

SIGNAL PROCESSING BASED SOLUTIONS FOR
HOLOGRAPHIC DISPLAYS THAT USE BINARY
SPATIAL LIGHT MODULATORS

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND
ELECTRONICS ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By

Erdem Ulusoy

January 2012

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Haldun M. Özaktaş (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Levent Onural

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Orhan Arıkan

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Aydın Alatan

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Assoc. Prof. Dr. Uğur GÜDÜKBAY

Approved for the Graduate School of Engineering and Science:

Prof. Dr. Levent Onural
Director of Graduate School of Engineering and Science

ABSTRACT

SIGNAL PROCESSING BASED SOLUTIONS FOR HOLOGRAPHIC DISPLAYS THAT USE BINARY SPATIAL LIGHT MODULATORS

Erdem Ulusoy

Ph.D. in Electrical and Electronics Engineering

Supervisor: Prof. Dr. Haldun M. Özaktas

January 2012

Holography is a promising method to realize satisfactory quality three-dimensional (3D) video displays. Spatial light modulators (SLM) are used in holographic video displays. Usually SLMs with higher dynamic ranges are preferred. But currently existing multilevel SLMs have important drawbacks. Some of the associated problems can be avoided by using binary SLMs, if their low dynamic range is compensated for by using appropriate signal processing techniques. In the first solution, the complex-valued gray level SLM patterns that synthesize light fields specified in the non-far-field range are halftoned into binary SLM patterns by solving two decoupled real-valued constrained halftoning problems. As the synthesis region, a sufficiently small sub-region of the central diffraction order region of the SLM is chosen such that the halftoning error is acceptable. The light fields are synthesized merely after free space propagation from the SLM plane and no other complicated optical setups are needed. In this respect, the theory of halftoning for ordinary real-valued gray scale images is extended to complex-valued holograms. Simulation results indicate that light fields that are given either on a plane or within a volume can be successfully synthesized

by our approach. In the second solution, a new full complex-valued combined SLM is effectively created by forming a properly weighted superposition of a number of binary SLMs where the superposition weights can be complex-valued. The method is a generalization of the well known concepts of bit plane decomposition and representation for ordinary images and actually involves a trade-off between dynamic range and pixel count. The coverage of the complex plane by the complex values that can be generated is much more satisfactory than that is achieved by those methods available in the literature. The design is also easy to customize for any operation wavelength. As a result, we show that binary SLMs, with their robust nature, can be used for holographic video display designs.

Keywords: Three-Dimensional Holographic Video Displays, Binary Spatial Light Modulators, Light Field Synthesis, Computer Generated Holography, Halftoning, Full-Complex Modulation

ÖZET

İKİLİ UZAMSAL IŞIK MODÜLATÖRLERİ KULLANAN HOLOGRAFİK EKРАНLAR İÇİN SİNYAL İŞLEME TABANLI ÇÖZÜMLER

Erdem Ulusoy

Elektrik ve Elektronik Mühendisliği Bölümü Doktora

Tez Yöneticisi: Prof. Dr. Haldun Özaktaş

Ocak 2012

Holografi, yeterli kalitedeki üç boyutlu video gösterimi için ümit verici bir yöntemdir. Holografik video gösterimi için uzamsal ışık modülatörleri (SLM) kullanılır. Genellikle pikselleri geniş bir aralıktaki pek çok ayırık değeri alabilen SLMler tercih edilmektedir. Ancak mevcut bu tür çok seviyeli SLMlerin önemli sorunları vardır. Pikselleri ikili değer alan SLMler kullanarak, bu sorunların bir kısmından kaçınılabılır. Ancak bunun için, uygun sinyal işleme tekniklerinin kullanılması gerekir. İlk çözümde, büyük olması gerekmeyen uzaklıklarda belirtilmiş ışık alanlarını sentezleyen karmaşık değerli gri SLM örüntüleri, ikili SLM örüntüleri içine kodlanmaktadır. Bu kodlama sırasında, birbirinden ayrı iki reel değerli yarım tonlama problemi çözülmüştür. Işık alanı, SLM'in merkezi kırınım bölgesinin yeterince küçük bir alt bölgesinde sentezlenmektedir. Bu sayede, yarım tonlama hatası kabul edilebilir düzeyde kalmaktadır. Işık alanları sentezlenirken, doğrudan uzaya yayılım yeterli olmakta ve başka bir optik düzeneğe gerek duyulmamaktadır. Bu bağlamda, sıradan reel değerli gri resimler için kullanılagelen yarım tonlama teknikleri, karmaşık değerli hologramlar için de kullanılmak üzere genişletilmektedir. Simülasyon sonuçları, bir düzlem

üzerinde ya da bir hacim içerisinde verilen ışık alanlarının, bizim yaklaşımımız ile başarıyla sentezlenebildiğini göstermektedir. İkinci çözümde, filen, karmaşık değerler üretebilen yeni bir SLM oluşturulmaktadır. Bu kombine SLM, bir dizi ikili değerler alan SLM'in uygun ağırlıklı toplamı oluşturularak elde edilmektedir. Buradaki ağırlıklar karmaşık değerli de olabilmektedir. Bu yöntem, sıradan görüntüler için iyi bilinen bit düzlemi ayrıştırımı ve gösterimleri kavramlarının genelleştirilmiş biçimidir ve seviye sayısı ile piksel sayısı arasındaki bir ödünleşimi içerir. Elde edilen karmaşık değerlerin karmaşık düzlemdeki yayılımı, literatürdeki yöntemlerle elde edilenlere göre çok daha iyidir. Tasarım, istenilen herhangi bir optik dalga boyu için de kolayca uyarlanabilir. Sonuç olarak, gürbüz yapıli ikili SLMlerin holografik gösterim sistemlerinde kullanılabilceđi gösterilmiştir.

Anahtar Kelimeler: Üç Boyutlu Holografik Video Ekranı, İkili Uzamsal Işık Modülatörleri, Işık Alanı Sentezi, Bilgisayarda Üretilmiş Holografı, Yarım Tonlama, Tam Karmaşık Modülasyon

ACKNOWLEDGMENTS

I would like to express my sincere thanks to my supervisor Prof. Dr. Haldun M. Özaktaş and to Prof. Dr. Levent Onural for their guidance, support and encouragement during every phase of this research.

I would also like to thank Prof. Dr. Orhan Arıkan, Prof. Dr. Aydın Alatan and Assoc. Prof. Dr. Uğur Güdükbay for reading and commenting on the thesis.

This work is partially supported by European Commission within FP6 under Grant 511568 with acronym 3DTV.

I also acknowledge the financial support of TÜBİTAK (The Scientific and Technological Research Council of Turkey) between 2004-2009.

It was a hard time for me, but I always felt the sincere support of my dear mother Emine Ulusoy and my dear father Sefa Ulusoy behind me; and needless to say I owe every happiness that I have felt so far in my life to them. Also, my dear sister Banu Ulusoy, to whom I dedicate this thesis, has been my primary source of happiness since she was born and since the beginning of this thesis, and I thank her for being a perfect sister with whom I am so proud of. I know that stories of success within a family do not outcome only as a result of the work of a single generation, but they are the result of the tremendous efforts spent by many past generations. In this respect, I should thank my dear grandmothers İsmahan Yiğit, Zeliha Ulusoy and my dear grandfathers Hüseyin Ulusoy and Ramazan Yiğit (who watches me from heaven with proud I hope) with all my sincere respect. I should also include my dear aunt, or my second mother, Müzeyyen Arslan with whom I spent most of my childhood, which was

a marvelous one. I am very lucky that I have many excellent friends, most of which I will not be able to list here due to space limitations. They have been the color of my life and I should thank them for sharing every joy or pain with me. Among them, I should mention Erdem Şahin (who is closer than a brother to me), Zehra Görgel, Gökhan Bora Esmer, Kıvanç Köse, Bilge Kaşlı, Fahri Yaraş, Ferda Yaraş, Alexander Suhre, Ali Özgür Yöntem, Alper Ünal, Ayça Özçellikkale, Mehmet Köseoğlu, Onur Tan, Yiğitcan Eryaman and Reşit Teksin. I should also thank my long term friends Onur Önder, Emre Ün, Çiğdem Yurttutan and Eda Coşkuner from my high school years, and my dear friends Yasemin Önder, Doruk Eker, Sinan Ercan and Güray Dinçol from the best and the most joyful project I have ever been involved in. I should also thank my dear students from the 2005-2009, 2006-2010, 2008-2011 and 2009-2012 periods; it was really a pleasure for me to be their assistant. I wish the best of success to all of them. I know that the greatest joy that a teacher can feel is to see his or her students achieve a success. In this respect, I am also indebted to all my teachers starting from the primary school. Again due to space limitations, I do not have the chance to mention most of them here, but I should express my sincere thanks to my primary school teacher Şehrinaz Coçkunçay, who made me a good person, and my high school teachers Ayfer Gürsoy, Ender Çakir, Hatice Ersözülü and Ünal Özmen, who definitely taught me much more than the course contents and who showed me the joy that a person can feel by learning and teaching. Finally, I should thank our department secretary Mürüvet Parlakay for all her help during the long years I spent in Bilkent. Definitely, many more names should appear here, but I do not have the chance to list them all, and I apologize for this. I close by stating my love and passion to my country, Turkey. I feel very happy to be born in this country and I feel proud to be a citizen of Turkey. I consider this thesis as a form of servitude to my country, and I am proud of that.

Contents

1	INTRODUCTION	1
1.2	Three-Dimensional Displays	2
1.3	Binary Spatial Light Modulators	13
1.4	Organization of the Thesis	15
2	PRELIMINARIES	18
2.1	Basics of Scalar Wave Optics	18
2.2	Basics of Scalar Diffraction Theory	20
2.3	Analysis of Light Field Generated by a Spatial Light Modulator	28
3	SYNTHESIS OF THREE-DIMENSIONAL LIGHT FIELDS WITH BINARY SPATIAL LIGHT MODULATORS	43
3.1	Effects of Applying a Low-Pass Filter to the SLM Pattern	46
3.2	Encoding Complex-Valued Oversampled Holograms on Binary SLMs	57
3.2.1	Thin Mask Based Solution	61
3.2.2	Oblique Illumination Based Solution	64

3.2.3	Small Deviations in the Incidence Angle of the Illumination Wave	76
3.2.4	Binary Pixel Values Other Than ± 1	77
3.2.5	Volumetric Synthesis Examples	79
3.3	Conclusion	81
4	FULL COMPLEX SPATIAL LIGHT MODULATORS OB- TAINED FROM BINARY SPATIAL LIGHT MODULATORS	90
4.1	Generic Method	93
4.2	Practical Implementation Using a $4f$ System	102
4.2.1	An LSI System to Form the Weighted Superposition of Binary SLMs	103
4.2.2	Imposing a Bandwidth Limitation	115
4.2.3	Implementation with a $4f$ System	120
4.2.4	Discussion About the $4f$ Setup	122
4.3	Pixellated and Quantized Fourier Plane Masks	125
4.4	Conclusion	130
5	SUMMARY AND CONCLUSIONS	134
5.1	Summary	134
5.2	Conclusions	136
	Bibliography	141

List of Figures

2.1	Impulse response of free-space propagation under Fresnel approximation (real part).	24
2.2	Low pass filtered version of $h_z(x, y)$ (real part).	27
2.3	Ratio of $u_z(x, y)$ to $h_z(x, y)$ (real part).	28
2.4	1D cross-section of the ratio of $u_z(x, y)$ to $h_z(x, y)$ (real part). . .	29
2.5	The sampling and interpolation scheme for pixellated SLMs. . . .	32
2.6	Spatial regions occupied by the diffraction orders of a finite size SLM.	37
2.7	An SLM pattern (real part).	38
2.8	Output field produced by the SLM pattern in Fig. 2.7 (magnitude). Pixels are assumed to be impulsive.	39
2.9	Output field produced by the SLM pattern in Fig. 2.7 (magnitude). Pixels are assumed to be rectangular.	42
3.1	Modified scheme for the application of a discrete filter to the SLM pattern.	47
3.2	Equivalent scheme to the one displayed in Fig. 3.1.	49

3.3	An SLM pattern (real part).	52
3.4	Output produced by the SLM pattern in Fig. 3.3 (magnitude). . .	53
3.5	Low pass filtered version of the SLM pattern in Fig. 3.3 (real part). . .	54
3.6	Output produced by the SLM pattern in Fig. 3.5 (magnitude). . .	55
3.7	Synthesis region (dashed region).	56
3.8	Desired light field (magnitude).	65
3.9	Ideal SLM pattern (real part).	66
3.10	Ideal SLM pattern (imaginary part).	67
3.11	Three level SLM pattern for the real part.	68
3.12	Three level SLM pattern for the imaginary part.	69
3.13	Binary SLM pattern obtained by adding the three level SLM patterns in Fig. 3.11 and Fig. 3.12.	70
3.14	Light field generated by the binary SLM pattern in Fig. 3.13. . . .	71
3.15	The new binary SLM pattern to be used with oblique illumination.	75
3.16	Light field generated by the binary SLM pattern in Fig. 3.15 under normally incident illumination.	78
3.17	Light field generated by the binary SLM pattern in Fig. 3.15 with black pixels being equal to 0 instead of -1	80
3.18	The ideal SLM pattern generating the light field depicted in Fig. 3.19 and Fig. 3.20 (real part).	82

3.19	Light field generated by the SLM pattern in Fig. 3.18 at $z = 0.8m$ (magnitude).	83
3.20	Light field generated by the SLM pattern in Fig. 3.18 at $z = 1m$ (magnitude).	84
3.21	Binary SLM pattern computed from the ideal SLM pattern in Fig. 3.18.	85
3.22	Light field generated by the binary SLM pattern in Fig. 3.21 at $z = 0.8m$ (magnitude).	86
3.23	Light field generated by the binary SLM pattern in Fig. 3.21 at $z = 1m$ (magnitude).	87
4.1	Complex numbers available for a pixel of the new SLM with the weights in Eq. 4.1.	95
4.2	Complex numbers available for a pixel of the new SLM with the weights in Eq. 4.2.	98
4.3	Division of the complex plane into bins to facilitate the search. . .	101
4.4	1D illustration of the process through which $s(x, y)$ is created out of $b(x, y)$	105
4.5	Binary SLM pattern.	108
4.6	Real part of $g(x, y)$	109
4.7	Imaginary part of $g(x, y)$	110
4.8	(a) Real part of $s(x, y)$. (b) Imaginary part of $s(x, y)$	111
4.9	Binary SLM pattern for the new weights.	112

4.10	Real part of $g(x, y)$	113
4.11	Imaginary part of $g(x, y)$	114
4.12	Binary SLM Pattern.	118
4.13	(a) Full-complex SLM pattern obtained by processing the binary SLM pattern in Fig. 4.12 with the LSI system described by $h(x, y)$. (b) Resulting diffraction field at $50cm$. (c) Full-complex SLM pattern obtained with the LSI system described by $h_B(x, y)$. (d) Resulting diffraction field at $50cm$	119
4.14	$4f$ setup.	120
4.15	(a) Fourier plane mask for the weights given in Eq. 4.11. (b) Pixellated Fourier plane mask that should be used for the weights given in Eq. 4.11.	122
4.16	(a) A 4-level Fourier plane mask. (b) Achievable complex numbers. (c) Another 4-level mask. (d) Achievable complex numbers.	131
4.17	(a) A 3-level Fourier plane mask. (b) Achievable complex numbers. (c) Another 3-level mask. (d) Achievable complex numbers.	132
4.18	(a) A binary Fourier plane mask. (b) Achievable complex numbers. (c) Another binary mask. (d) Achievable complex numbers.	133

List of Publications

This dissertation is based on the following publications.

[Publication-I] E. Ulusoy, L. Onural and H. M. Ozaktas, “Full-complex amplitude modulation with binary spatial light modulators,” *Journal of the Optical Society of America A*, vol. 28, no. 11, pp. 2310-2321, 2011.

[Publication-II] E. Ulusoy, L. Onural and H. M. Ozaktas, “Synthesis of three-dimensional light fields with binary spatial light modulators,” *Journal of the Optical Society of America A*, vol. 28, no. 6, pp. 1211-1223, 2011.

[Publication-III] E. Ulusoy, L. Onural and H. M. Ozaktas, “Signal processing for three-dimensional holographic television displays that use binary spatial light modulators,” in *Proceedings of IEEE Conference on Signal Processing and Communications Applications Conference*, pp. 41–44, 2010. In Turkish.

[Publication-IV] E. Ulusoy, G.B. Esmer, H. M. Ozaktas, L. Onural, A. Gotchev and V. Uzunov, “Signal Processing Problems and Algorithms in Display Side of 3DTV,” in *Proceedings of 2006 IEEE Conference on Image Processing*, pp. 2985-2988, 2006.

[Publication-V] E. Ulusoy, V. Uzunov, L. Onural, H. M. Ozaktas and A. Gotchev, “Three-dimensional monochromatic light field synthesis with a deflectable mirror array device,” in *Proceedings of SPIE Volume 6187*, 2006.

[Publication-VI] E. Ulusoy, L. Onural and H. M. Ozaktas, “Analysis of the complex light field generated by a deflectable mirror array device,” in *Proceedings of SPIE Volume 6252*, 2006.

[Publication-VII] E. Ulusoy, L. Onural, H. M. Ozaktas, V. Uzunov and A. Gotchev, “Three-dimensional complex scalar light field synthesis with a deflectable mirror array device,” in *Proceedings of 2nd Workshop on Immersive Communication and Broadcast Systems, ICOB 2005, Berlin, Germany, 2005*.

The contributions of the author of this thesis to the publications listed above were as follows: In Publications-I, -II, -III and -VI, the author designed and implemented all of the proposed algorithms and methods, performed the mathematical derivations, conducted and reported the simulations; and prepared the manuscripts. In Publications-V and -VII, the author designed and implemented all of the proposed algorithms and methods except for those related to the matching pursuit algorithm. He performed all of the mathematical derivations; and conducted and reported the simulations related to the algorithms that he developed. In Publication-IV, the algorithm related to the light field synthesis problem was designed and implemented by the author. The related manuscript and simulations were also prepared by the author.

Dedicated to my sister, Banu Ulusoy, to remind her
that my happiness is impossible without her happiness...

Chapter 1

INTRODUCTION

In this thesis, we study the signal processing problems related to the holographic three-dimensional (3D) video displays constructed using binary spatial light modulators (SLM). Binary SLMs have some important advantages over existing multilevel SLMs that make them quite attractive to use in holographic displays. However, minor work has been done to provide satisfactory solutions to the related signal processing problems, perhaps due to the fact that these problems seem challenging at a first glance due to the binary nature of the SLMs. With this work, we try to fill this gap. In the first part of our work, we develop the true approach for synthesizing desired light fields with binary SLMs from a signal processing perspective [1]. This part can be considered as the extension of the classical halftoning theory for two dimensional gray scale real-valued digital images to gray level complex-valued digital holograms. We show through our simulations that, when properly configured, for an observer standing sufficiently far away, a binary SLM can be made indistinguishable from a lower resolution multilevel SLM, resembling the indistinguishability of a standard halftoned 2D image from its original. In the second part of our work, we propose a method for effectively creating a satisfactory full-complex SLM out of binary SLMs [2, 3]. The proposed method is developed using simple signal processing concepts and

can be considered as the generalization of the familiar concepts of bit plane representation and decomposition for ordinary images to holograms. Again we show through our simulations that with the proposed method, it is possible to effectively obtain new SLMs the pixels of which have a better coverage of the complex plane than any existing multilevel SLM, where the new SLMs can function over a broad range of wavelengths within the visual spectrum. We believe that our solutions will provide strong guidance to designers who wish to exploit the advantages of the binary SLMs for constructing a satisfactory quality 3D display. The advantages of binary SLMs are discussed in Sec. 1.3 and the organization of the thesis is detailed in Sec. 1.4. Here, we continue with a review of different approaches for a 3D display and elaborate on the importance of SLMs within this framework.

1.2 Three-Dimensional Displays

To look into images that invoke a depth perception has proven to be an exciting experience for many people. Different methods for capturing and displaying still 3D images have been developed since 1840s [4, 5, 6, 7, 8, 9], and now, thanks to the advances in display and computing technology, it is the time for satisfactory quality 3D video [10, 11, 12, 13, 14, 15], which has already become the rising trend in the visual entertainment industry: every day, an increasing fraction of movies are captured in 3D, more movie theaters improve their infrastructure to deliver these movies, virtually all popular computer games are being designed to deliver 3D graphics, and 3D liquid crystal device (LCD) screens constitute the fashion product on the shelves of consumer electronics stores with the premise that TV channels will be broadcasting 3D content in a few years. All these developments are based mainly on the simple technique of stereoscopy, whose details we briefly explain in the next paragraph.

Natural depth perception in humans is based on various depth cues, such as binocular disparity, motion parallax, convergence, accommodation, occlusion, blurring [16, 17]. 3D display techniques make use of some or all of these depth cues. The simplest method to display a 3D image is stereoscopy, and in its raw form, it makes use only of the binocular disparity: naturally, there is a slight disparity between the images captured by the left and the right eyes of the same object, and stereoscopy takes advantage of this fact [18]. In particular, in stereoscopy, two images of a 3D scene are captured using two cameras placed side by side so as to mimic the two eyes of an observer, and then the image captured by the left (right) camera is presented to the left (right) eye of the observer. In this way, the observer perceives a depth variation in the displayed scene [19, 20, 21, 22]. Mostly, the left and right images are placed on the same screen, but some type of goggles (such as complementary color anaglyphs, polarized glasses or liquid crystal shutter glasses) are worn by the observers to separate the images [23]. Or in some cases, the images are displayed on two different screens placed right in front of each eye, such as in head mounted displays [24]. Stereoscopy is quite easy to implement, and that is why it is the foregoing 3D display technology for the time being. However, there are some limitations and drawbacks of it. To begin with, the necessity to wear the goggles distorts the comfort of the observer. It is possible to avoid the use of goggles at least within a restricted viewing zone by using techniques such as placing a parallax barrier or a lenticular sheet in front of the display. Such techniques are named autostereoscopy [25, 26]. Secondly, in its raw form, stereoscopy only provides binocular disparity, but other visual stimuli provided by it are incorrect. For instance, if a user moves while looking into a 3D object displayed on a standard stereoscopic system, she will recognize that the object will exhibit an unusual parallax. To remedy such problems, researchers developed techniques such as eye-tracking stereoscopy [27], in which the 2D images that are displayed are updated as the position of the observer changes; and multi-view stereoscopy [28], in which several 2D images are

displayed on the same screen where each image is seen only from a specific angle. These techniques provide an improvement to motion-parallax related problems of classical stereoscopy to a certain extent. Yet, there are some intrinsic problems associated with stereoscopy that can never be remedied. For instance, while looking at a stereoscopic display, the lenses of the eyes of an observer are focused at the screen whereas her eyeballs are converged towards the perceived position of the object. This phenomenon is called the accommodation-convergence conflict. It may cause a dizziness after watching for a while, and it will be inevitably present in any stereoscopic system [29, 30].

Actually, a striking peculiarity about stereoscopy that worths mentioning is that it is an already tried and abandoned technique in 3D display history [31]. Actually, the history of stereoscopy is nearly as old as that of conventional 2D photography. Stereoscopic photography was invented in 1838, and the stereoscopic 3D cinema was available in the early 1900s, but the interest in stereoscopic video was significantly lost after 1950s [32]. Therefore, in a sense, nowadays we are experiencing the rebirth of a once-closed era. The advancement in the conventional 2D display technology and the emergence of powerful computers are no doubt the primary reasons for this rebirth. However, the reason for the first failure of stereoscopic video was more than the inconvenient displays and poor computation technology of the day: stereoscopy itself has intrinsic drawbacks and limitations as explained above. Therefore, it is not possible to claim that the ultimate stage has been achieved in 3D display technology yet. In particular, it is well known by scientists for over 60 years that there is an alternative technique which has the potential to vastly eliminate the intrinsic drawbacks and limitations of stereoscopy, namely the holography.

As we mentioned above, the stereoscopic approach for 3D displays mainly takes advantage of a particular depth cue (binocular disparity), and tries to remedy the problems associated with other depth cues as much as possible. The

holographic approach, which has been developed and investigated as a strong candidate to realize the ultimate 3D display following the invention of holography by Dennis Gabor in 1948 [33, 34], takes a route different than focusing on one or more of the depth cues. We can broadly describe the holographic approach for 3D displays as follows: replicate the light that is emanating from the original 3D scene, such that any observer interacting with that light will see the original 3D scene with all its natural depth cues even if the scene is not there physically [31]. Therefore, neither the presence of the observer nor the properties of her visual system are of primary concern. In other words, all of the natural depth cues are aimed to be provided at the same time. As palpably seen, the goals that the holographic approach sets forth are incomparably more challenging than those set forth by the stereoscopic approach. This is why we have to wait more before holographic televisions enter our living rooms. Currently, a holographic TV is possible in principle, but in practice, both the display technology and the associated algorithms need to be improved along a long path before we start seeing satisfactory quality holographic images. But when that path is traveled, there is no doubt that the rewards will worth it.

In its initial form, holography involves the process of making a coherent recording (that is, a recording under illumination by laser light) of the interference pattern between a known reference wave and the object wave emanating from physically existing objects on a photographic film, which is named afterwards as the hologram [35, 36, 37, 38, 39, 40]. In this way, all the necessary physical properties of the object wave (both intensity and phase) can be captured. During replay, the hologram is illuminated by the same reference wave, and along with several side beams that propagate to different directions, the object wave is reconstructed. In this manner, a quite realistic 3D still image is obtained with all the natural depth cues. In other words, any observer intercepting this reconstructed wave sees the corresponding object as if it is physically there. Following this line, many scientist devoted their life to the improvement of

the basic technique and development of better photosensitive materials so that better quality images are obtained [41]. Thanks to their efforts, still holography has enormously excelled in the years that have passed, as can be seen from the amazingly high quality and quite natural-looking color holographic images displayed in several museums and exhibitions [42, 43]. The succession in this area continues to inspire many researchers to channelize their efforts to bring holographic video into reality. Yet, dynamic holography has not yet managed to reach the visual quality offered by still holography. But there is no doubt that when finalized the holographic video display will give a much more fascinating visual experience to observers than any other present technology, as verified by the enormous excitement that people feel when they see a still hologram for the first time.

Though the initial form of holography constitutes the basis on which these exciting ideas about the ultimate 3D display has flourished, there are many modifications that need to be considered on that initial form [44]. To understand the nature of those modifications, let us examine the current trend for the design of a holographic 3D television system [45]. Though from the capture to the display end the conventional 2D television systems evolved from a fully analog system (recording by analog cameras, broadcasting analog TV signals, display using analog cathode ray tube (CRT) screens) to a fully digital system (recording by digital cameras, broadcasting over internet or through satellite systems, display on digital LCD monitors), it seems that the holographic 3D television will emerge as a full digital system; simply due to the reason that analog holographic recording and display facilities are inconvenient to support dynamic operation, whereas digital counterparts are much more suitable and flexible. Several recently conducted large scale research projects carried on the topic roughly envision the following sequence of operations in an end-to-end holographic 3D television system [11, 46, 47, 48]:

- Capture the necessary visual information about a 3D scene using multi-camera systems
- Prepare an abstract computer graphics representation of the 3D scene
- Transmit that information to the display end
- From the abstract description of the 3D scene, digitally compute the hologram of the 3D scene
- Write this hologram on a convenient dynamic holographic screen and display it

From the above list, we recognize that most possibly, the principles of holography will be utilized merely at the display end for delivering the necessary visual signal to an observer. In particular, examining the first three items of the list, we recognize that most probably the depth information about a 3D scene will not be captured through an ordinary holographic recording. Rather, that information will be captured and conveyed to the display end using straightforward extensions of already existing capture, computer graphics and communications techniques. (Actually, we should mention that direct capturing of holographic fringe patterns is still a living alternative [49]. However, for the time being, this alternative seems to be somewhat inconvenient compared to the scheme presented above. One of the main reasons of this inconvenience is that, in this alternative, the 3D scene must be illuminated by coherent light, and this is impractical especially for large scale scenes. Also, the recording camera should be accompanied by an optical setup so that a reference wave is provided. Moreover, current resolution of digital cameras is much lower than that of photographic films, and this places a severe limit on the sizes of the objects that are wished to be imaged [50]. In this respect, we will continue with the scheme mentioned above since it is a stronger alternative.) Then, at the display end, the hologram of the abstractly described 3D scene will be computed just because that hologram is necessary to

produce the final light signal that is to be intercepted by observers. (Actually, in strict terminology, we only need to compute the complex-valued object wave, not the hologram, which corresponds to the interference pattern between the object wave and a reference wave. Note that in the holographic method of Gabor, the interference pattern is formed because the photosensitive materials can only record intensity patterns, not complex amplitudes. But now, electro-holographic displays already have the capability to provide complex values. Therefore, there is no need to form the interference pattern any more, so computation of the object waves is sufficient. But despite making this remark, for the sake of convenience, we will keep on saying that the hologram of the 3D scene must be computed, where it is meant that the cross section of the object wave must be computed over the hologram plane.) From this discussion, we see the modifications that need to be considered over the initial form of holography within the context of a 3D display: in ordinary holography

- the hologram is obtained physically, but now it should be computed; and
- the hologram is ready on a piece of photographic material, but now it should be written on some kind of dynamic device.

Each one of these problems has a unique character and is challenging [51, 52]. Actually, these problems form the bottleneck of holographic video systems for the time being. It is not possible to claim yet that the developed methods for these problems fully fulfill the expectations. Regarding the first problem, the current computational techniques are neither accurate nor fast enough to enable satisfactory real-time operation [53, 54, 55, 56, 57, 58]. Regarding the second problem, the spatial resolution of current holographic video displays are much lower compared to that of photographic films. In addition, the dynamic range of these displays are much lower or restricted compared to that offered by photographic films. Such constraints and restrictions place a limit on the quality of holograms that can be displayed on these devices. These restrictions can be

partially compensated with computational techniques, but these again are not developed to their final form yet by any means [59, 60].

Actually, the two problems discussed above emerged long before the technology advanced to the level that made a modest holographic television possible. Beginning from early 1960s, following the formulation of scalar wave optics theory of light within a signals and systems framework under the name Fourier optics, holography has also been investigated using signals and systems theory concepts [61]. And beginning from late 1960s; following the invention of computers and the development of the digital signal processing tools, holography has been investigated within a digital signal processing framework [62]. Many new problems have been defined and studied in this context, and many new sub research fields emerged. To name the most important ones, in the so called digital holography, which emerged mainly following the invention of the charged coupled devices (CCD) and cameras built using them, instead of using a photographic film, the hologram of a 3D object is captured using a digital CCD camera, and then the resulting holographic image is processed by a computer to extract the object wave and perform some diagnostics afterwards [49, 50, 63]. And in the so called computer generated holography, instead of performing a physical recording, the holograms that synthesize certain desired beams (that possibly never existed before) are computed digitally using a computer [64, 65, 66, 67, 68, 69, 70]. Then, a physical transparency is prepared from these digitally computed holograms and optical reconstructions are performed. Research is ongoing with increasing pace in all these fields to improve the quality of optical reconstructions while increasing the computational performance of the related algorithms as much as possible. And as seen in conjunction with the previous paragraph, there is no doubt that this research (especially the one on computer generated holography) has quite direct contributions towards the development of a 3D holographic television system. Therefore, we continue with a review of the achievements in computer generated holography.

The research on computer generated holography was initiated by Brown and Lohmann in late 1960s [71]. The goal of these researchers was to compute and produce optical masks that synthesized certain desired light distributions. They faced two important problems. The first one was, at that time, they did not have the required computational power for calculating the hologram that would synthesize a light field specified over an arbitrarily shaped volume and at an arbitrary distance from the hologram. To overcome this difficulty, they concentrated on small planar regions of space that lied in the far field, so that the relation between the samples of the hologram and the desired light field is to a very good approximation given by a discrete Fourier transform, which was rather easy to compute. The second problem they faced was, the only available devices that had the spatial resolution required for the physical production of the mask were black and white ink-jet dot printers, so that the final mask could only be binary. They resolved this issue by developing their famous detour phase method, which is actually a method for encoding a complex-valued hologram pattern into a binary pattern (we briefly explain the details of this method in the introduction of Chapter 3) [72, 73]. For a quite long time (till early 1980s), only binary masks could be produced, so the detour phase method was used abundantly in computer generated holography related applications [74, 75, 76, 77, 78, 79]. Later, following the advancements in both computational and optical technologies, it become possible to efficiently compute holograms for non-far field light distributions and write these holograms on pixellated masks (usually named diffractive optical elements) where more than two values were available for a pixel of the mask. And following the advancement of spatial light modulator (SLM) technologies in late 1980s, computer generated holograms were started to be written on SLMs, which are basically optical masks that can be dynamically programmable [80, 81, 82, 83]. Actually, it is only after the advent of SLMs that holographic video has been considered as a state of technology that can be achieved within several decades. And today, SLMs continue to be the most promising devices

to be employed during the build-up of a satisfactory quality holographic display. Understanding the current state of the SLM technology is therefore crucial for understanding how far away we are from having satisfactory quality holographic displays.

