that scalar diffraction theory holds, the most general transfer function for propagation over a distance  $d$  is seen from Eq. (2.24).

$$H(f_x, f_y) = \exp\left(i2\pi d \sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2}\right) \quad (3.10)$$

as long as the spatial frequencies satisfy the condition  $(f_x^2 + f_y^2)^{1/2} < 1/\lambda$ . According to the system theory, that transfer function is applied in the Fourier domain between elements, and phase modulations caused by the elements themselves are applied in the space domain. So, for instance, the field after the  $i$  th element ( $\psi_i$ ) can be written in terms of the field after the  $i-1$  th element ( $\psi_{i-1}$ ) as,

$$\psi_i = \mathcal{S}^{-1}[\mathcal{S}(\psi_{i-1}) \times H] \times PE_i \quad (3.11)$$

After being modulated by successive elements, the wave is handled by a Fourier transform which may require a positive thin lens for robustness and improved field of

view (FOV). Finally, mathematical definition of the field at output after  $N$  elements is shown in Table 3.1 below as a compact formulation.

In general, PEs may also be characterized mathematically as phase modulation profiles. In this case, for the  $i$  th element, phase delay ( $\chi$ ) of an individual pixel at coordinates  $(m, n)$  is,

$$\chi_i(m, n) = \exp(ik'\Theta_i(m, n)) \quad (3.12)$$

where  $\Theta_i(m, n)$  is surface profile index showing the local thickness and,

$$k' = \frac{2\pi(n-1)}{\lambda}$$

The wave-number in Eq. (2.6) can also be expressed as above in terms of wavelength and refractive index of the material,  $n$ . The  $(n-1)$  factor comes from the fact that the phase delay is calculated with respect to air which has a refractive index of unity. Eq. (3.12) is used in place of PEs in Table 3.2 below as an explicit formulation. Here, summation symbols are used to express Fourier transform pairs. Also, Eq. (2.29) is used with  $M = N$ , saying that the image is square for our practical purposes. Finally, the transfer function,  $H$  for Fourier propagation may be seen with corresponding variables under summation. Note that, in the continuous case all the summation symbols turn into integrals.

**Table 3.1** Output formulation regarding the system in Fig. 3.4. In compact form, the transforms are represented by symbols. The transforms related to lens, are also denoted. In expressions,  $\times$  shows multiplication.

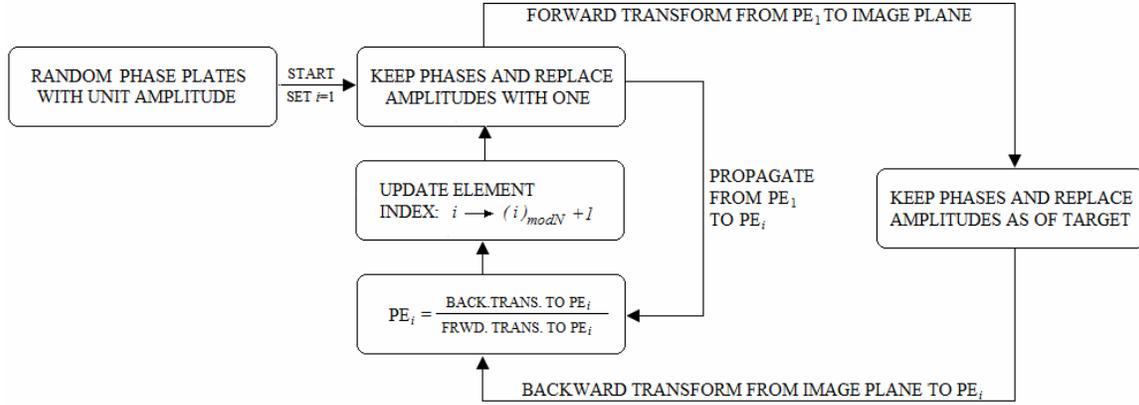
Number of PEs	Compact Output Formulation
1 PE	$\psi = \mathfrak{S}_{Lens} \{ PE_1 \}$
2 PEs	$\psi = \mathfrak{S}_{Lens} \{ \mathfrak{S}^{-1} [ \mathfrak{S} ( PE_1 ) \times H ] \times PE_2 \}$
⋮	⋮
N PEs	$\psi = \mathfrak{S}_{Lens} \{ \mathfrak{S}^{-1} [ \mathfrak{S} ( \dots \mathfrak{S}^{-1} [ \mathfrak{S} ( PE_1 ) \times H ] \times PE_2 \dots ) \times H ] \times PE_N \}$

**Table 3.2** Output formulation regarding to system in Fig. 3.4. In explicit form, the transforms are expressed clearly (referring to Chapter 2). Mathematically, there is no difference between Fourier transforms coming from lens and free space propagation. Note that  $\Theta_1$ ,  $\Theta_2$  and  $\Theta_n$  are phase delays of the related elements.

Number of PEs	Explicit Output Formulation
1 PE	$\psi(s, t) = \sum_{m=-M/2}^{M/2-1} \sum_{n=-M/2}^{M/2-1} \exp(ik'\Theta_1(m, n)) e^{-i2\pi(\frac{sm+tn}{M})}$
2 PEs	$\psi(s, t) = \frac{1}{M^2} \sum_{k=-M/2}^{M/2-1} \sum_{l=-M/2}^{M/2-1} \left\{ \sum_{p=-M/2}^{M/2-1} \sum_{q=-M/2}^{M/2-1} \left[ \left( \sum_{m=-M/2}^{M/2-1} \sum_{n=-M/2}^{M/2-1} \exp(ik'\Theta_1(m, n)) e^{-i2\pi(\frac{pm+qn}{M})} \right) \right. \right.$ $\left. \left. x \exp\left( i2\pi d \sqrt{\frac{1}{\lambda^2} - (p^2 - q^2)} \right) e^{i2\pi(\frac{kp+lq}{M})} \right] x \exp\left( \frac{2\pi}{\lambda} \Theta_2(m, n) \right) \right\} e^{-i2\pi(\frac{sk+tl}{M})}$
• • •	• • •
N PEs	$\psi(s, t) = \left( \frac{1}{M^2} \right)^{N-1} \sum_{f=-M/2}^{M/2-1} \sum_{g=-M/2}^{M/2-1} \left\{ \sum_{f=-M/2}^{M/2-1} \sum_{g=-M/2}^{M/2-1} \left( \left[ \sum_{u=-M/2}^{M/2-1} \sum_{v=-M/2}^{M/2-1} \left( \dots \right. \right. \right. \right.$ $\left. \left. \left. \sum_{p=-M/2}^{M/2-1} \sum_{q=-M/2}^{M/2-1} \left[ \left( \sum_{m=-M/2}^{M/2-1} \sum_{n=-M/2}^{M/2-1} \exp(ik'\Theta_1(m, n)) e^{-i2\pi(\frac{pm+qn}{M})} \right) \right. \right. \right. \right.$ $\left. \left. \left. x \exp\left( i2\pi d \sqrt{\frac{1}{\lambda^2} - (p^2 - q^2)} \right) e^{i2\pi(\frac{sp+qg}{M})} \right] x \exp(ik'\Theta_2(m, n)) \dots \right] e^{-i2\pi(\frac{fu+gv}{M})} \right]$ $\left. \left. \left. x \exp\left( i2\pi d \sqrt{\frac{1}{\lambda^2} - (f^2 - g^2)} \right) e^{i2\pi(\frac{xf+yg}{M})} x \exp(ik'\Theta_N(m, n)) \right\} e^{-i2\pi(\frac{sx+ty}{M})} \right.$

After expressing what the result should be after being processed by cascaded PEs, now, the design stages of such structures are discussed. After all, what is desired by the ultimate user is to input a unit-amplitude plane wave and reconstruct a predefined image at output by using the proposed system. In the design process, phase modulation profiles ( $\chi$ ) of PEs are found in order to optimize correct local thicknesses of surface ( $\Theta$ ) for correct etching (this topic will be revisited with the fabrication topic in Chapter 5).

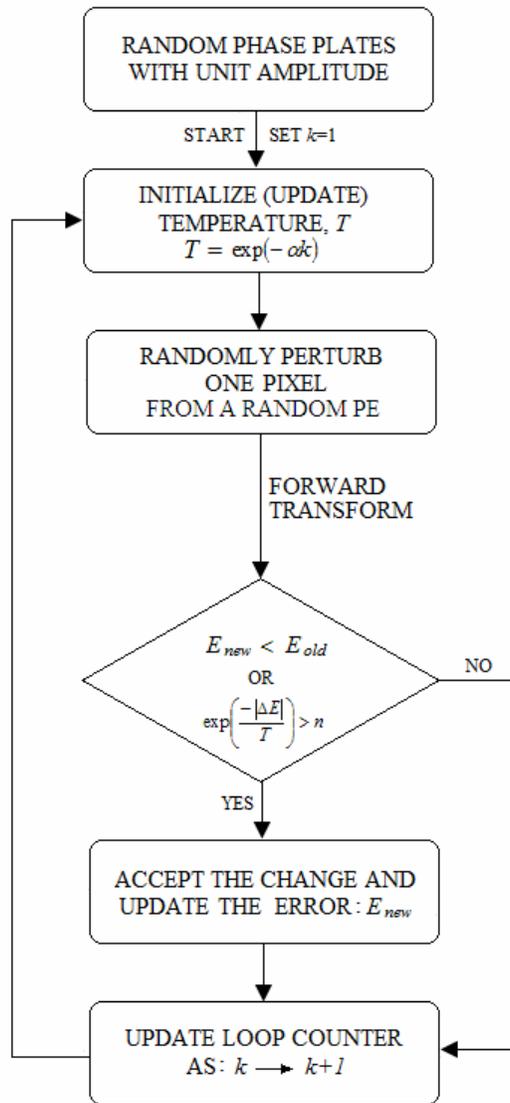
Similar cascaded structures have been studied for various reasons, such as multi-wavelength reconstruction and multiplexing, or diffraction efficiency. One way to optimize the system under consideration is to use IFTA type algorithm, where iterations go back and forth. In this case, though, each iteration consists of  $N$  sub-iterations, in which every PE is computed individually in loops, as shown in flow diagram in Fig. 3.5. After propagating all the way from the first PE to the last one and lens transform, leading to image formation as shown in tables, the field is manipulated such that its phases are kept and amplitudes are replaced with the target. After that, backward transform is applied by evaluating inverse Fourier transform coming from the lens. Then, a sequence of back propagations is performed, all the way back to the element that is being calculated. Apart from that, a sequence of forward propagations is calculated up to the same element. The resulting back propagated wave is then divided by the forward propagated wave to yield the complex values for that PE. Finally, amplitude information is discarded and phase information is saved. Next sub-iteration starts as before in order to determine the next PE. When every PE is found, one complete iteration ends and another one starts, until the MSE becomes constant. This format is the multi-CGH correspondent of the usual IFTA, applied with using the same image in every iteration.



**Fig. 3.5** One sub-iteration cycle is shown to optimize one PE in Fig. 3.4. Each element is calculated with an operation where the field coming from the image plane is divided with the field propagated up to that element. At the end, the element index is updated and another sub-iteration starts for the subsequent PE. When every component in the stack is finished, iteration is completed and the next one starts. This is controlled by mod ( $N$ ) factor.

As mentioned before, there are advantages and disadvantages of a deterministic method. On the other side, a probabilistic approach, SA method can also be applied after an expansion to include extra PEs. As shown in Fig. 3.6, our algorithm starts with random phase plates. Then, one random pixel is perturbed from a randomly chosen PE. The performance of this new configuration is tested after calculation of image formation (by using the formulations given in tables) and accepted to be permanent if its error is lower and conditionally accepted if it is not. Then another loop starts. Program stops when error converges. The exponentially decreasing definition of temperature and condition for acceptance in Eq. (3.9) are still valid and in use as before.

The fact that only one pixel is changed makes the flow slower since now the number of possible pixels to be changed is multiplied by the number of PEs. In spite of its reduced speed, that approach maximizes the possibility of finding a better solution which is our sole purpose of using an SA algorithm. Instead of one pixel, a group of them can be changed leading to a faster but less precise outcome. The conditional acceptance step is more important now with respect to single PE due to a larger solution space. And again because of the same reason,  $T$  should decrease slower depending on the number of PEs. This can be achieved by changing the value of constant  $\alpha$ .



**Fig. 3.6** Flow-chart for the SA algorithm in optimization of cascaded PEs for single-plane imaging. ' $k$ ' shows the loop number. The pixel is randomly chosen from a random element for a perturbation.

Now, computational experimentation stage is starting with some sample images. These samples will generally be 64 x 64 pixels size. But sometimes smaller ones will be used in order to reduce the computation time in some lengthy operations and to increase variety.

Fig. 3.7 shows the original images that are intended to be reconstructed by using the setup based on Fig. 3.4. In part (a) USC logo-type figure with a miniature Trojan statue and in part (b) USC Viterbi School of Engineering logo are displayed. These two examples have different characteristics which will be emphasized more. For now, suffice it to say that, the first one has more black space whereas the second resembles much to an ordinary image, with miscellaneous gray tones. They represent two main types of images from the point of this work.



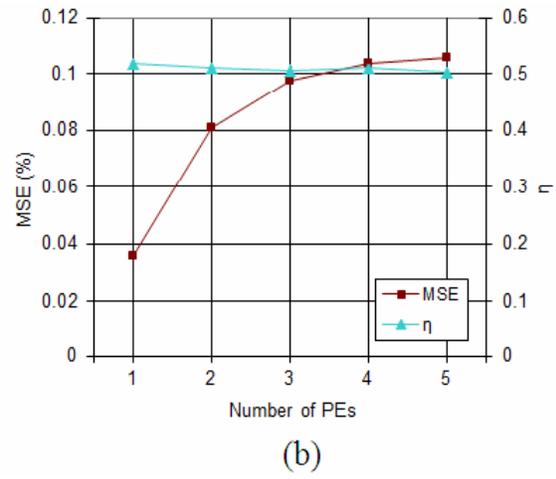
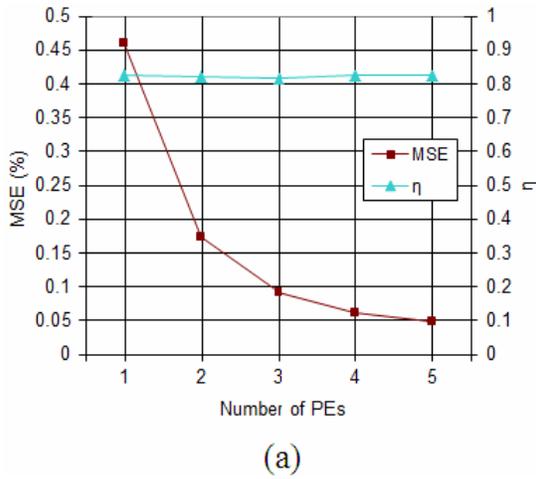
**Fig. 3.7** Two different sample images have been used in experiments. (a) First image is a USC, Trojan mascot and (b) Second one is a Viterbi School of Engineering logo.

Simulation results are shown below in Figs. 3.8, 3.9 and 3.10, where above outlined procedures have been applied to two different images by changing the number of cascaded CGHs. In addition, two types of phase quantization levels are considered. In the former, 64 levels are processed with IFTA-type only and in the latter, 4 levels are examined with both IFTA and SA-type methods. The parameters related to setup are as follows: wavelength is 0.632  $\mu\text{m}$ , image sizes are 64 x 64 pixels, size of one pixel is 5

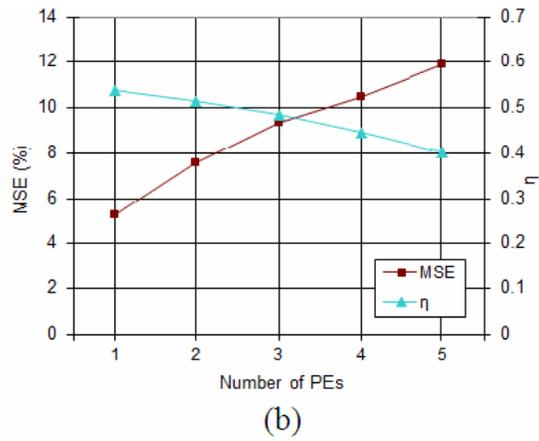
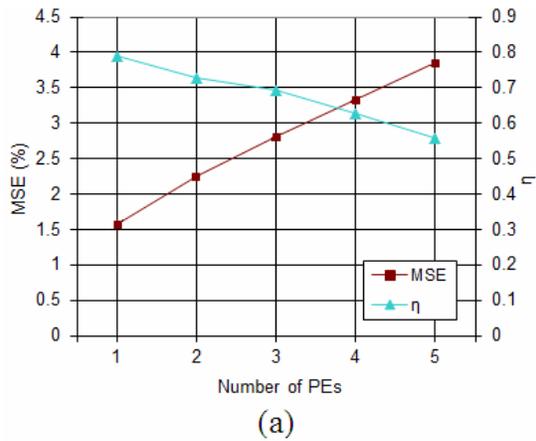
$\mu\text{m}$  (thus  $L$  is  $320 \mu\text{m}$ ), sampling size is  $1.25 \mu\text{m}$ , distance  $d$  between adjacent PEs is  $0.3 \text{ mm}$ , and focal length  $f$  of the lens has been chosen to be  $25 \text{ mm}$ . The image size can be estimated based on chosen figures. The pixel size of the output which equals to  $\lambda f/L$ , gives  $50 \mu\text{m}$  approximately (from Chapter 2, using  $f_x = \xi/\lambda f$  and  $\Delta f_x = 1/L_x$ ; The smallest  $f_x$  is used for the entire image dimensions). Final image sizes are  $3.2 \text{ mm}$ , which are of practical limits. In addition, sampling requirements are satisfied. By looking at Eq. (2.37) and using the system parameters, the left hand side ( $\Delta x$ ) is  $1.25 \mu\text{m}$  and the right hand side of the equation gives  $0.67 \mu\text{m}$ .

Fig. 3.8 shows effects of adding more PEs in a systematical way in graphical formats when there are 64 allowed phase quantization levels with iterative technique. In part (a), the first image is processed. We see that MSE decreases in an exponential-like manner, while  $\eta$  stays almost the same. From the classical 1 PE case to 5 PEs, error reduction is almost 10 times. It is reasonable to expect that, this decrease in error will stop and approach to an asymptote after certain number of elements. On the other hand, in part (b), results of the second image with the same parameters can be seen. According to the part (b), MSE starts from a relatively low value and increases with additional PEs.

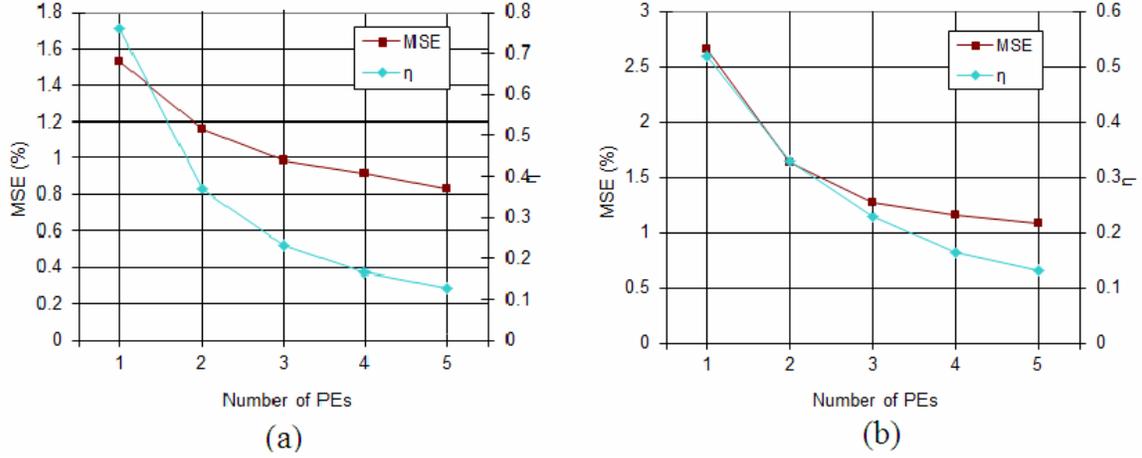
We also observe this phenomenon with 4 allowed phase quantization levels with iterative type method. In that case as shown in Fig. 3.9, similar trends can be seen for both types of images: a constant linear increase in MSE and decrease in  $\eta$  for both parts. This increase in error should be caused by the fact that the iterative algorithm couldn't handle limited levels of phase values. As an attempt to decrease the error, the SA-type approach can be applied. For low quantization levels this approach is relatively easy and less time consuming. Note that this can also be done using 64 levels as above but it would take too much computation power and time without any significant benefit. Results coming from application of SA -type method are seen below in Fig. 3.10. When the SA-type approach is applied, similar trends for both images appear; so that both MSE and  $\eta$  decrease smoothly. Before some discussion of these results, though, the reader can see visual outputs which are displayed in Figs. 3.11 through 3.13. They basically compare results of 1 PE as a classical case and 5 PEs as an advanced case.



**Fig. 3.8** Results of computer experiments with 64 allowed phase quantization levels. Iterative algorithm is used. Error and diffraction efficiency are displayed versus number of phase elements for (a) first image and (b) second image.



**Fig. 3.9** Results of computer experiments with 4 allowed phase quantization levels. Iterative algorithm is used. (a) Error and diffraction efficiency are displayed versus number of phase elements for (a) first image and (b) second image.



**Fig. 3.10** Results of computer experiments with 4 allowed phase quantization levels. SA type algorithm is used. (a) Error and diffraction efficiency are displayed versus number of phase elements for (a) first image and (b) second image.

There are a few points that need to be discussed. First of all, results in Fig. 3.8 (a) are encouraging such that as more elements are added, noise appears to decrease in a smooth way. But when another image is used, noise level increases. Actually, this can be understood better by creating a concept. For that purpose, we name images as optically “easy” and “hard”. Hard images have black pixels dominantly and thus hard to recover (it is due to the fact that when iterative transform is applied, phases are kept and amplitudes are exchanged. However, a black pixel in the image means zero multiplier, so it implicitly affects phase recovery by killing information in pixels). Most images are in between easy and hard, generally close to easy. According to that reasoning, the hard image in Fig. 3.8 (a) starts from a relatively high noise level due to the screening of black pixels and expansion of the device as a form of cascaded PEs yields improvement. However, the easy image in Fig. 3.8 (b) already starts from a very low MSE value and as more PEs are added, complexity increases and propagation between planes does not enhance quality. When Fig. 3.9 is considered, on the other hand, iterative method is applied with 4 quantization levels this time and both error and diffraction efficiency get worse. This makes sense because in iterative method there is not a special mechanism to handle low number of quantization levels. Basically adding more PEs cause more noise due to the hardness in adjusting phases in a limited scheme. On the other side, when stochastic approach is used as in Fig. 3.10 to these images, noise decreases. The decrease

in MSE can be explained by the ability of SA to handle low quantization condition since it proceeds pixel by pixel in nature. If Fig. 3.8 and Fig. 3.10 are compared, it is understandable that the total error gets bigger in the latter due to limited number of allowed phase levels but a more interesting point is seen when parts (a) and (b) are compared; in Fig. 3.8, the starting magnitude of MSE is way higher in part (a) than part (b) with respect to Fig. 3.10. So, easy and hard images exhibit different characteristics. Indeed this observation supports the hypothesis that in iterative transform approach black pixels kill the phase information causing more noise in hard image with respect to ordinary one. This phenomenon is not valid for SA due to difference in its flow.

In visual outputs, these effects can be observed too. In Fig. 3.11, the improvement coming from extra PEs is apparent from (a) to (c) for the first image and a slight deterioration can also be seen from (b) to (d). In the next figure, Fig. 3.12, one may notice quite noisy images due to low number of allowed quantization levels, and this is corrected in the Fig. 3.13 with stochastic SA approach where every pixel is treated individually. In the figure, 5 PEs improvement is also seen on the final result clearly. A brief quantitative discussion may also be required to explain why the results in Fig. 3.12 (b) and (d) are quite noisy. The main reason might be the fact that low number of quantization constraint is not able to handle miscellaneous gray tones appearing in the image. And since the environment is relatively pale, noise is apparent visually. This fact is reflected in numerical MSE value as well since the difference of target and reconstructed fields are calculated.

Note that in results of SA shown in Fig. 3.12 (c) and in Fig. 3.13 (c), (d), final images coming from 5 PEs have significantly low diffraction efficiency. Because of that, image intensities have been equaled just for comparison convenience. This can also be seen from the graphs above. Especially in SA type approach, efficiency is substantially low due mainly to the fact that system is constantly pushed to decrease MSE from the acceptance condition. The acceptance condition can be modified [38] so that it also contains diffraction efficiency. In this case error reduction would partly be sacrificed though.



(a)



(b)



(c)



(d)

**Fig. 3.11** Resultant Images of computer experiments with 64 allowed phase quantization levels, referring to Fig.3.8. Iterative algorithm is used. (a) first image processed by 1 PE (b) second image processed by 1 PE (c) first image processed by 5 PEs (d) second image processed by 5 PEs.



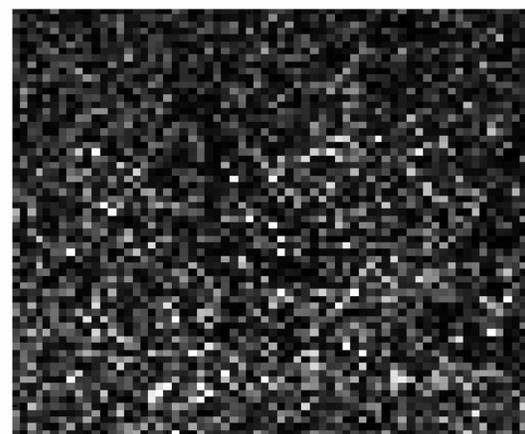
(a)



(b)



(c)



(d)

**Fig. 3.12** Resultant Images of computer experiments with 4 allowed phase quantization levels, referring to Fig.3.9. Iterative algorithm is used. (a) first image processed by 1 PE (b) second image processed by 1 PE (c) first image processed by 5 PEs (d) second image processed by 5 PEs.



(a)



(b)



(c)



(d)

**Fig. 3.13** Resultant Images of computer experiments with 4 allowed phase quantization levels, referring to Fig.3.10. SA type algorithm is used. (a) first image processed by 1 PE (b) second image processed by 1 PE (c) first image processed by 5 PEs (d) second image processed by 5 PEs.

From 2D imaging point of view, benefits coming from cascaded phase elements concept is limited. The results show that the concept works either by sacrificing too much diffraction efficiency for low number of quantization levels or for specific class of images for high number of quantization levels. We will now examine 3D applications for this concept. It will be seen that, since both phase and amplitude must be defined on the image side in the 3D case, results are much noisier and cascaded phase elements concept helps more in noise removal.

### **3.4 Three Dimensional Images**

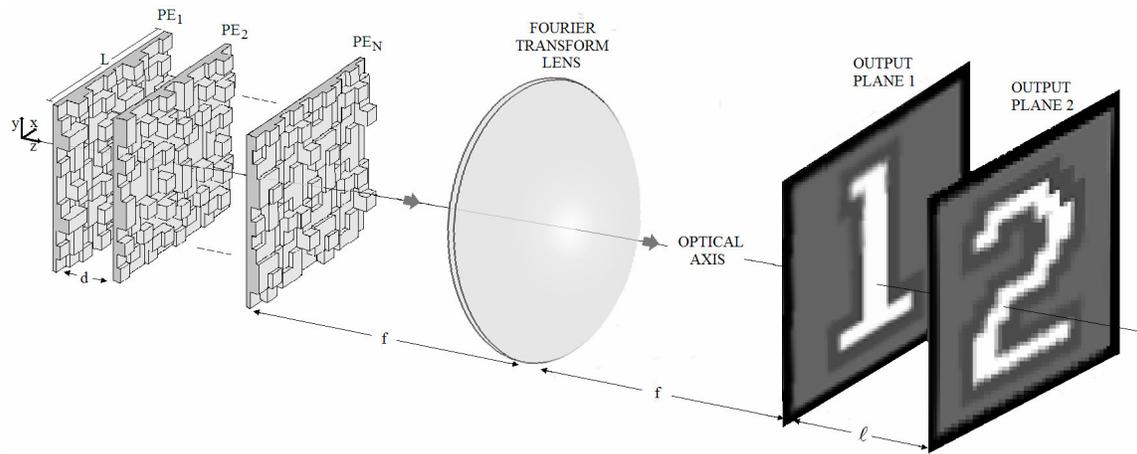
As mentioned in the introduction of this chapter, in the reconstruction of 3D images by using diffractive elements, there are methods that have been used. In these methods, one of them seems the most feasible according to ease of application and performance. In this method, a 3D target image is considered to consist of several 2D planar images. Iterations are applied between successive image planes from one to another and thus it's named as 'Ping-Pong' algorithm [24]. Although this algorithm has some different versions, its essence and final performance are similar. In the classical case, one diffractive element and two image planes are considered. Information on target image planes propagates, bounces back and forth from each other and finally, the obtained phase and amplitude distribution coming from this process is back transformed to device plane. As a result, when the device is illuminated by a coherent light, it should be able to reconstruct desired images at planes with different depth. Different from the 2D case, now diffractive element must create a desired phase distribution in addition to amplitude at the first image plane for further propagation to second image plane. These images do not necessarily have to be entirely optically compatible. This means that a perfect reconstruction might be impossible for both planes whatever device or devices are used.