There are many different types of SLMs working with different mechanisms [61, 84, 85, 86]. A large number of SLMs exploit the anisotropic nature of twisted nematic and ferroelectric liquid crystals to modulate the phase of light in a controlled manner. Some SLMs are mirror based. These SLMs consist of an array of quite small mirrors which can be tilted or deformed, so that light is spatially modulated. Yet some other SLMs make use of quantum mechanical or acoustical effects. We will not delve into a detailed discussion of the physics behind SLMs, but rather focus on the behavioral properties of SLMs. From a behavioral perspective, there are a number of different classifications that needs to be mentioned. One such classification is made according to the structure of the hologram that can be written on the SLMs: on some SLMs, a continuous function of position is written. In this case, the SLM is said to be analog. An example is the surface acoustic wave SLMs. On the other side are SLMs the hologram on which is of pixellated structure. Such SLMs are named pixellated or digital SLMs. Today, most SLMs belong to this group. Compared to analog SLMs, digital SLMs are usually much convenient to use; however, the pixellated nature of digital SLMs create some undesired effects such as diffraction orders, i.e., replicas of the main desired beam that flow in different directions. These effects may be disturbing especially within the context of holographic displays. Another classification is according to the nature of the control signals. Some SLMs are electrically addressed while some others are optically addressed. Though the optically addressed SLMs have some advantages over the electrically addressed ones (such as having a shorter response time), it seems that the latter will be more convenient to use in holographic displays, since with the former, the optical setups get quite complicated. Yet another classification is made according to the

way the illumination light is sent to the SLM. Some SLMs are of reflective type, that is, the SLM modulates and reflects back the incoming light. In this case, the illumination wave is sent from the front side of the SLM. The other SLMs are of transmissive type, in which case the illumination wave comes from behind, passes through the SLM while getting modulated, and then leaves the SLM from the front side. Though it seems at a first glance that optical setups would be less complicated with transmissive SLMs, some reflective SLMs (especially mirror based SLMs) have certain advantages that prevent them from getting excluded from consideration.

Perhaps, from the standpoint of holographic displays, the most important behavioral classification among the SLMs is made according to the type of modulation that they can provide. This classification is especially important because the richness of the light fields that can be synthesized with an SLM directly depends on the modulation capabilities of the SLM. Note that in general hologram patterns are complex-valued functions of position, meaning that at each spatial position, we require simultaneous and independent amplitude and phase control. Therefore, ideally SLMs should provide full-complex modulation. However, a satisfactory full complex SLM has not been developed yet. Virtually all SLMs provide only some restricted type of modulation. For instance, some SLMs can only provide phase modulation on the incoming light, and they are called phase-only SLMs. A phase-only SLM therefore places on a designer the restriction that all holograms must be phase-only. That is, if the hologram that generates a desired light field is complex-valued, the designer must find a way to encode that hologram into a phase-only pattern. Similarly, some SLMs provide only amplitude modulation and called amplitude only SLMs. But as mentioned, it is nearly impossible to find an SLM that provides both types of modulation simultaneously in a satisfactory manner. For the digital SLMs, usually, there is also the quantization constraint: pixels of most SLMs can be set only to a finite number of different values. Therefore, it is not possible to cover the complex plane

in a continuous manner. Rather, hologram values are quantized to the nearest available level. Altogether, these constraints impose a limit on the range of the light fields that can be synthesized with these devices. Given a desired field, determination of the best hologram pattern subject to the SLM constraints is a widely studied signal processing problem [51, 52, 87, 88, 89, 90, 91, 92, 93, 94]. For SLMs that have a large number of available values for a pixel (such as an 8-bit phase only or amplitude only SLM), this problem is relatively easy, but for some other SLMs (especially the ones on which the quantization constraint is harsh, such as binary SLMs), the problem becomes more interesting and challenging. Actually, in this thesis, we undertake the signal processing problems that arise in the case of binary (or harshly quantized) SLMs.

1.3 Binary Spatial Light Modulators

The most constrained SLMs are the binary ones. Pixels of binary SLMs can be set to only two possible distinct values such as (0,1) or (-1,1). In this case, the quantization constraint on the SLM is quite harsh. Not surprisingly, when other parameters such as number of pixels, pixel periods etc. are kept the same, the range of the light fields that can be synthesized is the most limited when a binary SLM is used. In addition, determination of a binary hologram that generates a desired light field is more difficult than determining a multilevel hologram. With these difficulties in mind, it is natural to question why a binary SLM might be preferred at all over already existing multilevel SLMs.

Binary SLMs have several important advantages over multilevel SLMs that make them attractive to be used in holographic displays. To begin with, most multilevel SLMs work in the prescribed manner around a certain wavelength. When the wavelength is changed, the pixel values usually change in a drastic

manner. These changes are hard to keep track of especially in multi-color applications such as 3D displays. On the other hand, amplitude-only binary SLMs (such as the digital micromirror devices (DMD) produced by Texas Instruments [95, 96, 97, 98]) provide the same (0,1) modulation independent of the wavelength of the illumination wave. In this respect, their behavior is much easier to keep track of in a multi-color application. (Actually, we show in Chapter 3 that even if the pixel values of a binary SLM change with the illumination wavelength, as long as the new values are still distinct, that change is inconsequential for observers.)

Secondly, most multilevel phase-only (or amplitude-only) SLMs are actually imperfect in the sense that in addition to the phase (amplitude) modulation that they provide, they perform an uncontrollable amplitude (phase) modulation, as can be seen from their operation curves. Therefore, they are not robust in this sense. On the other hand, since binary SLMs only provide two different complex values (and these values are not important as long as they are distinct), they are much more robust.

As a third factor, miniaturization of binary SLMs, that is, manufacturing binary SLMs with small pixel pitches and high pixel counts seems to have a higher potential compared to other types of SLMs, especially when the emerging micro-electro-mechanical systems (MEMS) and micro-opto-electro-mechanical systems (MOEMS) technologies are considered [99]. Actually, it seems that the development in the binary SLM technology will most likely be in the form of improvisations of the DMD concept. DMDs are most popular in current projection systems for their high light throughput and high contrast ratio. (Actually, DMDs provide perfect contrast in the sense that a blank pixel sends no light to the output image. This is quite impossible to achieve with an LCD SLM for instance, the blank pixels of such SLMs always transmit a certain nonzero optical power.) Therefore,

as additional advantages of binary SLMs, we can list their high contrast ratio and light throughput.

Despite these important advantages of binary SLMs, we see that until today the signal processing problems related to the usage of binary SLMs in holographic displays have not been solved to a satisfactory degree. Especially when holographic displays are of concern, almost all of the methods developed so far turn out to be somewhat inadequate, both from a theoretical and a practical perspective. And to our belief, the poor results obtained with those inadequate methods caused the binary SLMs to have a bad reputation among scientists and practitioners. Many designers have mistakenly thought that because of their low dynamic range, binary SLMs are intrinsically inadequate for holographic display purposes, especially in terms of the quality and richness of light fields that can be synthesized. As a result, they mostly abandoned binary SLM based display designs, and shifted their interest into multilevel SLMs. While doing so, they also lost the chance to enjoy the many advantages of binary SLMs over multilevel SLMs. This point stroke our attention when we decided to undertake this thesis [100, 101, 102, 103]. And we believe that with this thesis, we develop the correct approaches for the mentioned signal processing problems. We expect that after this study, researchers will dismiss their negative attitudes towards binary SLMs, and consider the usage of them more seriously in holographic displays.

1.4 Organization of the Thesis

In accordance with the usual practice in holographic research, we also develop our solutions in this thesis using the scalar wave optics theory of light. In this respect, in Chapter 2, we review the basics of scalar wave optics theory and scalar diffraction theory. Moreover, in that chapter we carry out an analysis of the light field generated by a finite size SLM. The results of this analysis is important for

both understanding the solutions we develop for binary SLMs and understanding the usage of SLMs in a holographic display.

In Chapter 3, we undertake the problem of the computation of a SLM pattern that generates a desired light field. We start by a review of the proposed methods and algorithms, which are mainly designed for far field or Fourier plane reconstructions. Then, we develop our approach assuming that the desired light field is synthesized within a volumetric region in the non-far field range merely after free space propagation from the SLM plane. In particular, we show that when the desired field is confined to a sufficiently narrow region of space, the ideal gray level complex-valued SLM pattern generating it becomes sufficiently low-pass (oversampled) so it can be successfully halftoned into a binary SLM pattern by solving two decoupled real-valued constrained halftoning problems, which can be solved using already existing algorithms for classical image halftoning. Our simulation results indicate that when the synthesis region is considered, the binary SLM is indistinguishable from a lower resolution full complex gray level SLM. In an other sense, by the end of that chapter, we will have extended the theory of halftoning for classical gray level images to holograms.

In Chapter 4, we take a different route and develop a solution for effectively creating a full complex SLM using a number of binary SLMs. Again, we start by reviewing several full complex modulation schemes developed using multilevel SLMs. Then we discuss our solution which is based on simple signal processing concepts. We first propose a generic method, by which, out of K binary (or 1-bit) SLMs of size $M \times N$, we effectively create a new 2^K -level (or K -bit) SLM of size $M \times N$. The method is a generalization of the well-known concepts of bit plane representation and decomposition for ordinary gray scale digital images and relies on forming a properly weighted superposition of binary SLMs. When K is sufficiently large, the effective SLM can be regarded as a full-complex one. Then, we discuss a $4f$ system as a possible and promising optical implementation.

That $4f$ system also provides a means for eliminating the undesirable higher diffraction orders, and the components of the $4f$ system can easily be customized for different production technologies. By the end of this chapter, we will have developed a method for utilizing binary SLMs as efficiently as possible from an information theoretical perspective towards achieving full complex modulation.

In Chapter 5, we state the conclusions of our work.

Chapter 2

PRELIMINARIES

2.1 Basics of Scalar Wave Optics

Until today, four major theories have been developed to explain optical phenomena [104]. From the simplest to the most complex, these theories are listed as

- *geometric optics theory*, in which light is described by rays that travel according to Fermat's principle,
- *scalar wave optics theory*, in which light is described by a scalar function of position and time that satisfies the wave equation
- *classical electrodynamic theory*, in which light is described by four vector-valued functions of position and time (electric field, magnetic field, electric flux density, magnetic flux density) that satisfy the Maxwell equations
- *quantum electrodynamic theory*, in which light is postulated to consist of photons which exhibit both wave-like and particle like characteristics. The behavior of photons are governed by quantum mechanical principles.

The geometric optics theory is adequate in partially explaining phenomena such as reflection, refraction, imaging; and the behavior of simple optical components such as lenses, prisms, mirrors etc. However, it fails to explain phenomena such as interference and diffraction (“any deviation of light rays from rectilinear paths which cannot be interpreted as reflection or refraction”, as defined by Sommerfeld) in which the wave nature of light becomes predominant. Scalar wave optics theory enters the picture at this stage and it accounts for most phenomena related to wave nature of light. Yet, it is not sufficient to fully explain the nature of energy transport by light waves, it does not accurately describe diffraction at high angles, it fails to account for the behavior of more advanced optical components such as polarizers and optical crystals, and it does not establish the link between the sources of radiation and the radiated waves in a precise manner. Therefore, in the course of history, it is replaced by the classical electrodynamic theory, which virtually explains all the optical phenomena within classical confines. Following the development in quantum mechanics, classical electrodynamic theory is refined to the quantum electrodynamic theory in order to account for light-matter interactions that occur at the quantum scale as well.

In this thesis, we are only interested in optical phenomena that arise within the context of holography. Historically, the holographic method was developed using the scalar wave theory of light. And until today, scalar wave theory has proven to be sufficient for most holographic applications, making it unnecessary to resort to a more advanced theory of light. Therefore, here, we use the scalar wave theory of light as well, and we continue with a brief summary of that theory.

According to the scalar wave theory, light is described by a scalar real-valued function of space and time that we denote by $\tilde{u}(x, y, z, t)$ with $x, y, z \in \mathbb{R}$ denoting the three spatial cartesian coordinates and $t \in \mathbb{R}$ denoting the time. It is postulated that \tilde{u} satisfies the so-called wave equation in free-space:

$$\nabla^2 \tilde{u} - \frac{1}{c^2} \frac{\partial^2 \tilde{u}}{\partial t^2} = 0 \quad (2.1)$$

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ and c denotes the speed of light in free space.

Mostly, we are interested in purely monochromatic light disturbances. If the light is monochromatic, at every point in space, \tilde{u} takes the form $\tilde{u}(x, y, z, t) = A(x, y, z) \cos \{2\pi \frac{c}{\lambda} t - \phi(x, y, z)\}$ where A and ϕ are real-valued functions of position and λ denotes the wavelength of the monochromatic light disturbance. In such cases, as well known, it is more convenient to work with phasors. If we define a complex-valued function of position that we denote by $u(x, y, z)$ such that $u = Ae^{j\phi}$, then we have $\tilde{u} = \mathcal{R} \{ue^{-j2\pi \frac{c}{\lambda} t}\}$ so that \tilde{u} can be recovered easily if we know u over the spatial region of interest. u is usually named the phasor that describes the light. In the rest of the thesis, we assume that we are dealing with monochromatic light disturbances and we use phasors. Note that if \tilde{u} is to satisfy Eq. 2.1, u has to satisfy the so called Helmholtz equation:

$$\nabla^2 u + k^2 u = 0 \tag{2.2}$$

where k is usually named the wave number and is given by

$$k = \frac{2\pi}{\lambda}. \tag{2.3}$$

2.2 Basics of Scalar Diffraction Theory

A well known problem investigated under scalar wave optics theory is the so called diffraction problem, in which it is assumed that $u(x, y, z)$ is known over a planar surface, and the objective is to determine it at some other region of space where there is only free-space propagation in between. Usually, the planar surface is taken as the $z = 0$ surface, the light waves are assumed to be propagating from the $z < 0$ half-space towards the $z > 0$ half-space, and the points of interest lie in the $z > 0$ half-space. Hence, given $u(x, y, 0)$ for all x, y ; the goal is to determine $u(x, y, z)$ for all x, y and $z > 0$ such that $u(x, y, z)$ satisfies the Helmholtz equation given by Eq. 2.2. Though there is fundamentally no difference in between, in most

references on the subject, the problem is posed as the determination of $u(x, y, z)$ as a two dimensional function of x and y for some constant $z > 0$. That form is also convenient for our purposes in this thesis. When the problem is posed in that form, it is more convenient to denote the light field over the $z = 0$ plane by $u_0(x, y)$ instead of $u(x, y, 0)$. Similarly, the light field over an arbitrary z -plane is more conveniently denoted by $u_z(x, y)$. Since this notation is more common, from now on we also switch to it. To summarize, in the diffraction problem, we assume that $u_0(x, y)$ is specified, and the goal is to determine $u_z(x, y)$ for some constant $z > 0$.

As explained in [61], the relation between $u_0(x, y)$ and $u_z(x, y)$ is given as a linear shift invariant (LSI) system:

$$\begin{aligned} u_z(x, y) &= u_0(x, y) ** h_z(x, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x', y') h_z(x - x', y - y') dx' dy' \end{aligned} \quad (2.4)$$

where $**$ denotes two-dimensional analog convolution operation and $h_z(x, y)$ denotes the impulse response of free space propagation. According to the Rayleigh-Sommerfeld (RS) theory, $h_z(x, y)$, i.e., the impulse response of free-space propagation, is given as:

$$h_z(x, y) = -\frac{1}{2\pi} \frac{z}{R} (1 - jkR) \frac{e^{jkR}}{R^2} \quad (2.5)$$

where $R = \sqrt{x^2 + y^2 + z^2}$. The Fourier transform (FT) of Eq. 2.5, i.e., the frequency response of free space propagation is given as [105, 106]:

$$\begin{aligned} H_z(\nu_x, \nu_y) &= \mathcal{F} \{h_z(x, y)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_z(x, y) \exp \{-j2\pi (x\nu_x + y\nu_y)\} dx dy \\ &= \exp \left\{ jkz \sqrt{1 - (\lambda\nu_x)^2 - (\lambda\nu_y)^2} \right\}. \end{aligned} \quad (2.6)$$

Note that if $U_0(\nu_x, \nu_y)$ and $U_z(\nu_x, \nu_y)$ respectively denote the Fourier transforms of $u_0(x, y)$ and $u_z(x, y)$, we have $U_z(\nu_x, \nu_y) = H_z(\nu_x, \nu_y)U_0(\nu_x, \nu_y)$.

Eq. 2.5 and Eq. 2.6 are exact solutions of the diffraction problem within the confines of the scalar wave optics theory. In these equations, in addition to

the propagating plane wave components, evanescent waves are also taken into account. Since evanescent waves decay very quickly, for $z \gg \lambda$, their contribution can be ignored and Eq. 2.5 and Eq. 2.6 can be taken as

$$h_z(x, y) = \frac{z}{j\lambda} \frac{e^{jkR}}{R^2} \quad (2.7)$$

and

$$H_z(\nu_x, \nu_y) = \begin{cases} \exp\left\{jkz\sqrt{1 - (\lambda\nu_x)^2 - (\lambda\nu_y)^2}\right\} & \text{for } (\lambda\nu_x)^2 + (\lambda\nu_y)^2 \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.8)$$

Ignoring the effects due to the slowly changing amplitude term $\frac{z}{j\lambda R^2}$, we can see that the instantaneous spatial frequencies of $h_z(x, y)$ along x and y directions are given by

$$\begin{aligned} \nu_X(x, y) &= \frac{1}{2\pi} \frac{\partial\{kR\}}{\partial x} \\ &= \frac{x}{\lambda R} \end{aligned} \quad (2.9)$$

and

$$\begin{aligned} \nu_Y(x, y) &= \frac{1}{2\pi} \frac{\partial\{kR\}}{\partial y} \\ &= \frac{y}{\lambda R}. \end{aligned} \quad (2.10)$$

Note that $\nu_X(x, y)^2 + \nu_Y(x, y)^2$ is always less than $\frac{1}{\lambda^2}$ in accordance with the propagating wave constraint.

Under paraxial cases, that is,

- either $u_0(x, y)$ is a sufficiently low-pass function,
- or $u_0(x, y)$ is confined to a narrow region around the optical axis and we are only interested in the portion of $u_z(x, y)$ lying close to the optical axis,

we can utilize the commonly used Fresnel diffraction theory, in which the impulse response is approximated with a chirp (quadratic phase exponential) function:

$$h_z(x, y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)}. \quad (2.11)$$

The corresponding frequency response becomes

$$H_z(\nu_x, \nu_y) = e^{jkz} \exp \{-j\pi\lambda z(\nu_x^2 + \nu_y^2)\} \quad (2.12)$$

which is also a chirp. Note that under the Fresnel approximation, the instantaneous frequencies are also approximated as:

$$\nu_X(x, y) = \frac{x}{\lambda z} \quad (2.13)$$

and

$$\nu_Y(x, y) = \frac{y}{\lambda z}. \quad (2.14)$$

Hence, under the Fresnel approximation, the instantaneous frequencies are assumed to vary linearly with the spatial coordinates. The impulse response given in Eq. 2.11 is illustrated in Fig. 2.1 ($\lambda = 632.9nm$, $z = 1m$). From this figure, we can clearly see that the instantaneous frequencies increase as we get away from the origin.

Having reviewed the basics of scalar diffraction theory, now we will examine the output field generated by a special input field in detail since that input field will be important for our developments in the rest of the thesis, especially during the analysis of the light fields generated by spatial light modulators (SLMs). In this respect, we will first concentrate on a useful approximate formula for the output fields generated by input fields that have a sufficiently narrow spatial support. We begin the discussion by considering the Fresnel diffraction case. Substituting $h_z(x, y)$ given by Eq. 2.11 into Eq. 2.4, we can explicitly write

$$u_z(x, y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x', y') e^{\frac{j\pi}{\lambda z}(x'^2+y'^2)} e^{-\frac{j2\pi}{\lambda z}(x'x+y'y)} dx' dy'$$

from which we see the well known fact that chirp convolution is equivalent to pre-multiplication with a chirp, taking a Fourier transform, and then post-multiplication with another chirp. Now, suppose the input $u_0(x, y)$ is concentrated around the origin and has a sufficiently narrow support. Also assume that z is sufficiently large. In this case, we can write

$$u_0(x', y') e^{\frac{j\pi}{\lambda z}(x'^2+y'^2)} \approx u_0(x', y')$$

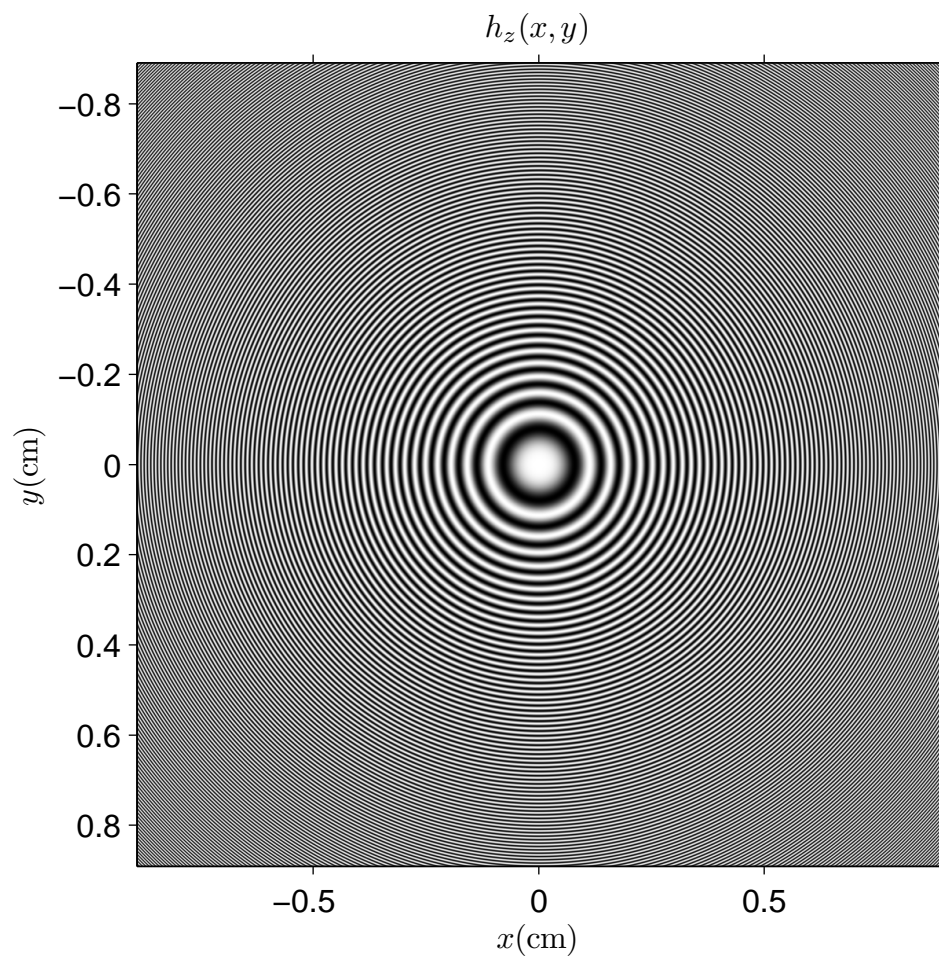


Figure 2.1: Impulse response of free-space propagation under Fresnel approximation (real part).

which is essentially equivalent to assuming that over the narrow spatial support of $u_0(x', y')$, the chirp term $e^{\frac{j\pi}{\lambda z}(x'^2+y'^2)}$ can be taken as unity. This assumption leads to the following approximate formula:

$$\begin{aligned} u_z(x, y) &\approx \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} U_0\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right) \\ &\approx h_z(x, y) U_0\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right). \end{aligned} \quad (2.15)$$

Actually, using Eq. 2.13 and Eq. 2.14, we can write Eq. 2.15 in the following form as well:

$$u_z(x, y) \approx h_z(x, y) U_0(\nu_X(x, y), \nu_Y(x, y)). \quad (2.16)$$

Note that we arrived at the above result by viewing $u_0(x, y)$ as the input field and $u_z(x, y)$ as the output field. However, sometimes, it is more convenient to view $u_0(x, y)$ as an arbitrary function (with a narrow spatial support as before) with which we convolve $h_z(x, y)$, and $u_z(x, y)$ as the result of that convolution. In that respect, we write Eq. 2.16 in the following form for easier reference in future:

$$u_0(x, y) * h_z(x, y) \approx h_z(x, y) U_0(\nu_X(x, y), \nu_Y(x, y)). \quad (2.17)$$

We arrived at Eq. 2.16 using the Fresnel diffraction formula. However, as long as $u_0(x, y)$ is sufficiently narrow, Eq. 2.16 can be used with the Rayleigh-Sommerfeld diffraction formulas as well, giving

$$u_z(x, y) \approx \frac{z}{j\lambda} \frac{e^{jkR}}{R^2} U_0\left(\frac{x}{\lambda R}, \frac{y}{\lambda R}\right). \quad (2.18)$$

Now using these approximate formulas, let us consider the input given as

$$u_0(x, y) = B_x B_y \text{sinc}(xB_x) \text{sinc}(yB_y) \quad (2.19)$$

with

$$\text{sinc}(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x} & \text{for } x \neq 0 \\ 1 & \text{for } x = 0 \end{cases}.$$

We assume that $B_x^2 + B_y^2 < \frac{1}{\lambda^2}$ and B_x^{-1} and B_y^{-1} range between several λ and several tens of λ (so that the narrow input assumption is valid). As well known,

the Fourier transform of $u_0(x, y)$ is given by

$$U_0(\nu_x, \nu_y) = \text{rect}\left(\frac{\nu_x}{B_x}\right) \text{rect}\left(\frac{\nu_y}{B_y}\right) \quad (2.20)$$

with

$$\text{rect}(x) = \begin{cases} 1 & \text{for } -0.5 < x < 0.5 \\ 0.5 & \text{for } x = \pm 0.5 \\ 0 & \text{otherwise} \end{cases} .$$

Note that $u_0(x, y)$ in Eq. 2.19 can also be viewed as the impulse response of an ideal low-pass filter with a rectangular frequency support and bandwidths B_x and B_y . In this respect, we can also view $u_z(x, y)$ as the low-pass filtered version of $h_z(x, y)$. In Fig. 2.2, $u_z(x, y)$ is displayed for $z = 1m$, $\lambda = 632.9nm$, $B_x = B_y = \frac{0.01}{\lambda}$, and $h_z(x, y)$ as given in Eq. 2.11. As seen, $u_z(x, y)$ is approximately equal to a windowed version of $h_z(x, y)$ (please compare Fig. 2.2 to Fig. 2.1). Actually, the approximation in Eq. 2.16 also predicts this result, i.e., when we apply that approximation, we get

$$u_z(x, y) \approx h_z(x, y) \text{rect}\left(\frac{\nu_X(x, y)}{B_x}\right) \text{rect}\left(\frac{\nu_Y(x, y)}{B_y}\right). \quad (2.21)$$

For a better comparison, in Fig. 2.3, the ratio of $u_z(x, y)$ to $h_z(x, y)$ is displayed, while in Fig. 2.4, a 1D cross-sections of that ratio is shown. From these figures, we see that the rectangular window predicted by the approximate formula is in fact not a perfect one: there are ripples within the window and there are nonzero terms outside the window. However, we see that the approximate formula is quite successful in predicting the support and overall shape of $u_z(x, y)$. Actually, when B_x , B_y and z are kept within suitable ranges, the approximation in Eq. 2.21 works fine for our purposes. For instance, the approximation holds with a normalized mean squared error that is less than 5% when B_x^{-1} and B_y^{-1} are between λ and 100λ , and z is greater than about $7.5 \times 10^5 \lambda$. These ranges for B_x , B_y and z are of interest to us in this thesis, and the indicated approximation

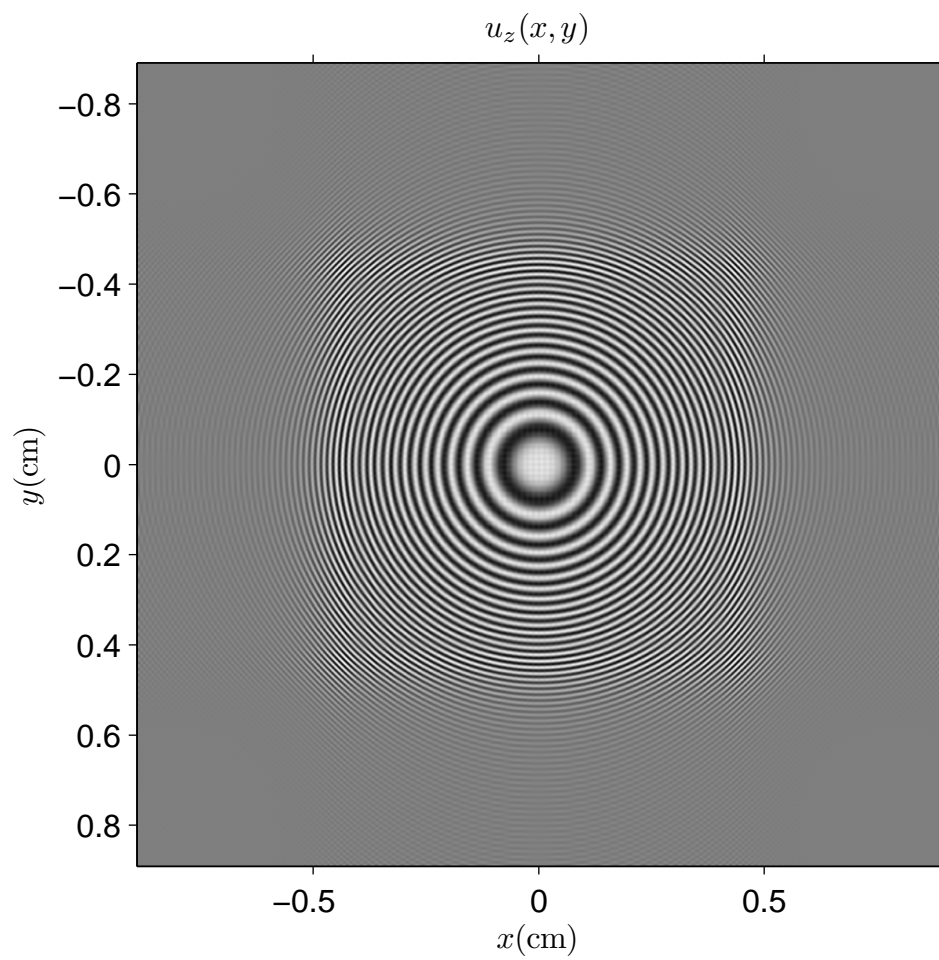


Figure 2.2: Low pass filtered version of $h_z(x, y)$ (real part). First appeared in [1].

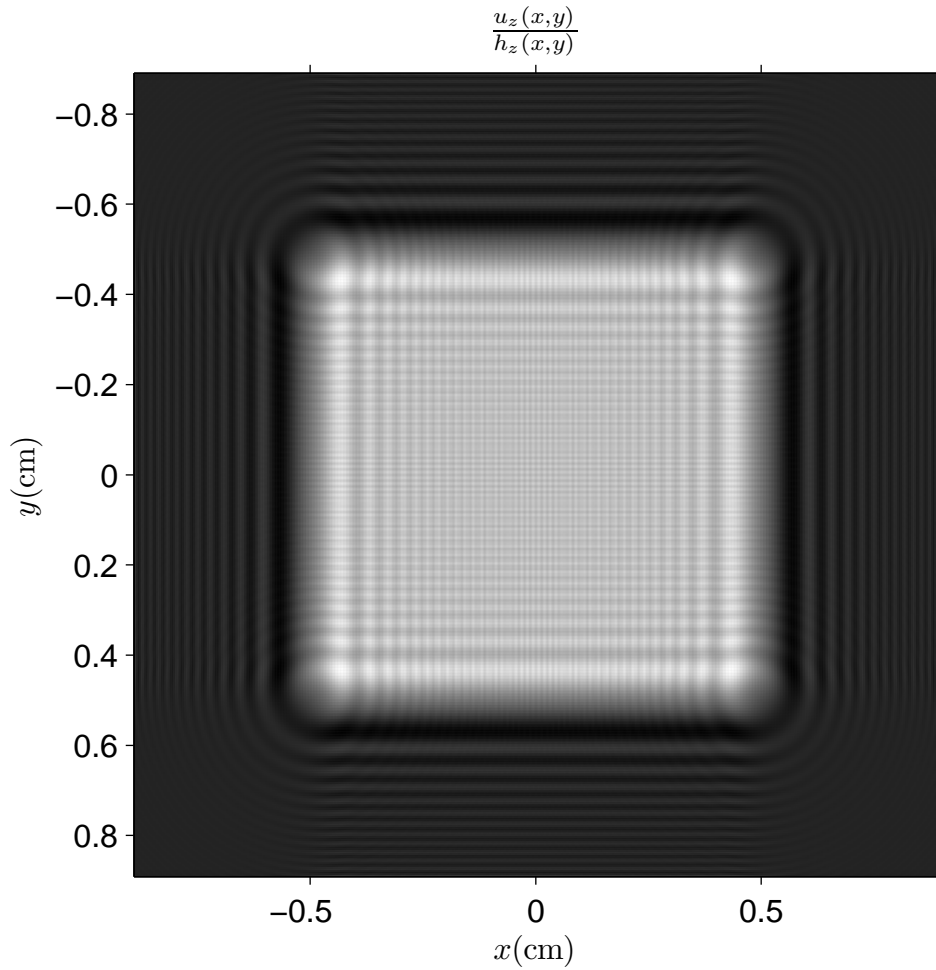


Figure 2.3: Ratio of $u_z(x, y)$ to $h_z(x, y)$ (real part).

error is acceptable. Hence, we assume that the approximation is successful and we will use it frequently from now on.