In our work, we use two and three image planes to represent any 3D entity. It would be expected that these two approaches give similar results (or trends) if not exactly identical. Actually, these separate considerations turn out to be beneficial also in the search for a general formulation of trends when an ordinary 3D image is to be processed by cascaded PEs.

### 3.4.1 3D abstraction with two image planes

As an application of cascaded CGHs, one might use the idea in the previous section to create 3D image as a compilation of two 2D planes. The configuration is depicted in Fig. 3.14. On the left, a group of CGHs is seen, and on the right, two targets, namely, images ‘1’ and ‘2’, are placed at output planes. The aim here is to find a correct structure of phase elements for the best target reconstructions. Since both device and image domains consist of multiple elements, we require a brand new algorithm to get the best possible image quality. This algorithm is supposed to unify classical IFTA, which is used for one image plane, with Ping-Pong in an iterative manner. We will call it extended IFTA.

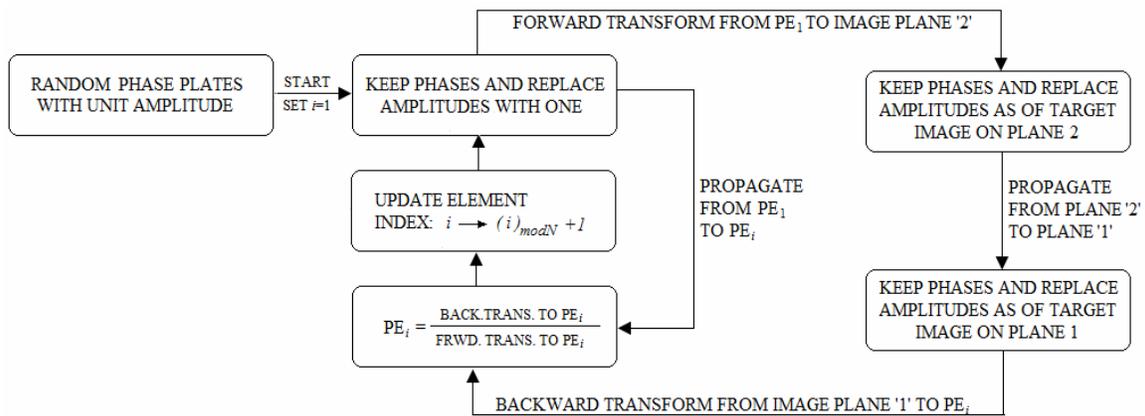
First of all, apart from design stage, the output coming from the device, i.e., physical processing through PEs, is the same as single output plane case. This is already shown in table 3.1 and table 3.2. Final result, two or three dimensional, is formed after the relevant design steps. Note also that for  $N = 1$ , the classical case with one PE is obtained again.



**Fig. 3.14** Hypothetical system configuration.  $N$  thin PEs, each  $L$  by  $L$ , with a separation  $d$ , are concatenated in order to create output, namely output (target) planes 1 and 2 with separation  $\ell$ .

The flow-chart diagram in Fig. 3.15 summarizes design steps, which is basically an extension of previous methods. As has already been described briefly in our articles [39, 40], the algorithm starts with random phase plates and proceeds with iterations. In one

iteration, every element is found one by one and the algorithm then passes to the next iteration until convergence. In an iteration, there are  $N$  sub-iterations. Basically, each sub-iteration is a bidirectional loop to calculate a specific element in the stack and consists of two steps. The first step flows in the forward direction. Initially, a unit plane wave is first affected by  $PE_1$  then  $PE_2$  until  $PE_N$ . Between elements, Eq. (3.10) is used again leading to complex fields shown in Tables 3.1 and 3.2. Different from that, this time, the field continues to propagate to image planes '1' and '2' as a far field output, thus forming two output planes instead of just one. Propagations between target planes are again controlled by the transfer function in Eq. (3.10), but with replacement of  $\ell$  instead of  $d$ . The second step in the sub-iteration flows primarily in the backward direction. First, intensity enforcement at target plane 2 is fulfilled by keeping phases and replacing amplitudes. The same is done for target plane 1 after back propagation between planes. This is followed by an inverse Fourier transform. Then, a sequence of back propagations is performed, all the way back to the element that is being calculated. Apart from that, a sequence of forward propagations is calculated up to the same element. The resulting back propagated wave is then divided by the forward propagated wave to yield the complex values for that PE. Finally, amplitude information is discarded and phase information is saved. Next sub-iteration starts as before in order to determine the next PE. When every PE is found, one iteration is completed.

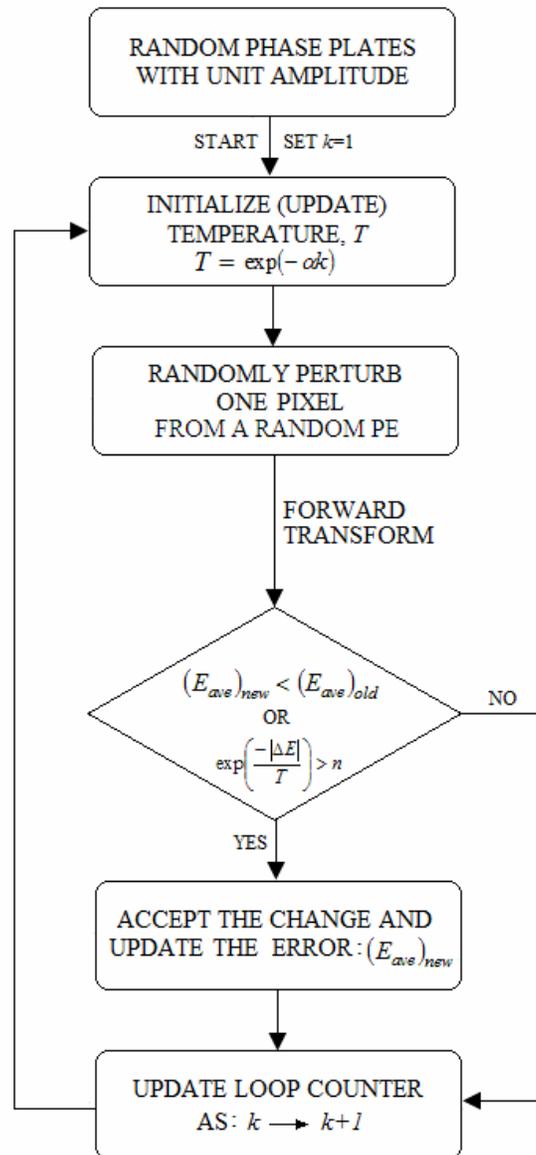


**Fig. 3.15** One iteration cycle of the extended IFTA which is used to optimize the system in Fig. 3.14. This flow scheme is capable of handling successive 2D planes for both input and output in a general sense.

As seen from the above explanation, the design steps are similar to that of single output plane. The only difference is the inclusion of the second output plane in an iterative manner. This version of the algorithm is the most general form of any kind of iterative algorithm where both 3D object and 3D image domains are scanned. As a result of multi-plane target images, phases on image domain gain importance, so while simulating the propagation from the device to the first image plane, a phase term is added to the usual Fourier transform as shown in Eq. (2.18). In order to remove this term, one needs to take the distance between the last element,  $PE_N$ , and the lens plane into consideration. Briefly, combining the expressions for the output of a lens in Eq. (2.18) and Fresnel propagation between  $PE_N$  and the lens by assuming it is valid for convenience, as in Eq. (2.25), it would be seen that one needs to have a separation equal to the focal length  $f$ , for the distance between the device and the lens. This fact is used in calculations and shown in figures as well.

That problem at hand might also be considered by using stochastic approach as before. In this case, there is more than one option for optimization. Flow chart is shown in Fig. 3.16. As usual, it starts with random phase plates and perturbation of one random pixel from a random PE. The performance of this new configuration can be tested in various ways. Since there are two targets, the weights of these targets can be adjusted according to importance. In ideal case though, quality expectations are the same and equal importance are given to output images. Hence, the best strategy is to average the two MSE values coming from images and use that value for comparisons throughout the computation. This causes noise to be distributed equally to two output planes. The rest of the algorithm is the same; new configuration is accepted to be permanent if its error is lower and conditionally accepted if it is not. The structure of temperature is the same as well. That approach will be named as extended SA.

Additional variations are possible on extended SA; a cluster of pixels can be chosen for perturbation making the algorithm converge faster in return of quality loss. One thing should be emphasized again that, as more PE is used, convergence is expected to be longer so parameters are adjusted accordingly due to a larger solution space.



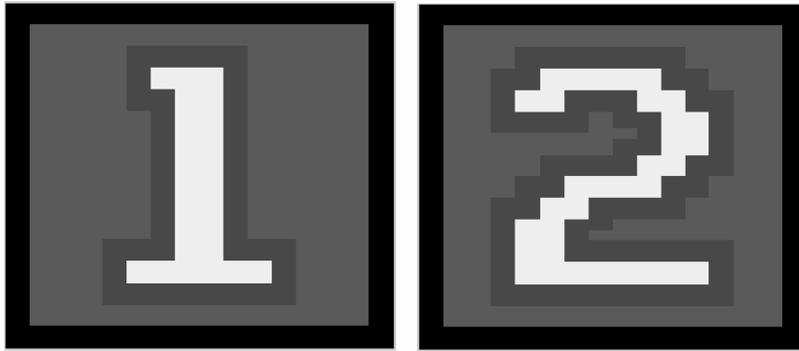
**Fig. 3.16** Flow-chart for the extended SA algorithm in optimization of cascaded PEs for multi-plane imaging. ‘ $k$ ’ shows the loop number. In addition to the last flow chart, this time errors for the two output planes are averaged for comparison.

Now that we pass to simulations part, some computer experiments with different qualities and quantities will be presented by using the above-explained algorithms. In addition, effects of different algorithms will be discussed as before. We used different image sets. It must be noted that no easy or hard image concept will be observed in this section. Every set of images shows similar behavior if not the same. In addition to amplitude formation at the output plane, phase reconstruction is also needed in 3D case and this makes the problem more difficult.

The setup in Fig. 3.14 has been simulated numerically. 64 quantization levels are handled with extended IFTA and 4 quantization levels are handled by extended SA. It is beneficial to state again that, high number of quantization levels cannot be handled by stochastic approach since it takes a lot of computation time, and the result can be expected to be not that different from the deterministic approach in practice. On the other hand, in cases of low number of quantization levels, deterministic method which is the extended IFTA in our case, does not give significant contribution. This fact can be seen in Fig. 3.9 for 2D imaging situation. We want to inform that this is also valid in 3D case so that MSE increases with addition of extra PEs.

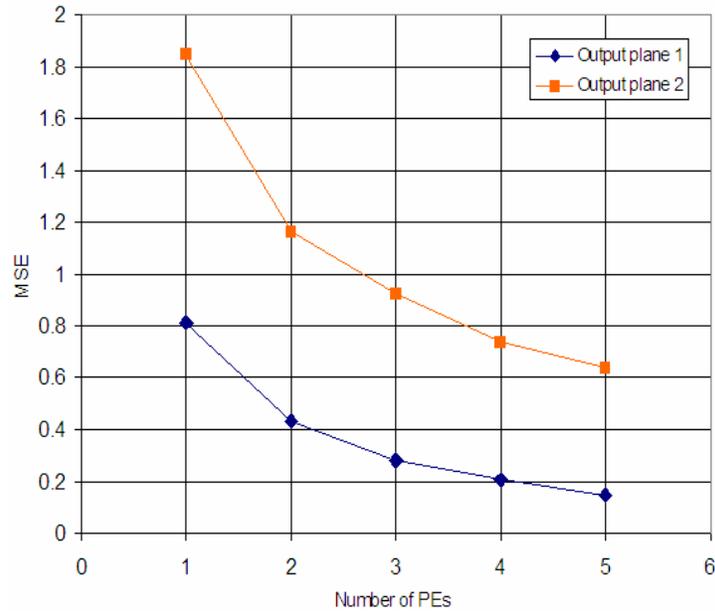
In computer experiments, images shown in Fig. 3.17 are used. The parameters are pretty much the same as before. As a reminder, wavelength is  $0.632 \mu\text{m}$ , size of one pixel is  $5 \times 5 \mu\text{m}$ , Sampling distance is  $1.25 \mu\text{m}$ , distance  $d$  between adjacent PEs is  $0.3 \text{ mm}$ , focal length of the lens is  $12.5 \text{ mm}$ , and distance  $\ell$  between two output planes is  $25 \text{ mm}$ . In the first experiment set, we use  $32 \times 32$  pixels of sample images this time. It contains one fourth as many pixels as our previous experiments due to lengthy computation times in SA approach. Resultant pixel sizes on the image side are the same as before, i.e.  $50 \mu\text{m}$  (since this time lens' focal length is halved, for smaller image sizes). For sampling requirements, using the current system parameters, on the device side,  $\Delta x$  is  $1.25 \mu\text{m}$  as before, and the right hand side of the Eq. (2.37) is  $1.23 \mu\text{m}$ . Similarly, in the image domain, these figures come out to be  $50 \mu\text{m}$  and  $9.9 \mu\text{m}$ . So, the conditions are satisfied.

The MSE graphs, when extended IFTA is used with 64 quantization levels and when extended SA is used with 4 quantization levels, are shown in Figs. 3.18 and 3.19, respectively. As seen from the figures, the MSE decreases smoothly as more PEs are added. This effect seems to be more dominant in the beginning and diminishes more slowly as  $N$  goes to infinity. In addition to general parameters for our setup, a constant  $\alpha$  exists in the expression of temperature in extended SA algorithm. The numerical value of it can be obtained by ad-hoc methods. In our case the values are  $1 \times 10^{-4}$ ,  $5 \times 10^{-5}$ ,  $2.5 \times 10^{-5}$ ,  $1.25 \times 10^{-5}$  and  $1 \times 10^{-5}$ , for  $N = 1, 2 \dots 5$  respectively. They are obtained by trial and seen that the lower  $\alpha$  parameter is, the slower temperature decreases thus allowing a better solution in expense of enormous computation time. However, after a certain limit decrease in  $\alpha$  does not give any significant benefit.

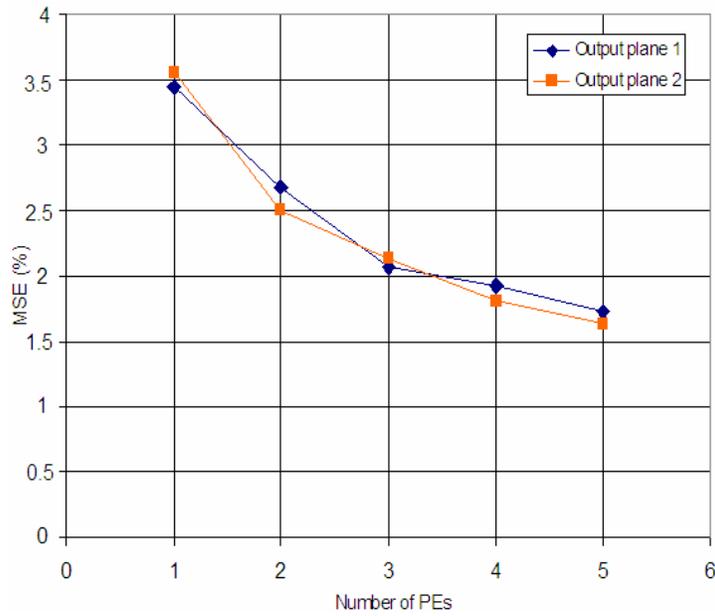


**Fig. 3.17** These gray scale images are used in the following experiments in place of the first and second output planes in order to demonstrate the effects of cascaded PEs in noise removal. The size of images used in experiments is 32 x 32 and 64 x 64 pixels.

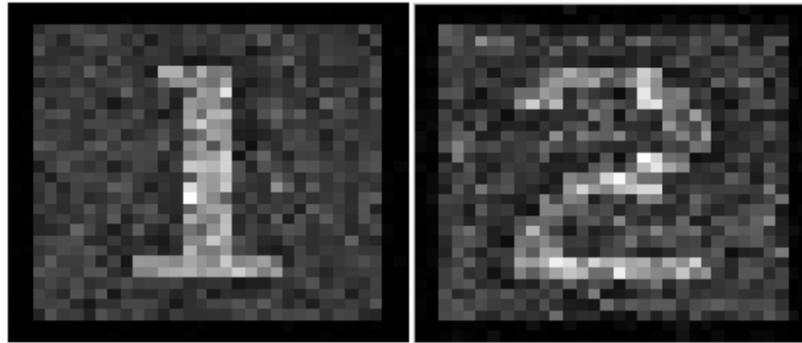
Careful inspection of MSE graphs shows that in Fig. 3.18 a significant amount of noise is removed from the targets '1' and '2'. Percentage decrease rate ( $R$ ) in the first output plane is slightly larger than the second output plane. And in general, first output plane is favored in terms of MSE. This is due to the fact that, in the algorithm, target plane amplitudes are imposed in order and first plane is the last one before back propagation. On the other hand when extended SA is involved there are significant changes in values and character. In general, MSE can be said to increase in value with respect to the former experiment as shown in Fig.3.19, although decrease trends of curves is still valid. That is obvious because of the involvement of lower quantization constraint,



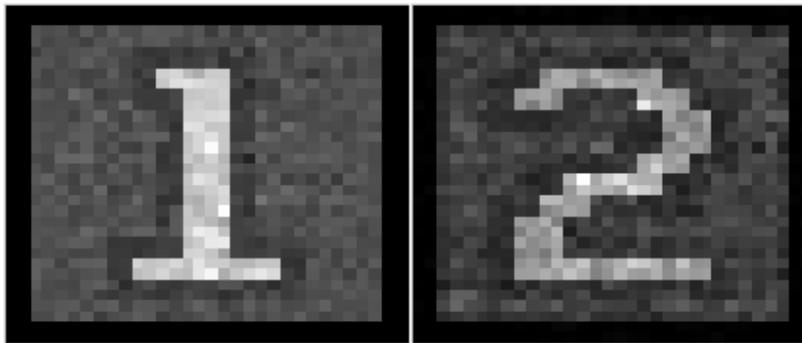
**Fig. 3.18** Results of computer experiments with 64 phase quantization levels allowed. Extended IFTA is used. MSE is seen to decrease as more PEs are added. Although percentage decrease rate,  $R$  and final values of the first output plane excel, the net amount of noise removed from the second output plane is more.



**Fig. 3.19** Results of computer experiments with 4 phase quantization levels allowed. Extended SA is used. MSE is seen to decrease as more PEs are added, similarly for both output planes.

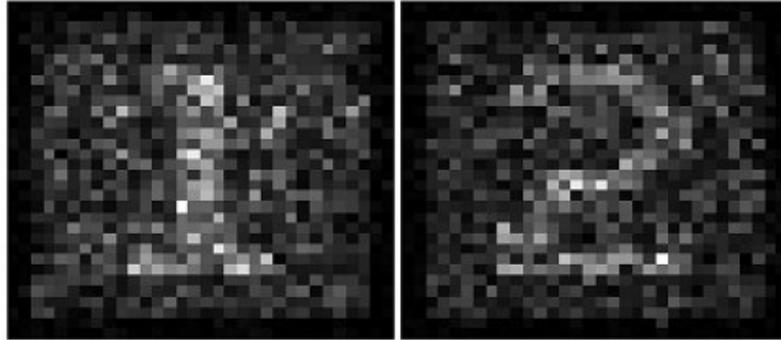


(a)

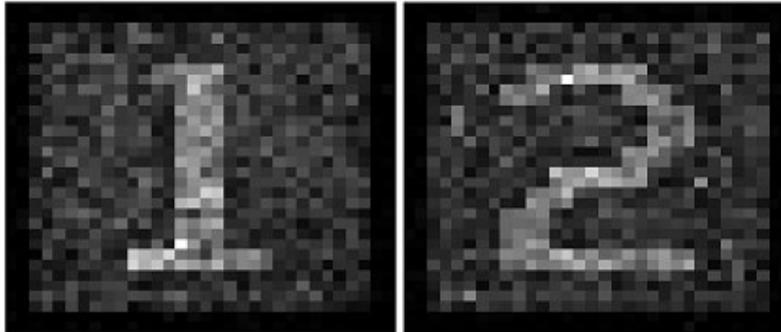


(b)

**Fig. 3.20** Resultant Images (32 x 32 pixels) of computer experiments with 64 allowed phase quantization levels when Extended IFTA is used, referring to Fig. 3.14. (a) The first image set (for output plane 1 and output plane 2) is processed by 1 PE (b) the second image set (for output plane 1 and output plane 2) is processed by 5 PEs. The improvement is clear. Note that max scaling is applied here. In other words, in printing the results, all pixels are divided by the maximum value for normalization.



(a)



(b)

**Fig. 3.21** Resultant Images (32 x 32 pixels) of computer experiments with 4 allowed phase quantization levels when Extended SA is used, referring to Fig. 3.14. (a) The first image set (for output plane 1 and output plane 2) is processed by 1 PE (b) the second image set (for output plane 1 and output plane 2) is processed by 5 PEs. Max scaling is applied here again.

which at some point, limits maximum quality achieved. Another point due probably to quantization constraint again is the fact that the  $R$  values in curves are slightly less in the latter experiment with SA. One other distinctive observation is the equalization of the curves. In fact, as mentioned before, in that approach the quality trade-off of targets can be adjusted and neither plane in terms of order is to be favored, as also seen from the flow-chart.

The improvement can also be noticed by just looking at the resultant images shown in Figs. 3.20 and 3.21. In parts (a), when one element is used as in the traditional case, images at two target locations are somewhat noisy and blurred; whereas, five PE planes improve the quality considerably. Even the pale shade around original figures above is noticeable in results coming from five planes. In addition to ordinary noise, cross-talk (in this case, the effect of first plane on the second) is reduced to some extent.

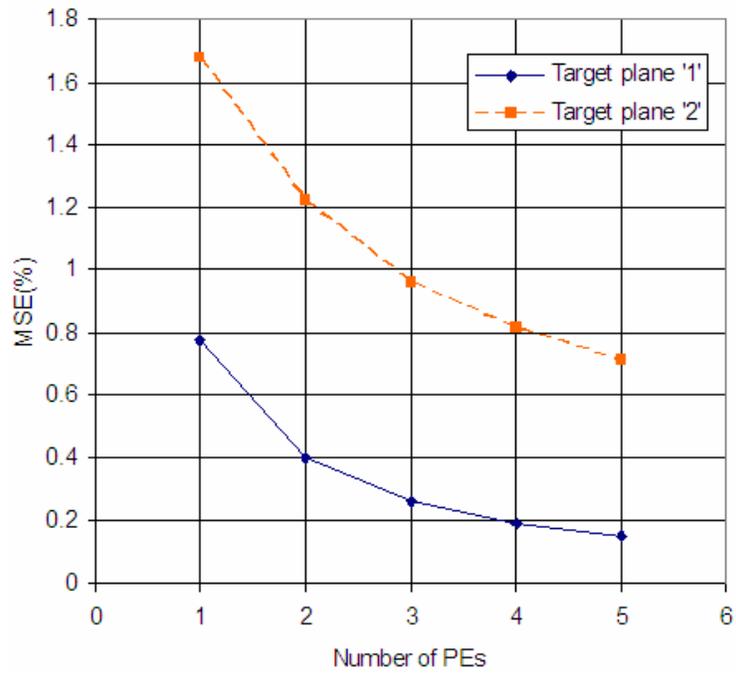
Note that, diffraction efficiency curves are not considered here since noise removal is the primary aim of this work. In addition, as a general trend, the results of 2D experiments may be applied to 3D as well. It can be said that when iterative algorithm is used for 64 levels,  $\eta$  is almost constant as more PEs are added (similar to those shown in Fig. 3.8 with an initial value depending on chosen image, furthermore the fact that more than one image plane is used, does not affect  $\eta$  as well). When annealing is used, however, for 4 levels of quantization, resultant  $\eta$  curves would be likely to drop in accordance with the acceptance condition (one image plane results are displayed in Fig. 3.10, multi-plane case will also show similar characteristics about diffraction efficiency).

One more effort has been spent on some advanced algorithms similar to extended IFTA to handle extremely low quantization constraints instead of annealing approach. In fact, once attempted with the current iterative approach (Fig. 3.15), the MSE would increase as more PEs are introduced (this case is similar to the graph in Fig. 3.9). So, for PEs with 4 levels designed with extended IFTA, our results showed MSE values of 4.2 and 4.9 for output planes 1 and 2, when one PE is used. However, MSE rise up to 7.5 and 8.5 when five elements are in the setup. So, this means that the trend has not changed;

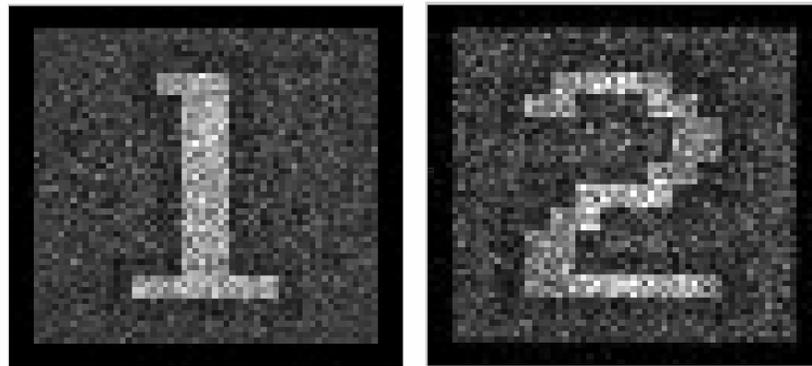
adding more PEs does not improve quality under strict quantization. Now, another method can be used, which is related to achieving gradual quantization. We will name it extended IFTA-2. Every phase element is found as usual without any constraint first by using the flow chart in Fig. 3.15, then only the first element is quantized and the entire algorithm proceeds followed by quantization of the second element, etc. This specific approach enables the user to control the constraint step by step and to maximize the performance. When applied, MSE values of 3.9 and 5.1 are obtained for output planes 1 and 2, in five PEs case. Thus, even in this situation, extra noise coming from strictly quantized phases for cascaded PEs accumulates and adding more elements does not increase image quality much. Even those improved values cannot be comparable with those coming from extended SA given in Fig. 3.19 for five PEs. So, after this discussion, one can infer that cascaded PEs with IFTA-type approach are useful only for moderate and high numbers of allowable phase levels; hence, an SA-type approach might be used to handle strict constraints and to exploit the benefits of the device being proposed.

These results basically show that expansion of the phase modulation concept as a form of cascaded phase elements, leads to significant amount of enhancement in image quality. In this process, novel algorithms which are capable of handling multi-planar concepts in both object and target domains, have been proposed with success. As a next step, we will examine the effects of allowed quantization levels more. This experiment will be similar to the one above in essence. In this case, SA MSE decrease curves will not be considered and larger images can be used. Our original 64 x 64 images are appropriate for the following trials.

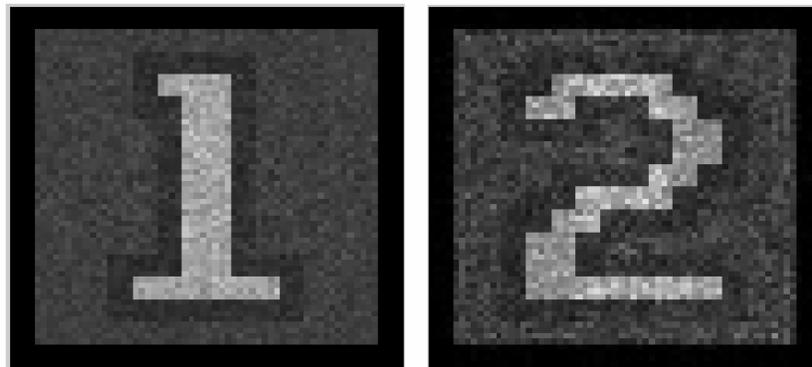
Undoubtedly, number of allowed phase quantization is one of the predominant factors in results. Now, a closer look will be taken to this issue. In the following computer experiment, 64 x 64 images are used with the same shape as in Fig. 3.17. Other system parameters are the same except the distance  $\ell$  between image planes have been extended to 40 mm since images are larger, and closer distances give results where cross-talk is dominant. We again start with 64 quantization levels which correspond to a very loose constraint on the micro-relief structure with results shown below.



**Fig. 3.22** MSE versus number of PEs used in two-output plane imaging, when extended IFTA is used with 64 allowed phase values.



(a)



(b)

**Fig. 3.23** Resultant Images (64 x 64 pixels) of computer experiments with 64 allowed phase quantization levels when Extended IFTA is used, referring to Fig.3.14. (a) The first image set (for output plane 1 and output plane 2) is processed by 1 PE. (b) The second image set (for output plane 1 and output plane 2) is processed by 5 PEs. Note that, no scaling is applied so reconstructed image intensities are unnormalized here and a little bit lower than the original due to other diffracted orders.

The graphical results given in Figs. 3.22 and 3.23 reveal the same facts as before. For the first target plane, almost half of the noise is cleared when progressing from 1 to 2 PEs and more than 80% when 5 PEs are used. Likewise the total noise removed from second plane is seen to be approximately 60% when changing from 1 to 5 PEs.

As discussed before, quantization of device is very important in practical applications. For iterative cases, we already saw that using very low quantization constraint leads to increase in noise as more PEs are added so that cascaded diffractive structures would no more give any benefit. Some extra efforts have been put on this topic. Other common values for allowed phase quantization levels used in diffractive optics are given in Table 3.3. These values have been applied using the extended IFTA and extended IFTA-2 methods. The latter can be considered as a derivative of the former and is useful especially when the number of phase levels is limited; furthermore their results get similar as number of phase levels increase, becoming unnoticeable after 16 levels. As expected, MSE increases as allowable phase levels decrease. Another remarkable result from the point of view of this work is the error reduction rate ( $R$ ) changing from top to bottom in the table. For the first row, as it was mentioned above, when four additional elements are added, 80% and 60% of the noise is removed from target planes ‘1’ and ‘2’ respectively. These values drop down to 65% and 50% for the second row. When it comes to 4 levels, there is no significant performance change according to the results.

**Table 3.3** Performance merits for 1 and 5 PEs according to allowable phase levels on device.

Number of Phase levels	MSE (1 PE)		MSE (5 PE)	
	Target plane 1	Target plane 2	Target plane 1	Target plane 2
64	0.77	1.68	0.15	0.71
32	0.78	1.72	0.28	0.85
16	1.00	1.95	0.44	1.06
8	1.67	2.69	1.36	2.02
4	3.98	4.91	3.91	4.79

On the other hand, when extended SA is applied to the problem at hand with 4 allowable phase levels and the same system parameters, minor and major effects appear on results. As a minor effect, for MSE coming from 1 PE, if 4 levels in Table 3.3 and

Table 3.4 are considered and compared, the one coming from SA excels slightly. This is due to the superiority of the stochastic approach in comparison with the deterministic one. However, the difference becomes much more apparent in the case of 5 PEs. Approximately half of the noise is seen to be eliminated with respect to the previous approach by looking at MSE values.

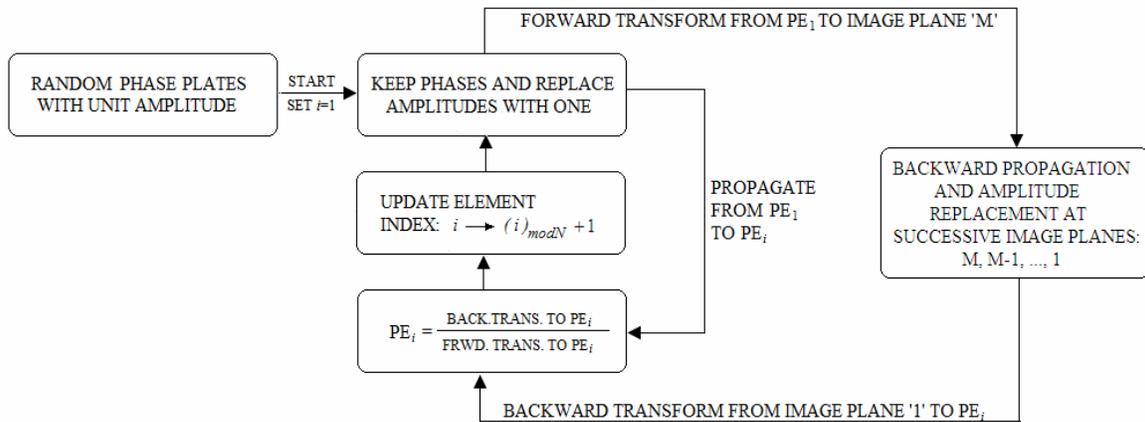
**Table 3.4** Performance merits for 1 and 5 PEs when allowed number of phase levels are 4, for extended SA case.

Number of Phase levels	MSE (1 PE)		MSE (5 PE)	
	Target plane 1	Target plane 2	Target plane 1	Target plane 2
4	3.41	3.39	2.00	1.97

### 3.4.2 3D abstraction with $M$ image planes

If MSE curves are considered like the ones in Fig. 3.22, a smooth decrease is observed for both output planes. Then a few questions arise such as whether or not these curves have a specific mathematical pattern, how does the number of image planes considered affect overall results of noise removal, or what are the sample image dependencies of these curves. Now that the analysis of the concept of cascaded CGHs has been examined for two successive image planes as a 3D representation, new attempts will be done to generalize this result to  $M$  image planes. Furthermore, MSE decrease trends will be examined and modeled in an attempt to find a generalization. For that purpose, no quantization scheme is applied, so the results will be free of any kind of phase quantization constraint error. With this approach one may investigate the pure mathematical contribution of these extra phase elements to the image quality thus the increment of information capacity of pure diffractive phase structures. In the implementation of the ideas expressed above, we will be using extended IFTA as before. An updated version of this flow chart is given below in Fig. 3.24. Note that, this ultimate version of extended IFTA enables user to compute  $N$  CGHs for  $M$  possible planes of targets. The general scheme for this most general form is that the field, after being propagated through PEs, is considered to continue to the last image plane ( $M$  th). Amplitude imposition stage is applied by keeping the phases and replacing the target

amplitudes pixel by pixel. Then back propagation is implemented step by step; first the field comes to the output plane  $M-1$  and amplitude replacement is applied, second it comes further back to the output plane  $M-2$  and so on. After the first output plane, the field passes to the device domain with a lens back-Fourier transform and continues again to the specific  $PE_i$  under consideration. The rest is the same; the  $PE_i$  is computed with a division where the numerator is that back transformed field which has traversed all the image planes in image domain and the denominator is again the forward transform up to that element.



**Fig. 3.24** One iteration cycle of the extended IFTA in handling  $N$  PEs and  $M$  image planes.

In summary, in this ultimate version of design, the logic is simple; the field propagates back and forth through image and device domains to convey several planar amplitude information regions onto the several planar phase surfaces (PEs in our case). The division operation at the end guarantees any specific element to behave as it is supposed to do in the real implementation where the field begins from the left, processed by all diffractive surfaces and forms several images with varying depth, as an expected operation.

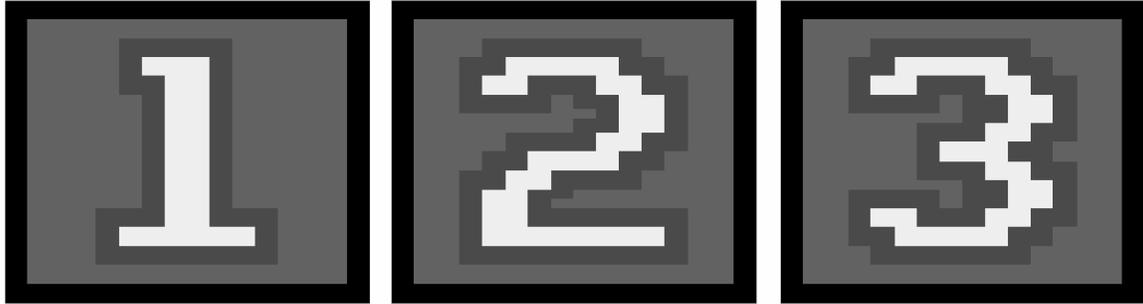
Note that, SA version of this approach also exists and is simple so that in the conditional statement of acceptance of new trial, an MSE will be found by using  $M$  output planes instead of 2. In numerical treatments in this part, we will only consider

design with iterative method and are content with clarifying that stochastic approach is also possible.

In the following experiments, there are two types of image sets to be investigated. Similar system parameters are used on 64 x 64 pixel images. Wavelength and sampling sizes as well as pixel sizes have not been changed throughout the examples here, so is the distance between PEs. Image sizes are directly related to the focal length thus any value is possible based on the desired target image sizes. We can still assume 25 mm for the focal length giving rise to 3.2 mm outputs separated by 40 mm. Note that, for a bigger  $f$ , one would obtain larger resultant images. This might cause the image planes after the first one a little bit more noisy if the separation distance is to be kept, since the field adjusts itself within that separation distance and making the images larger without changing their separation limits the flexibility of the system thus allowing error imbalance between output planes. This topic will be revisited again in Section 3.6 while examining another phenomenon and shortly mentioned in Chapter 6 again.

The first image set is a continuation of our last set, and is shown in Fig. 3.25. And the second set consists of 'U', 'S' and 'C' letters, given in Fig. 3.26.

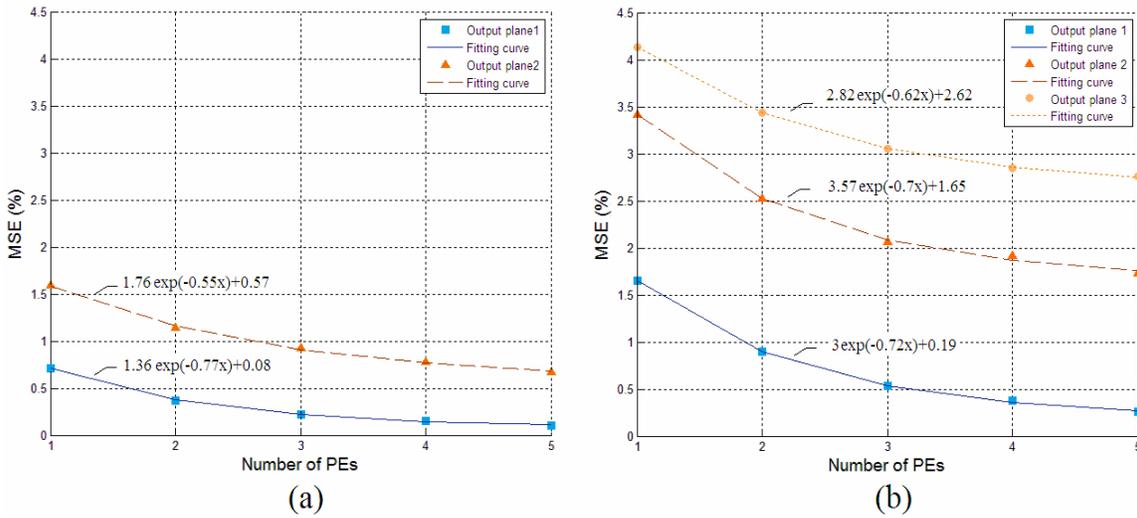
After the application of the deterministic algorithm in Fig. 3.24, the MSE curves and resultant images obtained are displayed below in Figs. 3.27 - 3.31. Several observations can be made based on these results. First of all, for Figs. 3.27 and 3.28, it can easily be said that all curves are smooth and clear. The main trends are independent of image used or number of output planes. The rates of exponential decay are not dependent on the image used as can be seen by comparing the two figures. If parts (a) belonging to two image planes and parts (b) belonging to three image planes are to be compared, one can see the net amount of noise removed in the case of three image planes is more. This result means that, as more image planes are added in 3D abstraction of the target, the amount of noise removal increases. In general, the starting error values increase in passing from two to three image planes. If individual curves are considered, the rate  $R$  does not seem to change much for the first output plane between the cases of



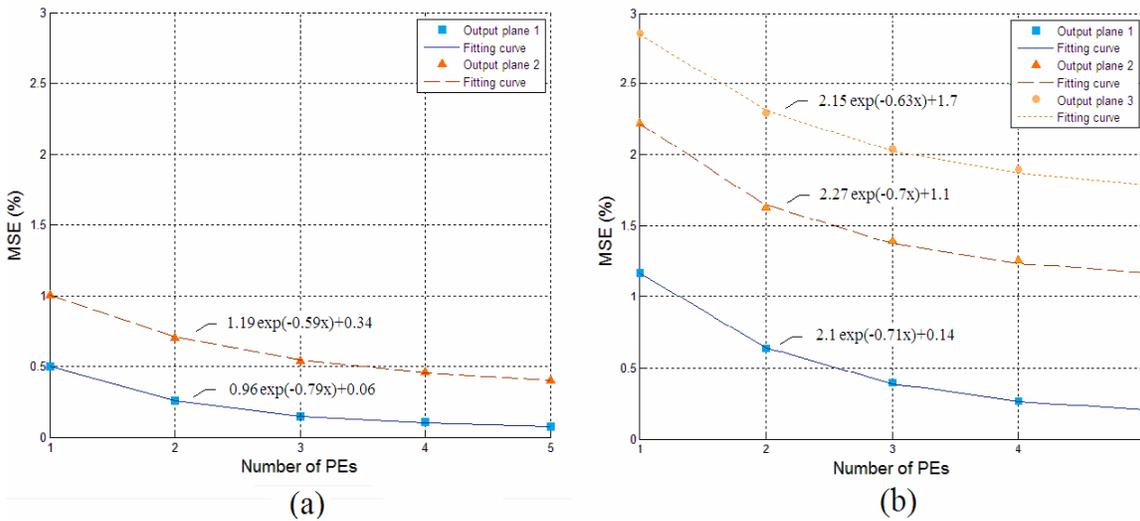
**Fig. 3.25** Image set 1, used in experiments.



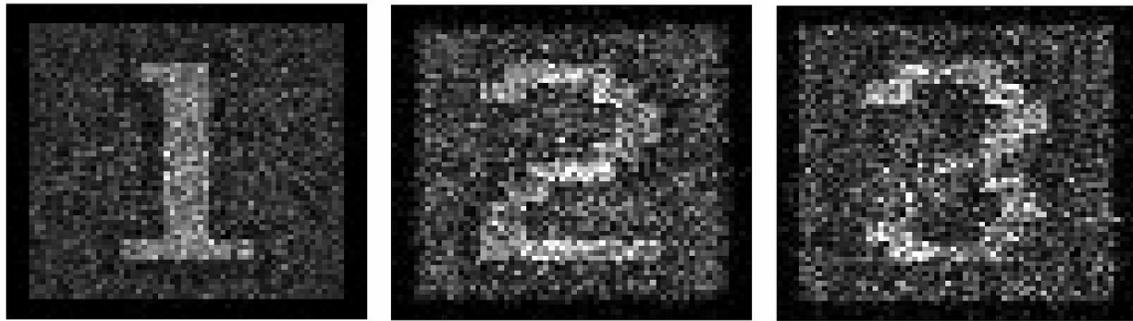
**Fig. 3.26** Image set 2, used in experiments.



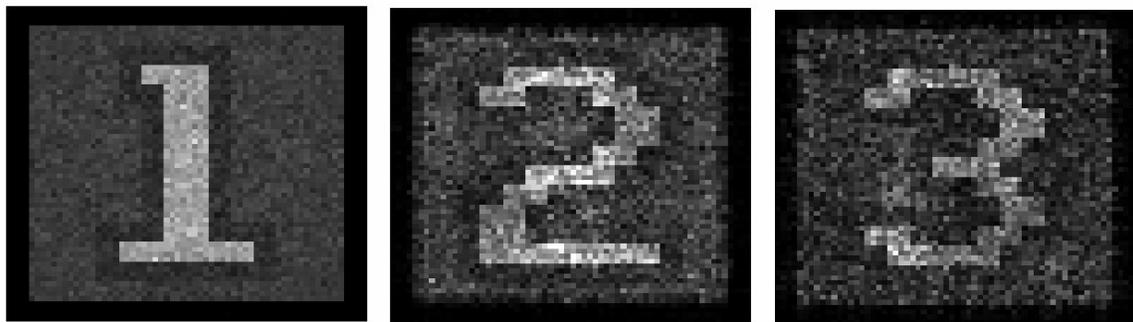
**Fig. 3.27** Graphs, corresponding to image set in Fig. 3.25. (a) Results for 2 image planes, '1' and '2'. Percentage decrease rates are 86% and 58% for output planes 1 and 2. (b) Results for 3 image planes, '1', '2' and '3'. Percentage decrease rates ( $R$ ) are 84%, 49% and 34% for output planes 1, 2 and 3.



**Fig. 3.28** Graphs, corresponding to image set in Fig. 3.26. (a) Results for 2 image planes, 'U' and 'S'. Percentage decrease rates are 85% and 60% for output planes 1 and 2. (b) Results for 3 image planes, 'U', 'S' and 'C'. Percentage decrease rates ( $R$ ) are 85%, 50% and 39% for output planes 1, 2 and 3.

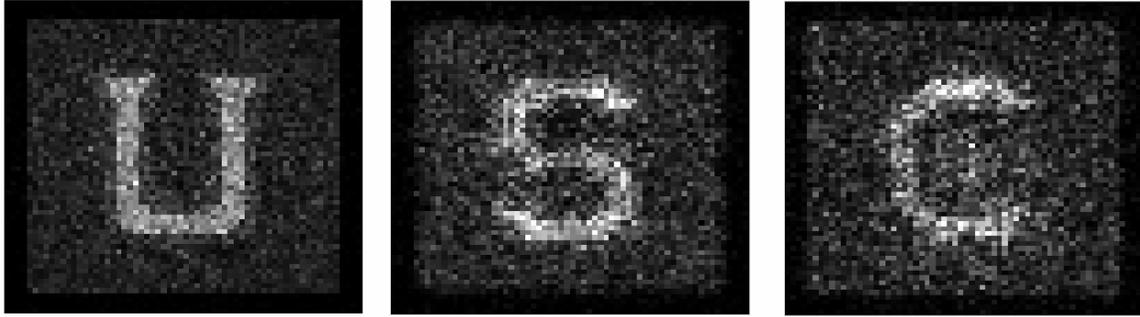


(a)



(b)

**Fig. 3.29** Resultant Images (64 x 64 pixels) of computer experiments coming from Fig. 3.25 when Extended IFTA is used. (a) The first image set is processed by 1 PE (b) the second image set is processed by 5 PEs.



(a)



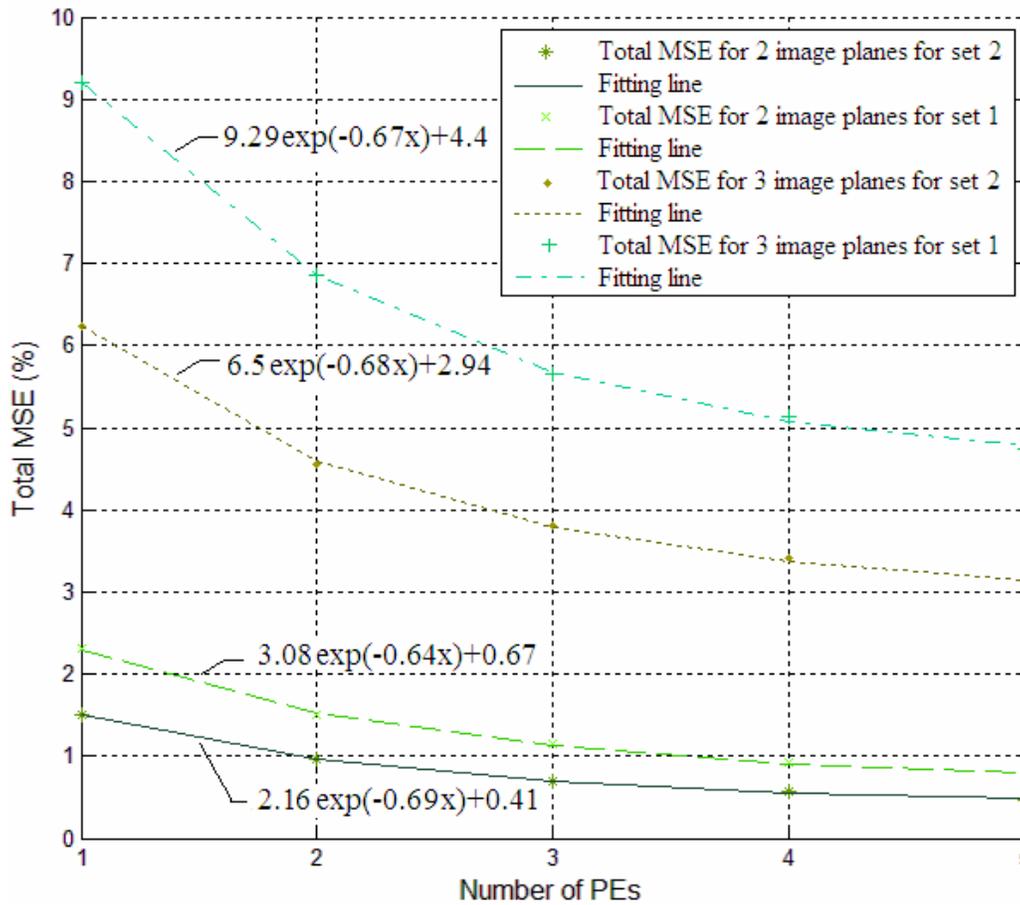
(b)

**Fig. 3.30** Resultant Images (64 x 64 pixels) of computer experiments coming from Fig. 3.26 when Extended IFTA is used. (a) The first image set is processed by 1 PE (b) the second image set is processed by 5 PEs.

two target images and three target images. Second output plane's rate is affected a little though. Another general trend is the decrease in those  $R$  values in going from the direction as output plane rank increases, which is understandable. In IFTA based algorithms the last assigned plane is favored in terms of quality during the amplitude imposition stage, so this also affects noise removal rates. After all, we may conclude that, appearance of the third MSE curve causes decrease in rates for first and second error curves in decreasing order.

A mathematical expression corresponding to these curves are determined to be in exponential form:  $A \exp(Bx) + C$ . This form was found to provide a good fit to the data at hand. Mathematical expression of the individual curves can be seen on figures. In the above expression,  $x$  corresponds to the number of PEs (i.e.  $x = 1, 2 \dots N$ ),  $C$  is the asymptote; the value of the MSE when  $N$  goes to infinity and  $B$  is a parameter that is related to decrease rate, like  $R$ .  $B$  parameter will be named as decay constant from now on. So, decay constant is another important measure of how a curve decrease and due to a nonzero asymptote  $C$ ,  $R$  in Eq. (3.3) and decay constant cannot be connected with a simple equation. The most distinctive feature of them is the fact that numerical value of  $R$  is highly affected from the starting value, i.e. MSE with 1 PE, and as the initial value gets higher,  $R$  increases as well even more than the decay constant. Actually, the decay constant is much more related to the general shape of the curve going from 1 to  $\infty$  PEs instead of just 5. We believe both measures are useful since they indicate similar trends in above graphs.

Apart from that, total MSE can also be examined. Fig. 3.31 shows the total MSE corresponding to image sets 1 and 2 when cascaded PEs are used in the reconstruction. The same form of exponential curves seems to be fitting well. This time the exponential decay constants are close for curves and it is in the range of [0.64-0.69]. It has been discovered that this number only depends on the separation distance  $\ell$  between image planes and number of phase quantization constraint. When  $\ell$  decreases, it will be hard for the field to adapt for the other images placed in depth, thus the total MSE can be expected to increase.



**Fig. 3.31** Total MSE curves belonging to image sets 1 and 2. For the curves the decay constant is constrained in an interval. In fact, experiments with more than 3 image planes show also that the constant is still found in the same interval  $[0.6-0.7]$  for  $\ell = 40$  mm.  $R$  values are 68, 67, 51 and 48 from bottom to top.

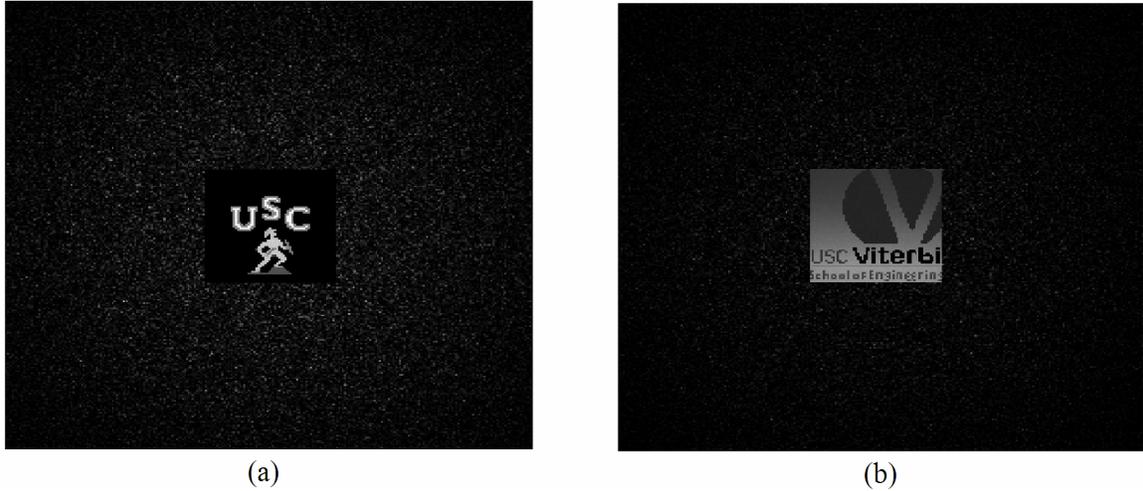
### 3.5 Other Ways of Mitigating Noise

As it was mentioned before, a dummy (noise) window may be placed around the target image(s) by increasing the number of pixels on the diffractive element in use. This basically causes noise to accumulate at the window by either shrinking the effective target image area in case of the same pixel size is kept on the CGH thus letting the element bigger, or keeping the effective target image area but shrinking the pixel size on the CGH. These methods are related to extension of the solution space in the  $x$ - $y$  plane, and one of the common methods for the noise removal process.

Noise window can be applied with 2D or 3D image reconstructions. The area allocated for noise can be anything although after a certain limit additional area does not affect the noise removal process much. In addition, phase quantization constraint affects the overall process such that when available phase values decrease, noise window does not provide significant benefit generally depending on the sample especially on hard images.

First we consider a setup with only one phase element like in Fig. 3.1 and samples as in Fig. 3.7. Below, Fig. 3.32 exemplifies results when the window concept is applied on a system with a very loose phase quantization constraint such as 64. In one output plane case, adding more phase elements as before would not cause any noise removal since almost all of the noise is already removed by the dummy window. Effective target areas which are the areas that the target would fit without anything extra, are shown in the figure; they are filled with removed noise plus the reconstruction result. The MSE values are very close to zero (0.007 and 0.010, for (a) and (b), respectively). In part (a), more noise is cleaned with respect to part (b) from numerical values when compared with the values without the noise window (they can be seen from Fig. 3.8 first data of related graphs). This is also seen from the images below.

In simulations, IFTA (Fig. 3.2) is used and MSE converges very fast. The diffraction efficiency, on the other hand, is extremely low for this application and normalized to those shown in Fig. 3.11 for comparison reasons.



**Fig. 3.32** Figure displays results for two different samples belonging to one phase element reconstruction using a noise window, with 64 quantization levels. The noise window used here is 4X the actual image.

In designing the diffractive element 256 x 256 pixels are used for 64 x 64 images. The image areas shown above are supposed to be the same as those displayed in previous sections in this chapter. Thus, although pixels on the device are the same, pixels at the image are smaller with a multiplier of 4.