2.3 Analysis of Light Field Generated by a Spatial Light Modulator

In this section, we analyze the light field generated by a finite size spatial light modulator (SLM). This analysis is particularly important for understanding the

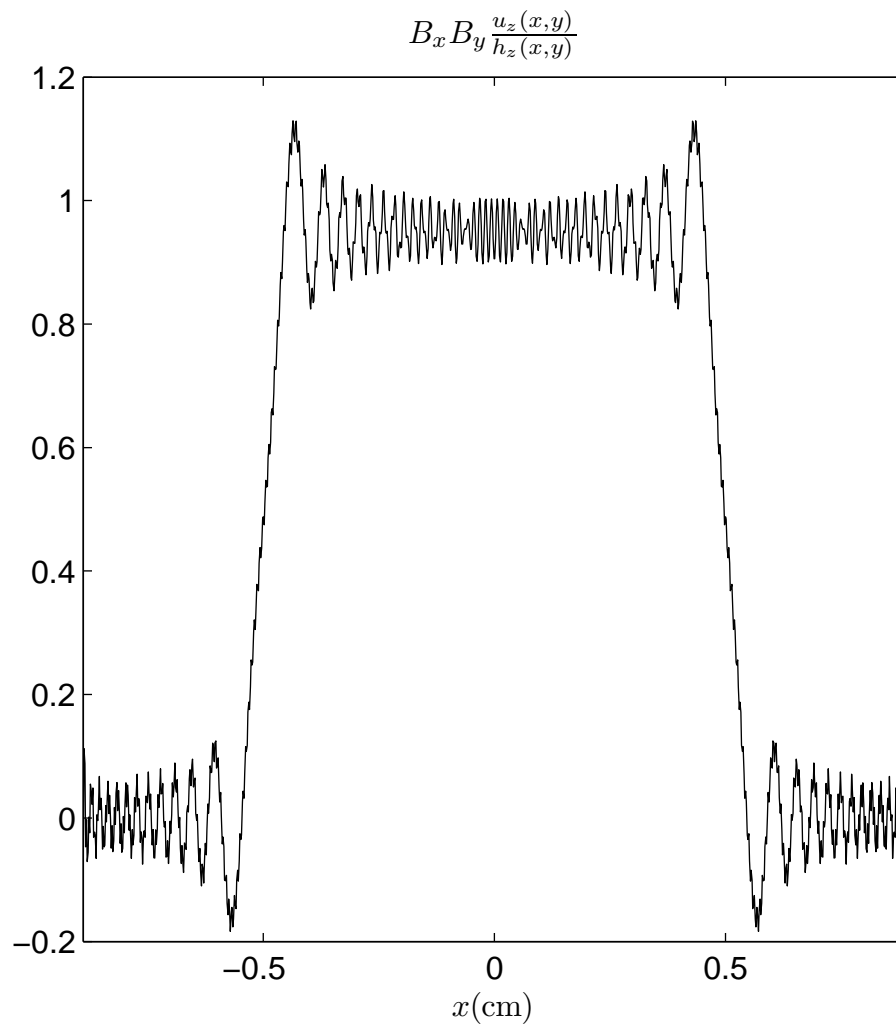


Figure 2.4: 1D cross-section of the ratio of $u_z(x, y)$ to $h_z(x, y)$ (real part).

restrictions that the SLM structure imposes on the light fields that are wished to be generated. This chapter is adapted from [1].

SLMs are usually modeled as programmable two-dimensional (2D) thin optical masks. Note that in scalar wave optics theory, a thin optical mask is represented by a 2D complex transmittance function $t(x, y)$ such that if the mask is placed to $z = 0$ plane and illuminated from left with a light wave whose profile on the $z = 0^-$ plane (that is, the left side of the mask) is $u_{0-}(x, y)$, the light field at the $z = 0^+$ plane (that is, the right side of the mask) is given as $u_{0+}(x, y) = t(x, y)u_{0-}(x, y)$. This simple model is of course only a mathematical idealization, a more accurate description of the physical behavior of the SLM requires a rigorous electrodynamic analysis. However, the accuracy offered by the mentioned model is sufficient for our purposes, and we will use it in this thesis. (Note: The thin mask model mentioned above is actually more suitable to the nature of transmissive SLMs, in which light really impinges on the SLM from one side and leaves the SLM from the other side. For reflective SLMs though, light impinges and outcomes from the same side. However, from a mathematical viewpoint, this difference is easy to manage. For sake of mathematical convenience, we will assume during the analysis that all the SLMs we are interested in are transmissive.)

Most of the SLMs today have pixellated structure and in this thesis we are interested only in such SLMs. Let Δ_x and Δ_y denote the pixel periods of a pixellated SLM. Typical values for Δ_x and Δ_y are $8\mu m$, $10\mu m$ etc. Let $a(x, y)$ denote the pixel aperture function of the SLM. For practical cases, $a(x, y) = 0$ for $|x| > \frac{\Delta_x}{2}$ or $|y| > \frac{\Delta_y}{2}$. Mostly, $a(x, y) = \text{rect}\left(\frac{x}{W_x}\right) \text{rect}\left(\frac{y}{W_y}\right)$ where $W_x \leq \Delta_x$, $W_y \leq \Delta_y$. Let $\bar{s}[m, n]$ ($m, n \in \mathbb{Z}$) denote the complex value of the $(m, n)^{\text{th}}$ SLM pixel. (In this section, we place no restriction on the values that $\bar{s}[m, n]$ can take. We assume that each $\bar{s}[m, n]$ can be adjusted to any complex number.) When viewed as a discrete function of m and n , $\bar{s}[m, n]$ denotes the 2D complex-valued

pattern that we write on the SLM. We will call $\bar{s}[m, n]$ as the SLM pattern from now on. We will assume that $\bar{s}[m, n]$ is defined for all $m, n \in \mathbb{Z}$, but that the SLM has only $M \times N$ pixels so that $\bar{s}[m, n] = 0$ for $m \notin [0, M-1]$ or $n \notin [0, N-1]$. For practical SLMs, M and N are around 1000 – 2000, so the physical dimensions of the SLM are around 1-2cm by 1-2cm. Suppose we place the SLM at the $z = 0$ plane in a symmetric manner around the origin. If we denote the complex transmittance of the SLM with $s^a(x, y)$, we have:

$$s^a(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] a(x - x_m, y - y_n) \quad (2.22)$$

with

$$x_m = \left(m - \frac{M-1}{2} \right) \Delta_x \quad (2.23)$$

and

$$y_n = \left(n - \frac{N-1}{2} \right) \Delta_y \quad (2.24)$$

denoting the location of the $(m, n)^{\text{th}}$ SLM pixel.

Let us assume that the SLM is illuminated by a normally incident plane wave of unit amplitude (that is, $u_{0-}(x, y) = 1$). Then, the light field just to the right of the SLM plane is given as $u_{0+}(x, y) = s^a(x, y)$. We assume that this field will propagate in free space and form the output field of the SLM at a distance z . If we denote the output field by $u_z^a(x, y)$, we have

$$u_z^a(x, y) = s^a(x, y) ** h_z(x, y) \quad (2.25)$$

where $h_z(x, y)$ denotes the impulse response of free-space propagation as in the previous section. Our purpose in this section is to understand the nature of $u_z^a(x, y)$.

Pixellated SLMs are inherently associated with sampling and re-interpolation of light fields. Fig. 2.5 depicts this association. As seen, the complex transmittance $s^a(x, y)$ of the SLM is viewed as being obtained by sampling and re-interpolating a light field denoted by $s(x, y)$. In particular, $s(x, y)$ is first sampled

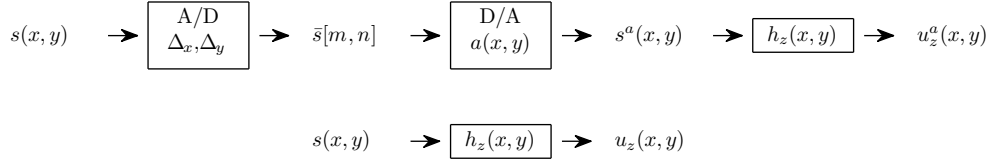


Figure 2.5: The sampling and interpolation scheme for pixellated SLMs.

at the pixel locations x_m and y_n of the SLM and as a result of this sampling, the SLM pattern $\bar{s}[m, n]$ is obtained. Then, $s^a(x, y)$ is obtained from $\bar{s}[m, n]$ by using a discrete to analog converter whose interpolating function is taken as the pixel aperture function $a(x, y)$ of the SLM. This scheme will facilitate our understanding of the nature of $u_z^a(x, y)$.

Actually, in the depicted scheme, the only constraint on $s(x, y)$ is that $s(x_m, y_n) = \bar{s}[m, n]$ for all $m, n \in \mathbb{Z}$. Therefore, infinitely many possibilities exist for the selection of $s(x, y)$. However, in order to facilitate our analysis of $u_z^a(x, y)$, we impose the additional constraint that $s(x, y)$ is to be bandlimited to the $|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$ band. In this respect, the only remaining choice for $s(x, y)$ turns out to be:

$$s(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] \text{sinc} \left(\frac{x - x_m}{\Delta_x} \right) \text{sinc} \left(\frac{y - y_n}{\Delta_y} \right). \quad (2.26)$$

Note that since $s(x, y)$ is bandlimited to the $|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$ band, during the analog to discrete conversion step in Fig., there is no aliasing. Actually, for an arbitrary $\bar{s}[m, n]$, $s(x, y)$ is sampled at the Nyquist rate.

Let $u_z(x, y)$ denote the light field produced by $s(x, y)$. Then we have

$$u_z(x, y) = h_z(x, y) ** s(x, y). \quad (2.27)$$

Note that since $s(x, y)$ is bandlimited, $u_z(x, y)$ is also bandlimited to the $|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$ band. $u_z(x, y)$ will be useful in our analysis of $u_z^a(x, y)$.

In mathematical terms, we can write the following relation between $s(x, y)$ and $s^a(x, y)$:

$$s^a(x, y) = a(x, y) ** \left\{ s(x, y) \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x - x_m, y - y_n) \right\} \quad (2.28)$$

Using the well known identity

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x - m\Delta_x, y - n\Delta_y) = \frac{1}{\Delta_x \Delta_y} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \exp \left\{ j2\pi \left(\frac{px}{\Delta_x} + \frac{qy}{\Delta_y} \right) \right\}$$

we can write Eq. 2.28 as:

$$\begin{aligned} s^a(x, y) &= c_{0,0} a(x, y) ** s(x, y) \\ &+ \sum_{\substack{p=-\infty \\ (p,q) \neq (0,0)}}^{\infty} \sum_{\substack{q=-\infty \\ (p,q) \neq (0,0)}}^{\infty} c_{p,q} a(x, y) ** \left[s(x, y) e^{j2\pi \left(\frac{px}{\Delta_x} + \frac{qy}{\Delta_y} \right)} \right] \end{aligned} \quad (2.29)$$

with

$$c_{p,q} = \frac{1}{\Delta_x \Delta_y} e^{j\pi [p(M-1) + q(N-1)]}. \quad (2.30)$$

Note that if we had $a(x, y) = \text{sinc} \left(\frac{x}{\Delta_x} \right) \text{sinc} \left(\frac{y}{\Delta_y} \right)$, the terms at the bottom line of Eq. 2.29 would disappear and we would have $s^a(x, y) = s(x, y)$. And thus, we would simply have $u_z^a(x, y) = u_z(x, y)$. However, this is not the case for the interpolating function $a(x, y)$ of a practical SLM. Thus, the terms at the bottom line remain and at the output they give rise to the well known diffraction orders.

Examining Eq. 2.29, we see that the term at the top line produces

$$\begin{aligned} u_z^{0,0}(x, y) &= c_{0,0} h_z(x, y) ** a(x, y) ** s(x, y) \\ &= c_{0,0} a(x, y) ** u_z(x, y) \end{aligned} \quad (2.31)$$

at the output plane. $u_z^{0,0}(x, y)$ is usually called the central diffraction order. The terms in the bottom line of Eq. 2.29 produce the so called higher diffraction orders. In particular, the $(p, q)^{\text{th}}$ diffraction order of the SLM output is given as

$$u_z^{p,q}(x, y) = c_{p,q} h_z(x, y) ** a(x, y) ** \left[s(x, y) e^{j2\pi \left(\frac{px}{\Delta_x} + \frac{qy}{\Delta_y} \right)} \right]. \quad (2.32)$$

Note that Eq. 2.31 is actually a special case of Eq. 2.32 with $(p, q) = (0, 0)$.

Since we chose $s(x, y)$ to be bandlimited to the $|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$ band, $u_z^{p,q}(x, y)$ is limited to the $\left|\nu_x - \frac{p}{\Delta_x}\right| \leq \frac{1}{2\Delta_x}$ and $\left|\nu_y - \frac{q}{\Delta_y}\right| \leq \frac{1}{2\Delta_y}$ band. Hence, with our convention for choosing $s(x, y)$, diffraction orders are strictly separated from each other in the frequency domain.

To facilitate the discussion, we will first assume that $a(x, y) = \delta(x, y)$. In this case, the $(p, q)^{\text{th}}$ diffraction order becomes

$$u_z^{p,q}(x, y) = c_{p,q} h_z(x, y) ** \left[s(x, y) e^{j2\pi\left(\frac{px}{\Delta_x} + \frac{qy}{\Delta_y}\right)} \right], \quad (2.33)$$

and in particular, the central order becomes

$$u_z^{0,0}(x, y) = c_{0,0} u_z(x, y). \quad (2.34)$$

In [107] it is shown that the $u_z^{p,q}(x, y)$ given in Eq. 2.33 essentially turns out to be the translated, modulated and dispersed version of the $u_z^{0,0}(x, y)$ given in Eq. 2.34 when RS diffraction model is used. In [108], it is shown that under the Fresnel approximation, $u_z^{p,q}(x, y)$ can be written in terms of $u_z^{0,0}(x, y)$ as

$$\frac{c_{p,q}}{c_{0,0}} e^{-j\pi\lambda z \left(\frac{p^2}{\Delta_x^2} + \frac{q^2}{\Delta_y^2}\right)} u_z^{0,0} \left(x - p \frac{\lambda z}{\Delta_x}, y - q \frac{\lambda z}{\Delta_y} \right) \exp \left\{ j2\pi \left(\frac{px}{\Delta_x} + \frac{qy}{\Delta_y} \right) \right\} \quad (2.35)$$

from which we see that $u_z^{p,q}(x, y)$ is essentially a shifted and modulated version of $u_z^{0,0}(x, y)$.

By Eq. 2.34 and Eq. 2.35, we see that the SLM output $u_z^a(x, y)$ consists of diffraction orders which are all related to $u_z(x, y)$. Thus, we can get more insight about $u_z^a(x, y)$ if we examine $u_z(x, y)$ in detail. Using Eq. 2.26 and Eq. 2.27, we can write

$$\begin{aligned} u_z(x, y) &= h_z(x, y) ** \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] \text{sinc} \left(\frac{x - x_m}{\Delta_x} \right) \text{sinc} \left(\frac{y - y_n}{\Delta_y} \right) \\ &= \left\{ h_z(x, y) ** \text{sinc} \left(\frac{x}{\Delta_x} \right) \text{sinc} \left(\frac{y}{\Delta_y} \right) \right\} \\ &\quad ** \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] \delta(x - x_m, y - y_n). \end{aligned} \quad (2.36)$$

Next, let us apply the approximation in Eq. 2.17 to the second line of the above equation, and in particular to the term within the curly brackets. Assuming that Fresnel diffraction model is applicable, we get

$$\begin{aligned}
u_z(x, y) &\approx \left\{ \Delta_x \Delta_y h_z(x, y) \text{rect} \left(\frac{x}{\lambda z / \Delta_x} \right) \text{rect} \left(\frac{y}{\lambda z / \Delta_y} \right) \right\} \\
&\quad ** \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] \delta(x - x_m, y - y_n) \\
&\approx \Delta_x \Delta_y \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] h_z(x - x_m, y - y_n) \\
&\quad \text{rect} \left(\frac{x - x_m}{\lambda z / \Delta_x} \right) \text{rect} \left(\frac{y - y_n}{\lambda z / \Delta_y} \right). \tag{2.37}
\end{aligned}$$

From Eq. 2.37, we see that $u_z(x, y)$ is approximately confined in space to the region given as $|x| < \frac{\lambda z}{2\Delta_x} + \frac{M\Delta_x}{2}$ and $|y| < \frac{\lambda z}{2\Delta_y} + \frac{N\Delta_y}{2}$. At this point, let us consider some practical values for λ , Δ_x , Δ_y , M , N and z to see some further simplification possibilities. Assume that we take $\lambda = 632.9nm$, $\Delta_x = \Delta_y = 8\mu m$, $M = N = 1024$ and $z = 1.5m$. Then, we get $\frac{\lambda z}{2\Delta_x} = 11.87cm$ and $\frac{M\Delta_x}{2} = 0.41cm$. To facilitate our analytical work, we can ignore the $\frac{M\Delta_x}{2}$ term in comparison to the $\frac{\lambda z}{2\Delta_x}$ term. Hence, we can approximate the support of $u_z(x, y)$ as

$$|x| < \frac{\lambda z}{2\Delta_x} \text{ and } |y| < \frac{\lambda z}{2\Delta_y} \tag{2.38}$$

and within the above region, we can simply write $u_z(x, y)$ as

$$u_z(x, y) \approx \Delta_x \Delta_y \text{rect} \left(\frac{x\Delta_x}{\lambda z} \right) \text{rect} \left(\frac{y\Delta_y}{\lambda z} \right) \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] h_z(x - x_m, y - y_n). \tag{2.39}$$

We will call the region specified by Eq. 2.38 as the central diffraction order region. When z is viewed as a varying parameter, Eq. 2.38 defines a 3D pyramid (whose tip is at origin and base expands in $+z$ direction), which we will name as the central diffraction order pyramid (see Fig. 2.6). However, since the approximations we made in arriving at Eq. 2.38 hold only after a certain distance from the SLM (approximately a meter), we will not be interested in the portion of the central order pyramid that lies too close to the SLM.

Having seen that the central diffraction order of the SLM output approximately lies in the region specified in Eq. 2.38, we turn our attention to the higher diffraction orders. By Eq. 2.35, we can see that $(p, q)^{\text{th}}$ diffraction order of the SLM is approximately centered around $\left(\frac{p\lambda z}{\Delta_x}, \frac{q\lambda z}{\Delta_y}\right)$, and has dimensions of $\frac{\lambda z}{\Delta_x}$ and $\frac{\lambda z}{\Delta_y}$. Therefore, for sufficiently large distances, diffraction orders of the SLM do not overlap in space (approximately), meaning that higher diffraction orders make no contribution to the the central diffraction order region (see Fig. 2.6). Hence, we can write

$$u_z^a(x, y) \approx \frac{1}{\Delta_x \Delta_y} u_z(x, y) \text{ for } |x| < \frac{\lambda z}{2\Delta_x} \text{ and } |y| < \frac{\lambda z}{2\Delta_y}. \quad (2.40)$$

Since higher diffraction orders do not contain any new information, it suffices to examine the SLM output only in the central diffraction order region. And we will do so from now.

As an illustration of the analysis until this point, Fig. 2.7 displays an SLM pattern which produces the output field displayed in Fig. 2.8 when written on an SLM with $M = N = 1024$, $\Delta_x = \Delta_y = 8\mu m$. The output field is displayed at $z = 1m$ and we took $\lambda = 632.9nm$. In Fig. 2.8, nine diffraction orders are displayed where the central order is the one in the middle. The output field is computed using the Fresnel diffraction model. Note that for the given parameters, Eq. 2.38 predicts the size of the central diffraction order region as $7.91cm \times 7.91cm$, which is also verified by Fig. 2.8.

Assuming that the Fresnel approximation is valid within the central diffraction order region, using Eq. 2.11, Eq. 2.39 and Eq. 2.40, we can write:

$$u_z^a(x, y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] e^{\frac{j\pi}{\lambda z}(x_m^2+y_n^2)} e^{-j\frac{2\pi}{\lambda z}(xx_m+yy_n)} \quad (2.41)$$

for $|x| < \frac{\lambda z}{2\Delta_x}$ and $|y| < \frac{\lambda z}{2\Delta_y}$. Since $\bar{s}[m, n]$ has $M \times N$ degrees of freedom, $u_z^a(x, y)$ also has $M \times N$ degrees of freedom as well. In fact, it can be shown that $u_z^a(x, y)$ of Eq. 2.41 can fully be represented by its $M \times N$ samples taken uniformly within

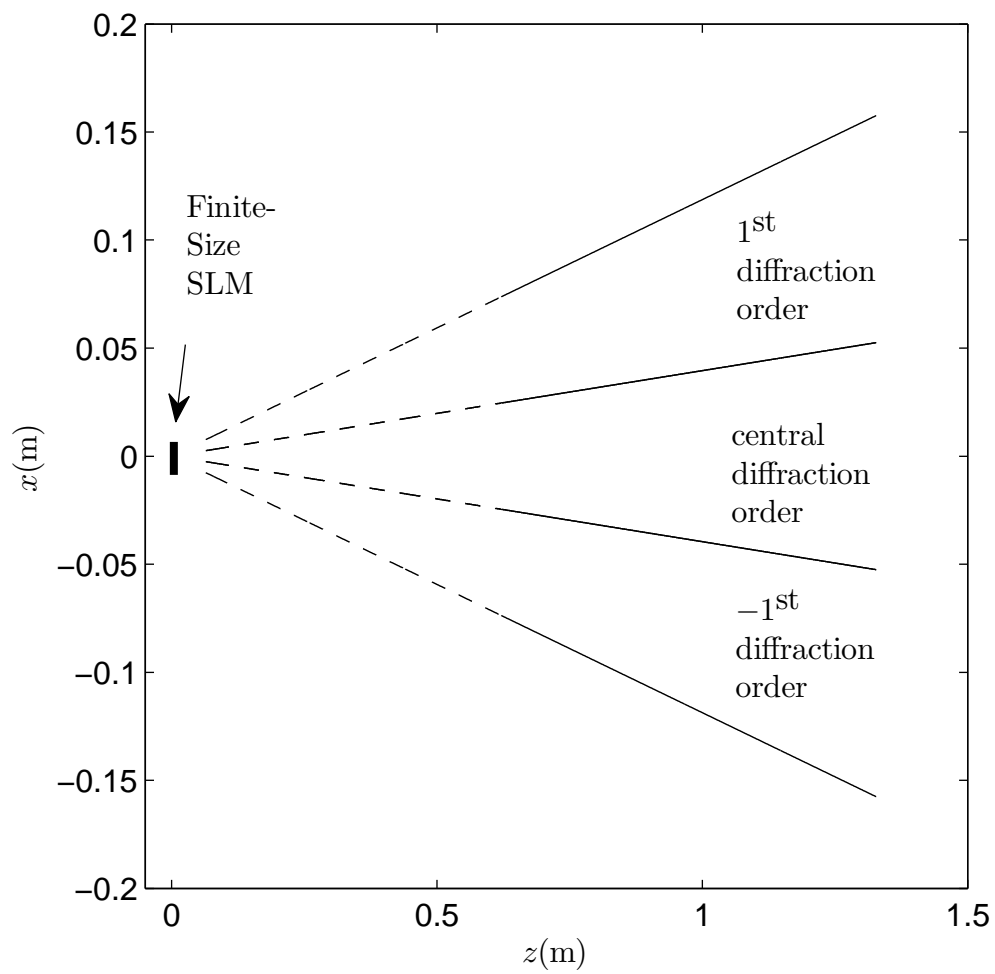


Figure 2.6: Spatial regions occupied by the diffraction orders of a finite size SLM.

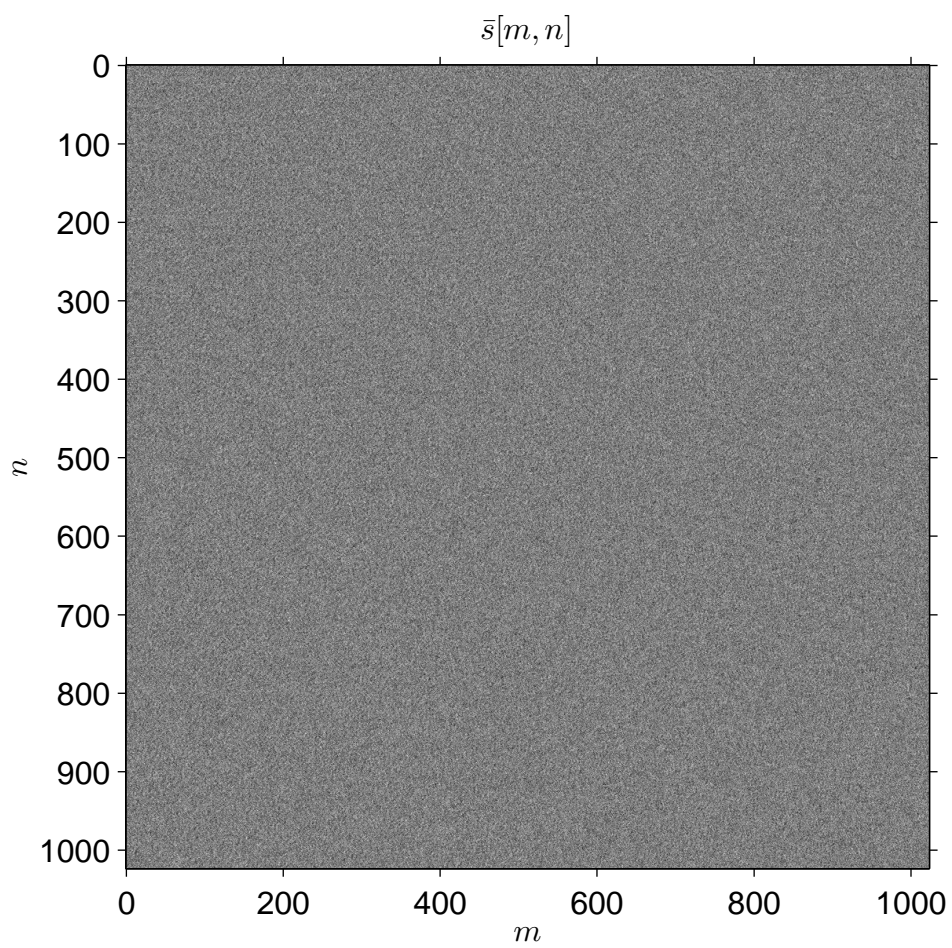


Figure 2.7: An SLM pattern (real part).

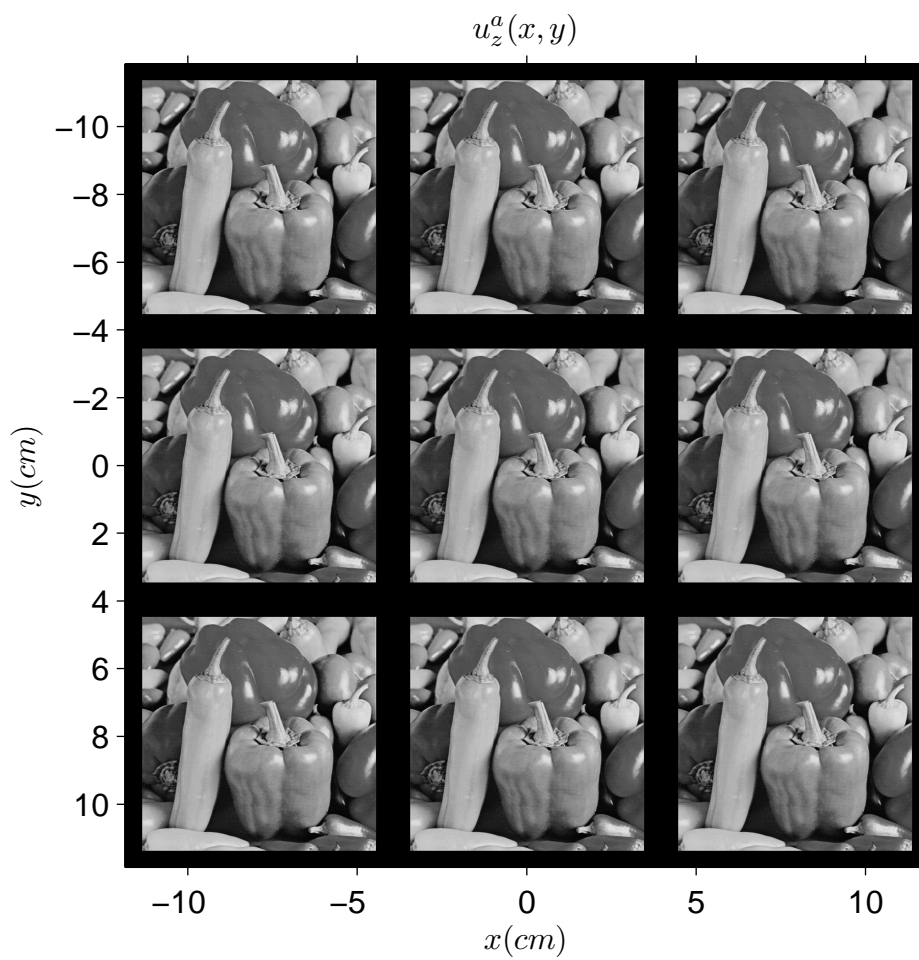


Figure 2.8: Output field produced by the SLM pattern in Fig. 2.7 (magnitude). Pixels are assumed to be impulsive.

the central diffraction order region. These samples can be computed as:

$$\begin{aligned}
\bar{u}_z^a[m, n] &= u_z^a(\tilde{x}_m, \tilde{y}_n) \\
&= \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(\tilde{x}_m^2 + \tilde{y}_n^2)} \\
&\quad \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \bar{s}[m', n'] e^{\frac{j\pi}{\lambda z}(x_{m'}^2 + y_{n'}^2)} e^{-j\frac{2\pi}{\lambda z}(\tilde{x}_m x_{m'} + \tilde{y}_n y_{n'})}
\end{aligned} \tag{2.42}$$

where $m, n \in \mathcal{Z}$; $0 \leq m \leq M - 1$, $0 \leq n \leq N - 1$ and

$$\tilde{x}_m = \left(m - \frac{M-1}{2} \right) \frac{\lambda z}{M\Delta_x}, \tag{2.43}$$

$$\tilde{y}_n = \left(n - \frac{N-1}{2} \right) \frac{\lambda z}{N\Delta_y}. \tag{2.44}$$

From Eq. 2.23, Eq. 2.24, Eq. 2.43, Eq. 2.44 and Eq. 2.42, we see that computation of the samples of SLM output within the central diffraction order region involves multiplying the SLM pattern $\bar{s}[m, n]$ with a discrete chirp, taking a centered 2D-DFT, and then multiplying with another discrete chirp [109, 110, 111]. Interpolation of $u_z^a(x, y)$ from $\bar{u}_z^a[m, n]$ is discussed in [112] and is slightly different than classical sinc interpolation. In light field synthesis problems, desired fields are usually specified through $\bar{u}_z^a[m, n]$. For a given $\bar{u}_z^a[m, n]$, the required SLM pattern $\bar{s}[m, n]$ can be computed as:

$$\begin{aligned}
\bar{s}[m, n] &= \frac{j\lambda z e^{-jkz}}{MN} e^{-\frac{j\pi}{\lambda z}(x_m^2 + y_n^2)} \\
&\quad \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \bar{u}_z^a[m', n'] e^{-\frac{j\pi}{\lambda z}(\tilde{x}_{m'}^2 + \tilde{y}_{n'}^2)} e^{j\frac{2\pi}{\lambda z}(x_m \tilde{x}_{m'} + y_n \tilde{y}_{n'})}
\end{aligned} \tag{2.45}$$

for $0 \leq m \leq M - 1$ and $0 \leq n \leq N - 1$. Indeed, Eq. 2.45 is just the inverse of Eq. 2.42.