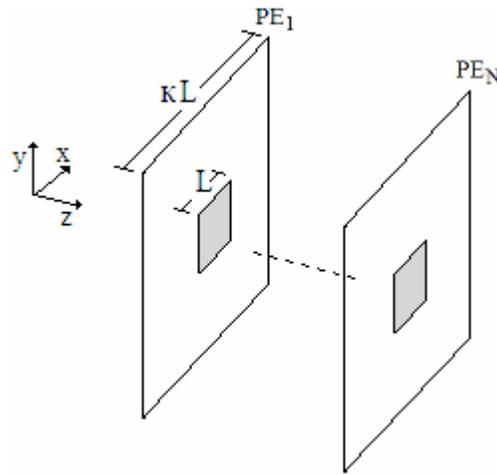
Noise window concept may also be applied to 3D images. Again, by using one phase element and two image planes, one can get a significant amount of improvement. In this situation, however, MSE does not come very close to zero due to the hardness of the problem of reconstructing both phase and amplitude (this will be examined as a special case of the application in the next section). In summary, in this section, noise window concept has been introduced. Instead, one can combine the two extension techniques (noise window and cascading) to obtain a hybrid approach.

### 3.6 A Hybrid Method

In this section 3D images which are considered to consist of two 2D planes again, will be analyzed further with modified diffractive elements as will be explained. Main goal here is to combine noise window method with cascaded diffractive surface concept such that a

CGH, in its most general sense, is enlarged in all spatial dimensions ( $x$ - $y$ - $z$ ) for extreme noise removal.

An abstract drawing is shown in Fig. 3.33. Gray toned areas in the middle have the same dimensions with originally designed elements before. Pixels in the extra area are responsible for noise removal. This new configuration contains more pixels, hence enlarged diffractive surfaces should either be larger than normal in which case we assume in this work, or pixel sizes are to be small. These two cases give similar physical results, except that in the latter some sampling conditions might cause problems so the former is preferred here. Therefore, the new feature size is defined by  $\kappa L$ , where  $\kappa$  is enlargement factor in  $x$ - $y$  dimension. In general, it can be in an interval of  $[1, \infty)$  and does not have to be an integer as long as it does not divide a pixel. In this work, we use,  $\kappa = 1, 1.25, 1.5, 1.75$  and  $2$ .



**Fig. 3.33** Ultimate space usage concept in CGHs is displayed. PEs are placed as before from 1 to  $N$ . Gray toned areas have the same dimensions with originally designed elements. Pixels in the extra area are responsible for noise removal. Since this new configuration contains more pixels, its feature size is defined by  $\kappa L$  and total area is  $\kappa LN$ .

For computer experiments, similar treatments have been applied to the system under consideration by using extended IFTA. Important parameters are as follows; size and number of pixels are  $5 \mu\text{m}$  and  $64\kappa$ , respectively for two target images of  $64 \times 64$  pixels shown in set 1. In the first case, it is assumed that lens focal length changes such

that the output pixel size remains to be  $50 \mu\text{m}$ . In the second case, lens focal length is kept constant at the previously used value (25 mm) and thus output pixel size becomes  $50/\kappa \mu\text{m}$ , distances between target planes are separated with a distance  $\ell$  which is equal to 15 mm. The reason for choosing relatively shorter separation distance will be discussed later. The separation between PEs, however, is still considered to be  $300 \mu\text{m}$ . The figures below show total MSE graphs for ‘1’ and ‘2’ target images with respect to  $N$  (number of PEs), and the  $\kappa$  (enlargement factor).

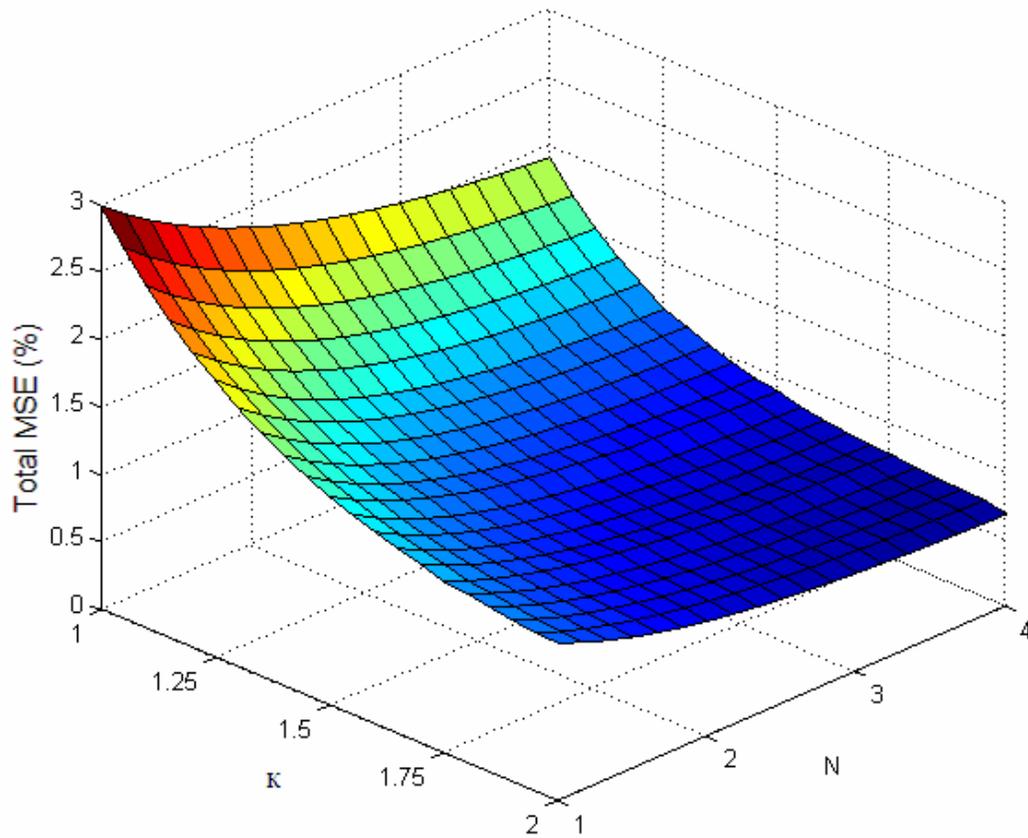
Fig. 3.34 displays relation between  $N$ ,  $\kappa$  and total MSE for the first case. The output image sizes are kept constant, and as a response to increment in  $\kappa$ , lens focal length is considered to be varying. In computing the values, exponential fitting curves of the format given above are used for both  $\kappa$  and  $N$  directions. Note also that, the maximal values for axes in the graphs are compatible (enlarging  $x$  and  $y$  dimension 2 times causes surface 4 times bigger which is the case in  $z$  dimension when 4 PEs are in use). These curves for 2 dimensions give almost the same values for the entire surface (although it is supposed to be the same). As it is seen, total noise level decreases as more enlargements involve either in  $x$ - $y$  or  $z$  domain. As an interesting point, the asymptote of this graph in the direction of diagonal unit vector ( $[\hat{N} + \hat{\kappa}]/\sqrt{2}$ ) approaches to a natural limit of noise. By saying natural limit we mean the minimum noise level that one could achieve under the same circumstances by using a device without any constraints like phase quantization or amplitude removal. So, if the problem is applied to one conceptual diffractive element with the ability of modulating phase and amplitude together, noise can be reduced up to a level that optical incompatibility allows, beyond this point is physically impossible.

By looking at Fig. 3.35 which depicts the second case, one can say that MSE decreases more swiftly. This is, of course, caused by the fact that the aspect ratio is different in the conceptual setup. The ratio between output pixel size and square of distance is almost a constant in multi-focal systems and is a consequence of the propagation transfer function (this relation could be seen best by examining Eq. (2.25) and Eq. (2.28)). Therefore when the noise window is introduced and output pixel sizes shrink, the processed field finds relatively more distance to adjust itself. The more

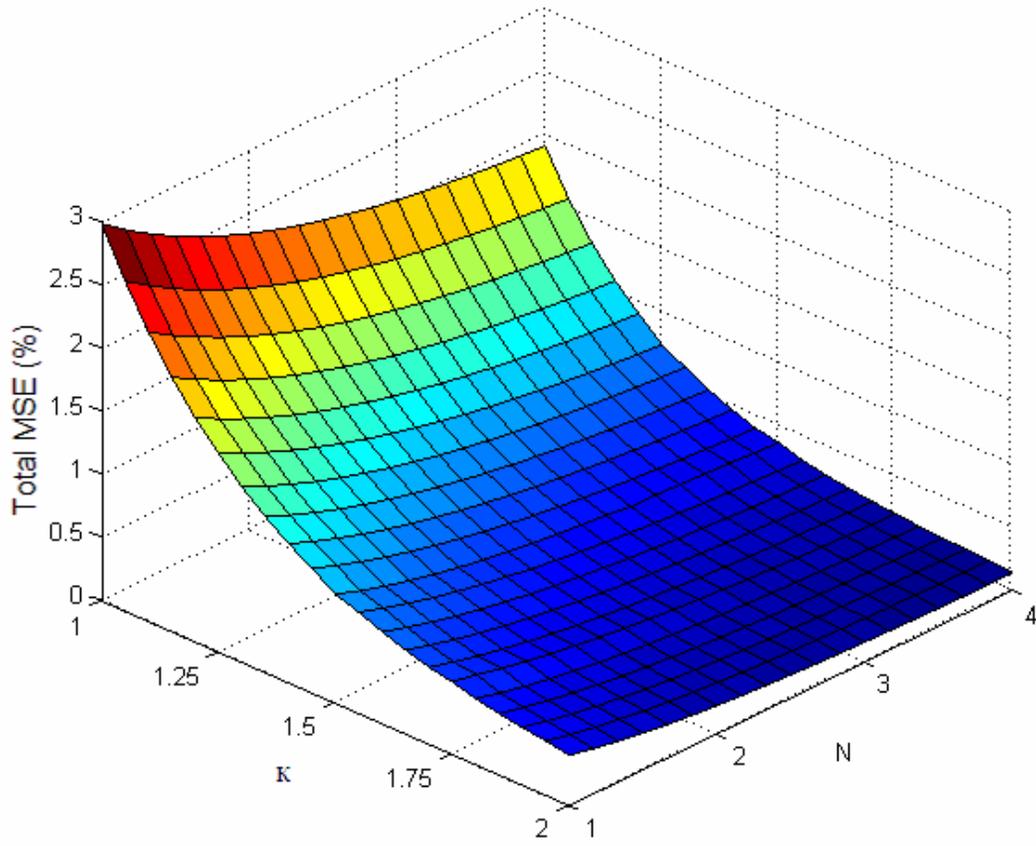
enlargements to increase  $\kappa$ , the smaller output pixels will be, thus yielding better adaptation for two target images on the other side. Certainly, the diagonal asymptote in this graph in case when  $\kappa$  and  $N$  goes to infinity is expected to be zero. Moreover that final graph tells another important fact; when  $\kappa$  goes to infinity, increasing  $N$  by adding more PEs lose its importance by a considerable amount. So, in the case of decrement in optical compatibility one needs less and less cascaded elements with respect to using noise window. This result is quite important.

The two axes, namely  $\kappa$  and  $N$ , can be analyzed more closely. Increase in  $\kappa$  seems to be more efficient in both cases, which makes sense because this gives more pixels thus more degree of freedom directly. As discussed before, adding more PEs also increases degrees of freedom and solution space but it happens in an indirect way so that propagation has to involve between elements. Another advantage of using noise window is the experimental easiness of application with respect to the act of cascading surfaces with ultimate accuracy. On the other hand, however, noise window yields shrunk outputs due to increased space-bandwidth product at the image side. In addition, it causes a relatively dirty environment on the image plane's noise area, which might be an issue for certain applications.

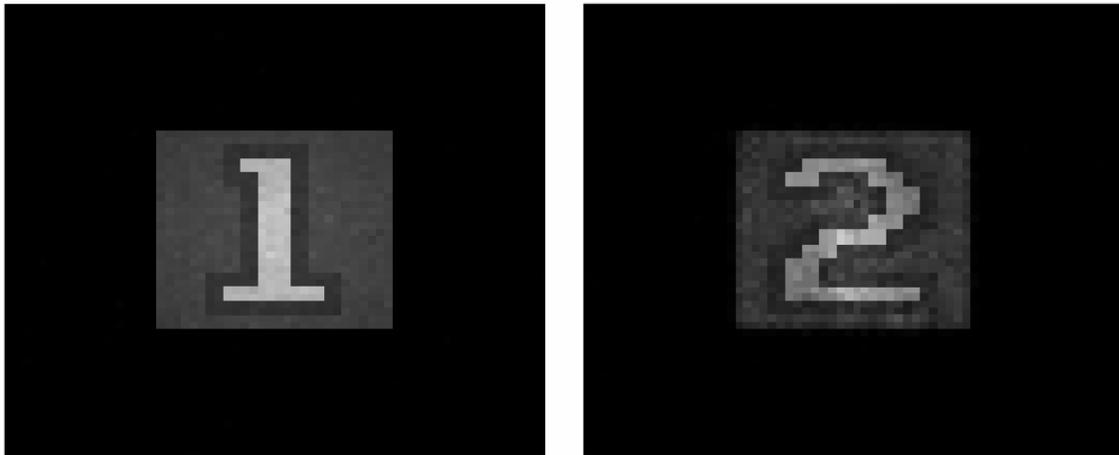
In multi-plane imaging applications, the separation distance  $\ell$  adjusts the optical incompatibility of the final 3D output. When this parameter is large, processed field finds more space to adjust itself between two (or more) planes. In this hybrid application, this parameter is chosen to be 15 mm which is smaller than before. That effect can be seen by looking at the below graphs and comparing the axis  $\kappa = 1$  and the graphs before. We increased optical incompatibility for especially the second graph in order to observe the decrease pattern for MSE. When the problem comes with less optical incompatibility, this graph drops down to zero very quickly, which is understandable but insufficient to analyze the phenomena. In addition, below graphs demonstrates results in case of only one PE and with variable noise window. That kind of analysis has also not been made before, to the best of our knowledge. So, if  $N = 1$  lines are taken into consideration in the graphs, the behavior solely caused by the noise window can be seen in detail.



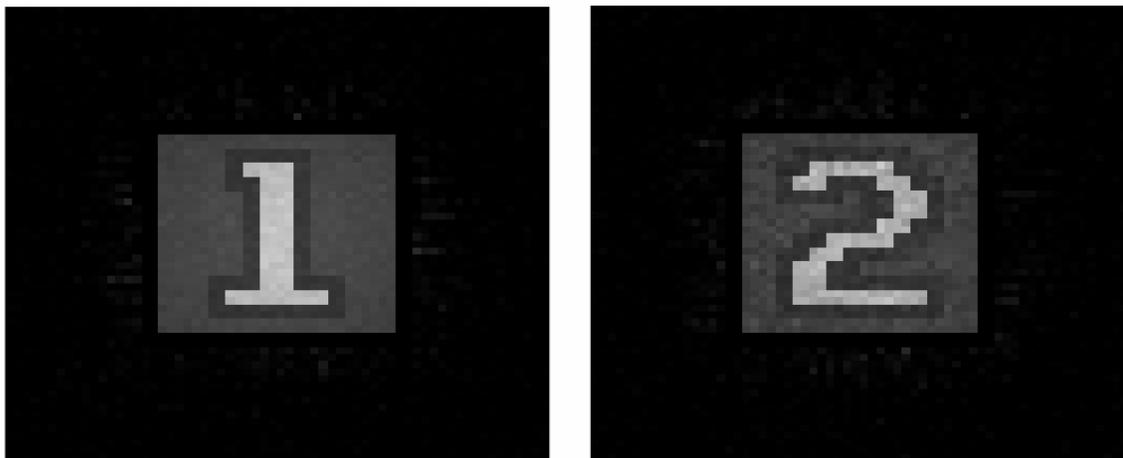
**Fig. 3.34** Result of cascaded PEs with also extended for dummy area. Two output planes are assumed. No phase quantization constraint is applied.  $N$  shows the number of PEs as before and  $\kappa$  denotes feature extension coefficient in  $x$  or  $y$  direction. The result of adding dummy area is compensated by changing the lens focal length in order to keep the same output dimensions. 3D graph reveals the combination of cascading and extending under these circumstances.



**Fig. 3.35** Result of cascaded PEs with also extended for dummy area. Two output planes are assumed. No phase quantization constraint is applied.  $N$  shows the number of PEs as before and  $\kappa$  denotes feature extension coefficient in  $x$  or  $y$  direction. The result of adding dummy area causes smaller output dimensions. 3D graph reveals the combination of cascading and extending under these circumstances.



(a)



(b)

**Fig. 3.36** Image outputs for the proposed system of ultimate expansion of a CGH when  $N = 4$  and  $\kappa = 2$ . (a) The actual image sizes are compensated by an imaging lens (b) Noise window cause decrease in output pixel size yielding a smaller image. For comparison reasons, both parts are shown in equal dimensions.

These results coming from above experiments are printed in Fig. 3.36. For both parts, if the image window is considered first, Noise, coming from amplitude elimination and cross-talk coming from multi-plane existence are seen to be improved considerably. Furthermore, adding cascaded PEs improve not only the image itself but also the noise confined in the noise window. Fig. 3.36 part (b) shows slightly better results with respect to part (a) with some more noise around the image, which reflects the facts in 3D plots and discussed before in detail. Between two parts, there are some other differences such as dimensions. In part (a), focal length changes (between 12.5 to 25 mm) to maintain a constant pixel size of 50  $\mu\text{m}$ . In this case, total image dimensions become 3.2 mm as before but target dimensions turn out to be 1.6 mm due to the dummy area. In part (b), on the other hand, a focal length of 12.5 mm is kept so that pixel dimensions at target locations take on values  $50/\kappa$   $\mu\text{m}$  as stated before. Finally, for  $\kappa = 2$ , total image dimensions become 1.6 mm and target dimensions become to be 0.8 mm, which is halved with respect to the former part. Apart from the topics discussed in this part, diffraction efficiency is slightly higher than expected amazingly. This may partly be caused by noise window area which fills sinc function edges and translates the existence of other orders.

As a summary, this chapter analyzes the error reduction phenomena for phase-only elements. Our original iterative type algorithm accepts input and output domains as multiple planar surfaces and calculates phases on input planes one by one by iterating over all the output target planes. As a result, the MSE values show that a significant amount of error reduction can be obtained by increasing the number of input PEs in the stack. Course quantization of phase levels degrades the results and limits the achievable quality drastically, and SA, which has not been considered much in such cascaded setups, handles pixels individually, thus being a remedy for cases with strict quantization constraints. The decay constant as a systematic addition of PEs are found to be confined in a range [0.6-0.7] and seems that this interval shrinks as number of experiments go to infinity making results average out. Of course, this constant depends upon the separation of individual image planes and independent of anything else, such as number of image planes or distance between PEs. In analyzing trends like that, mentioning the total MSE would be more meaningful just because the MSE for individual output planes might be

algorithm dependent. In the final stage, the total MSE of a system under hologram expansion via 3 spatial dimensions is analyzed. Also note that, in analyzing ultimate trends close to the end, no phase quantization is applied for better accuracy. However, if it was applied, the 64 levels would be approximately the same. On the contrary, strict constraints such as 4 would give different results. Reader may refer to Section 3.4.1 for a similar discussion but we can estimate that even with a very powerful algorithm and a computer, one might not be able to achieve an MSE decrease surface both in  $\kappa$  and  $N$  directions since degrees of freedom is confined under strict quantization constraints with iterative approaches. However, SA approach would work again although the amount of noise removal is debatable in this case by comparing the hardness of computation time and power in addition to other drawbacks discussed.

Basically this chapter's work sheds light on the relationship between phase-only elements and noise removal by using the most recent knowledge in literature and based upon the original research results. Now, a different kind of setup in which phase and amplitude are both modulated instead of just phase, will be investigated.

# CHAPTER 4

## IMAGING WITH PHASE & AMPLITUDE ELEMENTS

### 4.1 Introduction

In the previous chapter, we have shown that image reconstructions can be improved by using a multiple of CGHs successively modulating the incoming wavefronts. Those CGHs were phase elements which do not affect any amplitude information.

In this chapter, amplitude modulation will be investigated in addition to phase modulation. In fact, even though phase information is the most important factor in image formation, amplitude modulation may be important in several cases. For example, it can be used to modulate phase without using relief profiles as a most basic approach. Or, as shown later, amplitude and phase might be used together for enhanced performance. For the content of the chapter, traditional methods related to amplitude processing will be revised with some minor modifications first. Then, different set-ups will be proposed based on cascaded elements usage again but under different conditions. Methods can be applied for 2D or 3D image reconstruction. Results are promising after certain drawbacks are mitigated.

## 4.2 Encoding Techniques

### 4.2.1 Overview

There are lots of coding techniques to build a CGH for phase and amplitude modulation. They started to appear around late 1960s. In 1966, Brown and Lohmann described a method called detour phase for making binary CGHs where hologram plane is divided into cells to create phase and amplitude modulation effects [4]. Every cell has an individual rectangular aperture to adjust phase and amplitude for each pixel. Another method similar to detour phase method is Lee's delayed sampling method which was first published in 1970 [5]. This is based on decomposing the Fourier transform of the object into four quadrature components, which were represented by the real non-negative functions. The four functions are represented by apertures at four laterally displaced, or shifted, positions within each cell of the hologram plane (Fig. 4.1 (a)). For example, suppose the inverse Fourier transform of the target is taken and the complex amplitude to be encoded for one pixel is,

$$Ae^{i\theta} = A\cos\theta + iA\sin\theta \quad (4.1)$$

In encoding process, first the positive real part of the wave front is calculated at the center of the first section, and a rectangle proportional to this value is placed there. Rectangles having areas proportional to the positive imaginary, negative real, and negative imaginary parts are placed in the second, third, and fourth sections, respectively [41]. So, there will be two transparent rectangles with different sizes in two different cells at the end. In addition, instead of using apertures in cells, different transmittance values can also be used (Fig. 4.1 (b)).

After all, cells are created for hologram plane, a forward Fourier transform is taken to analyze the performance like in Fig. 3.1, except that the cell encoded CGH instead of pure phase modulating one. Next sub-section shows our results about cell encoding technique.



**Fig. 4.1** (a) According to above encoding scheme, cells represent pixel values and each cell is further divided to four sub-cells with a transparent rectangle inside. (b) Different gray tones may also be used instead of rectangles.

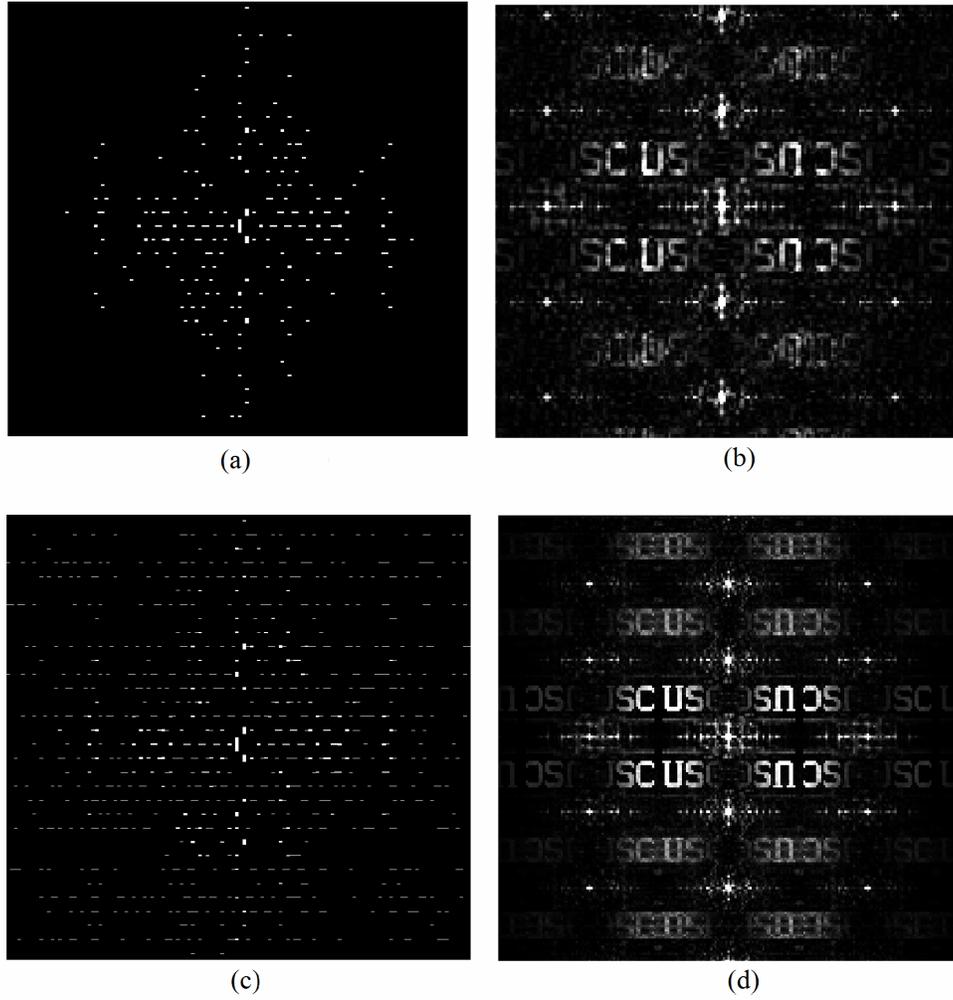
#### 4.2.2 Simulations

Lee's delayed sampling method has been produced and analyzed computationally. Both transparent rectangle version and grayscale version in Fig. 4.1 (a) and (b) are simulated with the help of a lens similar to a setup like in Fig. 3.1. The chosen image is displayed in Fig. 4.2 and the complete device look as well as results are shown in Fig. 4.3 and Fig. 4.4.

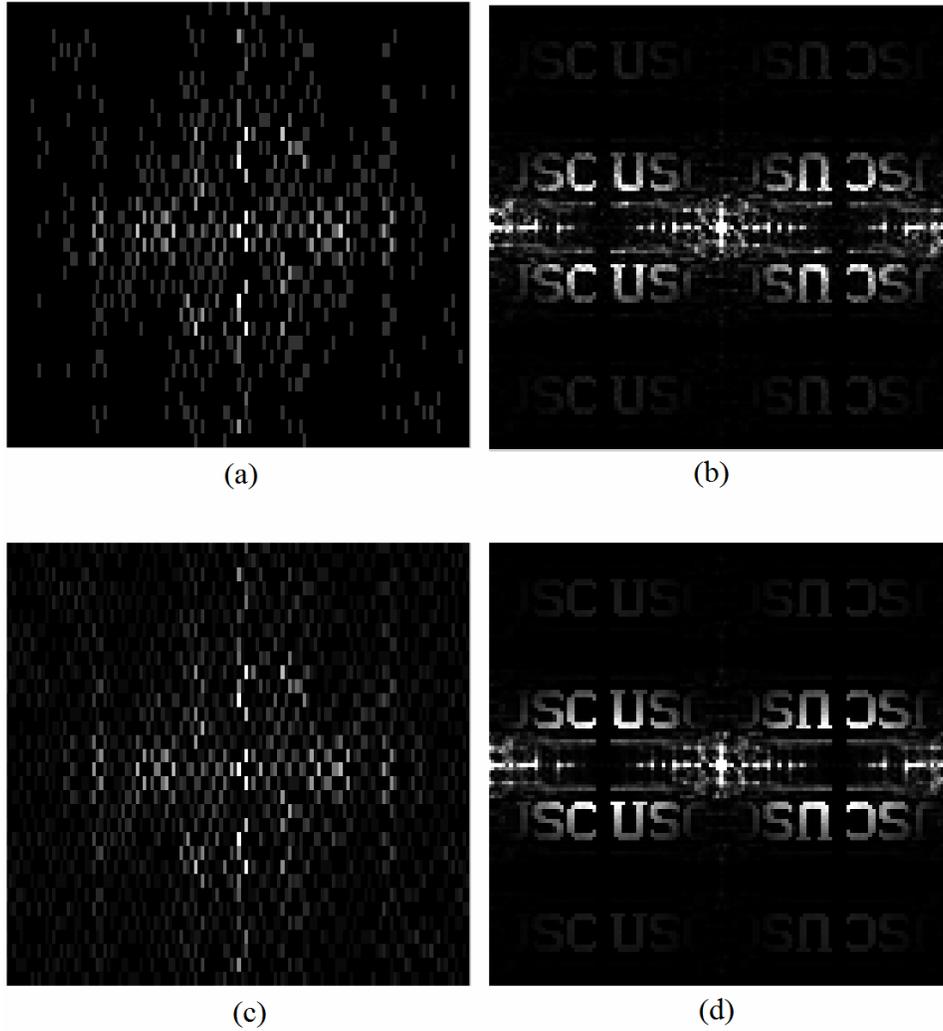
In Fig. 4.3, rectangular aperture type cells are used (Fig. 4.1 (a)). We have defined sensitivity parameter ( $S$ ), in order to adjust quantization, and prepared two types of CGHs; in the first case,  $S$  is 1 and a certain number of pixels are assigned to create one cell. In the second case,  $S$  is 5 and each cell is created by using  $5^2$  times more pixels.



**Fig. 4.2** Sample 64 x 64 image for computer experiments.



**Fig. 4.3** Results from CGHs that are capable of modulating both phase and amplitude. Sinc modulation is seen at results and cannot be eliminated easily in this application since there is no pure phase modulating element. Lee's method is used. (a) CGH pattern (b) reconstruction (c) CGH pattern with 5 times more precise (d) reconstruction.



**Fig. 4.4** Results from CGHs that are capable of modulating both phase and amplitude. Sinc modulation is seen at results and cannot be eliminated easily in this application since there is no pure phase modulating element. Lee's method is used with grayscale levels instead of rectangular apertures. (a) CGH pattern (b) reconstruction (c) CGH pattern with 5 times more accurate (d) reconstruction.

The device pictures are shown in Figs. 4.3 (a) and (c). And the corresponding results are given in Figs. 4.3 (b) and (d). The last row depicts the more precise configuration. So, the CGH device in (c) is much bigger than (a), which are all shown as the same size for convenience. As expected, the performance of the CGH device with bigger sensitivity parameter shown in (d) is better than the one with lower sensitivity parameter which is shown in (b). However, precise device output must either be smaller or, the resolution on hologram plane must be increased.

The same discussion holds for Fig. 4.4, where gray level toning is used to modulate phase and amplitude (Fig. 4.1 (b)). Sensitivity parameter is still the same, i.e., 1 and 5. Now for more sensitivity, larger number of gray scale level tones has been used. So, there is no sacrifice of resolution or space. Again, the results show that, better image quality is possible by increasing the  $S$ .

Note that there are no parameters defined for lens focal length, or pixel sizes. The reason for not specifying any spatial information is due mainly to the lack of need in this case since free space propagation is not in use, such that the pixel size will only affect the image size, but the image size can be arbitrary since we are interested only in image quality. In addition, a larger diffraction pattern has been given in figures, in order to show DC component of light (undiffracted). The intensity of this component is huge with respect to diffracted orders (images do not show this effect).

As it's seen up to now, when we make both phase and amplitude modulation on the same plane, resolution limitations become stricter. One must spend too many device pixels to encode one image pixel. In addition, using this method causes low diffraction efficiency as well as a very noisy environment; the undiffracted light coming from the DC component of the device gives a very strong peak at the center. Also, as it is seen from the images, the screen cannot be shifted for centralization of orders so there is no zero diffracted order achieved. Thus, this method like its similar others with cell method such as [42, 43], uses amplitude variations to affect both phase and amplitude for image formation. At that point it would be clever to combine both direct phase and amplitude

variation to increase image quality performance. This approach might be considered as an ultimate attempt to remove any noise residuals from results since all the constituents of the light are in the game now.

### **4.3 Implementation with Cascaded CGHs**

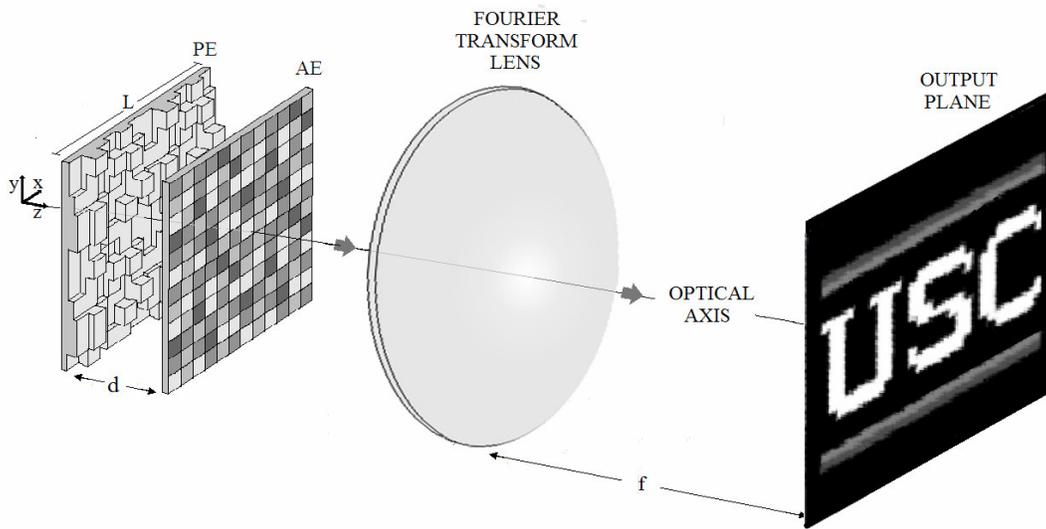
#### *4.3.1 Two dimensional images*

Phase and amplitude modulation may also be achieved at different planes with a specific distance between them. The proposed scheme is shown on the left hand side of Fig. 4.5. In this new method no cell encoding is used. We have two planes separated with distance  $d$ . One of the planes is responsible for phase modulation and the other is responsible for amplitude modulation. The phase modulating plane is assumed to have different relief profiles corresponding to each pixel in order to make it function. Likewise, amplitude modulating plane has different gray scale values for its pixels. Since there is no cell construction stage, resolution requirements may stay the same.

Now, the question to ask is; how to design this system such that it gives high quality images when it is used with Fourier transform lens as shown. It is obvious that when  $d$  is zero, a perfect result is obtained. However, it is physically impossible to modulate both phase and amplitude independently on the same plane. In fact, it has been observed that, the distance between associated planes is very important. As the distance gets larger, error increases abruptly.

In implementing the extended IFTA for Fig. 4.5 some slight changes need to be done to the flow scheme of pure phase structures (given in Fig. 3.5). In computing the form of amplitude element (AE), instead of discarding amplitude information like in a system of pure phase elements, now just phase information has to be removed and amplitude information is to be kept. Another difference is the distance between planes; unlike the separation of pure phase CGHs which is kind of arbitrary, phase & amplitude CGHs' separation is taken as in the order of microns. In this case a nearly perfect reconstruction might be possible. For realistic separation distances, some modifications should be required. Basically, when Eq. 3.10 is inspected carefully, some parameters are

seen to be related. These are separation distance, pixel size and refractive index of the separating medium (the last two are buried in spatial frequencies and wavelength). In fact, it turns out that, pixel size and (or) refractive index can be increased for compensation of the extremely small separation distances of phase and amplitude elements. The below results are obtained for our usual separation distance, i.e.,  $d = 300 \mu\text{m}$ . Also used are,  $0.632 \mu\text{m}$  wavelength on  $64 \times 64$  pixels where size of one pixel is  $50 \mu\text{m}$ . Sampling size is  $1.25 \mu\text{m}$  again. Focal length  $f$  for the lens is  $250 \text{ mm}$  yielding  $3.2 \text{ mm}$  of output as in the example of Section 3.3.2. In addition, in place of free space propagation, a material with  $n = 3.5$  is assumed to be ubiquitous between phase and amplitude modulation. Note that, large pixels cause smaller output which can be overcome by making the lens focal length larger with the same proportionality.



**Fig. 4.5** Phase and amplitude modulating planes with separation  $d$ . Phase element and amplitude element (AE) are placed successively.

Iterations saturate very quick yielding a result shown in Fig. 4.6 which is a result of the setup above. As it is seen, this is a nearly perfect image. For comparison, some numbers can be given. The MSE value of the below result is 0.022. If just one phase element would be employed like in Fig. 3.1, the MSE would be 0.71 and 0.11 in case of even 5 PEs.



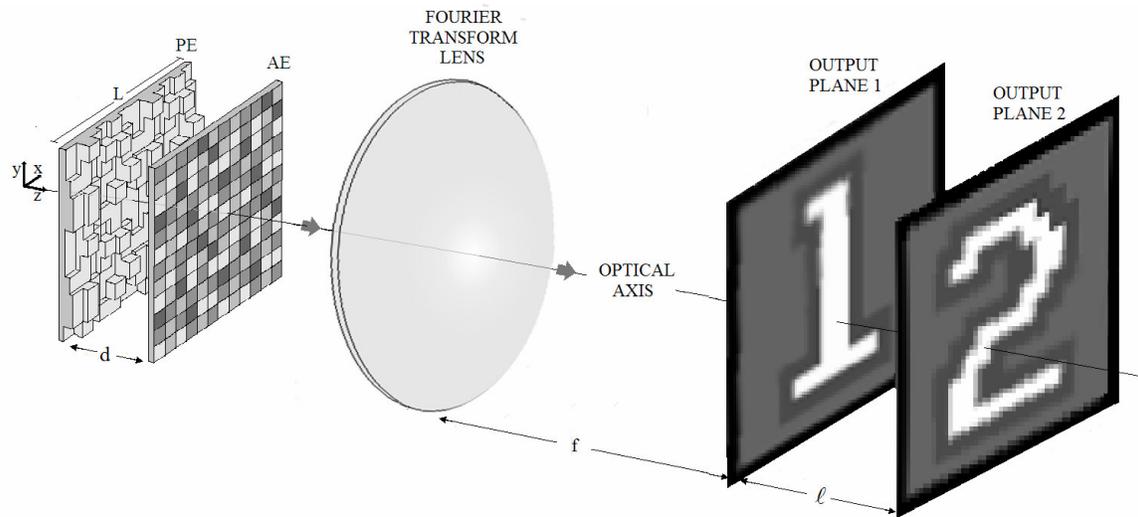
**Fig. 4.6** Result of phase-amplitude modulation with cascaded CGHs. An almost perfect (noise-free) reconstruction is possible when proper adjustment of parameters and setup is implemented.

Note that 64 levels is used in both phase and amplitude quantization. In the latter, that is achieved by grayscale levels by setting 0 for black and 1 for white as end points. For the more strict constraints like 4 levels, extended SA type of approach might be more appropriate. In this case, the flow chart in Fig. 3.6 is applied. Perturbations on amplitude element are applied by changing the gray level arbitrarily for a randomly chosen pixel. In phase and amplitude modulation the trends are similar for any kind of image.

In this part, a nearly noise-free reconstruction has been shown to be achieved through the usage of cascaded phase and amplitude modulation simultaneously. Furthermore, in design stage, one only needs the algorithms similar to the phase-only elements discussed in the previous chapter. Now, similarly, this idea will be extended to 3D realization.

#### *4.3.2 Three dimensional images*

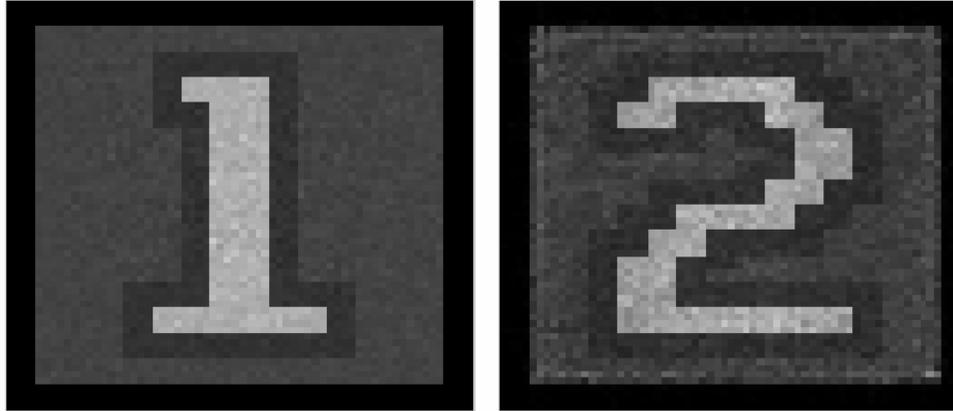
As an abstraction for 3D, two image planes will be used for reconstruction. These two planes are chosen to be like the ones in Fig. 3.17. In addition, the setup in Fig. 4.5 is extended for the new configuration. Finally, phase and amplitude modulating setup for 3D reconstruction might be plotted as in Fig. 4.7.



**Fig. 4.7** Phase and amplitude modulating planes with separation ' $d$ ' for a 3D reconstruction.

By extending the logic of multi-plane reconstruction of pure phase elements like in Section 3.4.1 and the flow chart in Fig. 3.15, one can easily optimize the system given above consisting of one PE and one AE. In this case, the phase information at the AE will be discarded instead; and the amplitude information will be saved. One point must be emphasized that, in finding the PE, there is a division with AE. This division may be a problem in case there are zero values at some pixels. The solution of this issue is to exchange all zero pixels with one, and let the system optimize itself according to this information and data.

Computer experiments have been implemented with the same parameters as in the Section 4.3.1. Pixel sizes on PE and AE planes are both adjusted to be  $50 \mu\text{m}$  with  $f = 250 \text{ mm}$ . Obtained results are outstanding and displayed in Fig. 4.8 below. One important thing to be emphasized here is the intensity calibrations of displayed results. Since amplitude adjustment is involved in computation stage, diffraction efficiency is supposed to be lower. Because the main goal here is to remove noise, intensity calibration has been done such that the total energy is the same as those given in Fig. 3.23 which displays results coming from cascaded PEs for the same size images.



**Fig. 4.8** Result of phase-amplitude modulation with cascaded CGHs. An almost perfect 3D reconstruction is possible when proper adjustment of parameters and setup is implemented. Numerical MSE values are 0.03 for the first plane and 0.39 for the second. Image intensity has been adjusted for comparison with those in Fig.3.23.

As seen from obtained results, there is remarkable increase in quality coming from the proposed setup. In the first plane, sample noise is barely noticeable and for the second plane, crosstalk and significant proportion of noise resulting from the distance  $d$  are eliminated and only a small amount of residual noise may be found. This result seems to surpass the results of 1 or even 5 phase elements. From a larger perspective this result makes sense since the two ingredients, namely phase and amplitude are both given to the system for imaging. Note that the separation between PE and AE can be made zero mathematically for computational simulations. In this case, minimum total noise accumulates at the second output plane due to image incompatibility. The importance of the phenomena was discussed in Section 3.6.

After all, usage of one phase and one amplitude modulator has changed the picture. Basically, one might get almost perfect targets after proper implementation and parameter selection in both 2D and 3D cases. The requirement for an extra large pixel size with respect to ordinary dimensions in diffractive optics seems to be a problem. The important point is to notice the relationship between separation distance  $d$  and pixel size. As  $d$  increases, pixel size should be increased exactly equals to the square root of the increment factor of  $d$ , to get the same result. Although small pixel sizes are one of

advantages of diffractive elements with respect to SLMs, this configuration might be used in SLMs too with time-varying pixels for real time outputs. By also considering the pixel size of today's typical LCD computer displays which are approximately 200  $\mu\text{m}$  [44], the idea in this chapter has a potential to be used in near-future displays under coherent laser illumination for enhanced multi-plane outputs. Some more systematic and elaborate study on phase and amplitude modulation may be required especially with SLMs by also taking variable-wavelength option into account.

# CHAPTER 5

## EXPERIMENTAL

### 5.1 Introduction

In the previous chapters some different micro-holographic-display schemes have been proposed. Mainly, they were related to either pure phase or phase & amplitude modulation. In this chapter, we will first look at fabrication possibilities of our proposed device, and then experimental errors are going to be examined briefly. These are mainly related to misalignment errors in experimental stage of the device. Finally, the reader may find a section about fabrication of CGHs using standard equipment like a laser printer, commercial overhead projector transparency and optical laboratory tools. Experimental results, which will be presented, have been fulfilled by using ubiquitous lab equipment. This part should be considered apart from the rest of topics given in this thesis and mainly related to practical aspects of CGHs.

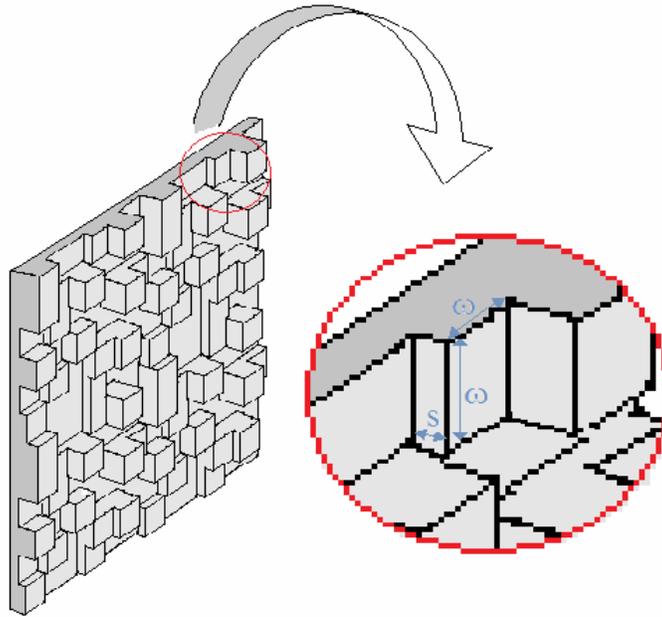
### 5.2 Fabrication of Cascaded CGHs

#### 5.2.1 General production methods

The fundamental question in DOE production is; how to fabricate these micron-scale structures with minimum possible error. Lots of different approaches and techniques are available for that [29, 45]. Lithographic techniques are commonly used in manufacturing of such components with some varieties. They are based on clean room processing and

employ the same technology as integrated circuit production. Since the features of diffractive optics are compatible in size with semiconductor devices, lithographic methods are optimal for fabrication.

In Fig. 5.1 one CGH as a form PE representation is shown. In addition, one portion of the phase relief structure is magnified to demonstrate system parameters. Each pixel has square shape with dimensions  $\omega \times \omega$  and depth  $s$ .



**Fig. 5.1** A CGH is shown with magnified section module. One square-shaped pixel is displayed with feature size  $\omega$  and etch-depth  $s$ .

The etch depth  $s$  (smallest feature that is fabricated) is the portion that is responsible for the relief profile and vital for the functionality of operation. The smallest phase difference due to that distance is,

$$\Delta\varphi = k_n s \tag{5.1}$$

Note that the unit for  $\Delta\varphi$  is radians.  $k_n$  is the wave-number within the material and by definition  $k_n = 2\pi m / \lambda$ . So for glass and air, we can write,

$$\Delta\varphi_{glass} = \frac{2\pi n_{glass}s}{\lambda} \quad (5.2a)$$

$$\Delta\varphi_{air} = \frac{2\pi n_{air}s}{\lambda} \quad (5.2b)$$

Thus the net phase difference is

$$\Delta\varphi = \Delta\varphi_{glass} - \Delta\varphi_{air} = \frac{2\pi(n_{glass} - n_{air})s}{\lambda} \quad (5.3)$$

Therefore,  $s$  can be found from Eq. (5.3)

$$\begin{aligned} s &= \frac{\Delta\varphi}{2\pi} \frac{\lambda}{(n_{glass} - n_{air})} \\ &= \frac{\lambda}{Q(n_{glass} - 1)} \end{aligned} \quad (5.4)$$

$Q$  denotes number of quantization levels in phase profile. As an example, the formula in Eq. (5.4) can be applied to simulations in this thesis; In Chapter 3, applications with  $Q$  equals to 4 and 64 exist with  $\lambda = 0.63 \mu\text{m}$  and  $n_{glass} = 1.5$ . Using this information,  $s$  is computed as  $0.315 \mu\text{m}$  for 4 levels and  $0.02 \mu\text{m}$  for 64 levels.

In binary optics fabrication process, each layer is formed individually. Then one needs to use masks depending on the number of quantization levels of the CGH. The general rule for that can be expressed mathematically as,

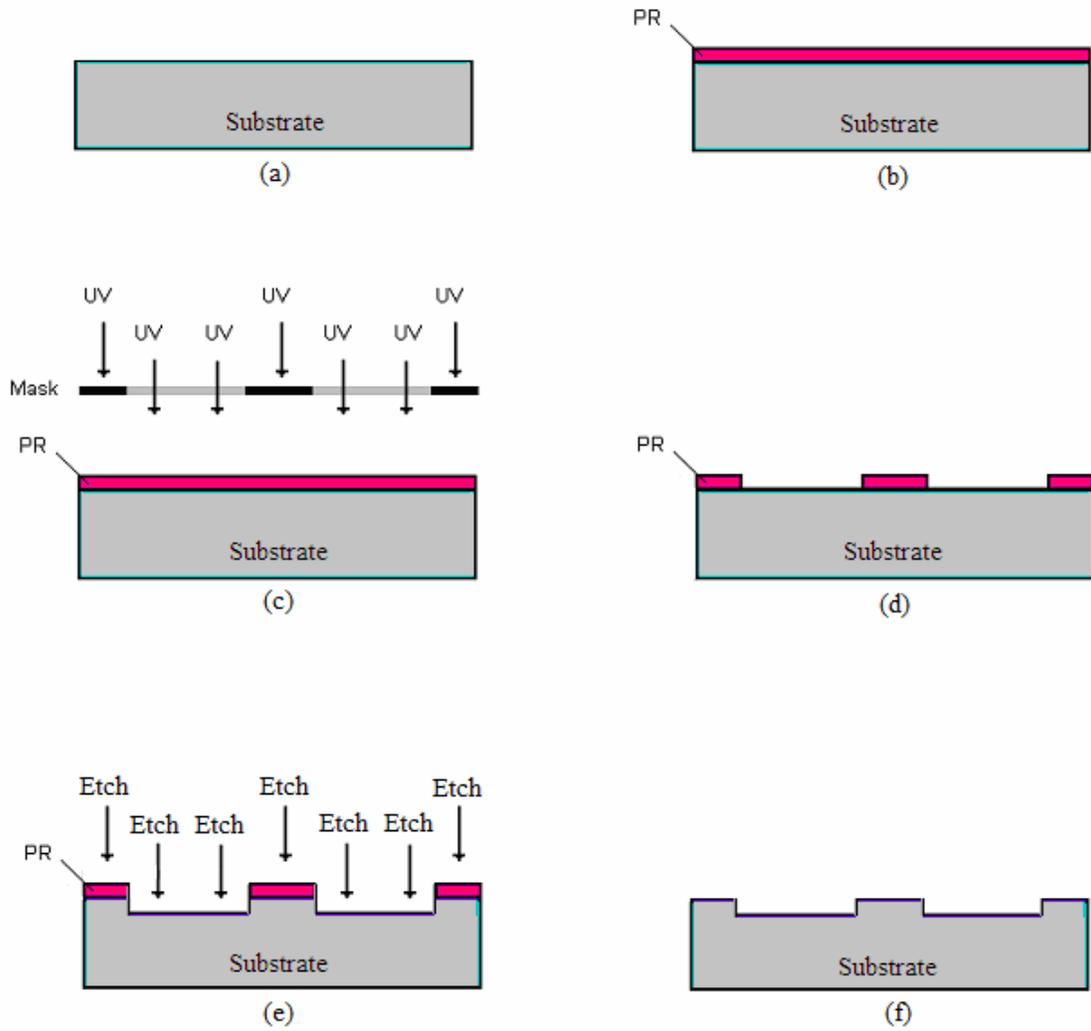
$$Q = 2^{\text{Number of Masks}} \quad (5.5)$$

So, every layer is fulfilled by a different mask. In case of 4 quantization levels, for instance, one needs 2 different masks in total. The process of fabrication is shown in Fig. 5.2. First, a substrate is taken and coated by a special light sensitive polymer called photoresist as in (a) and (b). A mask is needed to identify points to be etched. In masking process, the substrate is exposed to UV light and necessary parts defined by mask, are cleared from photoresist. These steps are shown in (c) and (d). Lastly, chemical etching removes the parts without photoresist and finally one layer is complete in (e) and (f). So, this selective etching enables the manufacturer to create different surface heights. This process can be continued with another mask for another layer in logarithmic mask

sequence. Each etching process involves errors due to the fact that accurate removal of material near sides is relatively hard. These processes must be implemented in a clean room for protection from dust and micro-organisms.

After all, mask application is the most important part since the pattern is created in this step. Most masks are binary transmission plates with transparent and opaque portions. In gray-scale lithography, though, a spatially varying transmission profile with gray tones is in use [46]. They are again prepared by photoresist and some lithographic stages similarly. In addition to that, in exposing the pattern on to the mask, some optical pattern generators or electron-beam machines are used. In etching step of CGHs, wet or dry etching might be chosen. In the former, some chemicals are used to create surface relief profile. On the other hand, in the latter, the sample is bombarded with high energy ions in order to knock off the molecules from required parts.

The method explained above is called photolithography since pattern on the mask is conveyed through UV light and it is the main technique in this area. Apart from that, there are some other lithographic techniques. One common approach is to write the pattern directly into the photoresist layer with the help of an electron-beam or laser beam [45, 47 and 48]. In this case, instead of a successive masking procedure for every layer, intensity of the beam is varied along the surface to create relief profile such that the local exposure is proportional to the required depth. This is called direct writing fabrication and possesses some advantages. One of the important advantages is the easiness in fabricating large number of phase levels, such as 256 or more. One disadvantage, however, is the fact that each pixel must be written one by one so it is not appropriate for serial processes. Overall this method is quite proper for the devices proposed here, especially the ones with 64 quantization levels. There are also holographic techniques which exploit interferometry for pattern generation on the photoresist [49].



**Fig 5.2** Production steps of a layer of phase relief structure in photolithography. (a) Substrate, usually glass. (b) Application of photoresist. (c) Masking application (d) Result (e) Etching (f) Final form

Whatever technique is used in lithography for photoresist carving, the next step is etching which has been explained before briefly. Other than that, both lithographic and etching steps can be combined for more direct approaches, as the next paragraph pinpoints.

In addition to lithographic techniques, there are direct machining methods (such as single point diamond turning) where substrate is carved through direct removal of substance. These processes can produce high-quality CGHs but might be relatively slow [50].

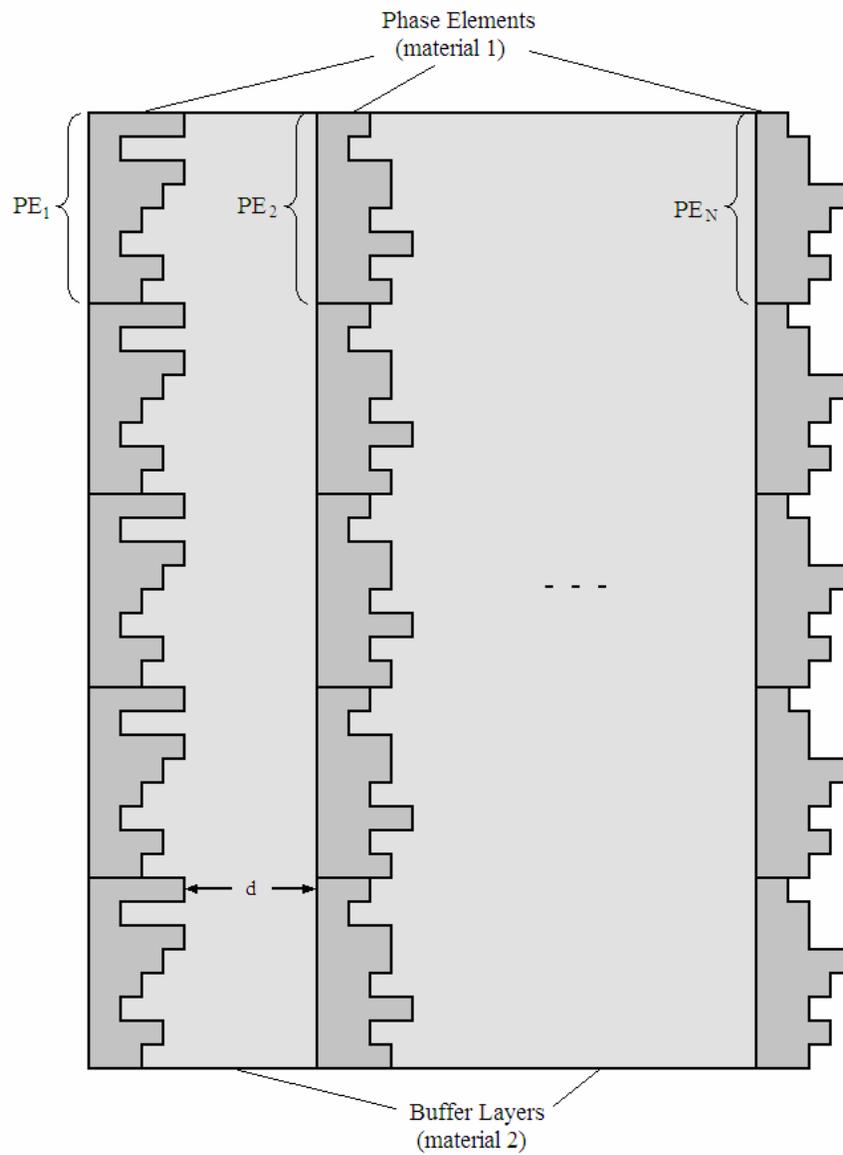
In operation of transmission CGHs, some of the light will be reflected back from parallel surfaces ( $\sim 4\%$  per surface). Furthermore, a little portion of that reflected light is re-reflected again, contributing to image formation thus causing some unwanted noise. In order to maximize the overall efficiency of a transmission CGH, it is desirable to minimize the amount of reflection. In many cases, this is achieved through overcoating the diffractive and flat surfaces with thin films which are called, antireflection coatings (AR). The usage of ARs can significantly reduce the amount of light reflected although the energy directed into a desired diffraction order may not be improved. Since in this work, the main concern is to reduce noise, ARs can perfectly serve that purpose since they are very efficient in suppressing successive reflections and eliminating noise coming from these reflections. One of the common coating materials for bare glass is magnesium fluoride,  $\text{MgF}_2$  [51].