Finally, up to now, we assumed that $a(x, y) = \delta(x, y)$. In practice, $a(x, y)$ extends over a nonzero area but is confined to the $|x| \leq \frac{\Delta_x}{2}$ and $|y| \leq \frac{\Delta_y}{2}$ region. Mostly, $a(x, y) = \text{rect}\left(\frac{x}{W_x}\right) \text{rect}\left(\frac{y}{W_y}\right)$ with $W_x \leq \Delta_x$ and $W_y \leq \Delta_y$. In such a case, in a strict analysis, convolution with $a(x, y)$ must be incorporated in Eq. 2.35, Eq. 2.39, Eq. 2.40 and Eq. 2.41. The result will be a blurring in the SLM

output relative to the $a(x, y) = \delta(x, y)$ case. Usually, $A(\nu_x, \nu_y) = \mathcal{F}\{a(x, y)\}$ is a decaying function of ν_x and ν_y , so that higher diffraction orders suffer more from this blurring and are attenuated relative to the central diffraction order. But the locations and spatial supports of diffraction orders will not change since $a(x, y)$ is narrow. Actually, a narrow $a(x, y)$ also implies that $A(\nu_x, \nu_y)$ can be considered as constant over the $|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$ band — which is the band occupied by the central diffraction order. Therefore, we can assume that the central diffraction order is only modified by a constant multiplicative factor. The conclusion of this discussion is that, a finite $a(x, y)$ has no drastic effect on our analysis. Therefore, we can (and will) assume that Eq. 2.38 through Eq. 2.45 are valid in the case of a practical $a(x, y)$ as well. (In Fig. 2.9, we display the output produced by the SLM pattern shown in Fig. 2.7 assuming that the pixels of the SLM are rectangular. Upon comparing this figure to Fig. 2.8, we indeed see that the central order is virtually not affected by the finite extent of the pixel aperture function.)

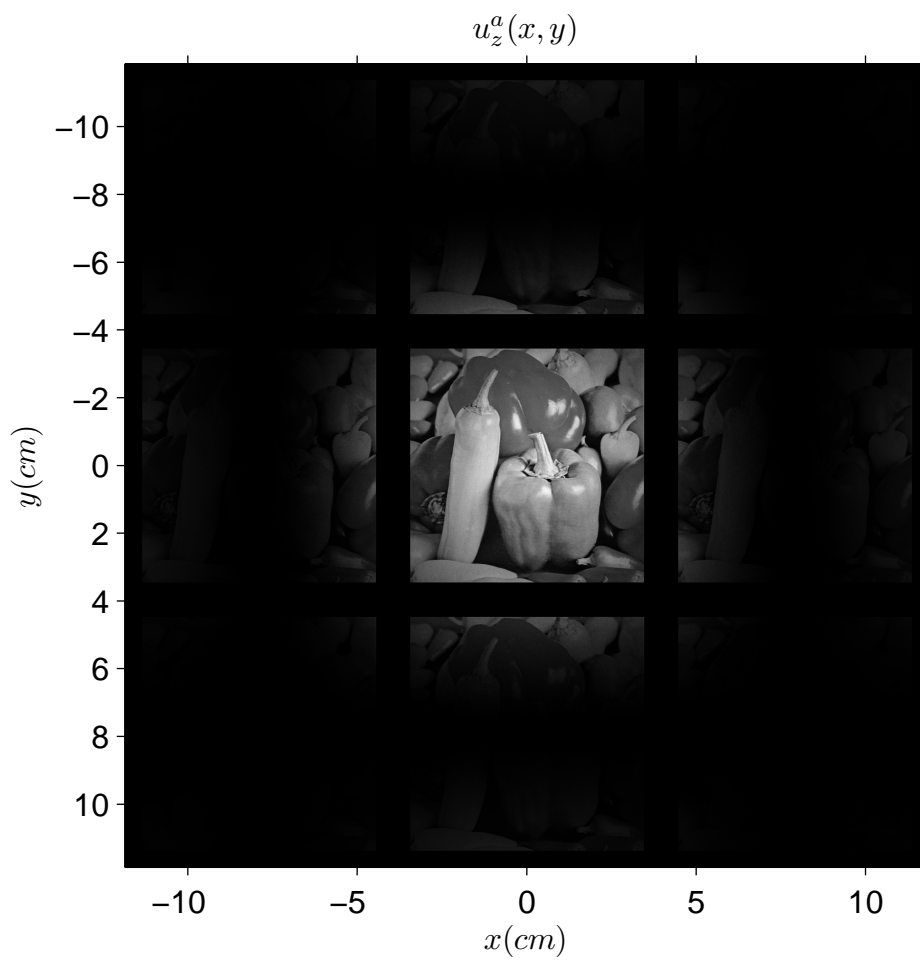


Figure 2.9: Output field produced by the SLM pattern in Fig. 2.7 (magnitude). Pixels are assumed to be rectangular.

Chapter 3

SYNTHESIS OF THREE-DIMENSIONAL LIGHT FIELDS WITH BINARY SPATIAL LIGHT MODULATORS

In Chapter 1, we pointed that binary SLMs have certain properties that make them quite attractive to be used in holographic display applications. In this chapter, we develop the theory of three-dimensional light field synthesis with a finite-size binary SLM. From another perspective, we develop the theory of halftoning for complex-valued gray scale holograms. We start by reviewing the methods that have been proposed until now.

The research on the computation of binary patterns that synthesize desired light fields initially emerged within the context of computer generated holography

[64, 65, 69, 70]. In computer generated holography, one first computes the hologram of a mathematically described and possibly physically non-existing object, and then prepares an optical mask on which the computed hologram is written. Holograms obtained in this manner are named computer generated holograms (CGH). With CGHs, it has been possible to synthesize light fields whose holograms have never been recorded optically. In the early days in which dynamic SLMs were not available, the CGHs were printed on one-time fabricated optical masks named diffractive optical elements (DOEs) [113, 114]. Actually, the first DOEs were binary. These DOEs were prepared using the so called detour phase method [71, 72, 73, 74, 75, 76, 77, 78, 79]. Using this method, it was possible to synthesize desired complex-valued monochromatic light fields within a small region (centered around the optical axis) of the far field or on the Fourier plane of a $2f$ setup. In particular, in the detour phase method, the DOE is broken down into a number of cells and in each cell, a rectangular hole is placed. The position and the dimensions of this hole are adjusted such that when illuminated with an oblique wave, the cell behaves no different than a gray level complex-valued pixel when the synthesis region is considered. Hence, the entire DOE behaves like a gray level DOE. Later, the basic method was improved and modified to operate in the non-far field range where Fresnel diffraction model is applicable [115, 116]. Such methods are called cell oriented methods.

With the advancement in the pixellated SLM technologies, the research on binary CGHs shifted towards pixel oriented methods since direct application of cell oriented methods became difficult. In these methods, the SLM is taken as a collection of binary pixels and the goal is to determine the discrete binary CGH pattern to be written on the SLM. Similar to the cell oriented case, the research initially focused on reconstructions at the far field or on the Fourier plane of a $2f$ setup. The reason is that in the far field or the $2f$ setup case, the relation between the SLM pixels and output field samples are simply given by a discrete Fourier transform, which is easy to understand and manipulate. Many

iterative and non-iterative algorithms have been designed or adapted and applied to this problem [117, 118, 119, 120, 121, 122, 123, 124]. Exploiting the intrinsic connection to the classical halftoning problem of image processing, researchers also adapted and applied halftoning algorithms such as error diffusion [125, 126, 127, 128, 129] and direct binary search [130, 131, 132, 133, 134]. In addition, projection onto convex sets (POCS) or Gerchberg-Saxton like algorithms such as iterative Fourier transform algorithm have been proposed [135, 136, 137]. Such algorithms have been extensively analyzed in terms of reconstruction error, diffraction efficiency, computational performance, etc [138, 139, 140, 141, 142]. However, minor work has been done to develop algorithms for the non-far field range where Rayleigh-Sommerfeld or Fresnel diffraction models are valid, perhaps due to the difficulty in the involved analytical relations [143, 92].

We assume in the upcoming discussions that the binary SLM is illuminated by a plane wave and the desired field is synthesized within a volumetric region in the non-far field range after merely free space propagation from the SLM plane. We start the discussion with Sec. 3.1, in which we show that in the case of a finite-sized SLM, an arbitrary SLM pattern essentially produces the same light field with its low-pass filtered version within a certain region of space. We use this observation in Sec. 3.2 to find binary SLM patterns that generate desired light fields specified within an appropriately defined volumetric region. Using computer simulations, we show that binary SLM patterns computed with our approach successfully generate planar as well as volumetric (three-dimensional) light fields. What is presented in this chapter is a follow-up to the work presented in [1].

3.1 Effects of Applying a Low-Pass Filter to the SLM Pattern

In Sec. 2.3, we analyzed the light field generated by a finite sized SLM. In this section, we consider the effects of applying a discrete low-pass filter to the pattern that we write on the SLM. The results of this section will be quite useful for the next section in which we determine binary patterns which synthesize desired light fields.

Let $\bar{s}[m, n]$ denote some pattern that we write on the SLM and let $u_z^a(x, y)$ denote the resulting light field, as in Sec. 2.3. Recall that $u_z^a(x, y)$ consists of diffraction orders which are approximately non-overlapping in space for sufficiently large distances from the SLM plane (about a meter), and the central order (the order in which we are interested in) approximately lies within the region given by Eq. 2.38; where within this region, $u_z^a(x, y)$ is related to $\bar{s}[m, n]$ through the relation given in Eq. 2.41. (We still assume a finite sized SLM so that $\bar{s}[m, n] = 0$ for $m \notin [0, M-1]$ or $n \notin [0, N-1]$. Also, we continue to assume that each $\bar{s}[m, n]$ can be adjusted to any desired complex number.)

Now, suppose instead of $\bar{s}[m, n]$, we write on the same SLM (that is, we do not assume any change in the physical parameters of the SLM) a new pattern that we denote by $\bar{s}_L[m, n]$, such that

$$\begin{aligned} \bar{s}_L[m, n] &= \bar{s}[m, n] \star \star \bar{g}[m, n] \\ &= \sum_{m'=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} \bar{s}[m', n'] \bar{g}[m - m', n - n'] \\ &= \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \bar{s}[m', n'] \bar{g}[m - m', n - n'] \end{aligned} \quad (3.1)$$

where $\bar{g}[m, n]$ denotes the impulse response of a discrete low-pass filter and $\star \star$ denotes two-dimensional discrete convolution operation. In this respect, we modify the scheme depicted in Fig. 2.5 as in Fig. 3.1. As also illustrated in Fig. 3.1,

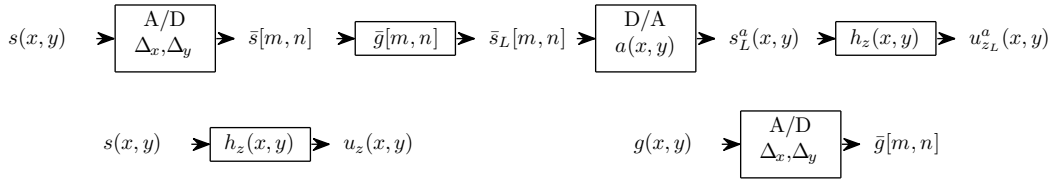


Figure 3.1: Modified scheme for the application of a discrete filter to the SLM pattern.

let $u_{zL}^a(x, y)$ denote the new SLM output, i.e., the SLM output generated by $\bar{s}_L[m, n]$. Our purpose in this section is to establish the relation between $u_z^a(x, y)$ and $u_{zL}^a(x, y)$.

Before going any further, we should mention that though on our SLM we can write $\bar{s}[m, n]$ without any problem, in strict sense, we will be unable to write $\bar{s}_L[m, n]$ exactly, since our SLM has (by our assumption) M by N pixels, but $\bar{s}_L[m, n]$ can theoretically have an infinite number of nonzero values due to the possible tails of $\bar{g}[m, n]$. Therefore, in reality, we can only write a clipped version of $\bar{s}_L[m, n]$. However, the particular $\bar{g}[m, n]$ that we will be interested in this section will have a small compact support around $(m, n) = (0, 0)$, so that the support of $\bar{s}_L[m, n]$ will be only slightly larger than the region specified as $m \in [0, M - 1]$ and $n \in [0, N - 1]$. Hence, we will assume that the clipping will not cause significant deviation from the presented results and we will not attempt any analysis of its effects in order not to complicate the discussion.

Now, recall that in Sec. 2.3, we viewed $\bar{s}[m, n]$ as being obtained by sampling a continuous signal $s(x, y)$ with the pixel periods Δ_x and Δ_y of the SLM (See Fig. 2.5 and Eq. 2.26). Similar to that, we can also view $\bar{g}[m, n]$ as being obtained by sampling a continuous signal $g(x, y)$ with the pixel periods Δ_x and Δ_y of the SLM (see Fig. 3.1). For sake of simplicity, let us take $g(x, y)$ as

$$g(x, y) = B_x B_y \text{sinc}(xB_x) \text{sinc}(yB_y) \quad (3.2)$$

which is actually the impulse response of the ideal low-pass filter (with rectangular frequency support) with bandwidths B_x and B_y . Note that the Fourier

transform of $g(x, y)$ is given as

$$G(f_x, f_y) = \text{rect}\left(\frac{f_x}{B_x}\right) \text{rect}\left(\frac{f_y}{B_y}\right). \quad (3.3)$$

With the current selection for $g(x, y)$, we get

$$\begin{aligned} \bar{g}[m, n] &= g(m\Delta_x, n\Delta_y) \\ &= B_x B_y \text{sinc}(m\Delta_x B_x) \text{sinc}(n\Delta_y B_y). \end{aligned} \quad (3.4)$$

As we did for $\bar{s}[m, n]$, let us also impose the no aliasing condition for $\bar{g}[m, n]$. In this respect, the bandwidth parameters B_x and B_y should satisfy

$$B_x < \frac{1}{\Delta_x}, \quad B_y < \frac{1}{\Delta_y}. \quad (3.5)$$

Under the above condition, $g(x, y)$ is bandlimited to the $|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$ band.

Now, let $s_L(x, y) = \frac{1}{\Delta_x \Delta_y} g(x, y) ** s(x, y)$ where $s(x, y)$ is as given in Eq. 2.26.

We can explicitly write

$$\begin{aligned} s_L(x, y) &= \frac{1}{\Delta_x \Delta_y} g(x, y) ** s(x, y) \\ &= \frac{1}{\Delta_x \Delta_y} g(x, y) ** \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] \text{sinc}\left(\frac{x - x_m}{\Delta_x}\right) \text{sinc}\left(\frac{y - y_n}{\Delta_y}\right) \\ &= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] g(x - x_m, y - y_n) \end{aligned} \quad (3.6)$$

where the last line follows from the fact that under the condition in Eq. 3.5, we have

$$g(x, y) ** \text{sinc}\left(\frac{x}{\Delta_x}\right) \text{sinc}\left(\frac{y}{\Delta_y}\right) = \Delta_x \Delta_y g(x, y).$$

Let us now sample $s_L(x, y)$ at (x_m, y_n) where x_m and y_n are as given in Eq. 2.23 and Eq. 2.24 (recall that x_m and y_n denote the location of the $(m, n)^{\text{th}}$ pixel of

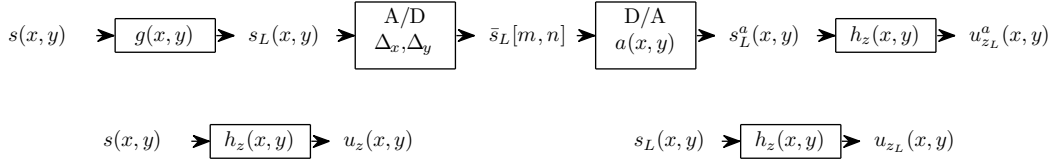


Figure 3.2: Equivalent scheme to the one displayed in Fig. 3.1.

our SLM). From Eq. 3.1, Eq. 3.4 and Eq. 3.6 , we can see that

$$\begin{aligned}
s_L(x_m, y_n) &= \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \bar{s}[m', n'] g(x_m - x_{m'}, y - y_{n'}) \\
&= \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \bar{s}[m', n'] g\{(m - m')\Delta_x, (n - n')\Delta_y\} \\
&= \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \bar{s}[m', n'] \bar{g}[m - m', n - n'] \\
&= \bar{s}_L[m, n].
\end{aligned} \tag{3.7}$$

Hence, we see that the new SLM pattern $\bar{s}_L[m, n]$ is obtained by sampling $s_L(x, y) = \frac{1}{\Delta_x \Delta_y} g(x, y) * * s(x, y)$. In this respect, we can replace the scheme depicted in Fig. 3.1 by the equivalent scheme depicted in Fig. 3.2. Note that just as $s(x, y)$ and $g(x, y)$, $s_L(x, y)$ is also bandlimited to the $|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$. Therefore, $\bar{s}_L[m, n]$ is also alias free. Now, let $u_{z_L}(x, y)$ denote the diffraction field produced by $s_L(x, y)$ such that

$$u_{z_L}(x, y) = s_L(x, y) * * h_z(x, y). \tag{3.8}$$

Then, by the analysis in the previous section, we know that within the central diffraction order region given by Eq. 2.38, we have

$$u_{z_L}^a(x, y) \approx \frac{1}{\Delta_x \Delta_y} u_{z_L}(x, y). \tag{3.9}$$

We can understand the relation between $u_{z_L}^a(x, y)$ and $u_{z_L}(x, y)$ if we examine $u_{z_L}(x, y)$ in detail.

Using Eq. 3.6 and Eq. 3.8, we can explicitly write

$$\begin{aligned}
u_{z_L}(x, y) &= h_z(x, y) ** \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] g(x - x_m, y - y_n) \\
&= \{h_z(x, y) ** g(x, y)\} \\
&\quad ** \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] \delta(x - x_m, y - y_n). \tag{3.10}
\end{aligned}$$

Next, let us apply the approximation in Eq. 2.17 to the second line of the above equation, and in particular to the term within the curly brackets. Assuming that Fresnel diffraction model is applicable, we get

$$\begin{aligned}
u_{z_L}(x, y) &\approx \left\{ h_z(x, y) \text{rect} \left(\frac{x}{\lambda z B_x} \right) \text{rect} \left(\frac{y}{\lambda z B_y} \right) \right\} \\
&\quad ** \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] \delta(x - x_m, y - y_n) \\
&\approx \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] h_z(x - x_m, y - y_n) \\
&\quad \text{rect} \left(\frac{x - x_m}{\lambda z B_x} \right) \text{rect} \left(\frac{y - y_n}{\lambda z B_y} \right). \tag{3.11}
\end{aligned}$$

From the above equation, we can see that $u_{z_L}(x, y)$ is approximately confined in space to the region given as $|x| < \frac{\lambda z B_x}{2} + \frac{M \Delta_x}{2}$ and $|y| < \frac{\lambda z B_y}{2} + \frac{N \Delta_y}{2}$. Again, when practical values are considered for λ , Δ_x , Δ_y , M , N and z , this region can be approximated as:

$$|x| < \frac{\lambda z B_x}{2} \text{ and } |y| < \frac{\lambda z B_y}{2}. \tag{3.12}$$

Note that because of Eq. 3.5, the region given above is a subregion of the central diffraction order region given in Eq. 2.38. With the recent approximation for its support, we can simply write $u_{z_L}(x, y)$ as

$$\begin{aligned}
u_{z_L}(x, y) &\approx \text{rect} \left(\frac{x}{\lambda z B_x} \right) \text{rect} \left(\frac{y}{\lambda z B_y} \right) \sum_{m=-M}^M \sum_{n=-N}^N \bar{s}[m, n] h_z(x - x_m, y - y_n) \\
&\approx \text{rect} \left(\frac{x}{\lambda z B_x} \right) \text{rect} \left(\frac{y}{\lambda z B_y} \right) \frac{1}{\Delta_x \Delta_y} u_z(x, y) \tag{3.13}
\end{aligned}$$

where the second line follows from Eq. 2.39. (Recall that $u_z(x, y)$ denotes the light field produced by $s(x, y)$, which is sampled to give $\bar{s}[m, n]$.)

Next, from Eq. 3.9 and Eq. 3.13, within the central diffraction order region, we can write

$$u_{z_L}^a(x, y) \approx \frac{1}{\Delta_x^2 \Delta_y^2} u_z(x, y) \text{rect}\left(\frac{x}{\lambda z B_x}\right) \text{rect}\left(\frac{y}{\lambda z B_y}\right). \quad (3.14)$$

Finally, by Eq. 3.14 and Eq. 2.40, within central diffraction order region we get the following relation:

$$u_{z_L}^a(x, y) \approx \frac{1}{\Delta_x \Delta_y} u_z^a(x, y) \text{rect}\left(\frac{x}{\lambda z B_x}\right) \text{rect}\left(\frac{y}{\lambda z B_y}\right). \quad (3.15)$$

Eq. 3.15 is the result we have been seeking for. This equation says that within the region specified by Eq. 3.12, $\bar{s}[m, n]$ and $\bar{s}_L[m, n]$ approximately produce the same field. Moreover, $\bar{s}_L[m, n]$ approximately produces nothing in the rest of central order diffraction region.

As an illustration of this effect, consider the SLM pattern $\bar{s}[m, n]$ shown in Fig. 3.3. This pattern produces the output field displayed in Fig. 3.4. (Here, the SLM size is 1024×1024 , $\Delta_x = \Delta_y = 8\mu m$, $\lambda = 632.9nm$ and $z = 1m$; so the physical size of the SLM is $8.2mm \times 8.2mm$ and the physical size of the central diffraction order is $7.91cm \times 7.91cm$. In Fig. 3.4, only the central diffraction order is displayed. Visually, other diffraction orders are essentially replicas of the central order.) As seen, the light field in Fig. 3.4 consists of an image in the middle surrounded by texts. If only the image were present, $\bar{s}[m, n]$ would be a low-pass pattern, because only low-angle rays from the SLM would be sufficient to produce the image. However, the presence of the texts, which require high-angle rays from the SLM, causes $\bar{s}[m, n]$ to be a full-band discrete signal. Next, consider Fig. 3.5, which shows the SLM pattern $\bar{s}_L[m, n]$ obtained with filtering the SLM pattern in Fig. 3.3 with $\bar{g}[m, n]$ (B_x and B_y are taken such that $B_x \Delta_x = B_y \Delta_y = 0.375$). The resulting output is shown in Fig. 3.6. Also shown in Fig. 3.6 are the borders of the region specified in Eq. 3.12. As seen, the image is preserved, while the texts are eliminated. Hence, as predicted, output approximately remains unchanged within the region specified in Eq. 3.12 and approximately vanishes outside this region.

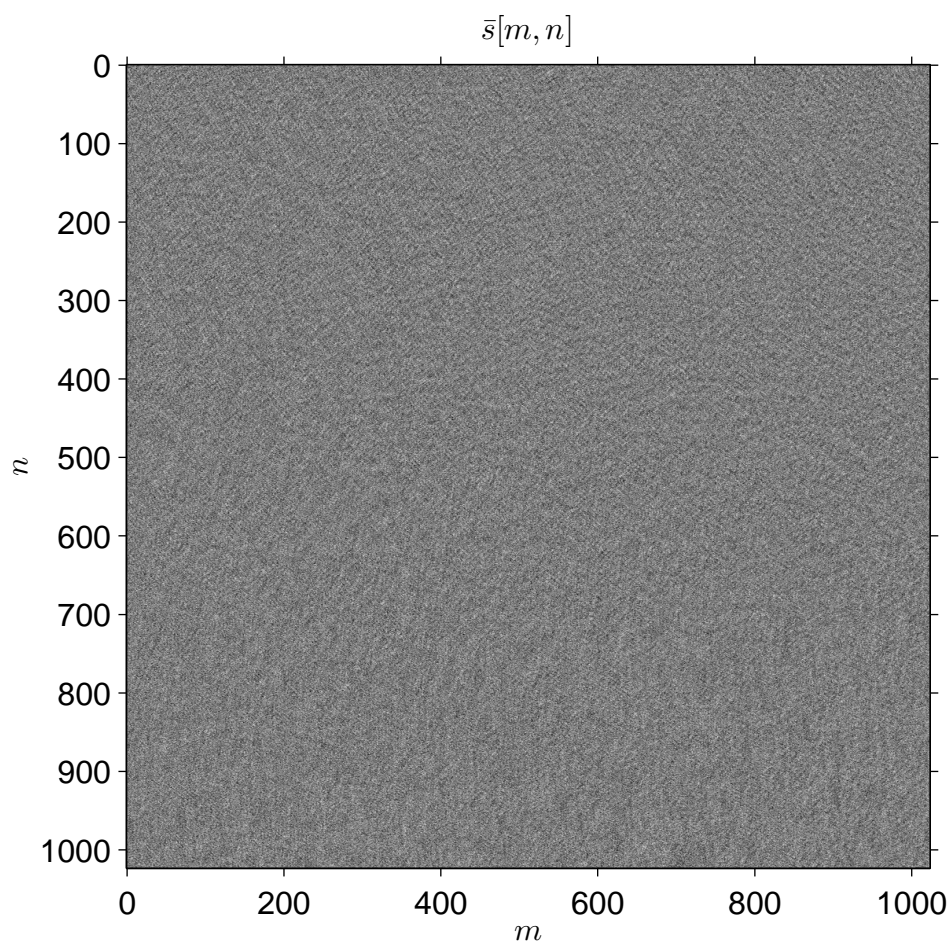


Figure 3.3: An SLM pattern (real part). First appeared in [1]

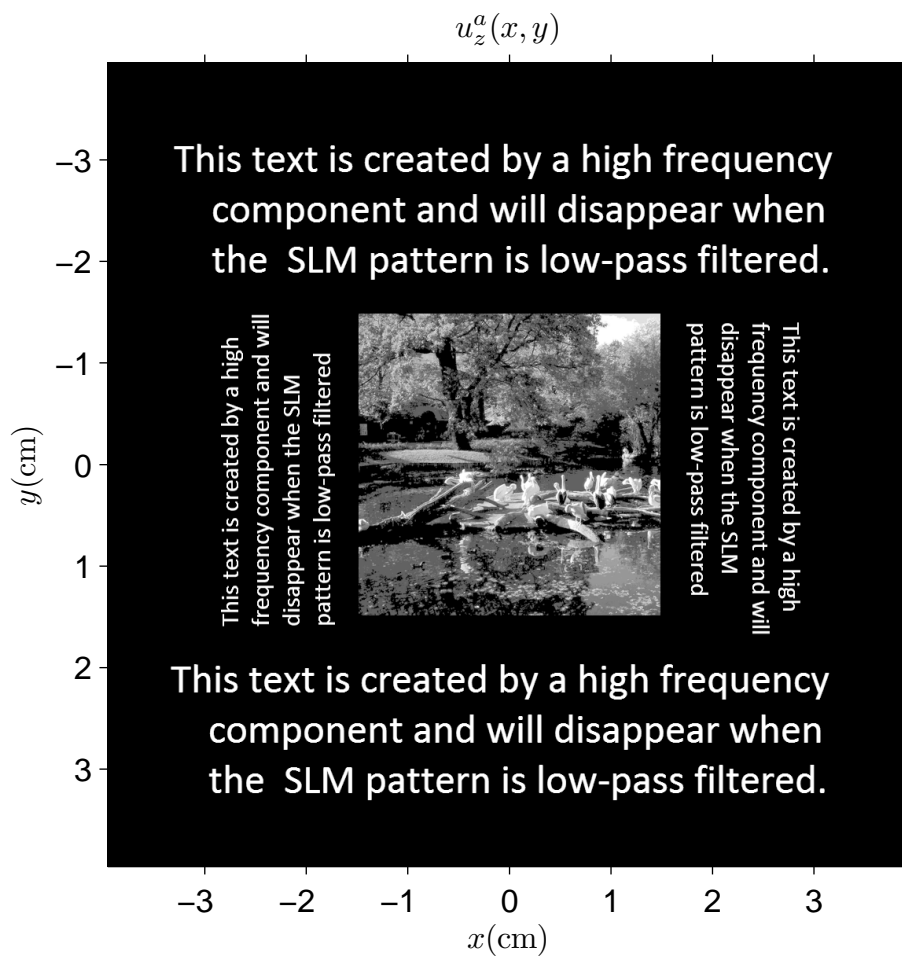


Figure 3.4: Output produced by the SLM pattern in Fig. 3.3 (magnitude). First appeared in [1].

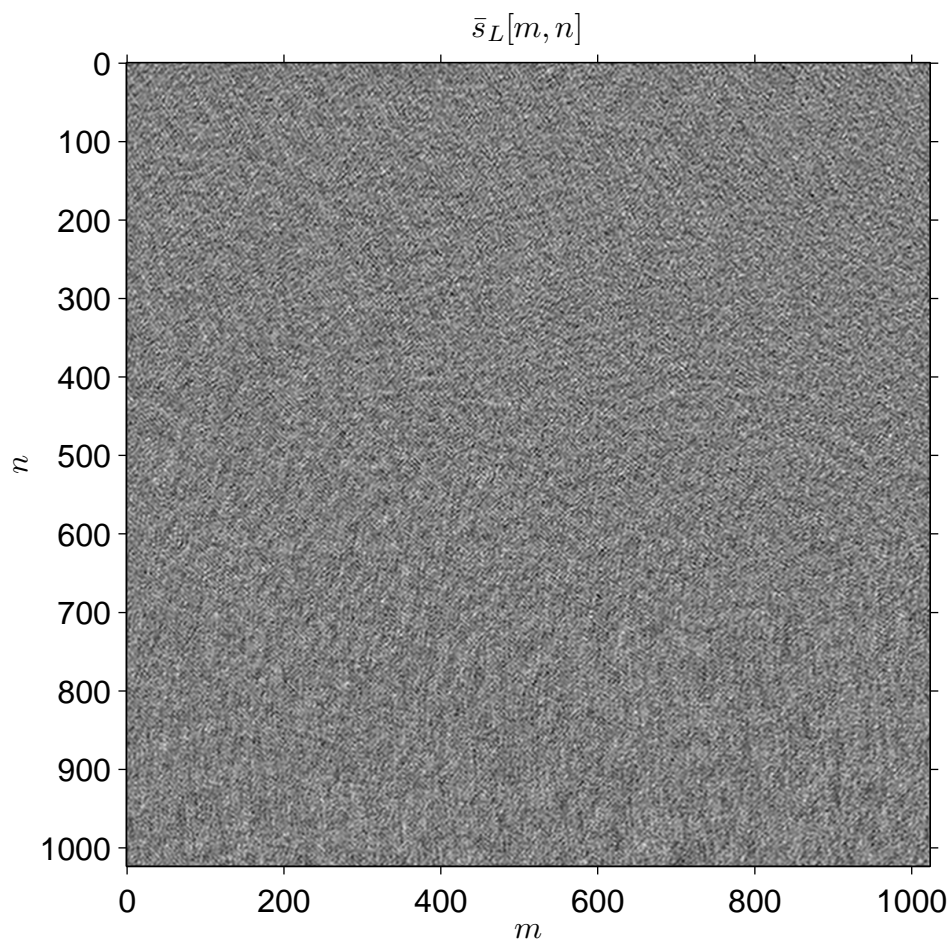


Figure 3.5: Low pass filtered version of the SLM pattern in Fig. 3.3 (real part). First appeared in [1].

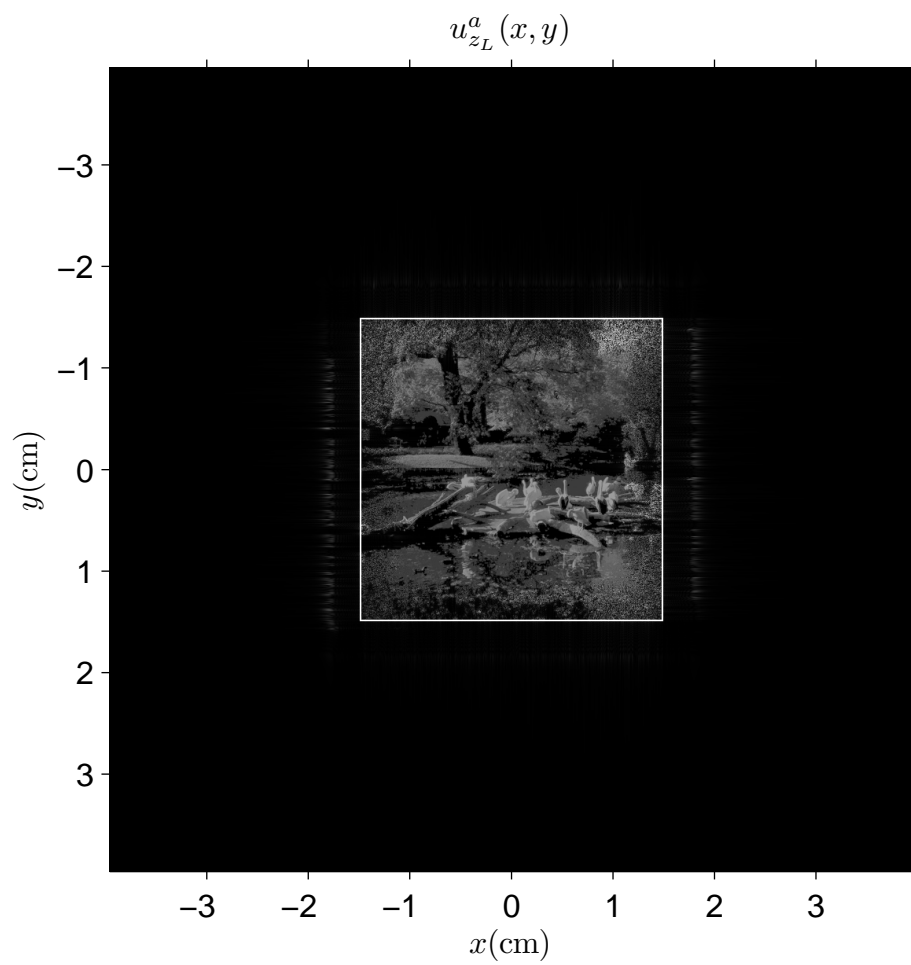


Figure 3.6: Output produced by the SLM pattern in Fig. 3.5 (magnitude). First appeared in [1].

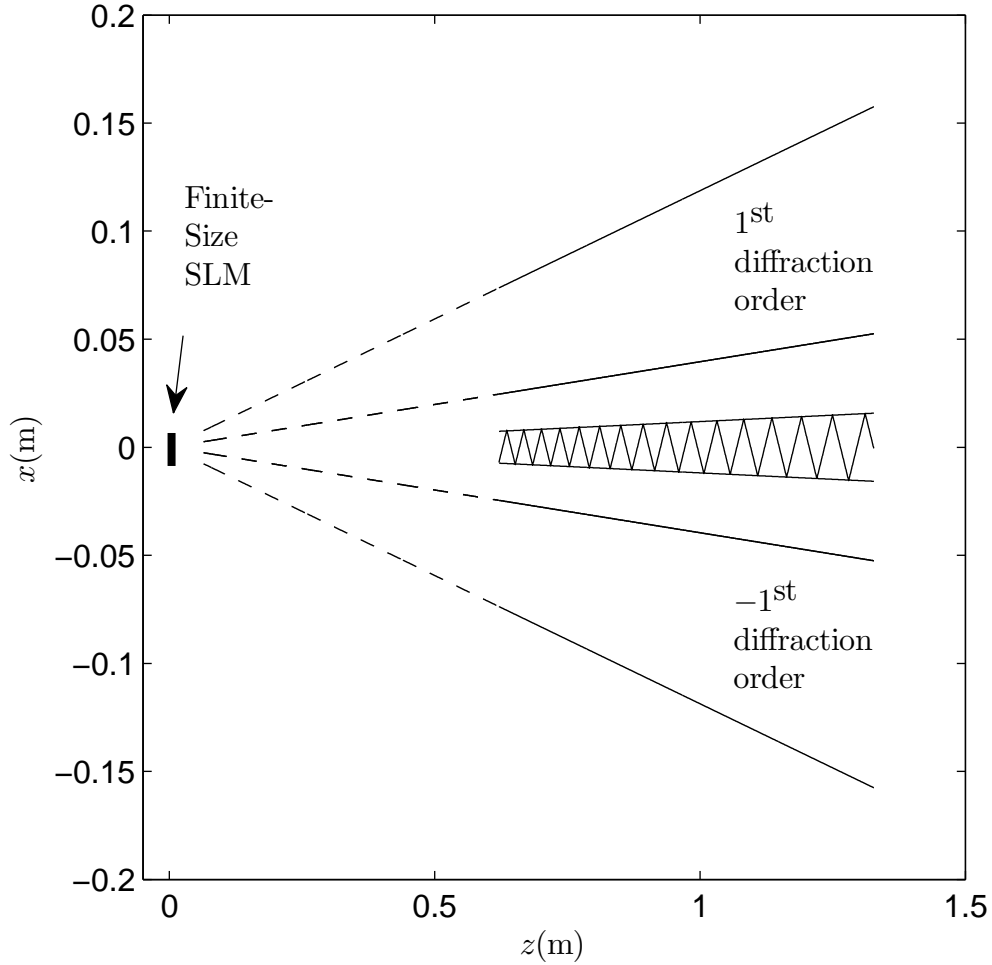


Figure 3.7: Synthesis region (dashed region).

In the next section, we exploit the result stated by Eq. 3.15 to find binary SLM patterns that generate desired light fields confined to the region specified in Eq. 3.12 for appropriate selections of the parameters B_x and B_y . For a single z , Eq. 3.12 defines a rectangular region, which we will call as the synthesis region. For varying z , Eq. 3.12 defines a pyramid lying inside the central diffraction order pyramid and we will call it as the synthesis pyramid (see Fig. 3.7). However, since the approximations we made in arriving at Eq. 3.12 hold only after a certain distance from the SLM (approximately a meter), we will not be interested in the portion of the synthesis pyramid that lies too close to the SLM (recall from Sec. 2.3 that we do the same for the central diffraction order pyramid as well).

The dimensions of the synthesis region increase as B_x and B_y increase. Therefore, as the low-pass effect of the filter gets stronger, the region that the stated result holds gets narrower. On the other side, as bandwidths approach the upper limits allowed by Eq. 3.5 ($B_x \rightarrow \frac{1}{\Delta_x}$ and $B_y \rightarrow \frac{1}{\Delta_y}$), the borders approach the borders of the central diffraction order given in Eq. 2.38. When light fields are to be synthesized with binary SLMs, a rational choice is to take $B_x \approx \frac{1}{4\Delta_x}$ and $B_y \approx \frac{1}{4\Delta_y}$ so that the area of the synthesis region is about $\frac{1}{16}$ th of the area of the central diffraction order region.

3.2 Encoding Complex-Valued Oversampled Holograms on Binary SLMs

Now we have all the necessary insights to tackle the main problem of this chapter: How can we find a binary SLM pattern that synthesizes a desired three-dimensional light field? Findings of the previous section will guide us.

First consider a gray level SLM pattern $\bar{s}[m, n]$ of size $M \times N$, that is, $\bar{s}[m, n]$ can be equal to any complex value for $m \in [0, M - 1]$ and $n \in [0, N - 1]$, but $\bar{s}[m, n] = 0$ for other (m, n) . Suppose we write this pattern on an SLM that has $M \times N$ pixels, and suppose we illuminate the SLM with a normally incident plane wave. Let $\bar{u}_z^a[m, n]$ for $m \in [0, M - 1]$ and $n \in [0, N - 1]$ denote the samples of the output field taken uniformly within the central diffraction order region. We know from Sec. 2.3 that the relation between $\bar{u}_z^a[m, n]$ and $\bar{s}[m, n]$ is as given by Eq. 2.42.

Now suppose we wish to determine $\bar{s}[m, n]$ such that the output samples within the synthesis region specified by Eq. 3.12 are equal to some desired discrete signal $\bar{d}[m, n]$. In this section, it is important for the synthesis region to be sufficiently small, and the reason will become evident as we continue. For now,

let us simply assume that B_x and B_y parameters in Eq. 3.12 are chosen as $B_x = \frac{1}{4\Delta_x}$ and $B_y = \frac{1}{4\Delta_y}$ so that the area of the synthesis region is about $\frac{1}{16}$ th of the area of the central diffraction order region. Hence, we wish to control about $\frac{1}{16}$ th of the output field samples. Suppose we are not concerned about the values of output samples that fall outside the synthesis region, i.e., they are “don’t care” samples. Let us call the part of central diffraction order outside the synthesis region as the don’t care region. Since the number of samples that we wish to control is less than the degrees of freedom that we have in $\bar{s}[m, n]$, this problem does not have a unique solution but has many solutions. An easy solution can be found simply by setting

$$\bar{u}_z^a[m, n] = \begin{cases} \bar{d}[m, n] & \text{for samples within the synthesis region} \\ 0 & \text{for samples within the don't care region} \end{cases} \quad (3.16)$$

and then finding $\bar{s}[m, n]$ according to Eq. 2.45. Let us denote this particular solution as $\bar{s}^i[m, n]$ and call it the ideal SLM pattern since it produces $\bar{d}[m, n]$ within the synthesis region with maximum efficiency, i.e., output samples within the don’t care region are zero so essentially no power is spent on the don’t care region.

The ideal SLM pattern $\bar{s}^i[m, n]$ found as above is a low-pass (oversampled) discrete signal. To see this, suppose we apply a real-valued discrete low-pass filter $\bar{g}[m, n]$ with bandwidth parameters $B_x = \frac{1}{4\Delta_x}$ and $B_y = \frac{1}{4\Delta_y}$ to $\bar{s}^i[m, n]$. As shown in the previous section, at the output, nothing will change in the synthesis region, and the field will vanish in the don’t care region. But the field produced by $\bar{s}^i[m, n]$ already vanishes in the don’t care region. Therefore, we should have $\bar{s}^i[m, n] \star \bar{g}[m, n] \approx \bar{s}^i[m, n]$, indicating that $\bar{s}^i[m, n]$ is the output of a low-pass filter, hence it is a low-pass (oversampled) SLM pattern.

Now let us try to reconstruct $\bar{d}[m, n]$ within the same synthesis region with a binary SLM pattern $\bar{s}_b[m, n]$ whose size is again $M \times N$. Let us assume for now that $\bar{s}_b[m, n] = \pm 1$ for $m \in [0, M - 1]$ and $n \in [0, N - 1]$, but $\bar{s}_b[m, n] = 0$

for other (m, n) . We now have a harsh constraint on the SLM pattern. Solving this problem is not as straightforward as for the gray level $\bar{s}[m, n]$. For instance, it is common experience that direct point-wise quantization of $\bar{s}^i[m, n]$ does not produce satisfactory results. Smarter strategies are necessary.

From the previous section, we know that $\bar{s}_b[m, n]$ and its low-pass filtered version $\bar{s}_b^L[m, n] = \bar{s}_b[m, n] \star \bar{g}[m, n]$ produce approximately the same field within the synthesis region. Hence, if $\bar{s}_b[m, n]$ is to produce $\bar{d}[m, n]$ within the synthesis region, $\bar{s}_b^L[m, n]$ should also produce $\bar{d}[m, n]$ within the synthesis region. In particular, suppose we find $\bar{s}_b[m, n]$ such that $\bar{s}_b^L[m, n]$ is equal to $\bar{s}^i[m, n]$. (This is possible since $\bar{s}^i[m, n]$ is already a low-pass pattern as we discussed above.) Since $\bar{s}^i[m, n]$ already generates the desired field within the synthesis region, we see that $\bar{s}_b[m, n]$ performs the desired synthesis as well. Therefore, if we can find $\bar{s}_b[m, n]$ such that

$$\bar{s}_b[m, n] \star \bar{g}[m, n] \approx \bar{s}^i[m, n] \quad (3.17)$$

we can achieve the desired synthesis.

Above, we use \approx instead of $=$ for two reasons. The first reason is to stress that in general, the problem may not have an exact solution, i.e, there may be no $\bar{s}_b[m, n]$ that exactly gives $\bar{s}^i[m, n]$ when low-pass filtered, so we may need to seek for the best solution instead of an exact solution. The second reason is, as also indicated in Sec. 3.1, according to our definitions $\bar{s}^i[m, n]$ and $\bar{s}_b[m, n]$ are finite sized patterns of size $M \times N$, but strictly speaking $\bar{s}_b[m, n] \star \bar{g}[m, n]$ is not because of the infinite extent tails of low-pass filter $\bar{g}[m, n]$. Therefore, it is not possible to strictly achieve $\bar{s}_b[m, n] \star \bar{g}[m, n] = \bar{s}^i[m, n]$. However, again as we indicated in Sec. 3.1, since $\bar{g}[m, n]$ has a compact support around $(m, n) = (0, 0)$, we will assume that it is sufficient for Eq. 3.17 to hold only over the support of $\bar{s}^i[m, n]$ and $\bar{s}_b[m, n]$, i.e., for $m \in [0, M - 1]$ and $n \in [0, N - 1]$.

From the discussion preceding Eq. 3.17, we can realize that actually the low-pass component of $\bar{s}_b[m, n]$ (that is, $\bar{s}_b^L[m, n]$) is responsible for generating $\bar{d}[m, n]$

within the synthesis region. The high-pass component $(\bar{s}_b[m, n] - \bar{s}_b^L[m, n])$ affects only the output samples in the don't care region. Note that when $\bar{g}[m, n]$ is applied to $\bar{s}_b[m, n]$, the high-pass component is eliminated, so nothing is generated within the don't care region. Actually, we can think that high-pass component is added to $\bar{s}_b^L[m, n]$ just to satisfy the binary SLM pattern constraint.

We recognize that the problem stated in Eq. 3.17 is quite similar to the well known halftoning problem of image processing in which one tries to compute a binary image which produces a desired low-pass (oversampled) gray level image when low-pass filtered. However, there is a slight difference in general. Notice that in Eq. 3.17, with our definitions, both $\bar{s}_b[m, n]$ and $\bar{g}[m, n]$ are real-valued, hence their convolution is also real-valued. If $\bar{s}^i[m, n]$ is also real-valued, then there is no difference with the classical halftoning problem and we can easily find $\bar{s}_b[m, n]$ using any of the well-established halftoning algorithms, such as ordered dither, error diffusion, direct binary search, etc [144]. The problem is that, in general, $\bar{s}^i[m, n]$ is complex-valued. Hence, the problem stated in Eq. 3.17 is not directly equivalent to a classical halftoning problem, and we need to find a method for handling the case of a complex-valued $\bar{s}^i[m, n]$.

The solution we propose in this thesis is to partition the pixels of $\bar{s}_b[m, n]$ into two groups, such that the first group pixels are responsible for halftoning the real part of $\bar{s}^i[m, n]$ and the second group pixels are responsible for halftoning the imaginary part of $\bar{s}^i[m, n]$. In the next three subsections, we propose three successively easier methods to achieve such partitions.

3.2.1 Thin Mask Based Solution

As a first solution, suppose that we place an optical thin mask with complex transmittance $t(x, y)$ just after the SLM such that $t(x, y)$ has the following form:

$$t(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{t}[m, n] \text{rect} \left(\frac{x - x_m}{\Delta_x} \right) \text{rect} \left(\frac{y - y_n}{\Delta_y} \right). \quad (3.18)$$

If we still assume a normally incident plane wave illumination ($u_{0-}(x, y) = 1$), then the light field just after the SLM can be written as:

$$u_{0+}(x, y) = t(x, y) s^a(x, y). \quad (3.19)$$

Using Eq. 2.22 and Eq. 3.18, we can explicitly write

$$\begin{aligned} u_{0+}(x, y) &= t(x, y) s^a(x, y) \\ &= \left\{ \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{t}[m, n] \text{rect} \left(\frac{x - x_m}{\Delta_x} \right) \text{rect} \left(\frac{y - y_n}{\Delta_y} \right) \right\} \\ &\quad \times \left\{ \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \bar{s}[m', n'] a(x - x_{m'}, y - y_{n'}) \right\} \\ &= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \bar{t}[m, n] \bar{s}[m', n'] \\ &\quad \text{rect} \left(\frac{x - x_m}{\Delta_x} \right) \text{rect} \left(\frac{y - y_n}{\Delta_y} \right) a(x - x_{m'}, y - y_{n'}). \end{aligned} \quad (3.20)$$

Noting that $a(x, y) = 0$ for $|x| > \frac{\Delta_x}{2}$ or $|y| > \frac{\Delta_y}{2}$, we have

$$\begin{aligned} &\text{rect} \left(\frac{x - x_m}{\Delta_x} \right) \text{rect} \left(\frac{y - y_n}{\Delta_y} \right) a(x - x_{m'}, y - y_{n'}) \\ &= \begin{cases} a(x - x_{m'}, y - y_{n'}) & \text{when } m = m', n = n' \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

from which we can rewrite Eq. 3.20 as:

$$u_{0+}(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n] \bar{t}[m, n] a(x - x_m, y - y_n). \quad (3.21)$$

Upon comparing Eq. 2.22 and Eq. 3.21, we see that placing the mask $t(x, y)$ is equivalent to writing on the SLM a new pattern $\bar{s}_t[m, n]$, which is given as

$$\bar{s}_t[m, n] = \bar{s}_b[m, n] \bar{t}[m, n], \quad (3.22)$$

which we call the effective SLM pattern.

Now suppose $\bar{t}[m, n]$ is given as:

$$\bar{t}[m, n] = \begin{cases} 1 & \text{when } m + n \text{ is even} \\ j & \text{when } m + n \text{ is odd} \end{cases}. \quad (3.23)$$

With this selection for $\bar{t}[m, n]$, we now have the following constraint on $\bar{s}_t[m, n]$:

$$\bar{s}_t[m, n] = \begin{cases} \pm 1 & \text{when } m + n \text{ is even} \\ \pm j & \text{when } m + n \text{ is odd} \end{cases}. \quad (3.24)$$

Therefore, though $\bar{s}_b[m, n]$ can only be adjusted to real binary values, i.e. ± 1 , we have more flexibility in $\bar{s}_t[m, n]$ since the even pixels of it can be adjusted to ± 1 while odd pixels can be adjusted to $\pm j$.

Note that we can write $\bar{s}_t[m, n]$ as $\bar{s}_t[m, n] = \bar{s}_t^R[m, n] + j\bar{s}_t^I[m, n]$ where

$$\begin{aligned} \bar{s}_t^R[m, n] &= \begin{cases} \pm 1 & \text{when } m + n \text{ is even} \\ 0 & \text{when } m + n \text{ is odd} \end{cases} \\ \bar{s}_t^I[m, n] &= \begin{cases} 0 & \text{when } m + n \text{ is even} \\ \pm 1 & \text{when } m + n \text{ is odd} \end{cases}. \end{aligned} \quad (3.25)$$

Actually, $\bar{s}_t^R[m, n]$ and $\bar{s}_t^I[m, n]$ can be considered as binary SLM patterns some of whose pixels are eliminated. In particular, odd pixels (pixels for which $m + n$ is odd) of $\bar{s}_t^R[m, n]$ and even pixels (pixels for which $m + n$ is even) of $\bar{s}_t^I[m, n]$ are eliminated.

Now, we can encode the real part of $\bar{s}^i[m, n]$ on $\bar{s}_t^R[m, n]$ and the imaginary part on $\bar{s}_t^I[m, n]$. In other words, we can try to find $\bar{s}_t^R[m, n]$ and $\bar{s}_t^I[m, n]$ such that

$$\begin{aligned} \bar{s}_t^R[m, n] \star \bar{g}[m, n] &\approx \mathcal{R} \{ \bar{s}^i[m, n] \} \\ \bar{s}_t^I[m, n] \star \bar{g}[m, n] &\approx \mathcal{I} \{ \bar{s}^i[m, n] \} \end{aligned} \quad (3.26)$$

where $\bar{s}_t^R[m, n]$ and $\bar{s}_t^I[m, n]$ are subject to the constraints of Eq. 3.25. Once we find $\bar{s}_t^R[m, n]$ and $\bar{s}_t^I[m, n]$, we can add them up to obtain a complete binary SLM

pattern such that $\bar{s}_b[m, n] = \bar{s}_t^R[m, n] + \bar{s}_t^I[m, n]$. This binary pattern performs the desired synthesis when written on an SLM just after which the mask $t(x, y)$ is placed. Note that in this manner, we have converted the complex-valued halftoning problem of Eq. 3.17 (which was problematic in that form) into two decoupled real-valued constrained halftoning problems as in Eq. 3.26 (which can now easily be solved with standard halftoning algorithms).

The eliminated pixel constraints on $\bar{s}_t^R[m, n]$ and $\bar{s}_t^I[m, n]$ given in Eq. 3.25 will not cause a significant halftoning error if $\bar{s}^i[m, n]$ (hence its real and imaginary components) is sufficiently low-pass. This is the case if the synthesis region is selected sufficiently small.

Let us illustrate these ideas. Suppose Fig. 3.8 shows the desired field within the central diffraction order region. Only the samples within the synthesis region are nonzero. Fig. 3.9 and Fig. 3.10 respectively show the real and imaginary parts of the ideal gray level SLM pattern $\bar{s}^i[m, n]$ that exactly reconstructs the desired field of Fig. 3.8. As discussed above, the ideal gray level SLM pattern is a low-pass (oversampled) discrete signal. (In this example, the SLM size is 1024×1024 , $\Delta_x, \Delta_y = 8\mu m$, $\lambda = 632.9nm$ and $z = 1m$; so the physical size of the SLM is $8.2mm \times 8.2mm$ and the physical size of the central diffraction order region is $7.91cm \times 7.91cm$. The synthesis region consists of 200×200 samples which corresponds to a physical size of $1.55cm \times 1.55cm$.)

Next, we considered the computation of a binary pattern that generates the desired field. As the halftoning algorithm, we used the standard error diffusion algorithm [144]. As in Eq. 3.25 and Eq. 3.26, we separately halftoned the real and imaginary parts of $\bar{s}^i[m, n]$ and computed $\bar{s}_t^R[m, n]$ and $\bar{s}_t^I[m, n]$, which are displayed in Fig. 3.11 and Fig. 3.12, respectively (black pixels have value -1 , gray pixels have value 0 , and white pixels have value 1). Note that as imposed by Eq. 3.25, in Fig. 3.11, odd pixels are 0 while even pixels are either -1 or 1 . Similarly, in Fig. 3.12, even pixels are 0 while odd pixels are either -1 or 1 .

Fig. 3.13 shows the binary pattern obtained as $\bar{s}_b[m, n] = \bar{s}_t^R[m, n] + \bar{s}_t^I[m, n]$. Note that this pattern contains only -1 or 1 ; as it should. Next we multiply this pattern by $\bar{t}[m, n]$ given by Eq. 3.23. The resulting output field is shown in Fig. 3.14. As seen, the desired field is generated successfully within the synthesis region, where we also see the noise components that appear in the don't care region due to high-pass component of $\bar{s}_b[m, n]$. We can see that the signal beam and the noise beam are well separated in space.

In the solution presented in this subsection, we obtained imaginary values using a mask placed just after the SLM. This option theoretically works, but complicates the optical setup in the sense that the mask should be physically produced and aligned properly with the SLM. Actually, now we are aware of the fact that the main purpose of using the mask is to obtain the effective SLM pattern of Eq. 3.24. But this pattern can be obtained with alternative physical arrangements. In the next subsection, we will consider an easier alternative.

3.2.2 Oblique Illumination Based Solution

In this subsection, we will not use any mask as in the previous subsection, but instead illuminate the SLM with an oblique (or tilted) plane wave instead of a normally incident plane wave. We will show that with a clever selection for the incidence angle of the oblique wave, it is possible to effectively accomplish the task managed by the mask of the previous subsection. We will start by examining the effects of tilted illumination on the SLM output.

We derived all the results in Sec. 2.3 (in which we analyzed the light field generated by an SLM) by assuming that the illumination wave of the SLM is normally incident. Now, instead of a normally incident wave, suppose the illumination wave is given as

$$i(x, y) = I_0 \exp \{ j2\pi (\nu_x^0 x + \nu_y^0 y) \} \quad (3.27)$$

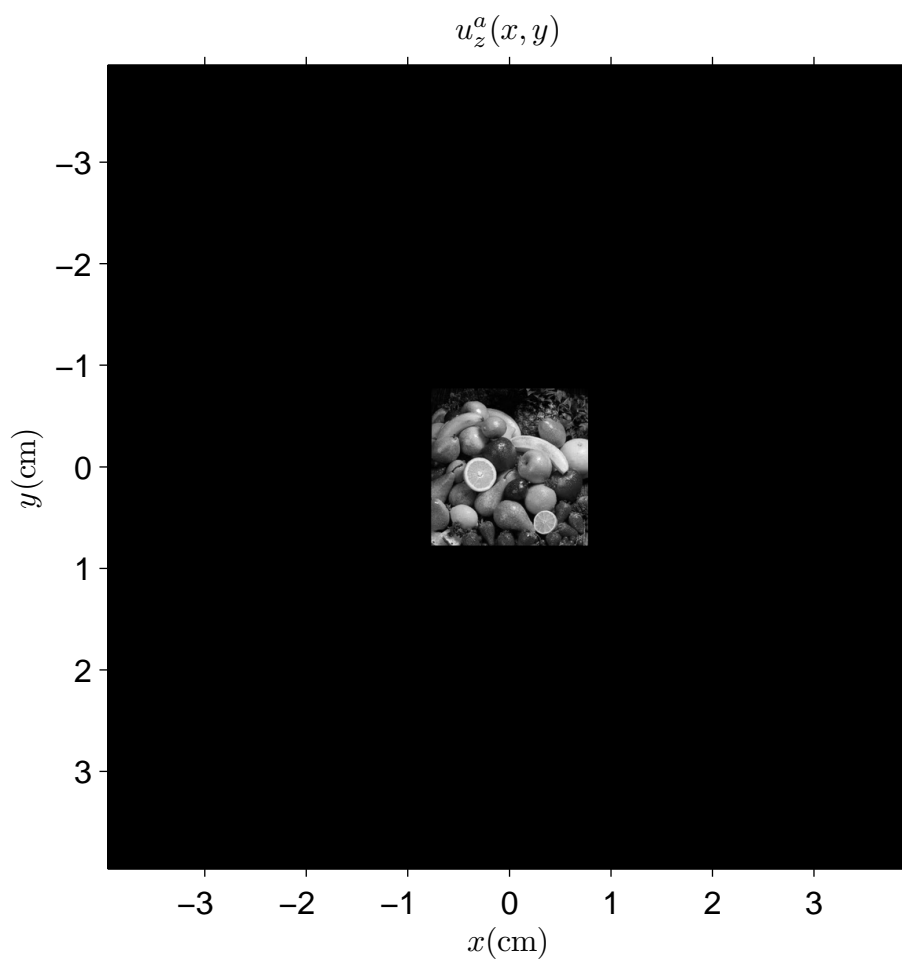


Figure 3.8: Desired light field (magnitude). First appeared in [1].

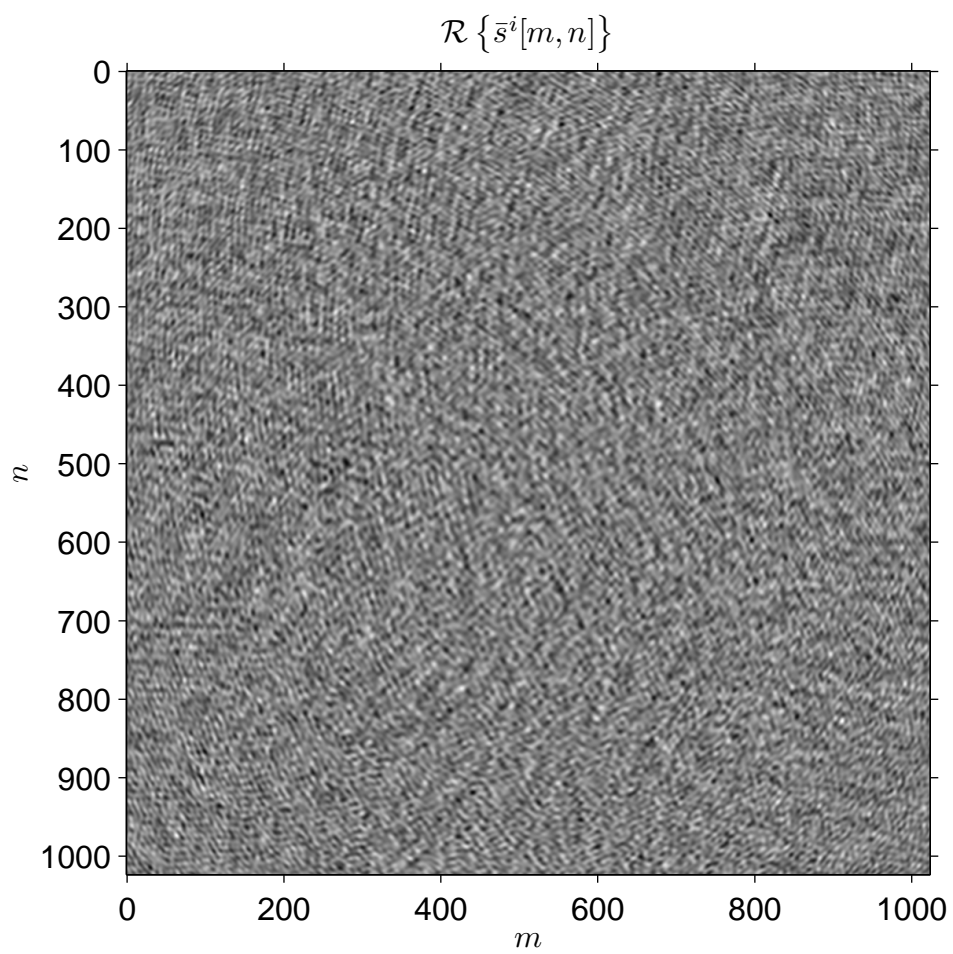


Figure 3.9: Ideal SLM pattern (real part). First appeared in [1].

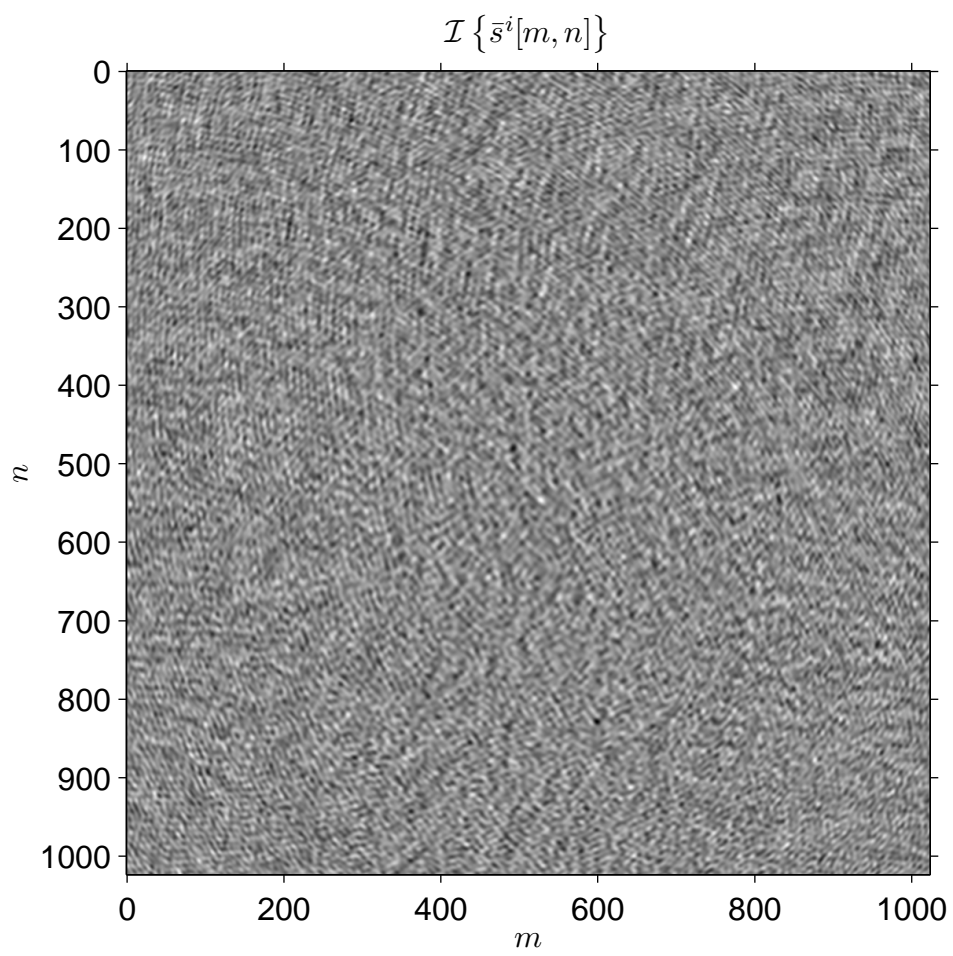


Figure 3.10: Ideal SLM pattern (imaginary part).