### *5.2.2 Production of proposed setup*

In our proposed system consisting of several stacked diffractive surfaces, a production material of fused silica, FK51A or BK7 glasses might be used for elements. On the other side, there are two options for fulfillment. One option is to fabricate elements separately and align them using x-y-z optical aligners (similar to the schemes in Chapter 3). In the second one, however, one could fabricate diffractive elements into a substrate one by one. Similar setups for different purposes can be found in references [18] and [52]. Now, these two techniques will be examined briefly.

In the former, elements can be fabricated separately by using electron-beam lithography for shaping the photoresist and then etching through ion beam (milling), or employing one of direct machining methods. In experimental phase, very precise alignment is required. In this case, one has to deal with three translational and three rotational degrees of freedom for tuning. A similar setup was used in [52] with piezoelectric stages.

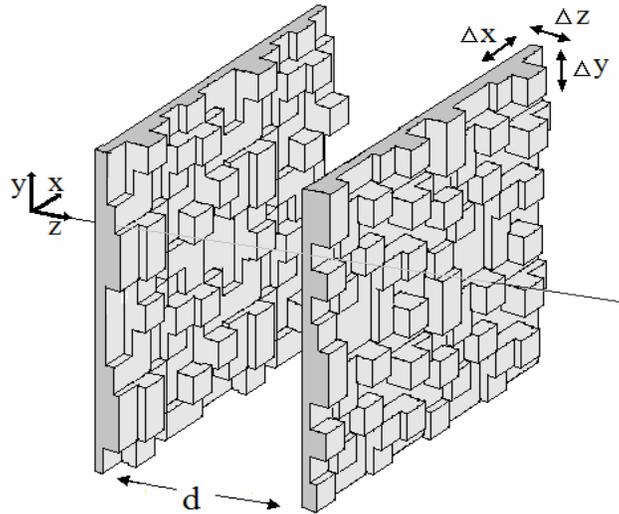
When the latter method is to be chosen, where cascaded CGHs are produced in only one entity, extreme care should be taken in the manufacturing stage. Once fabricated well, however, this method enables users to enjoy from a unique quality device. As seen in Fig. 5.3, there are diffractive surface layers as a matrix and buffer layers for propagation. CGHs serve as diffractive surfaces and replicated on the lateral extent in order to reduce speckle effects (Section 6.1.3). Steps are somewhat similar to the previously proposed method. First, substrate (material 1) is exposed to e-beam lithography and ion-beam etching, ending up with one layer of diffractive surface. Then, a buffer layer (material 2) is deposited using plasma enhanced chemical vapor deposition (PECVD). Next step is to deposit the material 1 by using PECVD, and to etch it for the formation of the second diffractive surface. These processes follow another until the end. The two important points here are, the alignment issue again, but this time in the production stage, and choice of these two materials for diffractive and buffer layers. The required alignment might be achieved in lithographic process, which is much easier than aligning in usual lab environment. Selection of materials is another point. If Eq. (5.4) is inspected, the etch depth 's' is seen to be inversely proportional to index difference between material 1 and material 2. Therefore, this means that, when different materials are to be chosen, the only thing that has to be changed is s from the diffractive surface point of view. On the other hand, since optical path length is  $nd$ , with  $n$  is the refractive index of the propagation medium, the propagation distance is affected and buffer layers' thicknesses should be adjusted accordingly.



**Fig. 5.3** Final view of the compact device. After PEs are computed, they are replicated for speckle reduction (at least 3 or 5 times). And fabricated successively with separation distance  $d$ . For the figure, there will be  $N \times 5$  PEs contributing to image formation. Material 1 and material 2 are intended to be used for PEs and buffer layers, respectively.

### 5.3 Misalignment Errors

In this section, some alignment errors in case of experiment will be examined. Up to now, it should be clear that, the proposed setups require micron-level precision and either manufactured together or separately, these CGHs need to be handled very carefully in the final experiment stage. So, we will just be examining alignment errors no matter where those elements are aligned, in production or experimentation stage. Actually, since high precision devices are used in clean-room processing, we can assume that most of the alignment errors would come in case of separately fabricated structures. One possible case is illustrated in Fig. 5.4.



**Fig. 5.4** Two PEs are shown. One of them is misaligned in three spatial coordinates, and this is denoted as  $\Delta x$ ,  $\Delta y$  and  $\Delta z$ .

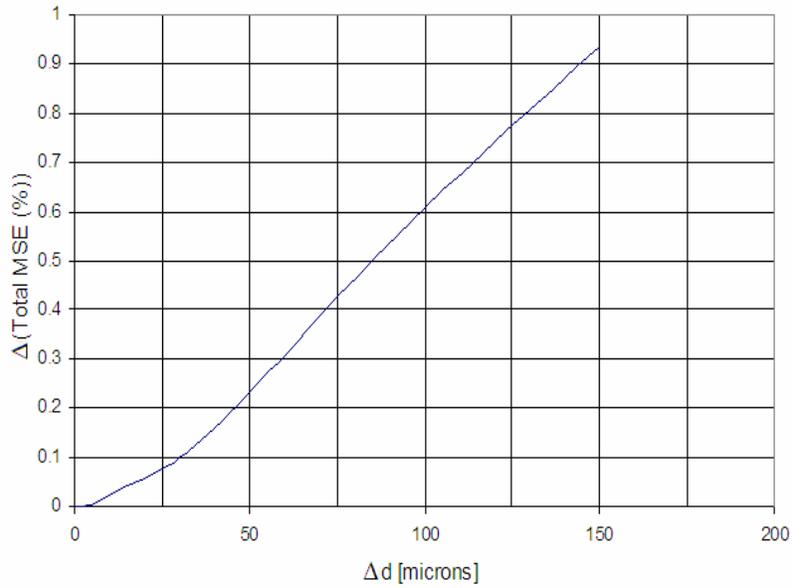
Basically, by misalignment error, we mean that the system is optimized for some spatial parameters and in the final reconstruction stage a change in these parameters is observed giving rise to a distorted result. As an evaluation tool, since our main concern here is to reconstruct 3D output, two image plane reconstruction scheme will be applied by using the flow chart in Fig. 3.15. The reader can assume that similar results will be obtained in case of a stochastic approach.

Simulations have been performed in order to analyze possible effects of spatial errors  $\Delta x$ ,  $\Delta y$  and  $\Delta z$  on final reconstructions by using two PEs, as a representation of a simplest case. For the below results shown in Figs. 5.5 and 5.6, images in Fig. 3.17 are used in a form of 64 x 64 pixels, with  $f = 25$  mm,  $d = 0.3$  mm and  $\ell = 40$  mm.

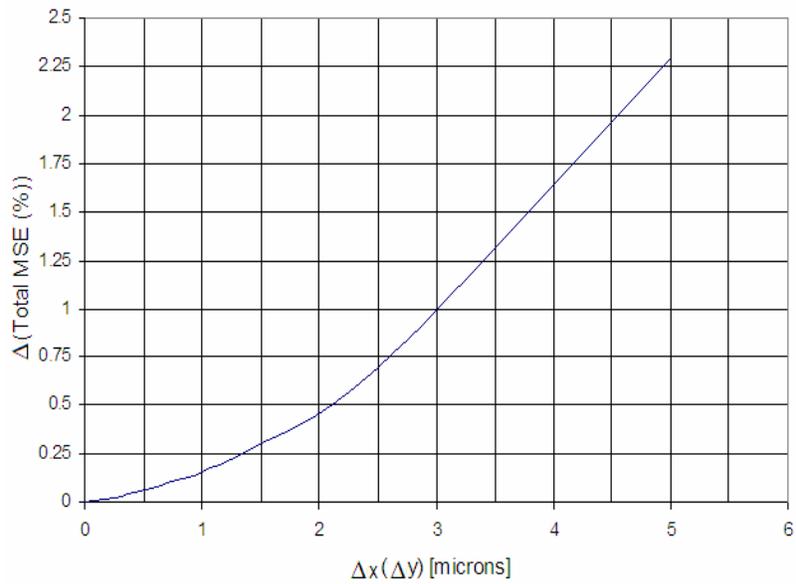
Fig. 5.5 shows the change in total reconstruction error in two plane imaging when the distance  $d$  is changed infinitesimally. Therefore, it reflects the results when one of the two PEs are misaligned as an experimental error in  $z$ -direction with magnitude  $\Delta z$ . According to the graph, extra noise caused by separation error, increases linearly after approximately 25  $\mu\text{m}$ . For longitudinal direction, error coming from  $\Delta z$  does not seem to affect the result much. Even for 100  $\mu\text{m}$ , increase in error is 0.6. In order to grasp what it means and to ease comparison, some numbers can be given referring to the results from Chapter 3. For example one would get total MSE of 2.5 approximately when just one PE is used and 1.6 in case of two PEs for reconstruction of the same image set under perfect conditions.

When Fig. 5.6 is considered, error increase trend can be noticed as similar from some aspects so that it becomes linear after some point but in this case system is much more sensitive to lateral displacement errors. Even a few microns may make the system's performance degrade down to the case of just one PE. This result seem to make sense because in the iterations optical system adjusts itself according to corresponding pixels on different PEs on the same propagation direction, and when disturbed, noise increases abruptly.

After all, we can summarize this section by expressing that cascaded CGH setups are much more tolerant for longitudinal displacement errors rather than lateral. Even for lateral displacement errors, an error of 500 nm, for example, causes 0.1% MSE increase. Therefore, placement and alignment issues are important but within achievable limits of sensitive x-y-z aligners. When rotational misalignments involve, it can be expected that similar magnitudes of MSE increase may occur. For more information on rotational manipulations in propagation, reference [53] is suggested.



**Fig. 5.5** This graph displays the change in total MSE versus the change in distance between two PEs in case of two image planes.



**Fig. 5.6** This graph displays the change in total MSE versus the change in lateral distance between two PEs in case of two image planes. This change can be in  $x$  or in  $y$  direction.

## 5.4 Experiments with Single Intensity-Modulated CGHs

The experimental results, which will be presented here, have been fulfilled by using existing lab equipment found in USC electrical engineering optical computing laboratory and printer in signal and image processing institute. One layer of intensity CGH is used. Before passing to the results, experimental setup is introduced.

Our experimental setup is shown in Fig. 5.7. And conceptual drawing is shown in Fig. 5.8. It basically includes, He-Ne laser, microscope lens (60X), spatial filter (25 $\mu$ m) and several lenses. Light output created in laser first. Then it passes through filter where spatially incoherent parts are cleared out, and a point source is created. After the filtering process, an aperture is used to limit excess light. Next, spherical wavefronts are converted to planar waves by using a lens with focal length 25 cm. Now, the light beam is ready to be processed by the diffractive element which is our CGH. Output coming from CGH passes through an imaging (Fourier transform) lens. This lens has a focal length of 75 cm. The focal length is chosen to be relatively large so that the final image's dimensions are relatively larger. The final result appears at one focal length behind the Fourier transform lens (lens 2). But, this image is still too small and therefore needs to be magnified. The last lens (lens 3) seen in Fig. 5.8 enlarges the image and gives result on white screen. For the magnifying lens ( $f = 22.5$ cm) at hand, 3x magnification is desired. In order to calculate appropriate image and object distances:  $s$  and  $s'$ , geometrical optics formulae are revisited. The imaging and magnification ( $M$ ) can be expressed as;

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{s'} \quad (5.1)$$

$$s' = M s \quad (5.2)$$

respectively with  $M = 3$  and  $f = 22.5$ cm. Since there are two equations with two unknowns, this set may be solved easily, yielding  $s = 30$ cm and  $s' = 90$ cm.

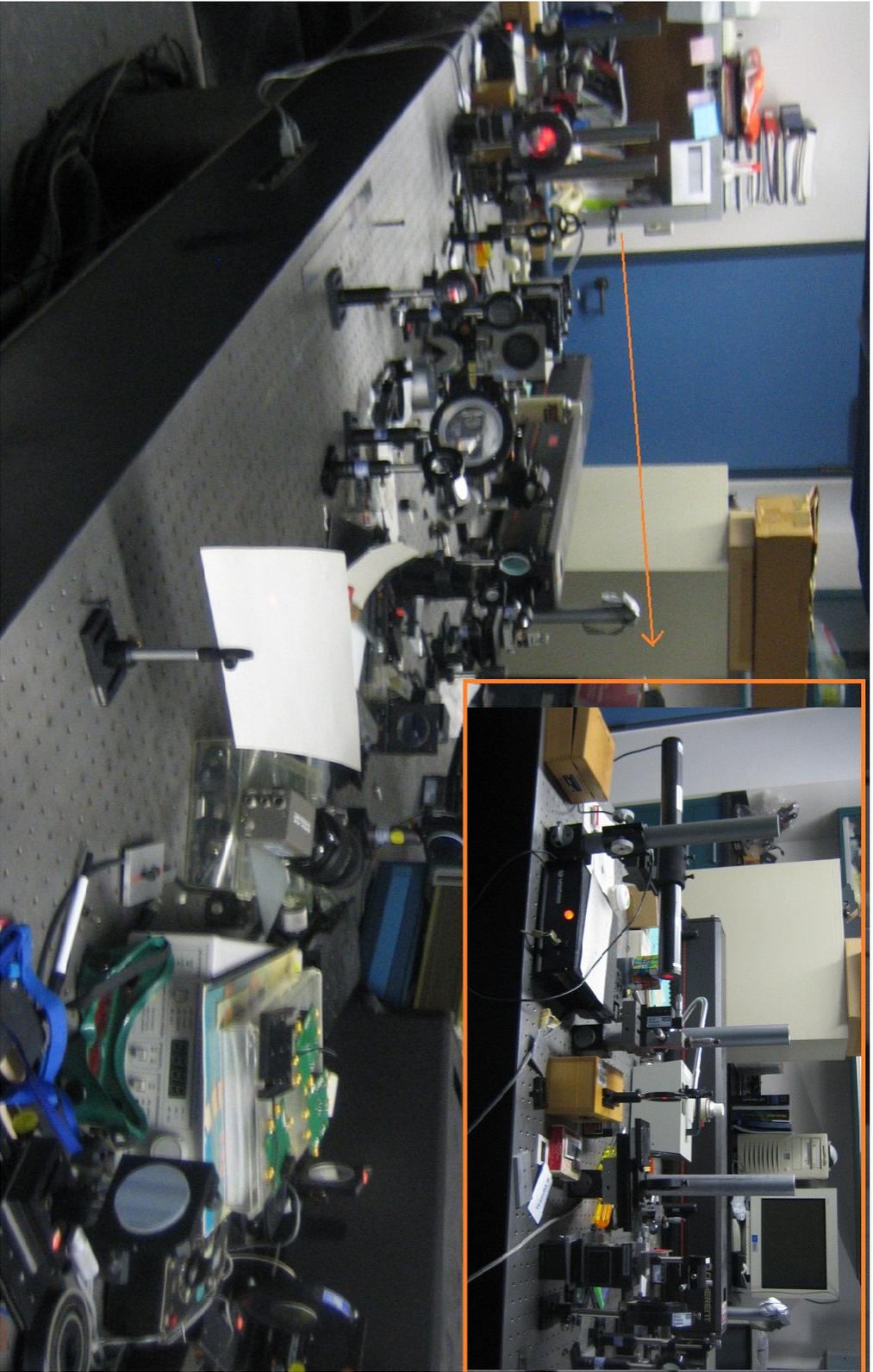
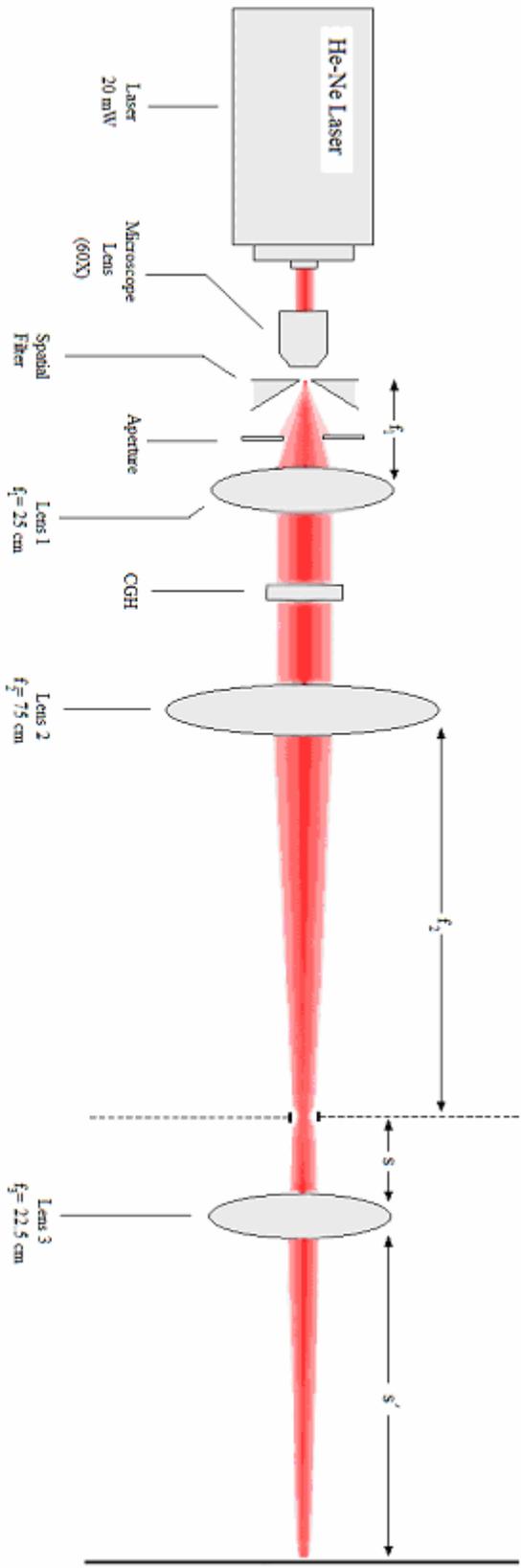


Fig. 5.7. Optical table



**Fig. 5.8.** Conceptual drawing of the experimental setup

### 5.4.1 *The He-Ne laser*

A helium–neon laser (He-Ne), is a type of gas laser whose gain medium consists of a mixture of helium and neon gases inside of a small bore capillary tube. The He-Ne laser operates at a wavelength of 632.8 nm in the red part of the visible spectrum. It was developed at Bell Telephone Laboratories in 1962 .

The gain medium of the laser, is a mixture of helium and neon gases, in approximately a 10:1 ratio, contained at low pressure in a glass envelope [54, 55]. The energy or pump source of the laser is provided by a high voltage electrical discharge passed through the gas between electrodes (anode and cathode) within the tube. The optical cavity of the laser usually consists of two concave mirrors, one having very high (typically 99.9%) reflectance and the output coupler mirror allowing approximately 99% reflectance. The basic working principle is as follows; The mechanism producing population inversion and light amplification in a He-Ne laser plasma originates with inelastic collision of energetic electrons with ground state helium atoms in the gas mixture. These collisions excite helium atoms from the ground state to higher energy excited states. Then, collisions between excited helium and ground state neon atoms results in a selective and efficient transfer of excitation energy from the helium to neon atoms due to near coincidence between the energy levels of those two elements. Metastable neon atoms lase with stimulated emission accompanying spontaneous emission [56]. The final light output goes out from the mirror with relatively less reflectance, and the reflected light contributes the energy inside cavity and helps ground state He atoms excite.

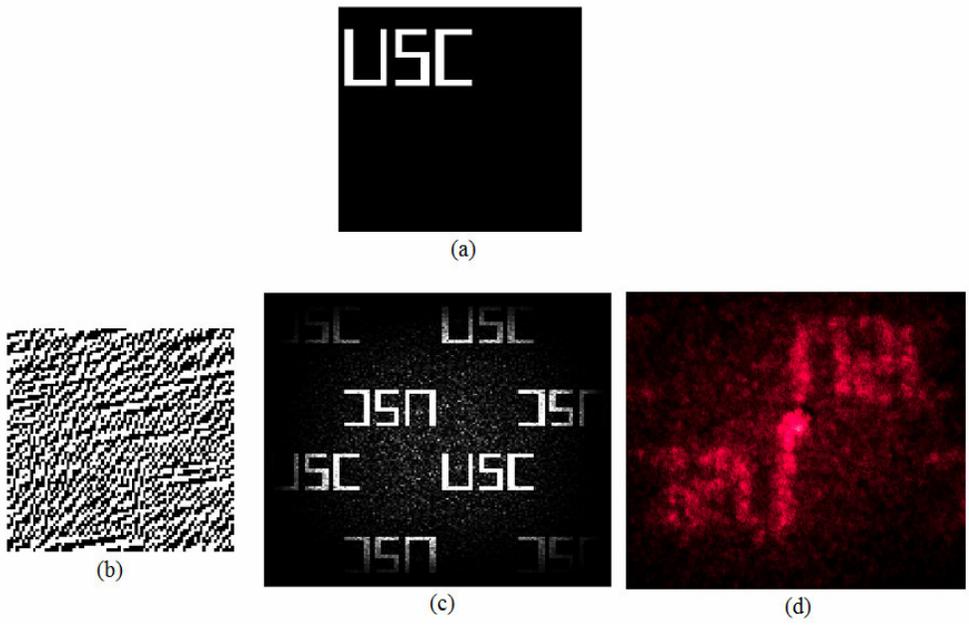
Commercial He-Ne lasers are relatively small devices, among gas lasers, with cavity lengths usually ranging from 15 cm to 50 cm and optical output power levels ranging from 0.5 to 50 mW. They are preferred due to cost, relatively long coherence length and ease of operation.

#### 5.4.2 Production of Holograms

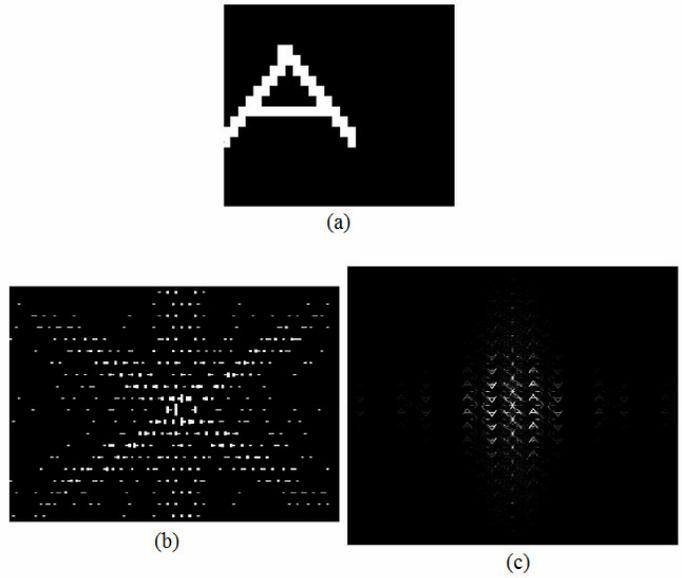
We have designed two types of CGHs shown in Figs. 5.9 and 5.10. One with the algorithm defined in Section 3.3 and the other with the method from Section 4.2.

For the former method, first algorithm in Fig. 3.2 is applied, then all the calculated phase values have been changed to binary intensity levels (black and white). Since cleanroom fabrication is not yet available, all pixels are engraved by using a commercial printer on plastic transparencies and employing halftoning technique. Pixel size is limited to dpi (dots per inch) value of the printer. Pixel size is taken to be  $150\ \mu\text{m}$  and number of pixels is  $80 \times 80$ . No dummy window is present. Final image dimension can be calculated by using the formula again (from Chapter 2) that relates spatial frequency and dimension:  $D_x = f_x \lambda f$  where  $D_x$  is the dimension of final image in  $x$  direction,  $\lambda$  is wavelength which is  $0.63\ \mu\text{m}$ ,  $f$  is the focal length which is  $75\ \text{cm}$  and finally  $f_x$  is the maximum spatial frequency of the object along  $x$ . Since pixel size is  $150\ \mu\text{m}$ ,  $f_x$  is equal to  $1/150\ \mu\text{m}^{-1}$ . Therefore  $D_x$  is found to be  $3.15\ \text{mm}$ . As stated before, this image is magnified 3 times from the last lens in setup. So, our final image is almost  $9.5\ \text{mm}$ . Since 'USC' letters covers one half in width and one third in height, it seems relatively smaller. These letters are placed off-centered due to twin image problem. Binarization (either in phase or amplitude) causes twin image problem in outputs. Twin images can be seen symmetrical around origin as may also be seen from the figures.

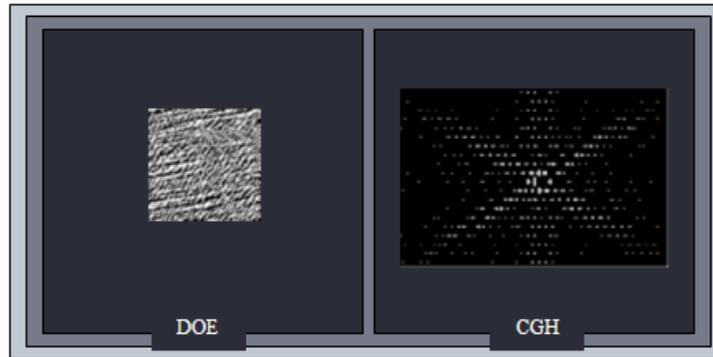
For the latter hologram, the technique which is explained in Section 4.2 has been used. Since the result is already binary no other process is applied at the end. The Fourier transform of the device gives desired output. Unfortunately, because of the extra small size of this image, features are hardly seen and digital camera is not powerful enough to capture quality pictures. So there is no experimental data coming from this device but suffice to say that the output is very similar to the predicted one given in part (c). We also display the devices that these holograms are mounted (Fig. 5.11). They are made in a non-sophisticated way; transparencies are placed between two microscope slides after a specially designed and cut paper. And the result is robust and professional look CGHs.



**Fig. 5.9** (a) Desired image (b) Hologram (c) Numerical output (d) Experiment



**Fig. 5.10** (a) Desired image (b) Hologram (c) Numerical output



**Fig. 5.11** Final device look fabricated and used in Optical Computing lab. Left hand side shows the device in Fig. 5.9 (b) and right hand side shows the device in Fig. 5.10 (b).

Based on this experiment, it is apparent that the first hologram is advantageous due to handy pixel size. So again, as pointed out in the beginning of Chapter 4, cell encoded elements suffer resolution problem commonly and that is one of the reasons that people use pure phase modulators. This fact increases the importance of the thesis from the point of view of noise removal for such structures.

This chapter summarizes and proposes fabrication methods of diffractive elements and the novel diffractive display treated in this thesis. In addition, we mention an inexpensive method of making simple diffractive elements which is not completely dependent on the rest of the work.

# CHAPTER 6

## DISCUSSION & CONCLUSION

### 6.1 Some Issues

Some important points of interest as well as future research directions will briefly be mentioned in this chapter. Possible drawbacks and ambiguities need to be discussed in the implementation of simulations and experiments as much as reviews of some important points. These points are intended to be emphasized so that a future continuation of this work yields better results. In addition, an elaborate discussion from the point of similar works in literature as a complementary part can be found in this chapter.

#### *6.1.1 Convergence*

In this thesis the main emphasis was on algorithms to design various configurations of diffractive elements. Error reduction process for these algorithms as iterations flow have not been mentioned in detail yet. The convergence behavior of both algorithms in this work is very similar to the traditional case where one uses a single entity for both image and object domains. Basically, MSE converges in a smooth manner and number of iterations roughly depends on the number of PEs used. When there are too many CGHs, however, these algorithms do not have any guarantee to converge, although in the stochastic case the chances are far higher with the expense of computation time.

### 6.1.2 Sampling

Sampling is another issue in computational optics. In the implementation of algorithms, sampling requirements should be considered carefully. In case of violation, for example when propagation distance exceeds limits, the system under consideration is forced to use a distorted propagation matrix. Hence the results would be approximate. Roughly, one can obtain reasonably accurate results unless the limits are violated extensively. Between phase elements, the sampling size has to be smaller than a pixel size (upper bound) in order to simulate higher diffraction orders in propagation [57]. The smaller the sampling size, the more precise the results, although after a certain limit finer sampling does not affect results much (one fourth of a pixel in our case). Moreover, the sampling theorem puts limits (lower bound) on the design parameters at both device and image domains when it comes to sampling of the transfer function in Eq. (3.10) [13]; this is due to the increasing slope of the phase term in the transfer function with respect to spatial frequency. For non-aliased results, maximum change in the absolute phase must be limited by a constant (Chapter 2). Therefore, this puts extra boundaries on the sampling problem, so that the sampling size has now a lower bound as shown in Eq. (2.37). As long as sampling conditions are satisfied and distances are far apart enough for scalar diffraction theory, any distance  $d$  between PEs may be used in the algorithms and setups described in this thesis. There is a tradeoff for the distance  $\ell$  between image planes such that it may not be allowed to be very small. In case of a short distance trials, all the image planes come out to be as the same as the first one. This is because of the fact that, light field cannot adjust itself in short propagation distances to form different images in planes after the first one.

When the sampling condition expressed in Eq. (2.37) or (2.38) is considered again, one concern is that any propagation simulation is limited by the distance and for relatively far distances our condition seems to cause imprecise values and fail gradually. In order to overcome this condition, one may derive sampling in spatial domain, instead [14]. Same procedure might be applied by using the Fresnel convolution kernel and then taking the Fourier transform of it. In that case, transfer function for propagation can be sampled appropriate only for distances farther than our condition. But this method brings problems

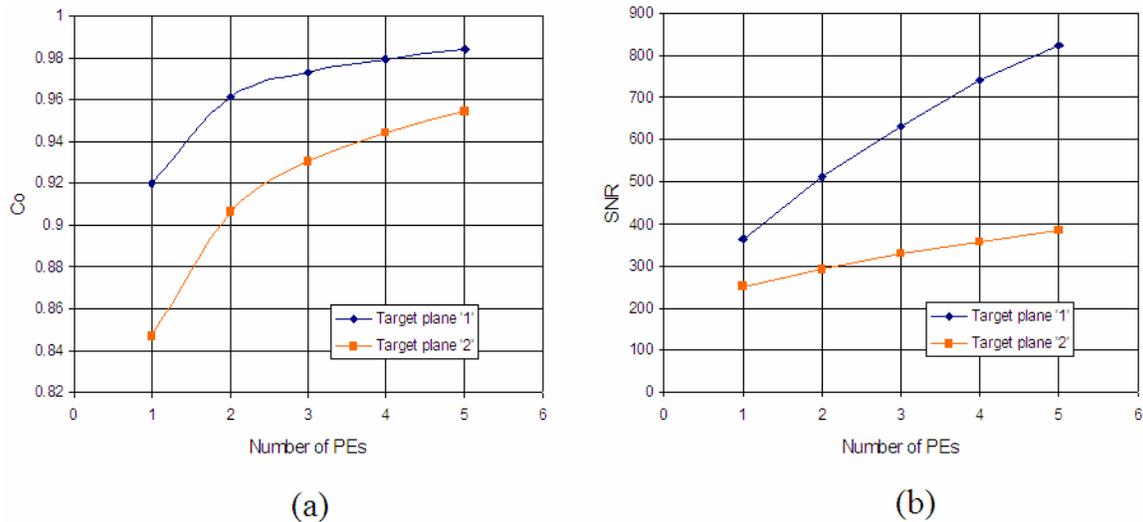
in the forward and backward transform, so that even if the sampling works fine in either direction alone, it does not work perfect when applied consecutively.

### *6.1.3 Speckle formation*

One other issue is speckle formation at the outputs of the proposed devices. If the bandwidth of the complex object exceeds the available hologram area, speckle formation (aliasing) is observed at the reconstruction planes [58, 59]. This problem can be solved by two main approaches computationally. One of them is to take the speckle pattern into account in design. The entire design process can be implemented with a zeropad array paved around the calculated element. That zeropad array enlarges frequency domain thus causing more pixels at the output. So, by assuming speckles in the computational design implicitly, one can reduce negative effects partially. One important issue in iterative algorithm stage is handling the zero pixels. They create problems in case of division. So, when division is involved, just phase freedom with unit amplitude may be employed. Note that this only occurs in design process, where backward propagation is involved, therefore when simulated at the end after the design, that point is no longer an issue since only forward propagations are involved. In addition simulated annealing type algorithms can be tried for this purpose which may yield better fit since no backward transform is applied in any step. Instead of this complicated design recipe, one may also replicate the final CGH at the end as a second method. Hence, replicates at different spatial locations disturb the phase matching and allow elimination of speckles. For practical applications, for example, usually three replicates are enough for one dimension yielding nine replicates for a planar CGH. Of course, when cascaded elements are used the same operation must be applied to all. Actually, speckle formation is closely related to the output planes' imaging capabilities. For instance in case of CCDs at target locations used in the experiment, the user is limited to the size of a pixel of the device. Similarly, human eye also has a resolution limit on the retina in case of direct observation. Therefore, detector resolution limits the speckle formation such that the coarser the resolution the less speckles that can be observed.

### 6.1.4 Other error metrics

In this work, mainly MSE metric, defined in Eq. (3.1), is used for image quality assessment. But, there are some other measures for the same purpose. Because of that, this section is devoted to sample demonstrations of these other measures in comparison with our standard metric, MSE. In fact it is hoped that the results in this section can especially be useful when assessing similar works in the literature in conjunction with this one. As mentioned and defined before in Eq. (3.4),  $C_o$  is a statistical metric for similarity relationship between two sets of data. Another useful metric would be SNR as in Eq. (3.7). In computation of SNR, RMS of total target image field is simply divided by its noise counterpart, which can be expressed as absolute value of target image field minus reconstructed image field. This metric is more popular in fiber-optic communications theory. We would like to give sample results when different metrics are in use in the evaluation of cascaded PEs' performance. One of the important results in this work was the demonstration of error reduction as a response to addition of phase modulating elements. So, for example, the results shown in Fig. 3.22 may be analyzed by employing different metrics. In Fig. 6.1 (a), the same result is expressed in terms of  $C_o$ , and in (b) it is expressed in terms of SNR.



**Fig. 6.1** (a)  $C_o$  value versus number of PEs (b) SNR versus number of PEs, for the experiment results shown in Fig. 3.22.

In case of non-similarity, the lowest value for Co is zero, whereas the highest is one when perfect match is observed. Due to the fact that Co is a sensitive parameter, values change in a relatively small range. On the contrary, SNR varies through a large range and yields almost linear (with slight curvature) results interestingly. Actually, it is not possible for SNR to go infinity as number of PEs increase, because in that case noise would have to go zero, which is impossible (at least for the second plane due to incompatibility). So, we can conclude that in the interval of 1 to 5 PEs, the trend looks more like linear but it is probable to become constant after a certain point. The mathematical limit for SNR is infinite in case of noise free reconstruction, the lower bound, however, is case dependent. The characteristics of curves above (linear or logarithmic) are different due to different expressions of definitions. For the parameters of the above experiment, the reader may again consult to Section 3.4.1.

#### 6.1.5 Binary phase quantization

In Chapter 3, while analyzing different phase quantization effects, binary levels, i.e. 2 phase quantization levels were not introduced. One reason for that is basically its physics is a little bit different from the rest. Other than that, the same trends can easily be observed with more degradation of results due to limited degrees of freedom. Regardless of the illumination wavelength, the outcome of a binary DOE is symmetric with respect to the origin with twin images at every diffracted order. A simple proof can be found in [38].

#### 6.1.6 Lensless Functionality

A lens is used in concepts and diagrams in order to ease imaging. In the lensless case, the distance from the last diffractive element to the first image plane must be sufficient enough to support Fraunhofer approximation as noted near the end of Section 2.2.1. Eq. (2.15) which describes Fresnel propagation can be re-written as follows,

$$\psi(\xi, \eta) = \frac{e^{ikz} e^{\frac{k}{2z}(\xi^2 + \eta^2)}}{i\lambda z} \iint_A \left[ \psi(x, y) e^{\frac{ik}{2z}(x^2 + y^2)} \right] e^{\frac{ik}{z}(\xi x + \eta y)} dx dy \quad (6.1)$$

According to Eq.(6.1), when the exponential factor near  $\psi$  goes to 1, a simple Fourier transform expression can be obtained. This fact can be seen when it is compared with Eq. (2.19a) by putting  $k = 2\pi / \lambda$  and  $z = f$ . Therefore, the necessary condition may be observed when,

$$z \gg \frac{k(x^2 + y^2)}{2} \quad (6.2)$$

In above equation, the distance  $z$  between device and image domains must be as much as possible to guarantee a correct approximation. In Eq. (6.2),  $(x, y)$  denotes source coordinates. As an extreme case,  $x_{\max}$  and  $y_{\max}$  must be used. When applied to our case, where the source is considered to be the cascaded structure with 64 by 64 pixels,  $x_{\max} = y_{\max}$  corresponds to 160  $\mu\text{m}$  (32 pixels, from the center to the edge multiplied by 5  $\mu\text{m}$  pixel size). Thus, the limiting value on  $z$  is calculated to be 256 mm, which is almost ten times larger than our usual lens focal length,  $f$ . Of course, in this case the output will be larger based on propagation distance. When the lens is not in the system, the FOV is limited.

## 6.2 Summary

### 6.2.1 Literature comparison

In this part, some recent literature information will be given in more detail with comparisons, pertinent to cascaded phase elements and phase retrieval techniques. In fact, there are also similar approaches that are used for various purposes in the literature.

In what follows, we summarize some of them in order to compare with our work. Iterative algorithms are often used in the process of phase retrieval. The typical phase retrieval problem that relates to our work is to infer the phase in a plane of interest, given intensity measurements in one or multiple planes. Intensity can be obtained experimentally in multiple planes, for example, by detecting in a succession of slightly

out-of-focus image planes [60], by employing a succession of different illumination techniques each of which gives rise to a different intensity pattern at the detector [61], or by using multiplexed optics that send images from different planes in depth to different spatial locations in a detector plane [62]. In all of these cases, a multiplane-like algorithm can be employed to infer the phase in the original plane of interest. Some of these algorithms use iterative IFTA-like algorithms based on propagation of the wave field, and others employ instead the transport of intensity equation (TIE). The TIE can be solved using a non-iterative method which proceeds by calculation of infinitesimal intensity variation along  $z$ . Either approach could in principle be applied to the image-plane side of our system, in order to find a phase distribution in one plane on the target-image side of our system. For our case, the iterative approach has advantages over the TIE approach in that it is more robust to phase discontinuities; also this method is not as common as the iterative approaches and a comparison is discussed in [60].

More generally, there are a number of key differences between the problem we are solving and the phase retrieval problem. First, our end goal is to find the phase modulation to implement in  $N$  cascaded planes; whereas the typical goal in phase retrieval is to find the physically existing phase in one plane. Second, in our case the given image intensity planes are not assumed to be optically compatible; but in the case of typical phase retrieval, the planes of measured intensity from a chosen object will indeed be optically compatible. And third, in either problem there may be many possible solutions for the phase in the plane(s) of interest. In our case, any one of the solutions will be sufficient, as long as it gives rise to a reasonable rendition of the target image plane intensities; in the phase retrieval case, the solution that is most similar to the original (unknown) phase needs to be found.

Also, there are some particularly relevant studies on cascaded CGHs for various purposes in literature, which need to be mentioned briefly. Deng et. al. [22] analyzed cascaded diffractive elements for multiwavelength interconnects. Their approach uses a weak-phase-deviations approximation technique [63] for the iterative design algorithm, and considers output as an indescriptive 3D continuum rather than slices of image planes.

The system is applied to wavelength-division demultiplexing using Gaussian incident beams. In addition to that, in a similar work, Shi et. al. [28] used two phase elements right after an incident key image that contains multiple hidden images. The phase elements serve to reconstruct the hidden images in the Fresnel domain. Their algorithm employs an iterative type of approach and computes these two phase elements in one loop, where the loop visits one image plane in each turn. On the other hand, in our work, the iterative algorithm flows all the way forward and each phase element is computed separately.

In the last part of this chapter, a more general approach is considered, in which one could expand diffractive surfaces by both cascading in  $z$ -dimension and supplementing with a window in  $x$ - $y$  domain. Although noise window concept is known, neither its effect on MSE with respect to its noise area (by considering just one PE as a classical case) has been analyzed systematically, nor an entire scheme is ever mentioned by taking cascade phenomena into account. In fact, the last part in Chapter 3 outlines noise removal process in the case of ultimate expansion in three spatial dimensions. Theoretical minimum can be reached in multi-plane imaging by enough expansion in a form of additional pixels or planes as PEs. One other important result tells that when the separation distance between image planes increases, optical incompatibility decreases. In those cases, the optical problem becomes like a 2D image reconstruction especially in the extreme values of separation distance. Again in such cases, the effect of cascade of elements are debatable and just the noise window concept might be applied.

### *6.2.2 Thesis Summary*

Finally, in this work some novel display applications are proposed. As unique features of our work, diffractive elements which have a capability to form multiple images at different depths have been proposed. Error reduction is investigated systematically as a function of ubiquitous phase and amplitude modulating elements. For phase only display, two mainstream design approaches are developed and proposed based on extensions of traditional methods. It has been shown that adding successive phase elements to the 3D image reconstruction problem might yield significant improvement under different constraints. Especially at the final stage of the work for phase-only elements, the total

MSE has also been examined because individual output planes in 3D representation may be algorithm dependent such that for iterative approach the noise is unequally distributed, however in annealing method it may be equaled. In addition, coarse quantization of phase levels degrades the results and limits achievable quality drastically; and simulated annealing, which has not been considered much in such cascaded setups in the literature, handles pixels individually thus being a remedy for cases with strict quantization constraints as explained in Section 3.4.1. Fundamentally, phase modulation phenomena with successive planes expands the system's solution space by adding more degrees of freedom, and causes error reduction in an inverse exponential-like pattern as a function of the number of phase modulators. This fact may also be exploited especially when the feature size is close to the wavelength and adding more pixels on the same spatial extent would be ineffective for one phase element. Phase and amplitude modulating display, on the other hand, might be more efficient as long as pixel dimensions are kept relatively large. Also note that, in the working scheme of the devices here, there is no one to one correspondence between input phase elements and target image planes in the longitudinal dimension; indeed, information in any output image plane is distributed among all the input phase elements.

From a more technical point of view, the concept can be applied to the construction of a versatile and light weight micro-holographic 3D display. It might also find usage in various areas such as medical imaging, head-up displays, digital image encryption or even in fiber optic interconnections. One way of achieving practical operation could be realized by fabricating relief structures sequentially onto the same substrate separated by buffer layers, which also minimizes alignment flaws and is explained at the end of Chapter 5. That proposed configuration can in principle be executable with current technology. Preliminary work indicates that precision is required especially in lateral dimensions. This requirement, however, has much more tolerance in longitudinal dimension.

As a result we proved that enlarging the CGH will cause noise reduction and if this enlargement is as a form of parallel cascades longitudinally, the outcome is reduced

optical incompatibility for multi-focal setups, whereas laterally expanded CGHs show improvement as well in expense of resolution and diffraction efficiency. For future work directions, physical realization of cascaded phase elements may be implemented by following the recipes given in this thesis. Furthermore, different techniques can be investigated for 3D reconstruction other than multi-planar representation. Also, some other algorithms might be used in optimization such as genetic algorithm or ant colony optimization. It would be beneficial to exploit parallel computing in order to speed up computation. Mathematical treatment of increment of information capacity [64] due to addition of extra elements could be done for a more theoretical and general view. Or, the concepts given here may be combined with some other work in diffractive optics realm such as the one in [65] where space-variant interconnections are implemented by successively placed optoelectronic components including diffractive elements.

## APPENDIX A

The formulation of diffraction problem is based on expressing the solution of the well known homogeneous wave equation at an arbitrary point in terms of the values of the solution and its first derivative on an arbitrary closed surface surrounding that point. In order to handle boundary conditions, Green's function formalism is used as a derivation tool [66]. It is a useful identity specifically employed in complex electromagnetic calculations. The identities due to George Green (1824) follow as simple applications of the divergence theorem. The divergence theorem basically relates the behavior of the flux of a vector field through a surface and its divergence inside that surface.

$$\iiint_V \vec{\nabla} \cdot \vec{C} dV = \oiint_S \vec{C} \cdot \vec{n} dA \quad (\text{A1})$$

Where  $V$  is a volume enclosed by a closed surface  $S$ . Also,  $\vec{n}$  is unit normal vector to the corresponding surface and  $\vec{C}$  is an arbitrary vector function. Now let this function is chosen to be  $\vec{C} = \psi \vec{\nabla} \phi$ , where  $\psi$  and  $\phi$  are arbitrary scalar fields. Then using vector identities, namely,

$$\vec{\nabla} \cdot (\psi \vec{\nabla} \phi) = \psi \nabla^2 \phi + \vec{\nabla} \psi \cdot \vec{\nabla} \phi \quad \text{and} \quad \psi \vec{\nabla} \phi \cdot \vec{n} = \psi \frac{\partial \phi}{\partial n}$$

When they are inserted into divergence theorem in (A1), Green's first identity is obtained.

$$\iiint_V (\psi \nabla^2 \phi + \vec{\nabla} \psi \cdot \vec{\nabla} \phi) dV = \oiint_S \psi \frac{\partial \phi}{\partial n} dA \quad (\text{A2})$$

If we write the above equation again but  $\phi$  and  $\psi$  are interchanged this time, then

$$\iiint_V (\phi \nabla^2 \psi + \vec{\nabla} \phi \cdot \vec{\nabla} \psi) dV = \oiint_S \phi \frac{\partial \psi}{\partial n} dA \quad (\text{A3})$$

Subtracting (A3) from (A2) side by side, Green's second identity is obtained.

$$\iiint_V (\psi \nabla^2 \phi - \phi \nabla^2 \psi) dV = \iint_S \left[ \psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] dA \quad (\text{A4})$$

When this theory is aimed to be applied to any diffraction problem, a well defined geometry is required, such as the one shown in Fig. A1. Consider an aperture on an opaque screen in front of a radiation field. Let  $P_1$  be any point on the aperture, the point of observation can be shown as  $P_0$ , and  $S_R$  denotes a closed surface surrounding  $P_0$ , as shown. Final aim of the formulation herewith, is to express optical disturbance at  $P_0$  in terms of its values on the surface  $S_R$ .

Consider an auxiliary function which is a unit-amplitude spherical wave expanding outwards from the point  $P_0$ . This is the same expression derived before and shown in Eq. (2.11).

$$\phi = \frac{\exp(ik |\vec{r}|)}{|\vec{r}|}$$

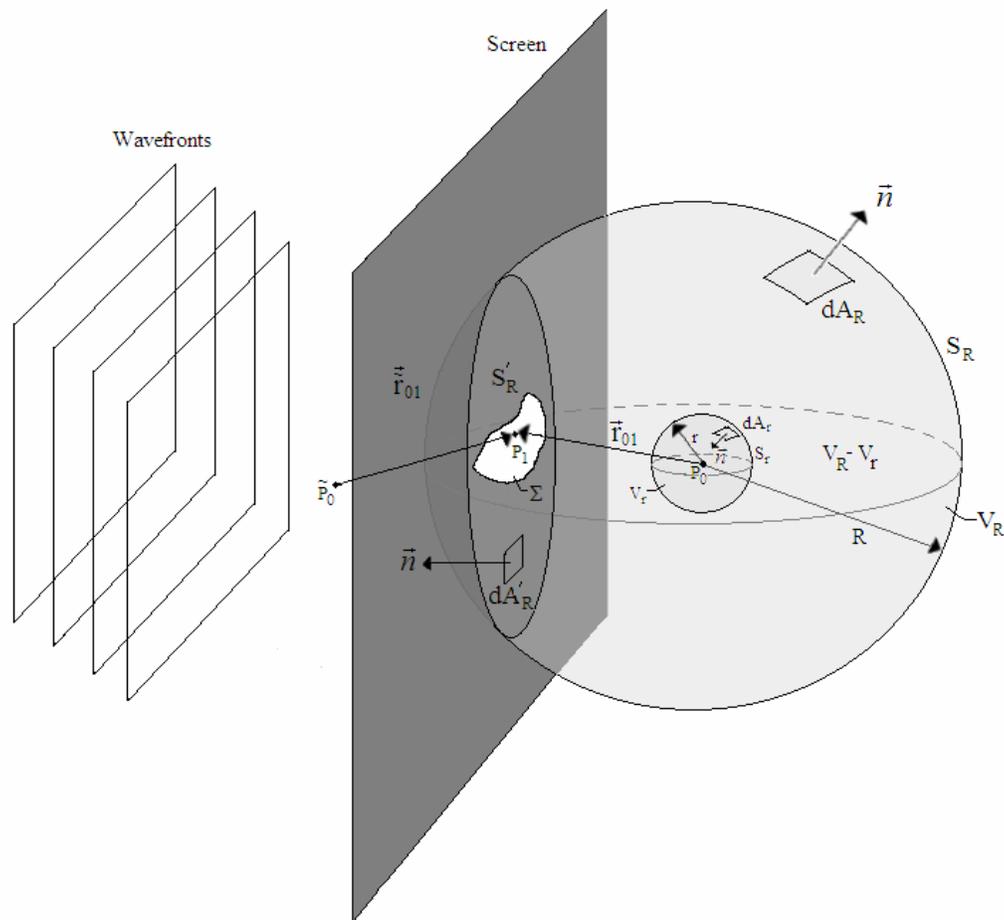
In order to exclude the discontinuity at  $P_0$  which blows up at origin, we assume a spherical surface with infinitesimally small  $r$  around  $P_0$ . Then Green's theorem is applied in volume between  $S_R$  and  $S_r$ , so the surface integration can be executed on composite surfaces, namely  $S_R + S_r$ , connected with an infinitesimally thin hypothetical channel and volume integration is implemented within  $V_R + V_r$ . Since desired solution  $\psi$  and the auxiliary function  $\phi$  both satisfy the Helmholtz equation in this new defined volume, we have from Eq. (2.6),

$$\nabla^2 \psi = -k^2 \psi \quad \text{and} \quad \nabla^2 \phi = -k^2 \phi$$

Left-hand side of Eq. (A4) simplifies to,

$$\iiint_{V_R + V_r} (-\psi k^2 \phi + \phi k^2 \psi) dV \quad (\text{A5})$$

This is basically zero. Thus the Green's theorem reduces to,



**Fig. A1** A screen with tiny aperture is shown. Surface normal vectors are denoted by  $\vec{n}$ . Hypothetical surfaces  $S_R$  and  $S_r$  are defined in the text and they serve the purpose of computing the field at  $P_0$ , assumed to be the origin in spherical coordinates.  $S'_R$  is a part of  $S_R$  that intersects the screen.  $V$  defines corresponding volumes. Finally,  $\tilde{P}_0$  is the mirror image of  $P_0$  and it is symmetric with respect to screen.

$$\iint_{S_R+S_r} \left[ \psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] dA = 0 \quad (\text{A6})$$

$$\iint_{S_R} \left[ \psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] dA_R = - \iint_{S_r} \left[ \psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] dA_r \quad (\text{A7})$$

For the derivative, by considering the surface  $S_r$ , we can derive,

$$\frac{\partial \phi}{\partial n} = \vec{n} \cdot \nabla \phi = \cos(\vec{n}, \vec{r}) \frac{\partial \phi}{\partial r} = - \frac{\partial}{\partial r} \left( \frac{\exp(ik|\vec{r}|)}{|\vec{r}|} \right) = \frac{\exp(ik|\vec{r}|)}{|\vec{r}|} \left( \frac{1}{|\vec{r}|} - ik \right)$$

where  $\cos(\vec{n}, \vec{r})$  factor represents the cosine of the angle between the normal and the radius of the small sphere  $r$ , and the normal  $\vec{n}$  is a unit vector for this sphere and points inwards (its because  $V_R + V_r$  composite surface has a normal pointing towards  $P_0$ ), and  $\vec{r}$  points outwards thus yielding cosine equals to -1. Substituting these in Eq. (A7) and letting the magnitude of  $r$  goes to zero gives a much simpler expression.

$$\begin{aligned} \iint_{S_R} \left[ \psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] dA_R &= - \lim_{r \rightarrow 0} \left\{ \iint_{S_r} \left[ \psi \frac{\exp(ik|\vec{r}|)}{|\vec{r}|} \left( \frac{1}{|\vec{r}|} - ik \right) - \frac{\exp(ik|\vec{r}|)}{|\vec{r}|} \frac{\partial \psi}{\partial n} \right] dA_r \right\} \\ &= - \lim_{r \rightarrow 0} \left\{ \left[ 4\pi r^2 \psi(P_0) \frac{\exp(ik|\vec{r}|)}{|\vec{r}|} \left( \frac{1}{|\vec{r}|} - ik \right) \right] - \right. \\ &\quad \left. \iint_{S_r} \left[ \frac{\exp(ik|\vec{r}|)}{|\vec{r}|} \frac{\partial \psi}{\partial n} \right] dA_r \right\} \end{aligned} \quad (\text{A8})$$

The second integral above is unknown since it includes normal derivative of  $\psi$ . Luckily, the entire part will vanish due to denominator which goes to zero.

$$\iint_{S_R} \left[ \psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] dA_R = -4\pi \psi(P_0) \quad (\text{A9})$$

Therefore,

$$\psi(P_0) = \frac{-1}{4\pi} \iint_{S_R} \left[ \psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] dA_R \quad (\text{A10})$$

Now suppose that  $R$  in Fig. A1 goes to infinity. Then, above integral vanishes for all values on  $S_R$  except the part cut by screen ( $S'_R$ ), as long as Sommerfeld radiation condition holds [12], due mainly to the functions  $\psi$  and  $\phi$ , which blows up at infinity. Thus,  $S_R$  can be replaced by  $S'_R$ .

$$\psi(P_0) = \frac{-1}{4\pi} \iint_{S'_R} \left[ \psi \frac{\partial \phi}{\partial n} - \phi \frac{\partial \psi}{\partial n} \right] dA'_R \quad (\text{A11})$$

This important result basically tells, in order to find the field at  $P_0$ , just the values on  $S'_R$ , namely values on transparent portion of the screen or aperture are required. At this point, another auxiliary function  $\phi$  can be generated not only by a point source at  $P_0$  but also simultaneously by a second point source at position  $\tilde{P}_0$  which is the mirror image of  $P_0$  on the opposite side of the screen. We assume the two sources are identical but a  $\pi$  phase difference. So, the expression for this new wave function at a point  $P_1$  within aperture is,

$$\phi(P_1) = \frac{\exp(ik |\vec{r}_{01}|)}{|\vec{r}_{01}|} + \frac{\exp(ik |\tilde{\vec{r}}_{01}|)}{|\tilde{\vec{r}}_{01}|} \exp(i\pi)$$

Where  $\vec{r}_{01}$  is a vector joining  $P_0$  to  $P_1$ . This function vanishes on the aperture. Note that the derivation done before also works for this auxiliary function as long as another hypothetical sphere pair (with infinitesimally small and large radii) are placed around  $\tilde{P}_0$ . Then, Eq. (A11) is simply written as,

$$\psi(P_0) = \frac{-1}{4\pi} \iint_{S'_R} \psi \frac{\partial \phi}{\partial n} dA'_R \quad (\text{A12})$$

The derivative can be found for more simplification,

$$\begin{aligned} \frac{\partial \phi}{\partial n} &= \cos(\vec{n}, \vec{r}_{01}) \left( ik - \frac{1}{|\vec{r}_{01}|} \right) \frac{\exp(ik |\vec{r}_{01}|)}{|\vec{r}_{01}|} - \cos(\vec{n}, \tilde{\vec{r}}_{01}) \left( ik - \frac{1}{|\tilde{\vec{r}}_{01}|} \right) \frac{\exp(ik |\tilde{\vec{r}}_{01}|)}{|\tilde{\vec{r}}_{01}|} \\ &= 2 \cos(\vec{n}, \vec{r}_{01}) \left( ik - \frac{1}{|\vec{r}_{01}|} \right) \frac{\exp(ik |\vec{r}_{01}|)}{|\vec{r}_{01}|} \end{aligned}$$

$$= 2ik \cos(\vec{n}, \vec{r}_{01}) \frac{\exp(ik |\vec{r}_{01}|)}{|\vec{r}_{01}|} \quad (\text{A13})$$

In above calculations, the angles between  $\vec{n}, \vec{r}_{01}$  and  $\vec{n}, \vec{\tilde{r}}_{01}$  are supplementary thus cosines are the same except for a minus sign, where  $\vec{n}$  here, is the normal to the surface  $S'_R$  and for magnitudes we have  $|\vec{\tilde{r}}_{01}| = |\vec{r}_{01}|$ . Furthermore, since  $|\vec{r}_{01}| \gg \lambda$ ,  $1/|\vec{r}_{01}|$  term may be neglected. By expressing the wave number  $k$ , in terms of  $\lambda$  we get,

$$\psi(P_0) = \frac{1}{i\lambda} \iint_{\Sigma} \psi \frac{\exp(ikr)}{r} \cos(\vec{n}, \vec{r}_{01}) dA_{\Sigma} \quad (\text{A14})$$

With  $\Sigma$  representing the transparent region (aperture) of  $S'_R$ . This result is known as Rayleigh-Sommerfeld theorem. It basically allows the field at any point to be expressed in terms of the boundary values.