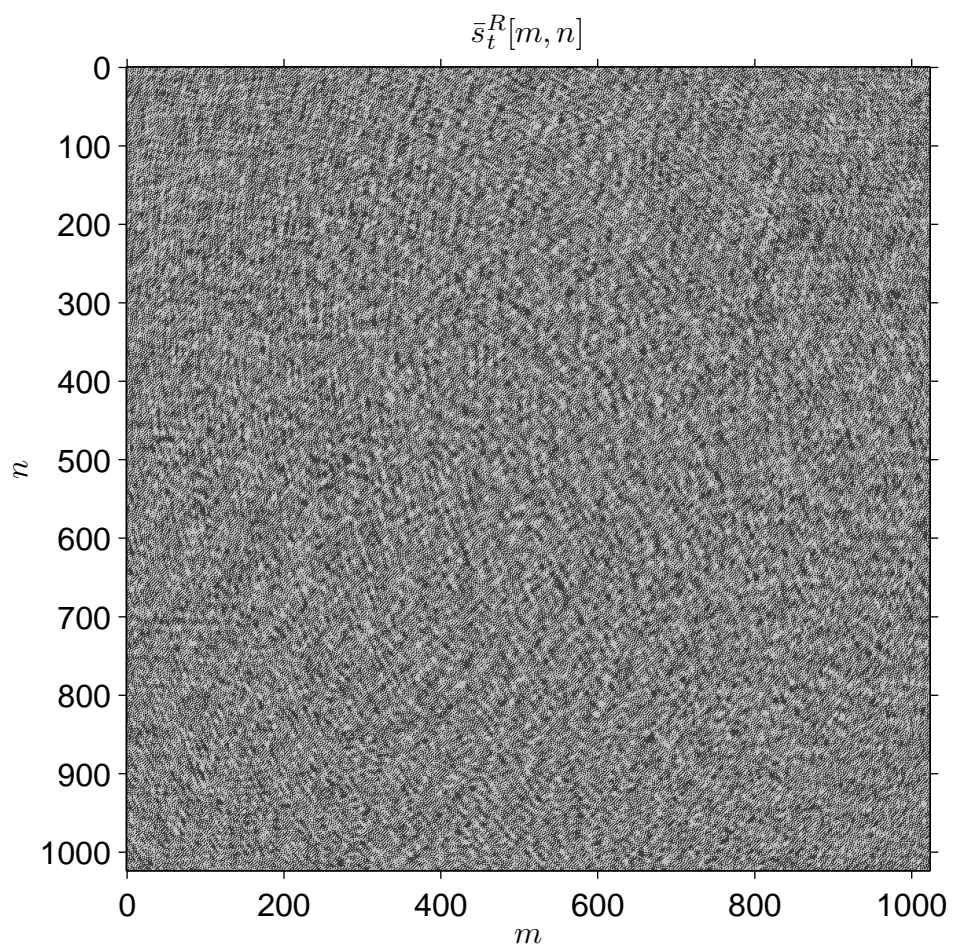


Figure 3.11: Three level SLM pattern for the real part. First appeared in [1].

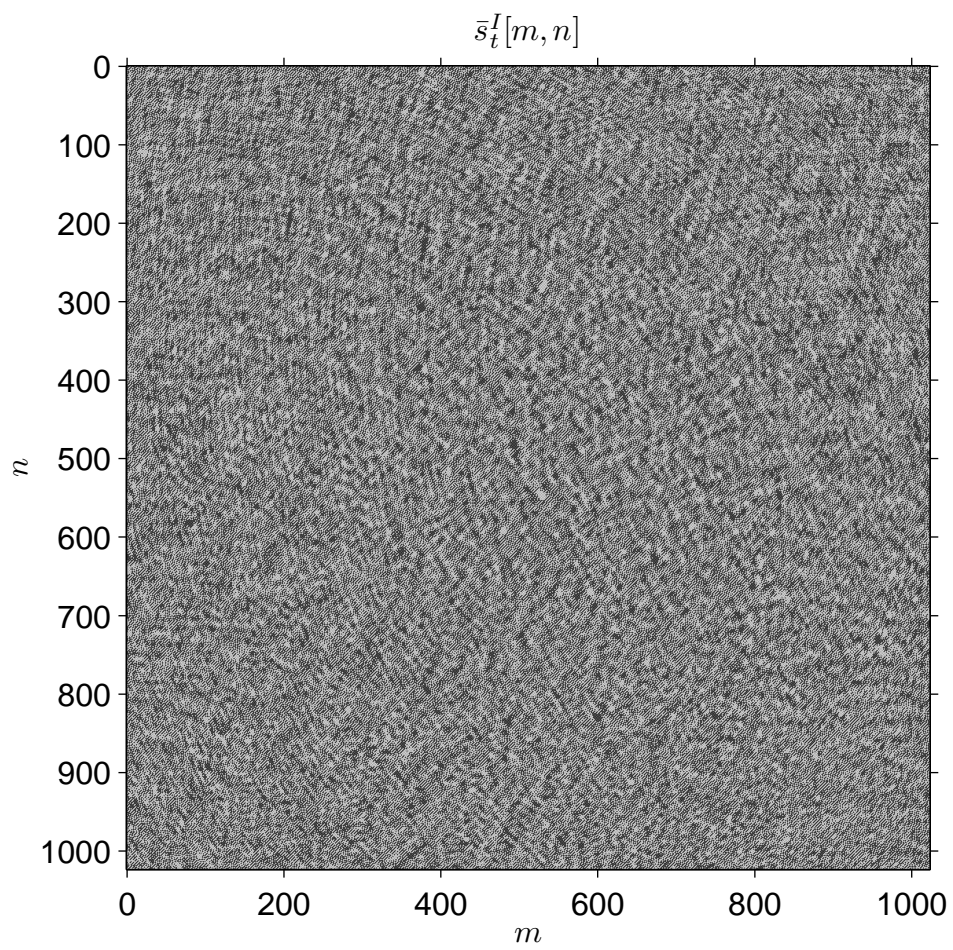


Figure 3.12: Three level SLM pattern for the imaginary part.

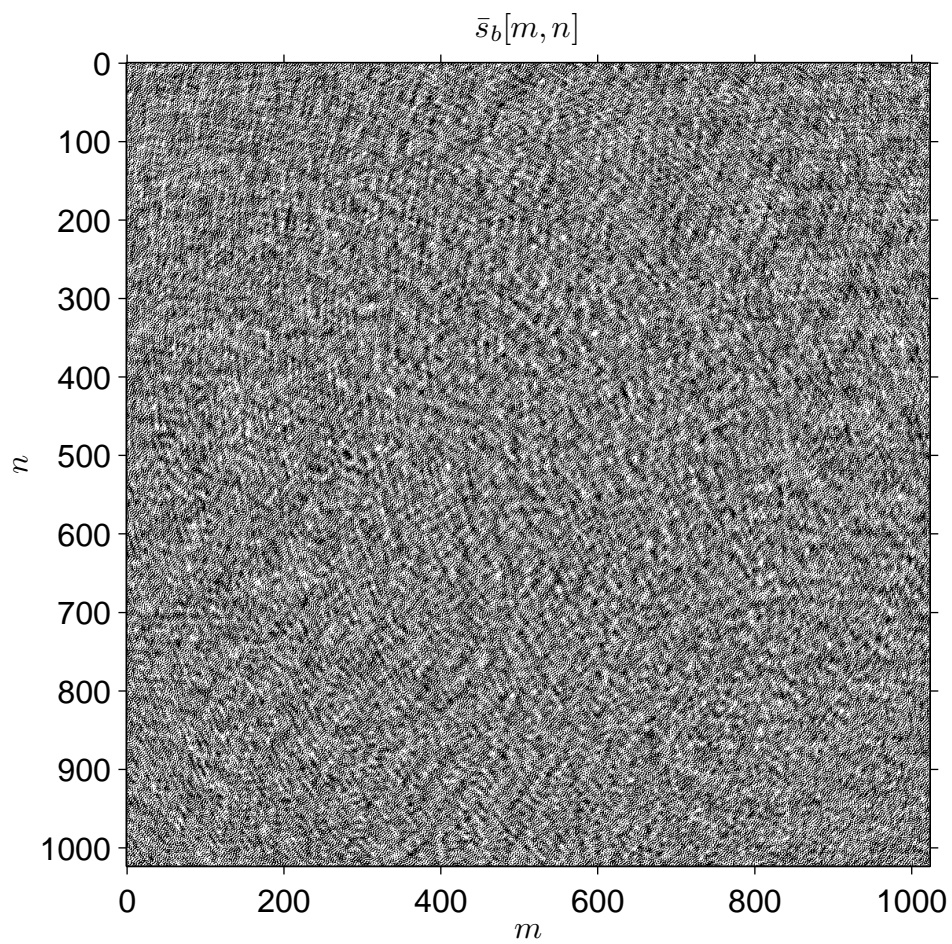


Figure 3.13: Binary SLM pattern obtained by adding the three level SLM patterns in Fig. 3.11 and Fig. 3.12. First appeared in [1].

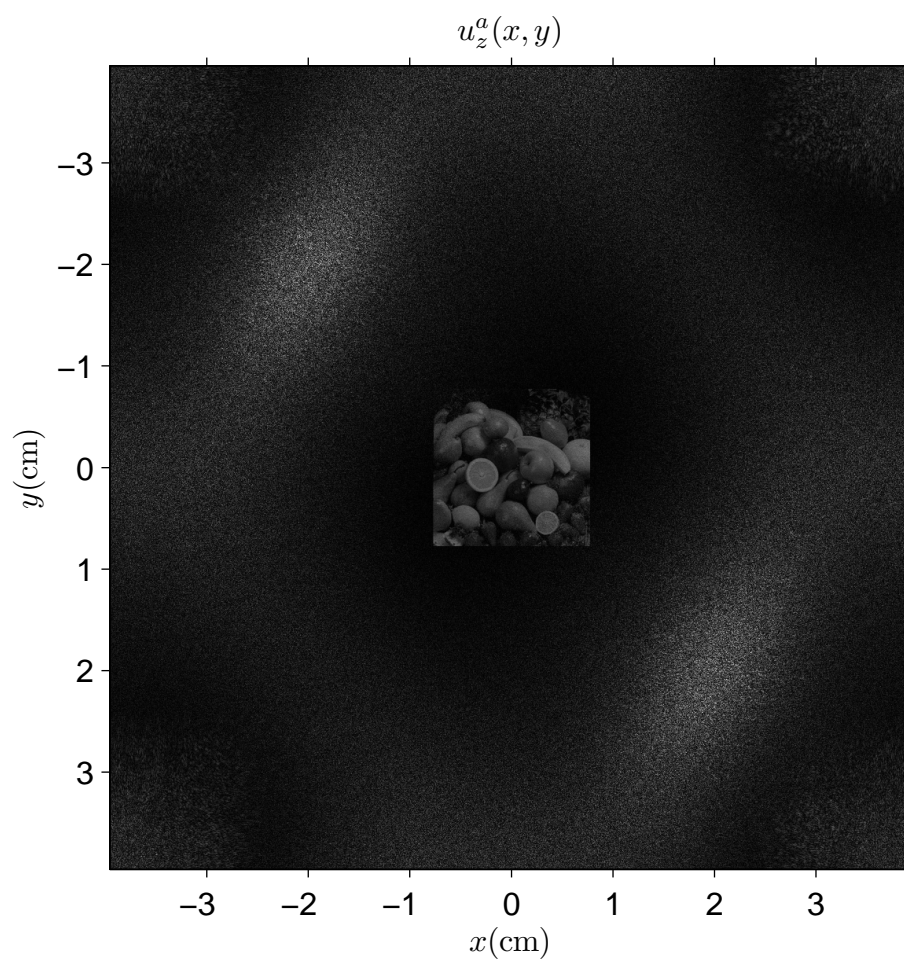


Figure 3.14: Light field generated by the binary SLM pattern in Fig. 3.13. First appeared in [1].

which is an oblique wave (of course unless $\nu_x^0 = \nu_y^0 = 0$). In this case, the field on the $z = 0+$ plane can be written as

$$\begin{aligned}
u_{0+}(x, y) &= i(x, y)s^a(x, y) \\
&= i(x, y) \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n]a(x - x_m, y - y_n) \\
&= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n]i(x, y)a(x - x_m, y - y_n) \\
&= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}[m, n]I_0 \exp \{j2\pi (\nu_x^0 x + \nu_y^0 y)\} \\
&\quad a(x - x_m, y - y_n) \\
&= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_0 \bar{s}[m, n] \exp \{j2\pi (\nu_x^0 x_m + \nu_y^0 y_n)\} \\
&\quad a(x - x_m, y - y_n) \\
&\quad \exp \{j2\pi (\nu_x^0 (x - x_m) + \nu_y^0 (y - y_n))\} \\
&= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{s}_1[m, n]a_1(x - x_m, y - y_n) \tag{3.28}
\end{aligned}$$

with

$$\bar{s}_1[m, n] = I_0 \bar{s}[m, n] \exp \{j2\pi (\nu_x^0 x_m + \nu_y^0 y_n)\} \tag{3.29}$$

and

$$a_1(x, y) = a(x, y) \exp \{j2\pi (\nu_x^0 x + \nu_y^0 y)\}. \tag{3.30}$$

Therefore, oblique illumination of an SLM whose SLM pattern is $\bar{s}[m, n]$ and aperture function is $a(x, y)$ is equivalent to a new SLM (which is illuminated with a normally incident wave) whose SLM pattern is $\bar{s}_1[m, n]$ and whose aperture function is $a_1(x, y)$. Let us now examine the meaning of these modifications.

Recall that until the very end of Sec. 2.3, we neglected the effect of a practical pixel aperture function (that is, we assumed $a(x, y) = \delta(x, y)$). We explained at the end of Sec. 2.3 that the modifications that we should make to the derived results due to a practical (finite extent) $a(x, y)$ can be neglected as long as we are only interested in the central diffraction order, and if $a(x, y)$ has a sufficiently narrow spatial support such that $A(\nu_x, \nu_y)$ is approximately constant over the

$|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$ band. Now, in this respect, consider the new SLM whose pixel aperture function is $a_1(x, y)$. If $a(x, y)$ is narrow, $a_1(x, y)$ is also narrow. Secondly, we can see from Eq. 3.30 that $A_1(\nu_x, \nu_y) = A(\nu_x - \nu_x^0, \nu_y - \nu_y^0)$. If ν_x^0 and ν_y^0 are sufficiently small (that is, the illumination wave is only slightly tilted) we can assume that $A_1(\nu_x, \nu_y)$ is also constant over the $|\nu_x| \leq \frac{1}{2\Delta_x}$ and $|\nu_y| \leq \frac{1}{2\Delta_y}$ band. In this case (and we will assume that this is the case), we can neglect the effects due to the modification of pixel aperture function from $a(x, y)$ to $a_1(x, y)$, as well.

Then, we can say that the only remaining effect of oblique illumination is that, the SLM pattern is changed from $\bar{s}[m, n]$ to $\bar{s}_1[m, n]$. That is, we can confidently use all the equations that we derived in Sec. 2.3 regarding the SLM output if we replace $\bar{s}[m, n]$ by $\bar{s}_1[m, n]$.

Having understood the effects of the oblique illumination, we now select the parameters of the illumination wave as

$$\nu_x^0 = \frac{1}{4\Delta_x}, \quad (3.31)$$

$$\nu_y^0 = \frac{1}{4\Delta_y} \quad (3.32)$$

and

$$I_0 = \exp \left\{ j \frac{\pi}{4} (M + N - 2) \right\}. \quad (3.33)$$

(Note that with these parameters, when practical values are considered for Δ_x , Δ_y and λ , the incidence angle of the corresponding wave - that is, the angle between the propagation direction of the wave and the normal vector of the SLM surface) becomes only a few degrees, which is in line with our assumption of slightly tilted illumination wave.) With the above selections, Eq. 3.29 becomes

$$\begin{aligned} \bar{s}_1[m, n] &= \bar{s}[m, n] \exp \left\{ j \frac{\pi}{2} (m + n) \right\} \\ &= \bar{s}[m, n] \bar{i}[m, n] \end{aligned} \quad (3.34)$$

where

$$\bar{i}[m, n] = \exp \left\{ j \frac{\pi}{2} (m + n) \right\}. \quad (3.35)$$

Now, suppose that $\bar{s}[m, n] = \bar{s}_{b_1}[m, n]$, that is, we write a binary pattern on the SLM. When we use the oblique illumination wave $i(x, y)$ with the selected parameters, the effective pattern on the SLM will become $\bar{s}_{b_1}[m, n]\bar{i}[m, n]$. Actually, we can easily see that $\bar{i}[m, n]$ is only a slightly different version of $\bar{t}[m, n]$ given by Eq. 3.23, and handling this difference is trivial. In fact, it is easy to show that $\frac{\bar{t}[m, n]}{\bar{i}[m, n]}$ is always -1 or 1 . Then, if we choose the new binary pattern $\bar{s}_{b_1}[m, n]$ as

$$\bar{s}_{b_1}[m, n] = \begin{cases} \bar{s}_b[m, n] & \text{when } \frac{\bar{t}[m, n]}{\bar{i}[m, n]} = 1 \\ -\bar{s}_b[m, n] & \text{when } \frac{\bar{t}[m, n]}{\bar{i}[m, n]} = -1 \end{cases} \quad (3.36)$$

(where $\bar{s}_b[m, n]$ denotes the binary pattern to be used with the thin mask of the previous subsection) we see that $\bar{s}_{b_1}[m, n]\bar{i}[m, n] = \bar{s}_b[m, n]\bar{t}[m, n]$. Therefore, in the oblique wave illumination option, we can first compute $\bar{s}_b[m, n]$ as in Eq. 3.25 and Eq. 3.26 (that is, as if we will use the mask $t(x, y)$), but then compute $\bar{s}_{b_1}[m, n]$ as in Eq. 3.36 and write it on the SLM which is illuminated by $i(x, y)$ (without using $t(x, y)$). At the end, we will get the same reconstruction.

Fig. 3.15 shows the updated SLM pattern obtained from the SLM pattern in Fig. 3.13. This SLM pattern also produces Fig. 3.14 when multiplied by $\bar{i}[m, n]$ of Eq. 3.35.

The oblique wave illumination option is much simpler than the mask based option since we do not need to produce any mask. However, at a first glance, it seems that we need to properly adjust the angle between the illumination wave and the SLM. In the next subsection, we will show that the mentioned proper adjustment is not that critical.

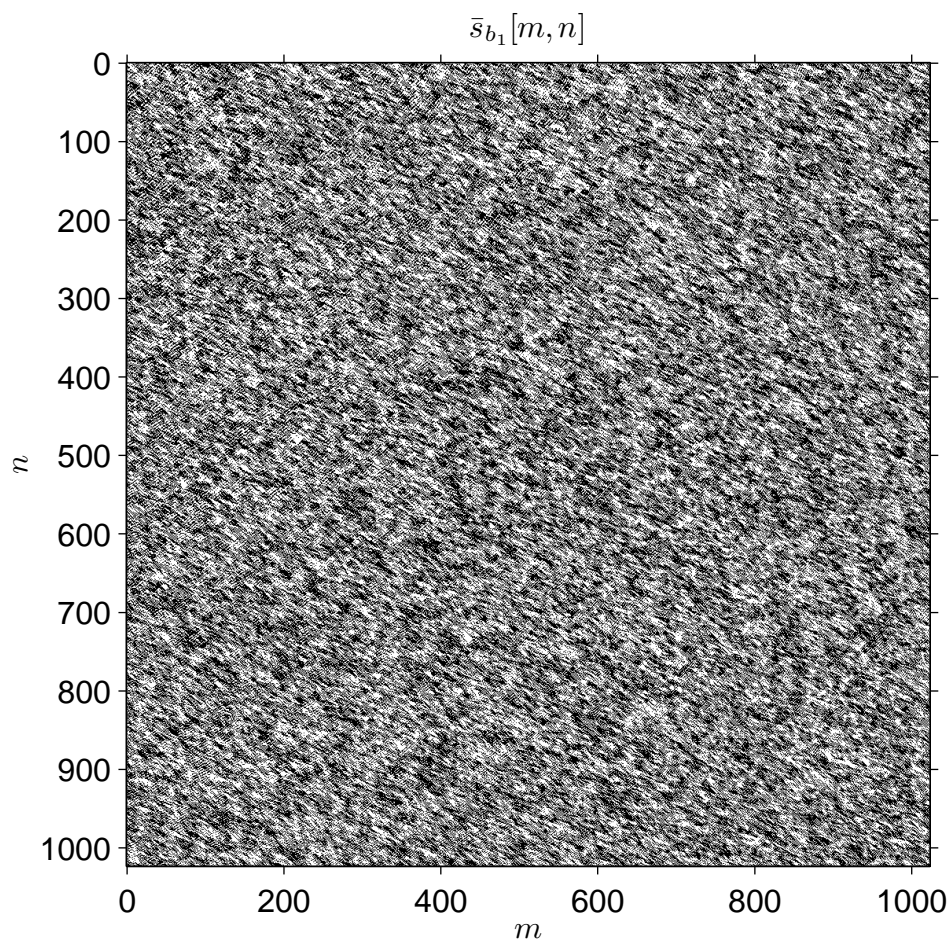


Figure 3.15: The new binary SLM pattern to be used with oblique illumination. First appeared in [1].

3.2.3 Small Deviations in the Incidence Angle of the Illumination Wave

In this subsection, we show that adjusting the incidence angle of the illumination wave used in the previous subsection is not critical at all. To understand this point, we will examine the effects of tilted illumination from another perspective.

As in Sec. 2.3, let $s^a(x, y)$ denote the complex transmittance of the SLM and $u_z^a(x, y)$ denote the diffraction field of the SLM when it is illuminated by normally incident illumination. Assuming that Fresnel diffraction model is valid, we can explicitly write

$$\begin{aligned}
u_z^a(x, y) &= h_z(x, y) ** s^a(x, y) \\
&= \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \\
&\quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s^a(x', y') e^{\frac{j\pi}{\lambda z}(x'^2+y'^2)} e^{-\frac{j2\pi}{\lambda z}(x'x+y'y)} dx' dy' \\
&= \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \mathcal{F} \left\{ s^a(x, y) e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \right\} \Big|_{(\nu_x, \nu_y) = (\frac{x}{\lambda z}, \frac{y}{\lambda z})}. \quad (3.37)
\end{aligned}$$

Let us now instead illuminate the SLM with $i(x, y)$ given by Eq. 3.27 with $I_0 = 1$ and let us call the new diffraction field as $u_z^{a,i}(x, y)$. Assuming that ν_x^0 and ν_y^0 are small enough that we can continue to use the Fresnel approximation, we can write

$$\begin{aligned}
u_z^{a,i}(x, y) &= h_z(x, y) ** \{s^a(x, y)i(x, y)\} \\
&= \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \mathcal{F} \left\{ s^a(x, y)i(x, y) e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \right\} \Big|_{(\nu_x, \nu_y) = (\frac{x}{\lambda z}, \frac{y}{\lambda z})} \\
&= \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \mathcal{F} \left\{ s^a(x, y) e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \right\} \Big|_{(\nu_x, \nu_y) = (\frac{x}{\lambda z} - \nu_x^0, \frac{y}{\lambda z} - \nu_y^0)} \quad (3.38)
\end{aligned}$$

where the last line follows from the well-known modulation property of Fourier transform. Then, from Eq. 3.37 and Eq. 3.38, we can write

$$u_z^{a,i}(x, y) = e^{-j\pi\lambda z(\nu_x^{02} + \nu_y^{02})} e^{j2\pi(\nu_x^0 x + \nu_y^0 y)} u_z^a(x - \nu_x^0 \lambda z, y - \nu_y^0 \lambda z) \quad (3.39)$$

Therefore, when we use oblique illumination, as seen from the output side, we get a modulated and shifted version of the diffraction field that we would obtain with the normally incident illumination.

In practice, from the perspective of an observer, this modulation and shift have no consequence (as long as the tilt angle of the wave is small). The modulation will only slightly change the direction of the beam that forms the reconstruction, and the shift will only slightly change the location of the reconstruction.

The implication of the above statement is that, it is not critical to use exactly the same illumination wave as given by Eq. 3.27. For instance, instead of that wave, suppose we use a normally incident plane wave. We have seen above that the oblique illumination causes a shift and modulation relative to the normally incident case. Hence, when we do not use the oblique illumination, the output field will shift back to its original position. This case is illustrated in Fig. 3.16. As seen, relative to the oblique illumination case, the desired field appears within a shifted version of the synthesis region. It can be shown that the center of the synthesis region is shifted to $\left(-\frac{\lambda z}{4\Delta_x}, -\frac{\lambda z}{4\Delta_y}\right)$. For typical practical values of Δ_x , Δ_y and λ , the amount of the shift is a few centimeters for z around a meter. Hence, as the third option, we can proceed as in the second option until we compute $\bar{s}_{b_1}[m, n]$, but then use a normally incident plane wave (or any other plane wave that makes a small angle with the wave in Eq. 3.27) and accept to obtain the desired field within a shifted version of the synthesis region.

3.2.4 Binary Pixel Values Other Than ± 1

Up to now, we assumed that the binary SLM pixels can be set to ± 1 . In a more general case, the pixels are set to two different complex numbers c_1 and c_2 rather than ± 1 . Such a case is fundamentally no different than the ± 1 case. To see this, suppose on the physical binary SLM we can write c_1 and c_2 where $c_1 \neq c_2$.

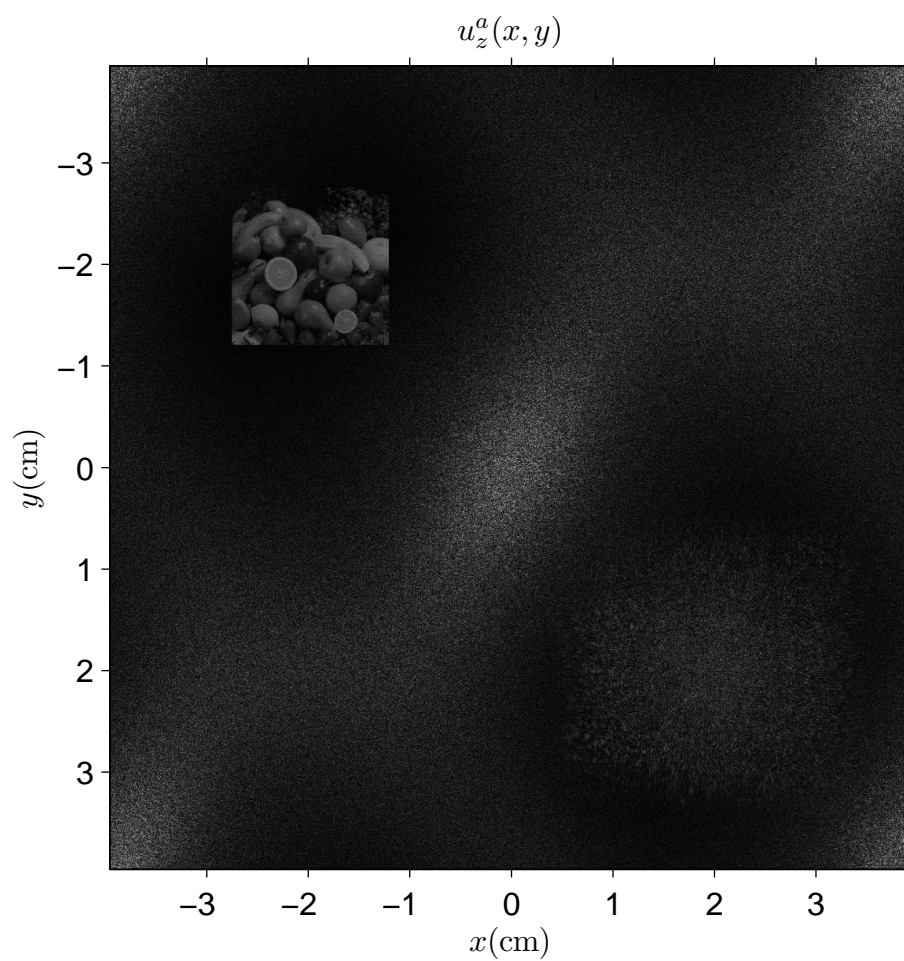


Figure 3.16: Light field generated by the binary SLM pattern in Fig. 3.15 under normally incident illumination. First appeared in [1].

Suppose given a desired field, we first compute $\bar{s}_{b_1}[m, n]$ which consists of ± 1 s as described in Sec. 3.2.2. Now, let $\bar{s}_{b_2}[m, n]$ denote the actual binary SLM pattern. Assume that we set $\bar{s}_{b_2}[m, n] = c_1$ when $\bar{s}_{b_1}[m, n] = -1$ and $\bar{s}_{b_2}[m, n] = c_2$ when $\bar{s}_{b_1}[m, n] = 1$. It is easy to show that $\bar{s}_{b_2}[m, n] = \frac{c_2 - c_1}{2} \bar{s}_{b_1}[m, n] + \frac{c_1 + c_2}{2}$. Assume the SLM is illuminated by a normally incident plane wave (so that the reconstruction will be observed in a shifted window). The effect of the $\frac{c_2 - c_1}{2}$ term that multiplies $\bar{s}_{b_1}[m, n]$ is a trivial change in the output complex amplitude. The additive $\frac{c_1 + c_2}{2}$ term, which is nonzero when $c_1 \neq -c_2$, is more problematic and creates the so called undiffracted DC beam which propagates around the optical axis and has dimensions approximately equal to that of the SLM for practical distances. However, when the reconstruction is performed sufficiently away from the SLM, this DC beam does not interfere with the reconstruction since the synthesis is performed in an off-axis window.

As an example, we display in Fig. 3.17 the light field that we would obtain if the black pixels of Fig. 3.15 were equal to 0 instead of -1 . (We assume that the white pixels in Fig. 3.15 are still equal to 1, and normally incident illumination is used.) We see the mentioned undiffracted DC beam of the SLM in the middle of the figure. Note that in this simulation, the physical size of the binary SLM is $8.2mm \times 8.2mm$, which is roughly equal to the size of the undiffracted DC beam as explained above.

3.2.5 Volumetric Synthesis Examples

We conclude this chapter noting that our approach can be used to compute binary SLM patterns for three-dimensional applications, for instance to synthesize 3D objects floating in air as in 3D display applications. Recall that in Sec. 3.1, we stated that within the pyramid given in Eq. 3.12 (which we named the synthesis pyramid), an SLM behaves the same as its low-pass filtered version. When this pyramid is sufficiently narrow, a desired field specified within it can be

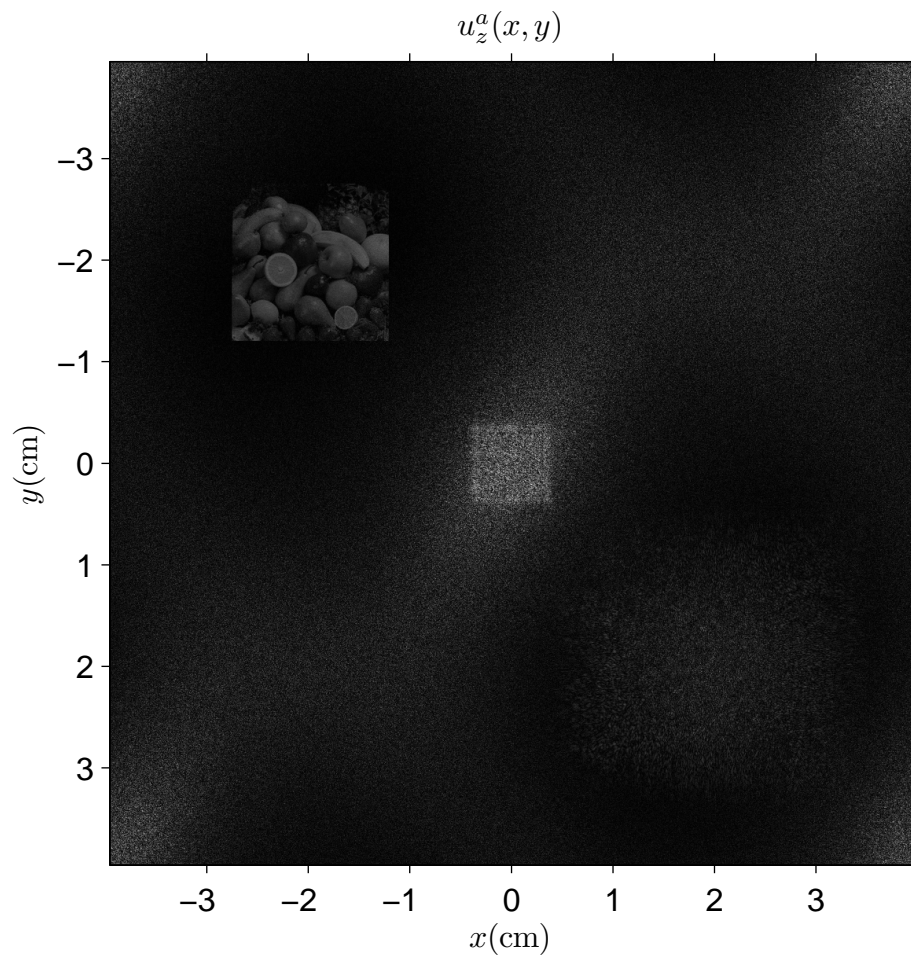


Figure 3.17: Light field generated by the binary SLM pattern in Fig. 3.15 with black pixels being equal to 0 instead of -1 . Image is enhanced for better visualization.

synthesized with a sufficiently low-pass gray level SLM pattern which can be halftoned with low error into a binary SLM pattern.

As an example, consider the 1024×1024 gray level SLM pattern (which is again a low pass pattern) shown in Fig. 3.18. When written on an SLM with pixel periods $\Delta_x, \Delta_y = 8\mu m$ and when $\lambda = 632.9nm$, this pattern generates the fields displayed in Fig. 3.19 and Fig. 3.20 at $z = 0.8m$ and $z = 1m$ respectively. (In this example, the physical size of the SLM is $8.2mm \times 8.2mm$, the physical size of the central diffraction order is $6.33cm \times 6.33cm$ at $z = 0.8m$ and $7.91cm \times 7.91cm$ at $z = 1m$. The physical sizes of the objects are around a centimeter.) As seen, a tomato is focused at $z = 0.8m$ while two peppers are focused at $z = 1m$. In Fig. 3.21, we display the binary SLM pattern computed from the gray level pattern in Fig. 3.18. Under normally incident illumination, this pattern produces the output fields shown in Fig. 3.22 and Fig. 3.23 at $z = 0.8m$ and $z = 1m$, respectively. We see that the images are successfully reconstructed at their respective depths, and that the quantization noise is successfully distributed over the don't care region. Note that since the objects (and therefore their images) were chosen small enough, the ideal gray level SLM pattern generating them was sufficiently low-pass, so we managed to successfully halftone it into the binary SLM pattern in Fig. 3.21.

3.3 Conclusion

In this chapter, we show that when the desired light fields are suitably specified, a binary SLM can be used to synthesize them merely after free space propagation for about a meter without the need to use any complicated optical setup. By “suitably specified”, we mean that the desired fields should obey the constraints due to pixellated SLM structure discussed in Sec. 2.3 and they should be confined within the synthesis region or pyramid specified by Eq. 3.12 where B_x and B_y are

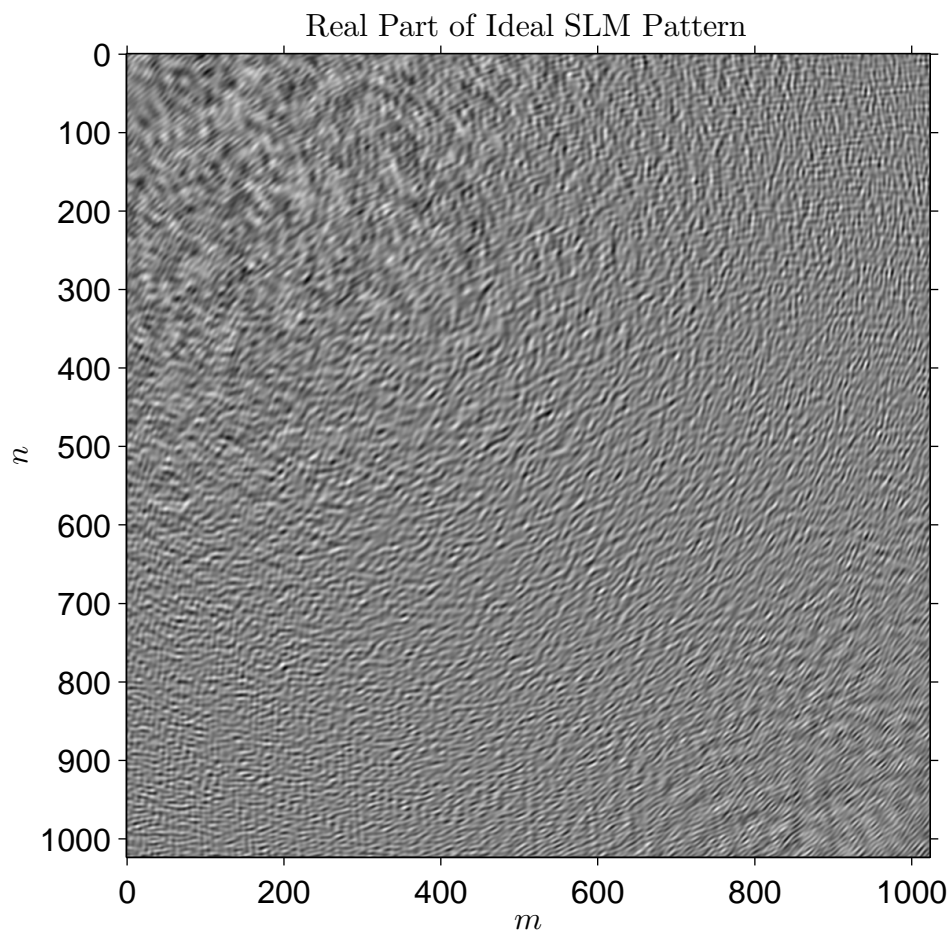


Figure 3.18: The ideal SLM pattern generating the light field depicted in Fig. 3.19 and Fig. 3.20 (real part).

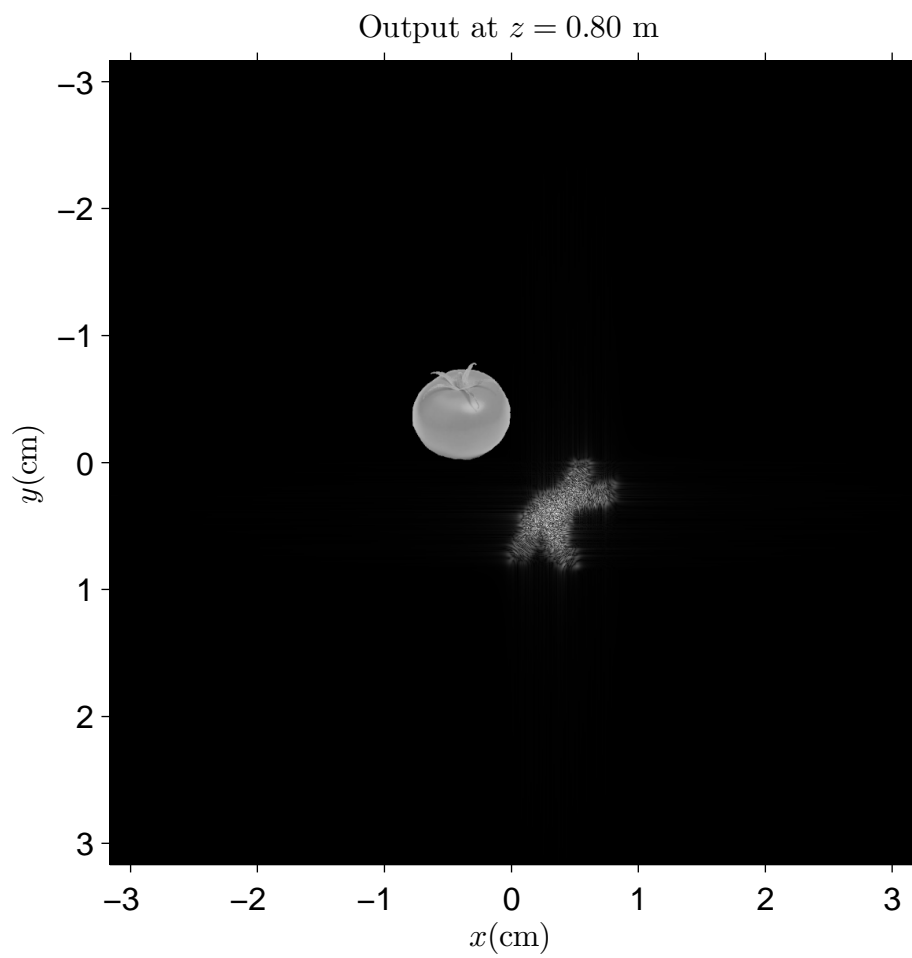


Figure 3.19: Light field generated by the SLM pattern in Fig. 3.18 at $z = 0.8m$ (magnitude). Image is enhanced for better visualization.

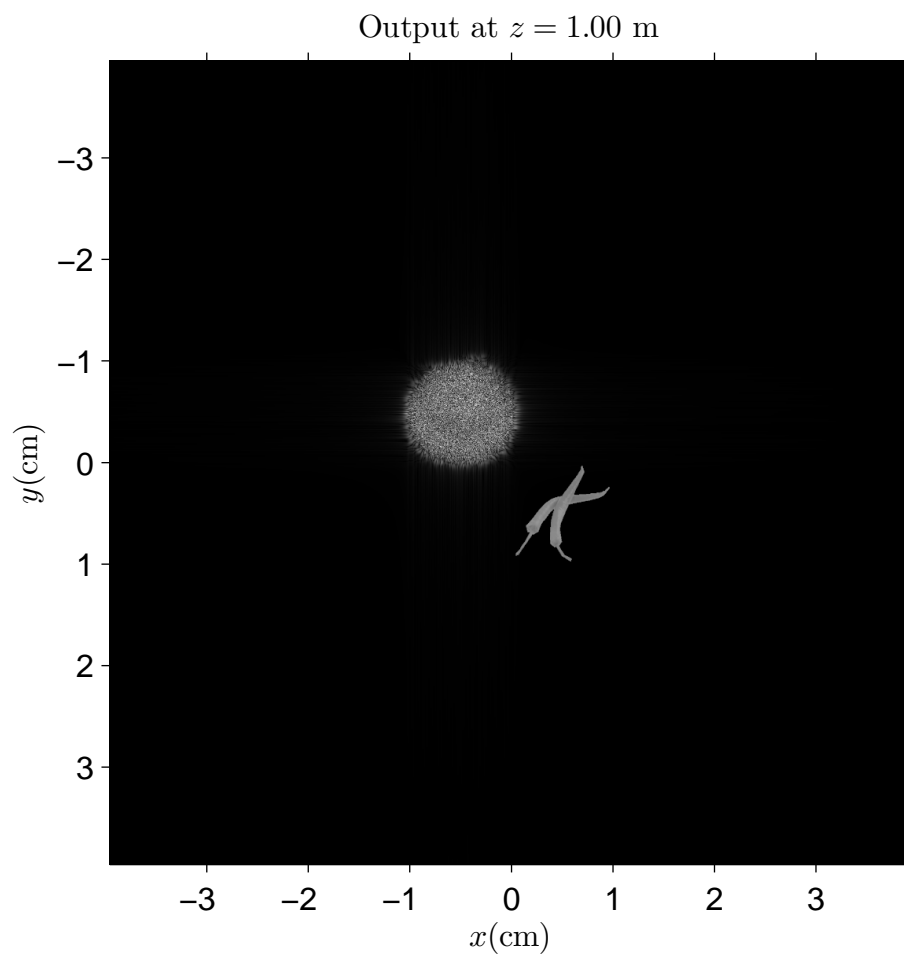


Figure 3.20: Light field generated by the SLM pattern in Fig. 3.18 at $z = 1m$ (magnitude). Image is enhanced for better visualization.

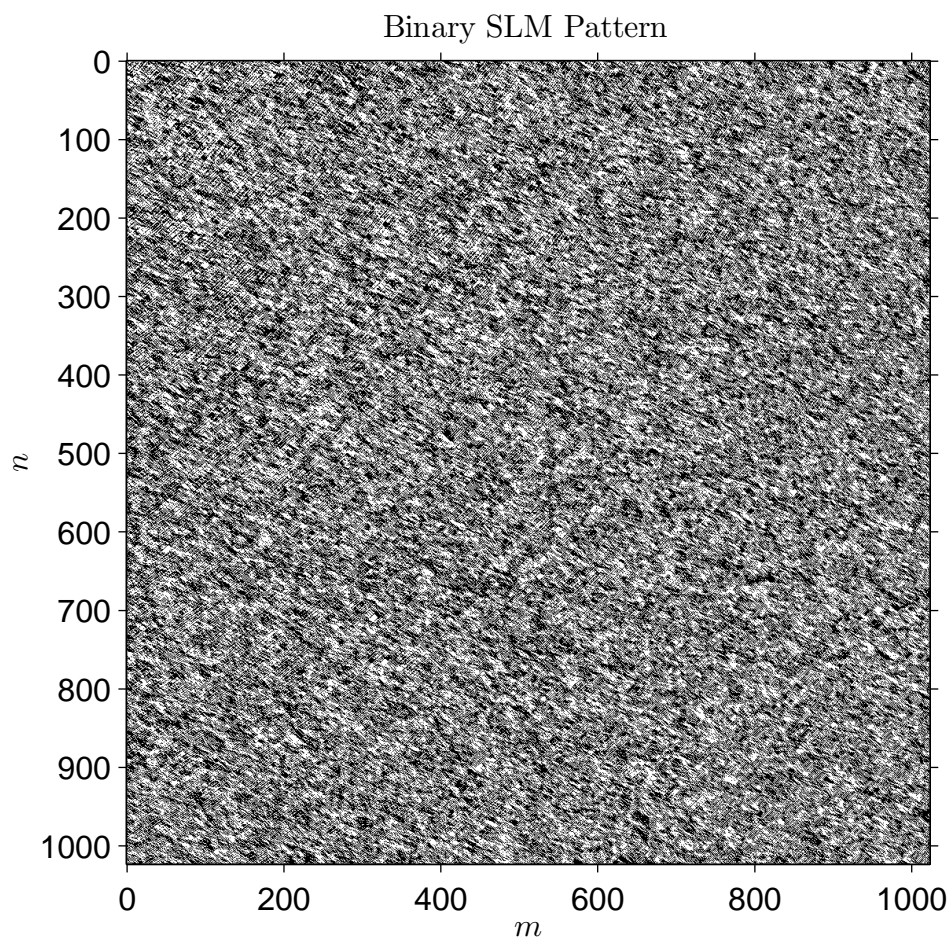


Figure 3.21: Binary SLM pattern computed from the ideal SLM pattern in Fig. 3.18. First appeared in [1].

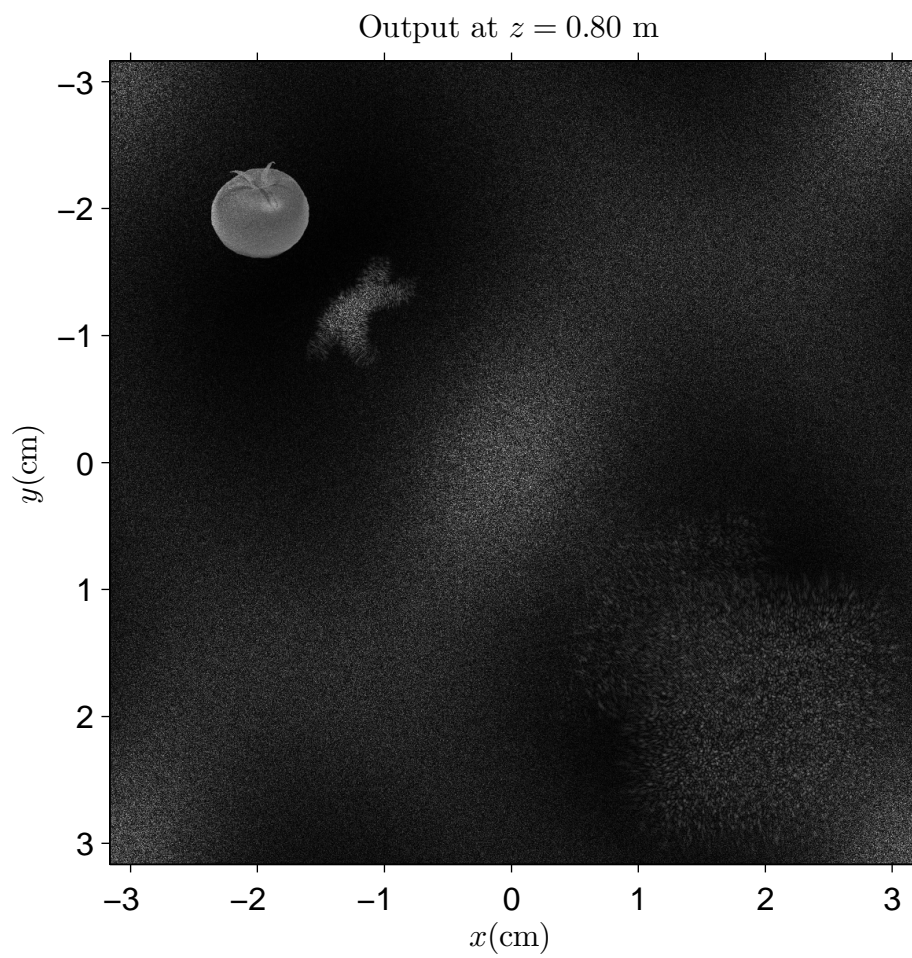


Figure 3.22: Light field generated by the binary SLM pattern in Fig. 3.21 at $z = 0.8\text{m}$ (magnitude). First appeared in [1].

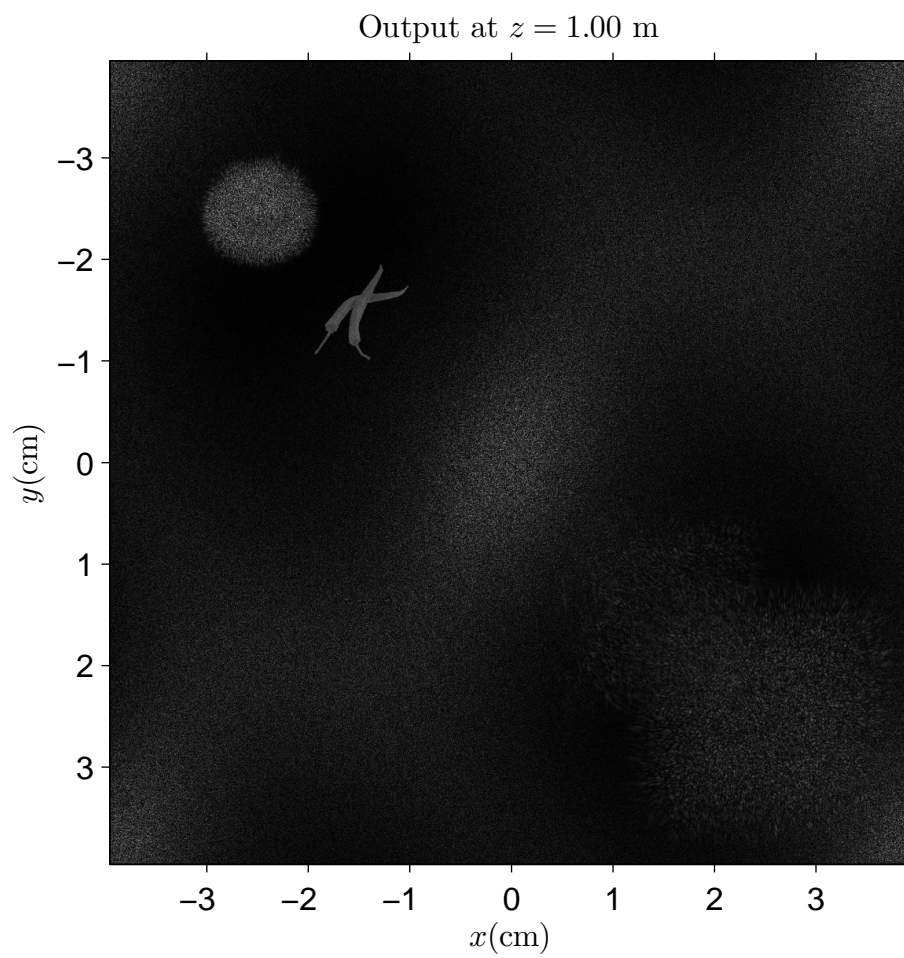


Figure 3.23: Light field generated by the binary SLM pattern in Fig. 3.21 at $z = 1\text{m}$ (magnitude). First appeared in [1].

selected sufficiently small. We showed that if these constraints are satisfied, the ideal gray level complex-valued SLM pattern becomes sufficiently oversampled, and it can be successfully halftoned with an acceptable error. Though at a first glance the halftoning issue seemed problematic due to the fact that a binary SLM pattern is essentially real-valued but the desired ideal SLM pattern is in general complex-valued, we showed that this problem can be overcome with the simple technique proposed in Sec. 3.2. Our technique essentially decomposes the complex-valued halftoning problem to two decoupled real-valued constrained halftoning problems for the real and imaginary parts of the ideal SLM pattern. Using our simulations, we show that the proposed method can be used to generate planar as well as volumetric light field distributions. Our results indicate that when ideal SLM patterns use about $\frac{1}{16}$ th of the available bandwidth (that is, $B_x = \frac{1}{4\Delta_x}$, $B_y = \frac{1}{4\Delta_y}$), quite satisfactory results are obtained.

An important property of our approach is that, as long as the desired light field is specified properly, computation of a suitable binary SLM pattern is reduced to solving the complex-valued halftoning problem in Eq. 3.17. That is, free space propagation related computations can be handled separately from the halftoning related computations. This is an important advantage over many existing algorithms, especially ones that use iterative POCS-like methods similar to the Gerschberg-Saxton algorithm [130, 131, 132, 133, 134, 135, 136, 137]. Usually in such algorithms, during a typical iteration, the output field produced by some current binary SLM pattern is computed, and then that binary SLM pattern is updated according to the error between the output and desired fields. Such calculations greatly increase the computational complexity of those iterative algorithms. On the other hand, in our approach, given the desired field, it is sufficient to compute the ideal gray level complex-valued SLM pattern only once. Then, all the computations can be carried out for solving the halftoning problem. Since any error that we have on the SLM surface after the halftoning process is directly reflected to the synthesis region, we do not need to separately incorporate

free space propagation in the optimization procedure for halftoning. To sum up, we can say that the overall complexity of our approach is mainly determined by the complexities of the algorithms used for computing the ideal complex-valued SLM pattern and for carrying out the halftoning. This decoupling gives designers the freedom to choose among the many existing algorithms for hologram computation and halftoning. To solve the real-valued halftoning problems given in Eq. 3.26, one can use any of the many existing halftoning algorithms, depending on the expectations about the computational performance, reconstruction accuracy, binarization efficiency etc [144]. In our simulations, we used the standard error diffusion algorithm, which is a simple non-iterative algorithm, and we found its performance quite satisfactory both in terms of reconstruction quality and computation speed.

Chapter 4

FULL COMPLEX SPATIAL LIGHT MODULATORS OBTAINED FROM BINARY SPATIAL LIGHT MODULATORS

In Chapter 3, we discussed our method for synthesizing light fields with binary SLMs. The solutions we proposed will be quite useful for those who wish to use binary SLMs in the simplest way for 3D display applications (that is, with the fewest number of optical components that are used). Recall that for a single binary SLM, the proposed technique performs the desired synthesis within a window that lies within the central diffraction order region, and the technique simply treats the region that lies outside the mentioned window as the “don’t care” region. We saw that the don’t care region is mainly occupied by the quantization related noise terms, undiffracted DC beams and higher diffraction orders. It is easy to eliminate these side beams in the case of a single SLM.

However, to build a satisfactory quality holographic display, a single SLM will not be sufficient because a single SLM can only provide a quite narrow viewing angle. Therefore, we need to place many SLMs side by side. But then, the removal of the aforementioned side beams becomes problematic, and their presence may be disturbing for some designers. To provide an alternative to those who do not want these disturbing beams but who still wish to enjoy the advantages of binary SLMs, we propose another solution in this chapter. Here, rather than focusing on the problem of the determination of a binary SLM pattern that synthesizes a desired field, we focus on the problem of effectively creating a new full-complex SLM out of binary SLMs, where the full-complex SLM is created in such a way that virtually all the disturbing beams mentioned above are eliminated as well.

Actually, there are already some methods that have been proposed to effectively create full-complex SLMs out of restricted type SLMs (which are mostly phase or amplitude only). As for some examples, in one of the proposed methods, an amplitude-only SLM is imaged on a phase-only SLM [145]. In this manner, light passing through the SLMs is modulated both in phase and amplitude as if it comes out of a single full-complex SLM. In an other method, the beams of two phase-only SLMs are added using a beamsplitter [146]. In that way, effectively, a new SLM is created, where a pixel of the new SLM is equal to the sum of two phase-only pixels, so it can be adjusted to a large number of different complex values. Similar methods are discussed in [147, 148, 149, 150, 151, 152]. These methods are successful in the sense that the new SLM provides a richer modulation compared to the component SLMs. But, a common problem in all these methods is that the set of complex values available for a pixel of the new SLM does not have a good coverage of the complex plane, so it is hard to regard the new SLM as a satisfactory full-complex SLM. The main reason of this problem is the imperfections of practical multilevel SLMs: for instance, most phase-only SLMs do not cover the $0 - 2\pi$ range for the phase, but cover only a restricted angular range. Similarly, most phase-only SLMs perform an undesired amplitude

modulation along with the phase modulation and vice versa, that is hard to keep track of and causes the new SLM to behave differently than intended. Moreover, the behavior of most multilevel SLMs depends strongly on the illumination wavelength, causing the proposed designs to operate in a satisfactory manner only for a narrow range of wavelengths. To avoid such problems, one can consider the creation of full-complex SLMs out of binary SLMs, as we do in this chapter.

We begin the discussion by explaining our generic method in Sec. 4.1. This method actually trades pixel count to dynamic range, and carries out this trade-off by generalizing the concepts of bit plane representation and decomposition for ordinary digital gray scale multilevel images [153]. In particular, we propose to effectively obtain a 2^K -level (or K -bit) SLM by forming a properly weighted superposition of K binary (or 1-bit) SLMs. When K is sufficiently large (such as $K = 16$), the new SLM can be regarded as a full-complex SLM. We show that in this way, information-wise, the binary SLMs are utilized in the most efficient manner that is possible. In Sec. 4.2, we propose a $4f$ system to optically implement our generic method. With this system, out of a binary SLM with $PM \times QN$ pixels, we can effectively create a PQ -bit (or 2^{PQ} level) SLM with $M \times N$ pixels. Again, when P and Q are sufficiently large (such as $P = Q = 4$), the new SLM can be regarded as a full-complex one. We show that the $4f$ system also provides a means for eliminating the disturbing higher orders from the diffraction field of the new SLM while preserving the central order with little distortion. The key element of the proposed $4f$ system is an optical thin mask that needs to be physically produced and placed to the Fourier plane. In Sec. 4.3, we discuss several alternatives for this mask depending on the production capabilities, and show that even very simple 4-level, 3-level or binary masks work. What is presented in this chapter is a follow-up to the work presented in [2].

4.1 Generic Method

In this section, we explain our generic method for effectively creating full-complex SLMs out of binary SLMs. The reason we call the method as generic is that, there can be many different possible optical implementations. In this respect, we will not adhere to a specific optical system in this section, but carry out the discussion at an abstract level. In the next section, we propose a possible optical implementation.

Our generic method is actually based on the well known concepts of the bit plane representation and decomposition for ordinary digital gray scale images [153]. Therefore, we start by briefly reviewing these concepts.

Let \mathbf{S} denote a 2^K -level (or K -bit) gray scale digital image of size $M \times N$ ($K, M, N \in \mathcal{Z}^+$). That is, the image has $M \times N$ pixels where each pixel can take 2^K different gray levels. Usually, the pixel values of \mathbf{S} can be set to one of $0, 1, 2, \dots, 2^K - 1$, and let us assume that this is also so in our case. Typically, taking $K = 8$ is sufficient for high quality images of daily life scenes, so the pixel values can be set to $0, 1, 2, \dots, 255$. It is well known that \mathbf{S} can be written as $\mathbf{S} = 2^0\mathbf{B}_0 + 2^1\mathbf{B}_1 + \dots + 2^{K-1}\mathbf{B}_{K-1}$; where $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{K-1}$ are all $M \times N$ binary (or 1-bit) images each pixel of which can be equal to either 0 or 1. The binary images $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{K-1}$ are called the bit planes of \mathbf{S} , where \mathbf{B}_0 is named the least significant bit plane and \mathbf{B}_{K-1} is named the most significant bit plane. Given \mathbf{S} , writing it as $\mathbf{S} = 2^0\mathbf{B}_0 + 2^1\mathbf{B}_1 + \dots + 2^{K-1}\mathbf{B}_{K-1}$ is called the bit plane representation of \mathbf{S} and the process of finding the appropriate $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{K-1}$ is called the bit plane decomposition of \mathbf{S} . Note that in its bit plane representation, \mathbf{S} is written as a particular weighted superposition of its bit planes, where the weights are given by $2^0, 2^1, \dots, 2^{K-1}$. Our generic method for creating full-complex SLMs out of binary SLMs amounts to a generalization of this weighted superposition concept to include complex-valued weighting coefficients as well.

In particular, in our method, we will form a properly weighted superposition of a number of identical binary SLMs (corresponding to the superposition of bit planes of an image) to obtain a full-complex SLM (corresponding to a gray level image).

Suppose we have K binary SLMs of size $M \times N$ at hand. Assume these SLMs are identical, i.e., the pixel periods, pixel geometries and other physical parameters of all the SLMs are the same. Let us denote these binary SLMs with $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{K-1}$. Let $\bar{b}_i[m, n]$ denote the value of the $(m, n)^{\text{th}}$ pixel of \mathbf{B}_i where $0 \leq m \leq M-1, 0 \leq n \leq N-1, 0 \leq i \leq K-1$ and $m, n, i \in \mathcal{Z}$. Suppose $\bar{b}_i[m, n]$ can be set to either -1 or 1 .

Now suppose that we have some kind of an optical system which processes all these binary SLMs and effectively creates a new $M \times N$ SLM that we denote by \mathbf{S} , such that $\mathbf{S} = w_0\mathbf{B}_0 + w_1\mathbf{B}_1 + \dots + w_{K-1}\mathbf{B}_{K-1}$ where $w_i \in \mathcal{C}$ for $0 \leq i \leq K-1, i \in \mathcal{Z}$. (Note that if $\bar{s}[m, n]$ denotes the value of the $(m, n)^{\text{th}}$ pixel of \mathbf{S} , we have $\bar{s}[m, n] = w_0\bar{b}_0[m, n] + w_1\bar{b}_1[m, n] + \dots + w_{K-1}\bar{b}_{K-1}[m, n]$.) That is, the optical system effectively creates the new SLM by forming a weighted superposition of the binary SLMs where w_i 's denote the possibly complex-valued weighting coefficients. Below, we show that selecting these coefficients wisely, and taking K sufficiently large, we can make \mathbf{S} a full-complex SLM.

As a first example, suppose $K = 16$, and suppose the weighting coefficients are taken as

$$w_i = \begin{cases} \frac{1}{255}2^i & \text{for } 0 \leq i \leq 7 \\ \frac{j}{255}2^{i-8} & \text{for } 8 \leq i \leq 15 \end{cases} \quad (4.1)$$

with $j = \sqrt{-1}$. Then, adjusting $\bar{b}_0[m, n], \bar{b}_1[m, n], \dots, \bar{b}_{15}[m, n]$; we can set each $\bar{s}[m, n]$ to any complex number of the form $\frac{1}{255}(R + jI)$ where $R, I = -255, -253, -251, \dots, 251, 253, 255$. Note that the number of different complex values available for each $\bar{s}[m, n]$ is 2^{16} , corresponding to an information content of 16 bits per pixel. With the current weights, we reserve 8 bits for the real part

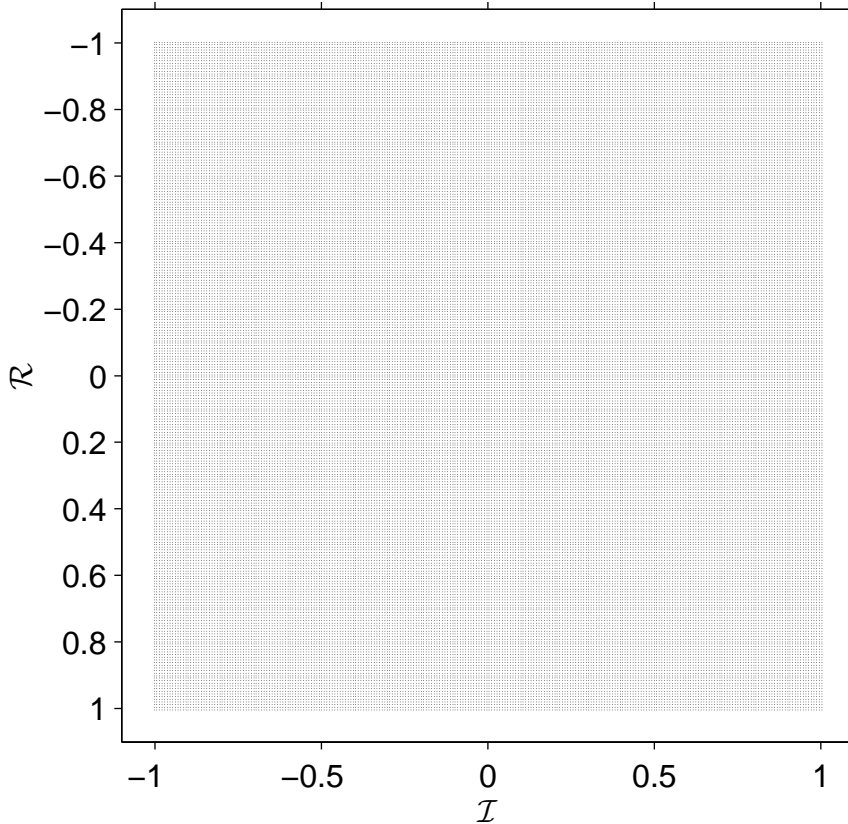


Figure 4.1: Complex numbers available for a pixel of the new SLM with the weights in Eq. 4.1. Each dot on the figure represents an achievable complex number. First appeared in [2].

and 8 bits for the imaginary part of the new SLM. These complex values are displayed in Fig. 4.1. These complex numbers also have a good coverage of the complex plane, as we can see from the figure. Therefore, we can regard \mathbf{S} as a full-complex SLM. Hence, out of 16 binary (1-bit) SLMs of size $M \times N$, we have created a single but full-complex (16-bit) SLM of size $M \times N$.