## APPENDIX B

For any  $n$ , the binomial can be expanded as,

$$(a + x)^n = a^n + na^{n-1}x + \frac{n(n-1)}{2}a^{n-2}x^2 + \dots x^n \quad (\text{B1})$$

When  $x \ll a$ ,

$$(a + x)^n \approx a^n + na^{n-1}x \quad (\text{B2})$$

In its most frequently used form, where  $a = 1$  and  $n = 1/2$ , Eq. (B2) becomes,

$$(1 + x)^{1/2} \approx 1 + \frac{1}{2}x \quad (\text{B3})$$

## APPENDIX C

Consider an arbitrary one dimensional bandlimited function  $g(x)$  and its sampled form  $g(x)_{sampled}$  from Eq.(2.27a). The sampling interval will be shown by  $\Delta x$  as in the text.

$$g(x)_{sampled} = comb\left(\frac{x}{\Delta x}\right) g(x) \quad (C1)$$

and its spectrum,

$$\mathfrak{S}[g(x)_{sampled}] = G(f_x)_{sampled} = \mathfrak{S}\left[comb\left(\frac{x}{\Delta x}\right)\right] \otimes G(f_x) \quad (C2)$$

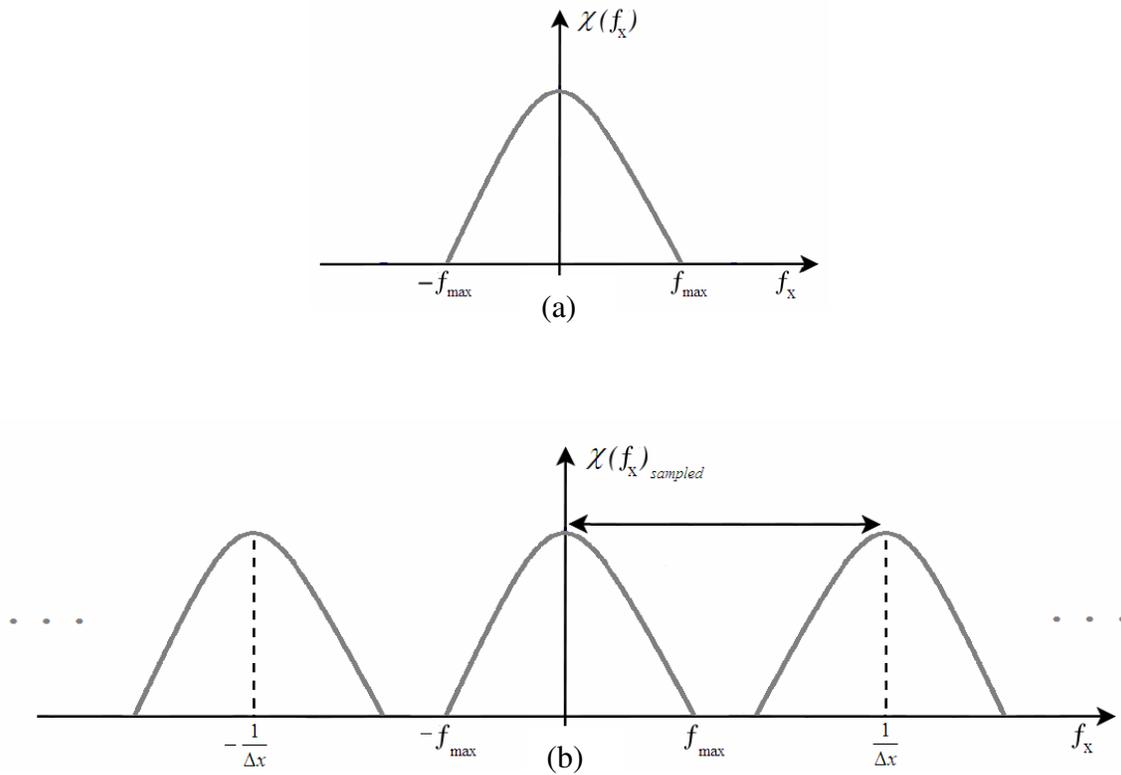
Where the sign  $\otimes$  shows convolution and  $G(f_x)$  is the transform of  $g(x)$ . The Fourier transform of comb function is still a comb. Thus, after using the definition of comb in Eq. (2.27c) and a little algebraic manipulations, we have,

$$G(f_x)_{sampled} = \sum_{m=-M/2}^{M/2-1} G\left(f_x - \frac{m}{\Delta x}\right) \quad (C3)$$

So, the spectrum of  $g(x)_{sampled}$  can be found by erecting the spectrum of  $g(x)$  about each point at locations,  $f_x = m/\Delta x$  with  $m$  is an integer given in Eq. (2.27b). It is apparent that the sampled function's spectrum  $G(f_x)_{sampled}$  is a repetitive form of the original spectrum of  $G(f_x)$  with interval  $1/\Delta x$ . These concepts are summarized in Fig. C1, where the amplitude ( $\chi$ ) of  $G(f_x)$  is shown in part (a) and the amplitude ( $\chi_{sampled}$ ) of  $G(f_x)_{sampled}$  is shown in part (b). Since the various terms in the spectrum of the sampled data are separated by distances  $1/\Delta x$ , it's guaranteed that no overlap happens between two adjacent terms in the spectrum as long as the below condition holds.

$$\frac{1}{\Delta x} \geq 2f_{\max} \quad (C4)$$

When the sampling condition holds then the separations of the various spectral islands will be large enough to assure that the adjacent regions do not overlap. Thus the recovery of the original spectrum  $G(f_x)$  from  $G(f_x)_{sampled}$  can be achieved exactly. Otherwise, some details disappear causing aliasing.



**Fig. C1** Amplitude versus frequency (a) Spectrum of a function (b) Spectrum of a sampled function. Separation of two terms is shown by an arrow. Amplitude is shown with  $\chi$ , so that  $G = \chi e^{i\zeta}$  and  $G_{sampled} = \chi_{sampled} e^{i\zeta_{sampled}}$  respectively.  $f_{max}$  shows the bandwidth.

## APPENDIX D

Consider the definition of mean squared error (Eq. (3.1)), where  $C$  is a reconstruction factor chosen to make MSE minimum for a given reconstructed image.

$$MSE(\%) = \frac{\sum_{All\ pixels} (|\psi_{target}| - C|\psi_{reconstructed}|)^2}{Number\ of\ pixels} \times 100$$

Mathematically the fact that MSE gets its minimum can be expressed as taking the derivative and equating to zero.

$$\frac{\partial}{\partial C} \sum_{All\ pixels} (|\psi_{target}| - C|\psi_{reconstructed}|)^2 = 0$$

$$\frac{\partial}{\partial C} \sum_{All\ Pixels} (|\psi_{target}|^2 - 2C|\psi_{target}||\psi_{reconstructed}| + C^2|\psi_{reconstructed}|^2) = 0$$

$$\sum_{All\ Pixels} (-2|\psi_{target}||\psi_{reconstructed}| + 2C|\psi_{reconstructed}|^2) = 0$$

$$\sum_{All\ Pixels} |\psi_{target}||\psi_{reconstructed}| - C \sum_{All\ Pixels} |\psi_{reconstructed}|^2 = 0$$

Finally,

$$C = \frac{\sum_{All\ Pixels} |\psi_{target}||\psi_{reconstructed}|}{\sum_{All\ Pixels} |\psi_{reconstructed}|^2} \tag{D1}$$

## APPENDIX E

Some pseudo-codes to implement certain functions in simulations for computational experiments will be presented here. These codes play important roles either in the main program body or as a user-defined-function. The logic and syntax are compatible with the programming language C and executed in Matlab®. Some image processing toolbox functions are also used when required. In addition, we also would like to mention some physical phenomena occurring in light propagation and imaging throughout the processes briefly so that the reader may find useful to take a look at this section other than codes which might come too technical to some.

### E.1 Image read

Images must be entered and read by computer before everything else starts. Below, 'I' denotes the image matrix that is represented by bitmap (bmp) format or sometimes tagged image file format (tiff). Actually every planar digital image is a three dimensional object by considering the color codes as well. So, for gray toned samples we just need to take one of these dimensions immediately after the sample is read. These steps are shown in the first two lines. Last line just adjusts the numerical format and normalizes by assigning unity as maximum.

```
I=imread('1.bmp', 'bmp');  
I=I(:,:,1);  
I=double(I)/255;
```

In computer's memory there are cells which consist of smaller units called bytes. A byte is the amount of storage required to store a single character. The number of bytes a memory cell can contain varies from computer to computer. A byte is composed of even smaller units of storage called bits which is an abbreviation of binary digit. One bit is the smallest element a computer can deal with and it is either 1 or 0. Generally there are 8 bits in a byte [67]. That allows  $2^8$  (256) configurations to be executed. Similarly in image processing there are 256 gray tones represented by a byte, maximum number is 255 as brightest by also considering 0 as darkest.

## E.2 Random Phase Plate (RPP)

Computer experiments start with elements that have random phase values in its pixels. So, a matrix with random phase entries is to be created.

```
for x=1:1:M
    for y=1:1:N
        RPP(x,y)=exp(i*2*pi*rand);
    end
end
```

Where, x and y are dummy variables, M and N are the number of pixels in x and y directions and 'rand' creates a random number between interval [0, 1].

## E.3 Upsampling (supersampling)

In simulating propagations between elements, it is necessary to create as much diffracted order as possible. For that purpose, one should take more than one sample per pixel. This will enlarge available frequency space.

```
for m=1:1:M;
    for n=1:1:N;
        for x=1:1:X
            for y=1:1:Y
                Mout(x+X*(m-1), y+Y*(n-1))=Min(m,n);
            end
        end
    end
end
```

In this script, x, y, m and n are dummy variables, X and Y are upsampling rates in the x and y directions, again, M and N are the number of pixels in x and y directions. Finally, Min is the input matrix and Mout is the output matrix.

The main goal in finding Mout is taking more samples without extending effective area. We actually upsample the field by shrinking the length between samples. This process can also be named as supersampling interchangeably.

### E.5 Downsampling

Another thing required is downsampling of the field. This is especially important in the stage of amplitude imposition stage of the target image. In fact, if the field is not shrunk through this process, incompatibility between original target would appear due to the fact that the field is processed in upsampled domain through elements. Note that upsampling of the target image as well, does not solve the problem since in this case experiments are executed as though there was a larger image instead of simulating the higher diffraction orders. For downsampling operation,

```
for m=1:1:M
    for n=1:1:N

        for x=1:1:X
            for y=1:1:Y

                g(x,y)=Min(x+X*(m-1),y+Y*(n-1));

            end
        end

        Mout(m,n)=mean(mean(g));

    end
end
```

Similar to the upsampling case, dummy variables, X and Y are the downsampling rates, M and N are the number of pixels again in x and y directions. Min is the input matrix and Mout is the output matrix. g(x,y) is a dummy matrix used in the script. So above pseudo-code segment takes a portion of data and simply finds its average over all data matrix. Note that, in these sampling schemes, the rates X and Y are usually powers of 2.

## E.6 Fourier Propagator

Fourier propagator engine is given in below script. It first creates spatial frequencies and uses the formula in Eq. (3.8)

```
fx=-1/(2*dx):1/L:1/(2*dx)-1/L;
fy=-1/(2*dy):1/L:1/(2*dy)-1/L;
[FX,FY]=meshgrid(fx,fy);
H=exp(-j*2*pi*z*sqrt((1./lambda^2)-(FX.^2+FY.^2)));
H=fftshift(H);
```

We show the sampling distance as dx (dy) and spatial frequencies as fx (fy). Note that, dx (dy) is different from number of pixels since every pixel is sampled more than once (4 times in our case). L is the length of the area being sampled from Chapter 2.

## E.7 Noise Window and Amplitude Imposition

In image imposition stage, target is placed by keeping propagated phases. When a dummy noise window is in use, however, one should be careful on this step. In application, target must be placed in the middle of available space and the rest of the area is used for noise accumulation by normalizing the amplitude and keeping the phase as it is.

```
C1=((k-1)/2)*M;
C2=((k-1)/2)*N;
Eout=abs(Ein)./max(max(abs(Ein)));

for x=1:1:M
    for y=1:1:N

        Eout(x+C1,y+C2)=sqrt(I(x,y));

    end
end
```

Where, Ein is the input field and Eout is the output field. C1 and C2 are internal parameters. Above segment just cuts the middle portion of the field and imposes the target amplitudes. Noise area remained fixed with normalized random amplitude in order to give freedom.

## E.8 Shifting

When evaluating the performance of the devices proposed, the last element must be multiplied with a simple phase mask in order to shift the entire diffraction pattern to the center. This phase mask plate is calculated by running the below script. It basically uses circular shifting theorem, so that an exponential shift in one Fourier pair causes spatial shift to the other. Below, M and N show the number of pixels in x and y directions, respectively.

```
for x=1:1:M
    for y=1:1:N
        shf(x,y)=exp(i*pi*x)*exp(i*pi*y);
    end
end
```

Additionally, 'fftshift' command should be used when it is needed to shift coordinate systems. Physically, the central point is in the middle of the device but in programming, we start from 1. This can be adjusted using the inherent shifting command.

## E.9 Sinc function

Sometimes it is required to use a sinc function ( $\text{sinc}(x)=\sin(\pi x)/\pi x$ ) during iterations and at the end. In the iterations one need to subtract sinc function from the target by division so that in the actual propagation sinc function effect disappears. This function is originated from the pixelated shape of the diffractive surfaces. In Matlab®, there is a built-in 'sinc' function.

```
for x=1:1:M
    for y=1:1:N
        m=(M/2)+1;
        n=(N/2)+1;
        G(x,y)=(sinc((x-m)/M).*sinc((y-n)/N));
    end
end
```

As a footnote, normally, sinc function removal from image is used in every step of iteration when amplitude imposition stage comes in every step of iteration. On the contrary, it is in use at the last iteration only for applications with noise window. The reason for that is deterioration of the reconstructions as program proceeds when sinc function removal is applied in every step in applications with noise window. In fact sinc function is not very efficient at all since there is extra window around the target image and the top of sinc coincides with the reconstruction of the desired image.

### **E.10 Phase Quantization**

It is generally required to quantize phase as mentioned in the text. Below script takes `phase_in` as an input initial phase matrix and gives `phase_out` as an output quantized phase matrix.

```
B1=angle(phase_in);  
if B1<0, B1=B1+2*pi; end  
B2=floor(B1/(2*pi/NL));  
phase_out=exp(i*B2*2*pi/NL);
```

Above, B1 and B2 are internal dummy variables. The command ‘floor’, rounds towards minus infinity.

### **E.11 Amplitude Quantization**

Similar to phase, amplitude quantization needs to be applied in the system mentioned in Chapter 4. Below line takes `phase_in` as an input initial amplitude matrix and gives `phase_out` as an output quantized amplitude matrix.

```
phase_out=(floor(phase_in/(1/NL)))*(1/NL);
```

## REFERENCES

- [1] D. Gabor, "A new microscope principle," *Nature*, 161:777, 1948.
- [2] R. J. Collier, C. B. Burckhardt and L. H. Lin, *Optical Holography* (Bell telephone Laboratories, Inc., 1971).
- [3] G. Tricoles "Computer generated holograms: an historical review" *Appl. Opt.* Vol. 26, No. 20, 4351-4360 (1987).
- [4] B. R. Brown and A. W. Lohmann, "Complex Spatial Filtering with Binary Masks," *Appl. Opt.* 5, 967 (1966).
- [5] W.-H. Lee, "Sampled Fourier Transform Holograms Generated by Computer," *Appl. Opt.* 9, 639 (1970).
- [6] Gregory P. Nordin, Richard V. Johnson, and Armand R. Tanguay, Jr., "Diffraction properties of stratified volume holographic optical elements," *J. Opt. Soc. Am. A* **9**, 2206-2217 (1992).
- [7] S. Borgsmüller, S. Noethe, C. Dietrich, T. Kresse, and R. Männer, "Computer-generated stratified diffractive optical elements, " *Appl. Opt.* 42, 5274–5283 (2003).
- [8] Fahri Yaras, Hoonjong Kang, and Levent Onural, *Journal of Display Technology*, Vol. 6, No.10 (2010).
- [9] A. A. Gulses, M.S. Thesis, Middle East Technical University, 2004.
- [10] E. Hecht, *Optics*, 4th ed. (Addison Wesley, 2002).
- [11] G. O. Reynolds, B. J. DeVelis, G. B. Parrent, Jr., B. J. Thompson, *The New Physical Optics Notebook: Tutorials in Fourier Optics* (SPIE, 1989).
- [12] J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts & Company, 2005).
- [13] David G. Voelz and Michael C. Roggemann, "Digital simulation of scalar optical diffraction: revisiting chirp function sampling criteria and consequences," *Appl. Opt.* 48, 6132-6142 (2009).
- [14] D. Voelz, *Computational Fourier Optics* (SPIE, 2010).
- [15] J. Shamir and R. Piestun, in: *Holography for the New Millennium*, Editors J. Ludman, H. J. Caulfield and J. Riccobono (Springer, 2002) pp. 121-155.

- [16] J. R. Fienup, "Iterative method applied to image reconstruction and to computer-generated holograms," *Optical Engineering* 19(3), 297-305 (1980).
- [17] H. Bartelt, "Computer-generated holographic component with optimum light efficiency," *Appl. Opt.* 23, 1499-1502 (1984).
- [18] Thomas Kämpfe, Ernst-Bernhard Kley, Andreas Tünnermann, and Peter Dannberg, "Design and fabrication of stacked, computer generated holograms for multicolor image generation," *Appl. Opt.* 46, 5482-5488 (2007).
- [19] Tim D. Gerke and Rafael Piestun, "Aperiodic computer-generated volume holograms improve the performance of amplitude volume gratings," *Opt. Express* 15, 14954-14960 (2007).
- [20] Rafael Piestun and Joseph Shamir, "Control of wave-front propagation with diffractive elements," *Opt. Lett.* 19, 771-773 (1994).
- [21] T. D. Gerke, R. Piestun, "Aperiodic volume optics," *Nature Photonics* 4, 188 - 193 (2010).
- [22] Xuegong Deng and Ray T. Chen, "Design of cascaded diffractive phase elements for three-dimensional multiwavelength optical interconnects," *Opt. Lett.* 25, 1046-1048 (2000).
- [23] O. Bryngdahl and F. Wyrowski, in *Progress in Optics*, E. Wolf, ed. (North Holland, Amsterdam, 1990), Vol. 28, pp. 3-86.
- [24] Rainer G. Dorsch, Adolf W. Lohmann, and Stefan Sinzinger, "Fresnel ping-pong algorithm for two-plane computer-generated hologram display," *Appl. Opt.* 33, 869-875 (1994).
- [25] T Haist, M. Schönleber, H.J. Tiziani., "Computer-generated holograms from 3D-objects written on twisted-nematic liquid crystal displays" *Optics Communications* 140, 299-308 (1997).
- [26] Michał Makowski, Maciej Sypek, Andrzej Kolodziejczyk and Grzegorz Miłkowska, "Three-plane phase-only computer hologram generated with iterative Fresnel algorithm", *Opt. Eng.* 44, 125805 (2005).
- [27] Chao-fu Ying, Hui Pang, Chang-jiang Fan and Wei-dong Zhou, "New method for the design of a phase-only computer hologram for multiplane reconstruction", *Opt. Eng.* 50, 055802 (2011).

- [28] Yishi Shi, Guohai Situ, and Jingjuan Zhang, "Multiple-image hiding in the Fresnel domain," *Opt. Lett.* 32, 1914-1916 (2007).
- [29] D. C. O'Shea, T. J. Suleski, A. D. Kathman, and D. W. Prather, *Diffractive Optics: Design, Fabrication and Test* (SPIE Press, 2004).
- [30] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik (Stuttgart)* 35, 237-246 (1972).
- [31] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.* 21, 2758-2769 (1982).
- [32] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.* 3, 27-29 (1978).
- [33] Emile Aarts and Jan Korst. *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing* (Wiley, New York, 1989).
- [34] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley, 2000).
- [35] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, Quantum annealing: A new method for minimizing multidimensional functions, *Chem. Phys. Lett.*, 219 (1994), pp. 343--348.
- [36] [http://en.wikipedia.org/wiki/Quantum\\_annealing](http://en.wikipedia.org/wiki/Quantum_annealing).
- [37] Joseph N. Mait, "Understanding diffractive optic design in the scalar domain," *J. Opt. Soc. Am. A* 12, 2145-2158 (1995).
- [38] C. C. Huang, Ph.D. Thesis, University of Southern California, 1997.
- [39] A. A. Gülses and B. K. Jenkins, "Multi-plane image reconstruction using cascaded phase elements," in *Digital Holography and Three-Dimensional Imaging*, OSA Technical Digest (Optical Society of America, 2012), paper DM4C.2.
- [40] A. Alkan Gülses and B. Keith Jenkins, "Cascaded diffractive optical elements for improved multiplane image reconstruction," *Appl. Opt.* 52, 3608-3616 (2013).
- [41] Cheng-Shan Guo, Zhen-Yu Rong, Hui-Tian Wang, Yurong Wang, and L. Z. Cai, "Phase-shifting with computer-generated holograms written on a spatial light modulator," *Appl. Opt.*, Vol. 42, No. 35 (2003).
- [42] C. B. Burckhardt, "A Simplification of Lee's Method of Generating Holograms by Computer," *Appl. Opt.* 9, 1949-1949 (1970).

- [43] C. K. Hsueh and A. A. Sawchuk, "Computer-generated double-phase holograms," *Appl. Opt.* 17, 3874-3883 (1978).
- [44] <http://www.prismo.ch/comparisons/desktop.php>.
- [45] H.P. Herzig (Ed.), *Micro-optics: Elements, Systems, and Applications* (Taylor and Francis, London, 1997).
- [46] Thomas J. Suleski and Donald C. O'Shea, "Gray-scale masks for diffractive-optics fabrication: I. Commercial slide imagers," *Appl. Opt.* 34, 7507-7517 (1995).
- [47] T. Fujita, H. Nishihara and J. Koyama, "Fabrication of micro-lenses using electron-beam lithography," *Opt. Lett.* 6, pp. 613-615 (1982).
- [48] S. Sinzinger, J. Jahns, *Microoptics*, 2nd ed. (Wiley, 2003).
- [49] Rolf C. Enger and Steven K. Case, "Optical elements with ultrahigh spatial-frequency surface corrugations," *Appl. Opt.* 22, 3220-3228 (1983).
- [50] C. Gary Blough, M. Rossi, Stephen K. Mack, and Robert L. Michaels, "Single-point diamond turning and replication of visible and near-infrared diffractive optical elements," *Appl. Opt.* 36, 4648-4654 (1997).
- [51] [http://en.wikipedia.org/wiki/Anti-reflective\\_coating](http://en.wikipedia.org/wiki/Anti-reflective_coating).
- [52] T. Kämpfe, E. Kley, and A. Tünnermann, "Creation of Multicolor Images by Diffractive Optical Elements, Arranged in a Stacked Setup," in *Adaptive Optics: Analysis and Methods/Computational Optical Sensing and Imaging/Information Photonics/Signal Recovery and Synthesis Topical Meetings on CD-ROM*, OSA Technical Digest (CD) (Optical Society of America, 2007), paper DTuD8.
- [53] Kyoji Matsushima, Hagen Schimmel, and Frank Wyrowski, "Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves," *J. Opt. Soc. Am. A* 20, 1755-1762 (2003).
- [54] A.D. White and J.D. Rigden, "Continuous Gas Maser Operation in the Visible". *Proc IRE* vol. 50, p1697: July 1962
- [55] [en.wikipedia.org/wiki/Helium-neon\\_laser](http://en.wikipedia.org/wiki/Helium-neon_laser).
- [56] <http://hyperphysics.phy-astr.gsu.edu/hbase/optmod/lasgas.html>.
- [57] Thomas Kämpfe, Ernst-Bernhard Kley, and Andreas Tünnermann, "Designing multiplane computer-generated holograms with consideration of the pixel shape and the illumination wave," *J. Opt. Soc. Am. A* 25, 1609-1622 (2008).

- [58] J. P. Allebach and B. Liu, "Minimax spectrum shaping with a bandwidth constraint," *Appl. Opt.* 14, 3062-3072 (1975).
- [59] J. P. Allebach, N. C. Gallagher, and B. Liu, "Aliasing error in digital holography," *Appl. Opt.* 15, 2183-2188 (1976).
- [60] L. J. Allen and M. P. Oxley, "Phase retrieval from series of images obtained by defocus variation," *Optics Communications* 199, 65–75 (2001).
- [61] Alexandre Mazine and Kevin Heggarty, "Phase mapping and wavefront analysis based on multi-illumination light fields generated by a spatial light modulator," *Appl. Opt.* 50, 2679-2691 (2011).
- [62] Laura Waller, Yuan Luo, Se Young Yang, and George Barbastathis, "Transport of intensity phase imaging in a volume holographic microscope", *Opt. Lett.* 35, 2961-2963 (2010).
- [63] Xuegong Deng, Yongping Li, Dianyuan Fan, and Yue Qiu, "Pure-phase plates for super-Gaussian focal-plane irradiance profile generations of extremely high order," *Opt. Lett.* 21, 1963-1965 (1996).
- [64] Mark A. Neifeld, "Information, resolution, and space-bandwidth product," *Opt. Lett.* 23, 1477-1479 (1998).
- [65] Ching-Chu Huang, B. Keith Jenkins, and Charles B. Kuznia, "Space-Variant Interconnections Based on Diffractive Optical Elements for Neural Networks: Architectures and Cross-Talk Reduction," *Appl. Opt.* 37, 889-911 (1998).
- [66] J. D. Jackson, *Classical Electrodynamics*, 3th ed. (Wiley, 1998).
- [67] J. R. Hanly, E. B. Koffman, *Problem Solving and Program Design in C* 3rd ed. (Addison Wesley Longman, 1999).