Let us continue to examine the current example in detail. Note that to start with, we have a total of $16MN$ pixels (from 16 binary SLMs where each binary SLM has MN pixels). At the end, we have only MN pixels (from a single SLM

with MN pixels). So, the pixel count is reduced. However; the pixels in the beginning are binary (1-bit), while the pixels of the new SLM are full-complex (16-bit). So, the dynamic range is increased. Therefore, we have essentially traded pixel count to dynamic range while obtaining the full-complex SLM out of the binary SLMs.

Actually, information-wise, the mentioned trade off is carried out in the most efficient manner that is possible. To see this, note that the information content of each of the binary SLMs is MN bits, and since we use 16 binary SLMs, the total information content in the beginning is $16MN$ bits. At the end, we obtain the full complex SLM, which has MN pixels where each pixel has an information content of 16 bits, so the information content of the full complex SLM is also $16MN$ bits. Hence, the full complex SLM is created out of the binary SLMs without any loss in the information capacity. Therefore, information-wise, binary SLMs are utilized in the most efficient manner.

The nice result stated above follows as a consequence of the fact that we create the binary SLM by forming a properly weighted superposition of the binary SLMs. That is, our choice of the weighting coefficients as in Eq. 4.1 is one of the wise choices. As an example, to see what happens when we do not choose these coefficients wisely, consider taking $w_i = 1$ for all i , which corresponds to directly adding all 16 binary SLMs without using any weighting. If this were the case, then each pixel of \mathbf{S} (obtained as $\mathbf{S} = \mathbf{B}_0 + \mathbf{B}_1 + \dots + \mathbf{B}_{15}$) could only be set to $-16, -14, -12, \dots, 12, 14, 16$. That is, it would not be possible to set a pixel of \mathbf{S} to some value other than the listed ones no matter how much we play with the values of the corresponding binary pixels that are superposed. So, only 17 different values would be available for each pixel of the new SLM, that is quite small compared to 2^{16} (which is the number of different values obtained with the former weights). Therefore, if we directly superposed the binary SLMs without any weighting, we would have used them in a quite inefficient manner

while creating the new SLM (and it would not be possible to regard the new SLM as a full-complex one at all since only 17 different values would be available for each pixel). This example clearly suggests that a wise selection of the weighting coefficients is crucial in the explained method.

As we have seen, our initial selection for the weighting coefficients is a wise selection, but it is not the only possible wise selection. For instance, again suppose that $K = 16$, and suppose the weighting coefficients are selected as

$$w_i = \left| \frac{1 - e^{\frac{j\pi}{16}}}{2} \right| e^{\frac{ji\pi}{16}} \quad (4.2)$$

for $0 \leq i \leq 15$, $i \in \mathcal{Z}$. Using these coefficients, each pixel of \mathbf{S} can be set to one of the complex numbers displayed in Fig. 4.2. These complex numbers are different from the ones shown in Fig. 4.1, but again there are 2^{16} different complex numbers on this figure and they also have a good coverage of the complex plane. So, the new coefficients can also be used to create a 16-bit full-complex SLM as well. Similarly, many other nice selections for the weighting coefficients can also be found.

Note that our choice of K and w_i 's determines the values that are available for a pixel of \mathbf{S} . That is, if during some application, the desired value for a pixel of \mathbf{S} is not among the available values, we should first perform a quantization. Therefore, our choice of K and w_i 's actually defines a quantizer on the complex plane. In order to achieve optimum performance, we should design this quantizer (i.e. choose K and w_i 's) by taking into account the statistical properties of the non-quantized source (that is, the complex-valued pattern that one wants to write on the full-complex SLM). Actually, in the above examples, since we chose $K = 16$ (and we chose the weights wisely), we had a quite large number of values with a good coverage of the complex plane, so that the quantization error will most probably be quite small and thus quantization related issues can be neglected. However, we should note that in holographic applications, smaller values of K can be sufficient. For instance, in [82, 154], it is discussed that even

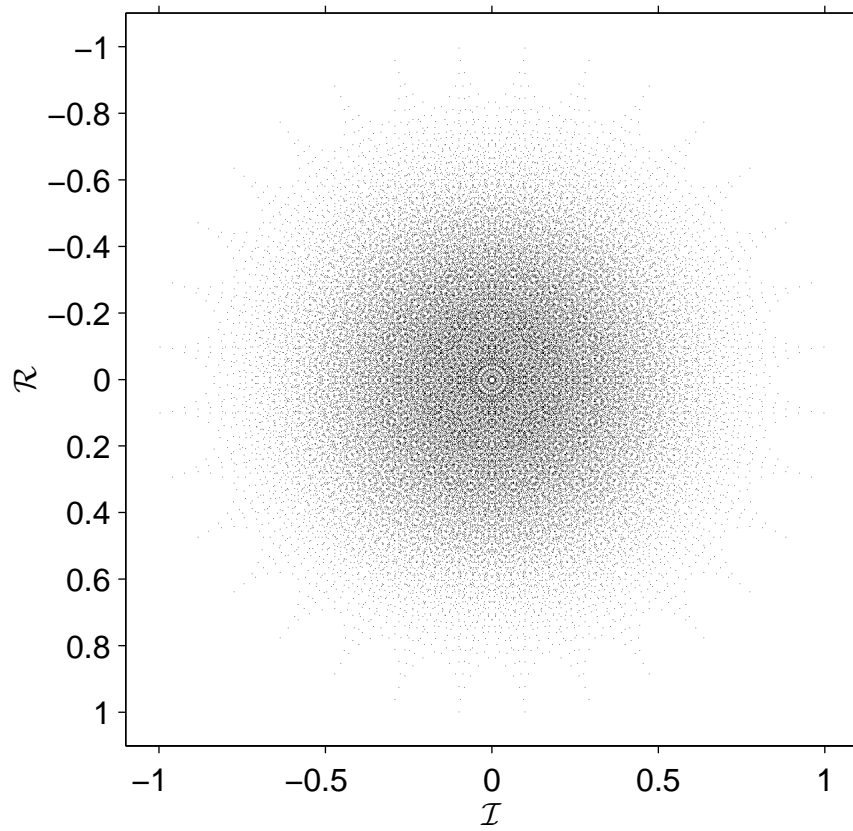


Figure 4.2: Complex numbers available for a pixel of the new SLM with the weights in Eq. 4.2. First appeared in [2].

4-bit quantization can be quite sufficient for certain holographic applications. In such cases, in order to make the overall system as economical as possible, one may wish to use only the needed number of binary SLMs. Then, the designer must make sure that his choice of w_i 's leads to a satisfactory distribution of achievable complex numbers on the complex plane.

In summary, our generic method for creating full-complex SLMs out of binary SLMs can be described as follows: Using some optical system, effectively form a weighted superposition of K binary SLMs to obtain a new SLM where K is sufficiently large. Select the weights such that each pixel of the new SLM can be set to 2^K different complex values where these values also have a satisfactory coverage of the complex plane. Under these conditions, the new SLM can be regarded as a full-complex one. In this way, information-wise, the binary SLMs are utilized in the most efficient manner that is possible while creating the full-complex SLM.

Before closing this section, let us discuss how to configure the binary SLMs in order to make the new SLM \mathbf{S} equal to some desired full-complex SLM denoted by \mathbf{S}_d . That is, we wish to determine $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{K-1}$ such that we achieve $\mathbf{S} = \mathbf{S}_d$ where $\mathbf{S} = w_0\mathbf{B}_0 + w_1\mathbf{B}_1 + \dots + w_{K-1}\mathbf{B}_{K-1}$. Let us assume that the weighting coefficients are selected wisely such as in Eq. 4.1 or in Eq. 4.2, so that each pixel of \mathbf{S} can be set to 2^K different complex values. Let us also assume that \mathbf{S}_d is already quantized (that is, the value of each pixel of \mathbf{S}_d is equal to one of the achievable complex values). Then, for each pixel of \mathbf{S}_d , we should solve the equation $\bar{s}_d[m, n] = \sum_{i=0}^{K-1} w_i \bar{b}_i[m, n]$ and determine $\bar{b}_i[m, n]$ under the constraint that $\bar{b}_i[m, n] = \pm 1$. A straightforward method is to prepare a look-up table that holds the mapping between the possible binary patterns of size $1 \times K$ and the complex numbers produced by them, and use this look-up table to determine $\bar{b}_0[m, n], \bar{b}_1[m, n], \dots, \bar{b}_{K-1}[m, n]$. Note that this look-up table will have 2^K entries. When $K \leq 16$, such a look-up table can be handled easily

with today's computation and storing technology (with $K \leq 16$, the number of entries in the table will be less than or equal to 65536). If the achievable complex numbers are listed in the look-up table in an intelligent manner, search times may be minimized or in certain cases no search may be needed at all.

As a first example, suppose we use the weights given in Eq. 4.1. In this case, actually we do not need to use any look-up table at all. Given a desired \mathbf{S}_d , we only need to perform separate bit plane decompositions of real and imaginary parts in the standard way (that is, as we do for ordinary images).

As a second example, suppose we take the weighting coefficient as in Eq. 4.2. In this case, the solution is not as easy as for the first example, because, referring to Fig. 4.2, we can see that the achievable complex numbers do not have an orderly distribution on the complex plane. Then, we can resort to the look-up table based solution. Suppose first that we list the achievable complex numbers and the corresponding binary patterns in a random order. In that case, given a pixel $\bar{s}_d[m, n]$, we need to search the entire look-up table to find out the complex number that is closest to $\bar{s}_d[m, n]$. However, this method is impractical especially when the size of the look-up table is considerable. A better method is illustrated in Fig. 4.3. Here, we first divide the complex plane covered by the achievable complex numbers into a number of bins such that

- each bin contains at most some manageable number of achievable complex numbers, and
- it is easy to find the bin in which $\bar{s}_d[m, n]$ lies.

For instance, in Fig. 4.3, we see that the bins are formed by a uniform sampling in the phase and magnitude domains. This makes it easy to find the bin in which the given $\bar{s}_d[m, n]$ lies. As also seen from the figure, in each bin, there are only a small number of achievable complex numbers. Then, given the value of $\bar{s}_d[m, n]$, after determining the bin inside which it lies, to find the closest

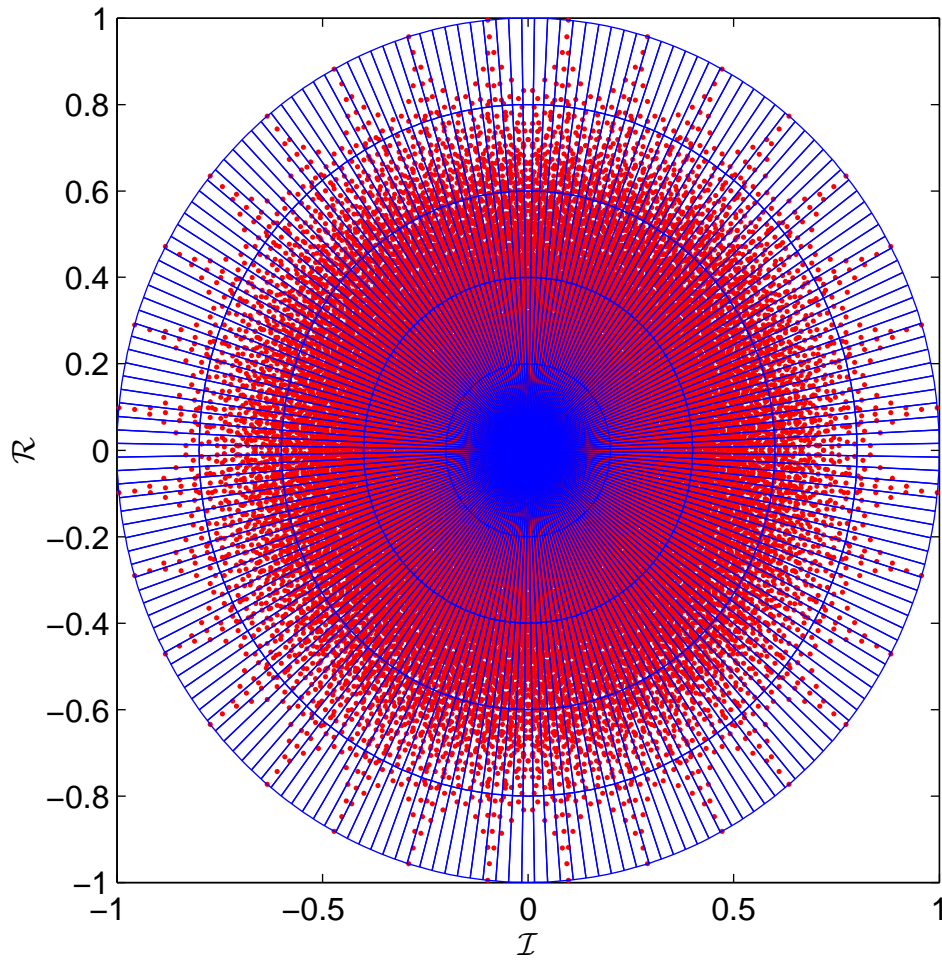


Figure 4.3: Division of the complex plane into bins to facilitate the search.

achievable complex number and the corresponding binary pattern, we only search among the alternatives that lie inside the bin. In this manner, search times are made quite short.

4.2 Practical Implementation Using a $4f$ System

In the previous section, we described our generic method for creating a K -bit SLM out of K binary SLMs. We carried that discussion at an abstract level without adhering to a specific optical system, and concluded that we can successfully create the full-complex SLM if we have some kind of an optical system which effectively forms a properly weighted superposition of K binary SLMs. In this section, we propose an optical system which forms this weighted superposition. But, rather than K separate but identical $M \times N$ binary SLMs, we will assume that we have a single binary SLM with $K \times M \times N$ pixels. We will partition this SLM into a K sub SLMs, and form the weighted superposition of these sub-SLMs to obtain a single full-complex $M \times N$ SLM. Though there is no fundamental difference in between, this new form makes it easier for us to present our work. Actually, we will further take K as $K = P \times Q$ where $P, Q \in \mathcal{Z}^+$, and assume that we have a binary SLM of size $PM \times QN$ at hand. Out of this binary SLM, we will create a PQ -bit SLM of size $M \times N$. Hence, we will consider the binary SLM as a collection of PQ sub-SLMs, where each sub-SLM is also binary and has size $M \times N$. We will form the weighted superposition of these sub-SLMs to obtain the PQ -bit SLM.

We explain our optical system in three subsections. In Sec. 4.2.1, we still stay on the conceptual ground and propose a simple linear shift invariant (LSI) system through which we obtain the weighted superposition of the sub-SLMs. In Sec. 4.2.2, we consider a bandlimited version of the LSI system we propose in Sec. 4.2.1. The reason is, we plan a $4f$ system based implementation, but the initial LSI system has infinite bandwidth, so it is not possible to practically implement. We will also see that the band-limitation will have the positive effect of eliminating the disturbing higher diffraction orders of the SLM output.

In Sec. 4.2.3, we propose a $4f$ system which implements the bandlimited LSI system considered in Sec. 4.2.2.

4.2.1 An LSI System to Form the Weighted Superposition of Binary SLMs

Consider a binary SLM that has $PM \times QN$ pixels. Let Δ_x and Δ_y denote the pixel periods of the SLM. Let $a(x, y)$ denote the pixel aperture function of the SLM such that $a(x, y) = 0$ for $x \notin [0, \Delta_x]$ or $y \notin [0, \Delta_y]$. In most cases, $a(x, y) = \text{rect}\left(\frac{x}{\Delta_x} - 0.5\right) \text{rect}\left(\frac{y}{\Delta_y} - 0.5\right)$ where $\text{rect}(x) = 1$ for $|x| < 0.5$, $\text{rect}(x) = 0.5$ for $|x| = 0.5$ and $\text{rect}(x) = 0$ for $|x| > 0.5$. Let $\bar{b}[m, n]$ denote the value of the $(m, n)^{\text{th}}$ SLM pixel ($m, n \in \mathcal{Z}$) for $0 \leq m \leq PM - 1$ and $0 \leq n \leq QN - 1$ such that $\bar{b}[m, n]$ can be set to either -1 or 1 . If $b(x, y)$ denotes the complex transmittance of the binary SLM, we have:

$$b(x, y) = a(x, y) ** \sum_{m=0}^{PM-1} \sum_{n=0}^{QN-1} \bar{b}[m, n] \delta(x - m\Delta_x, y - n\Delta_y). \quad (4.3)$$

Above, $**$ denotes the two dimensional convolution operation such that $f_1(x, y) ** f_2(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1(x', y') f_2(x - x', y - y') dx' dy'$. With the definitions above, the SLM is assumed to lie in the region $0 \leq x \leq PM\Delta_x$ and $0 \leq y \leq QN\Delta_y$.

Note that we can view our binary SLM as a collection of PQ sub-SLMs (which are also binary) where each sub-SLM consists of $M \times N$ pixels. In this respect, we can write $b(x, y)$ as

$$b(x, y) = \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} b_{p,q}(x - pM\Delta_x, y - qN\Delta_y) \quad (4.4)$$

where $b_{p,q}(x, y)$ denotes the complex transmittance of the $(p, q)^{\text{th}}$ sub-SLM and is given as

$$b_{p,q}(x, y) = b(x + pM\Delta_x, y + qN\Delta_y) \text{rect}\left(\frac{x - 0.5M\Delta_x}{M\Delta_x}\right) \text{rect}\left(\frac{y - 0.5N\Delta_y}{N\Delta_y}\right) \quad (4.5)$$

for $0 \leq p \leq P-1$ and $0 \leq q \leq Q-1$. Note that all sub-SLMs have pixel aperture function $a(x, y)$ and pixel periods Δ_x and Δ_y . When forming the binary SLM, the $(p, q)^{\text{th}}$ sub-SLM is placed in the region $pM\Delta_x \leq x \leq (p+1)M\Delta_x$ and $qN\Delta_y \leq y \leq (q+1)N\Delta_y$ (a 1D illustration is provided in Fig. 4.4 where $P = 4$ and the sub-SLMs are denoted by b_0, b_1, b_2 and b_3). Our purpose is to form a weighted superposition of these sub-SLMs. We will accomplish this by processing $b(x, y)$ with a suitably defined linear shift invariant (LSI) system.

Consider an LSI system whose impulse response $h(x, y)$ is given as

$$h(x, y) = \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \bar{w}[p, q] \delta(x - pM\Delta_x, y - qN\Delta_y) \quad (4.6)$$

where $\bar{w}[p, q] \in \mathcal{C}$ for $0 \leq p \leq P-1, 0 \leq q \leq Q-1, p, q \in \mathcal{Z}$. As seen, $h(x, y)$ consists of a $P \times Q$ grid of impulses which are spaced by $M\Delta_x$ and $N\Delta_y$. The $(p, q)^{\text{th}}$ impulse is located at $(pM\Delta_x, qN\Delta_y)$ and has strength $\bar{w}[p, q]$. If $H(\nu_x, \nu_y)$ denotes the frequency response of this LSI system, we have

$$\begin{aligned} H(\nu_x, \nu_y) &= \mathcal{F} \{h(x, y)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \exp \{-j2\pi(x\nu_x + y\nu_y)\} dx dy \\ &= \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \bar{w}[p, q] e^{-j2\pi\{\nu_x pM\Delta_x + \nu_y qN\Delta_y\}}. \end{aligned} \quad (4.7)$$

It is easy to see that $H(\nu_x, \nu_y)$ is periodic with periods $\frac{1}{M\Delta_x}$ and $\frac{1}{N\Delta_y}$.

Suppose $b(x, y)$ is processed by this LSI system. Let $g(x, y)$ denote the resulting output such that $g(x, y) = b(x, y) * h(x, y)$. Then we can write:

$$g(x, y) = \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \bar{w}[p, q] b(x - pM\Delta_x, y - qN\Delta_y). \quad (4.8)$$

The spatial support of $g(x, y)$ is given by the region $0 \leq x \leq (2P-1)M\Delta_x$ and $0 \leq y \leq (2Q-1)N\Delta_y$. Examining $g(x, y)$, we see that the LSI system actually forms a superposition of shifted and weighted replicas of $b(x, y)$. Such LSI systems are usually called echo systems in the signal processing literature,

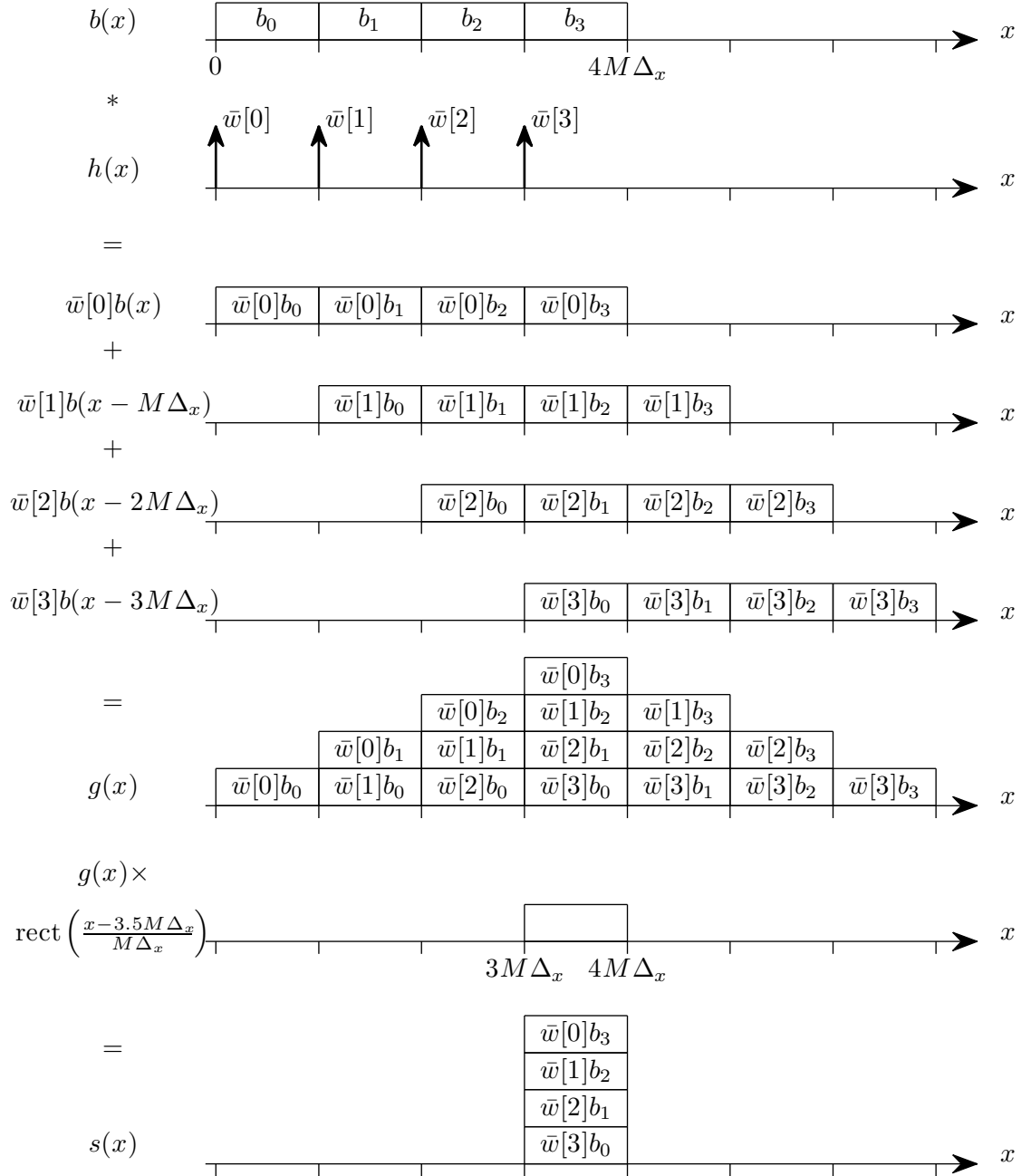


Figure 4.4: 1D illustration of the process through which $s(x, y)$ is created out of $b(x, y)$. $b(x)$, $h(x)$, $g(x)$, $s(x)$ and $\bar{w}[p]$ respectively denote the 1D counterparts of $b(x, y)$, $h(x, y)$, $g(x, y)$, $s(x, y)$ and $\bar{w}[p, q]$ which are discussed in the text. In the 1D case, we assumed that the binary SLM is divided into four sub-SLMs of size M , and denoted these sub-SLMs with b_0 , b_1 , b_2 and b_3 . First appeared in [2].

since one-dimensional (1D) versions of them are used to produce synthetically generated echoes of sound signals in audio processing.

Now, define $s(x, y)$ such that

$$s(x, y) = g(x, y) \text{rect} \left(\frac{x - (P - 0.5)M\Delta_x}{M\Delta_x} \right) \text{rect} \left(\frac{y - (Q - 0.5)N\Delta_y}{N\Delta_y} \right). \quad (4.9)$$

Hence, $s(x, y)$ is obtained by windowing $g(x, y)$ in space. The window selects the portion of $g(x, y)$ lying in the the region $(P - 1)M\Delta_x \leq x \leq PM\Delta_x$ and $(Q - 1)N\Delta_y \leq y \leq QN\Delta_y$. It is straightforward to show that

$$s(x, y) = \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \bar{w} [P - 1 - p, Q - 1 - q] b_{p,q}(x - x_0, y - y_0) \quad (4.10)$$

with $x_0 = (P - 1)M\Delta_x$ and $y_0 = (Q - 1)N\Delta_y$.

Eq. 4.10 is the result we have been seeking for. We see that $s(x, y)$ is obtained as the weighted superposition of $b_{p,q}(x, y)$'s where the weights are given by $\bar{w} [p, q]$ (a 1D illustration of the process through which $s(x, y)$ is obtained from $b(x, y)$ is provided in Fig. 4.4 for $P = 4$). Hence, $s(x, y)$ represents the complex transmittance of a new SLM that is obtained as the weighted superposition of the sub-SLMs of the binary SLM. Note that the new SLM also has the pixel aperture function $a(x, y)$ and pixel periods Δ_x and Δ_y . It lies in the region $(P - 1)M\Delta_x \leq x \leq PM\Delta_x$ and $(Q - 1)N\Delta_y \leq y \leq QN\Delta_y$. We know from Sec. 4.1 that, properly choosing $\bar{w} [p, q]$, we can make this new SLM a PQ -bit SLM. And we also know that when P and Q are chosen sufficiently large (such as $P = Q = 4$), the new SLM can be regarded as a full-complex one. Hence, using the proposed LSI system, we can create an $M \times N$ full-complex (PQ -bit) SLM out of a $PM \times QN$ binary SLM.

As a first illustration for the $P = Q = 4$ and $M = N = 256$ case, Fig. 4.5 shows a 1024×1024 binary SLM. Fig. 4.6 and Fig. 4.7 respectively show the real

and imaginary parts of $g(x, y) = h(x, y) ** b(x, y)$ when $\bar{w}[p, q]$ are taken as:

$$\begin{bmatrix} \bar{w}[0, 0] & \bar{w}[0, 1] & \bar{w}[0, 2] & \bar{w}[0, 3] \\ \bar{w}[1, 0] & \bar{w}[1, 1] & \bar{w}[1, 2] & \bar{w}[1, 3] \\ \bar{w}[2, 0] & \bar{w}[2, 1] & \bar{w}[2, 2] & \bar{w}[2, 3] \\ \bar{w}[3, 0] & \bar{w}[3, 1] & \bar{w}[3, 2] & \bar{w}[3, 3] \end{bmatrix} = \frac{1}{255} \begin{bmatrix} j2^7 & j2^6 & j2^5 & j2^4 \\ j2^3 & j2^2 & j2^1 & j2^0 \\ 2^7 & 2^6 & 2^5 & 2^4 \\ 2^3 & 2^2 & 2^1 & 2^0 \end{bmatrix}. \quad (4.11)$$

Fig. 4.8a and Fig. 4.8b, which are respectively the magnified versions of the signals within the windows in Fig. 4.6 and Fig. 4.7, show the real and imaginary parts of $s(x, y)$. As seen, a 256×256 full-complex SLM has been successfully created out of the binary SLM. (In this example, we took $\Delta_x = \Delta_y = 10\mu m$ and assumed rectangular pixels. In Fig. 4.5, black pixels have value -1 , and white pixels have value 1 .)

As a second illustration, suppose again that $P = Q = 4$ and $M = N = 256$, and our goal is to obtain the $s(x, y)$ depicted in Fig. 4.8. However, suppose that this time the weights are given as

$$\begin{bmatrix} \bar{w}[0, 0] & \bar{w}[0, 1] & \bar{w}[0, 2] & \bar{w}[0, 3] \\ \bar{w}[1, 0] & \bar{w}[1, 1] & \bar{w}[1, 2] & \bar{w}[1, 3] \\ \bar{w}[2, 0] & \bar{w}[2, 1] & \bar{w}[2, 2] & \bar{w}[2, 3] \\ \bar{w}[3, 0] & \bar{w}[3, 1] & \bar{w}[3, 2] & \bar{w}[3, 3] \end{bmatrix} = \left| \frac{1 - e^{\frac{j\pi}{16}}}{2} \right| \begin{bmatrix} e^{\frac{j0\pi}{16}} & e^{\frac{j1\pi}{16}} & e^{\frac{j2\pi}{16}} & e^{\frac{j3\pi}{16}} \\ e^{\frac{j4\pi}{16}} & e^{\frac{j5\pi}{16}} & e^{\frac{j6\pi}{16}} & e^{\frac{j7\pi}{16}} \\ e^{\frac{j8\pi}{16}} & e^{\frac{j9\pi}{16}} & e^{\frac{j10\pi}{16}} & e^{\frac{j11\pi}{16}} \\ e^{\frac{j12\pi}{16}} & e^{\frac{j13\pi}{16}} & e^{\frac{j14\pi}{16}} & e^{\frac{j15\pi}{16}} \end{bmatrix}. \quad (4.12)$$

In this case, the binary SLM pattern that we should be using is illustrated in Fig. 4.9. The real and imaginary parts of the corresponding $g(x, y)$ are illustrated respectively in Fig. 4.10 and Fig. 4.11. $s(x, y)$ is again drawn into a rectangular box, and we can see that the $s(x, y)$ illustrated in Fig. 4.8 is obtained successfully.

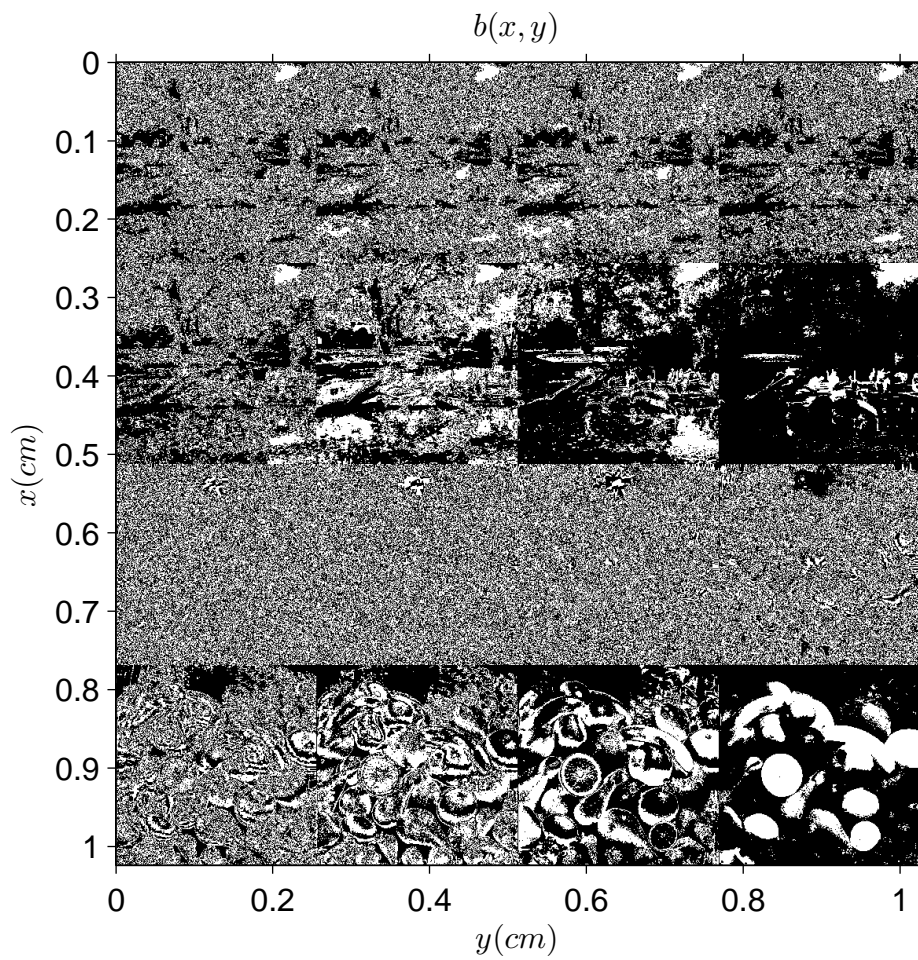


Figure 4.5: Binary SLM pattern. First appeared in [2].

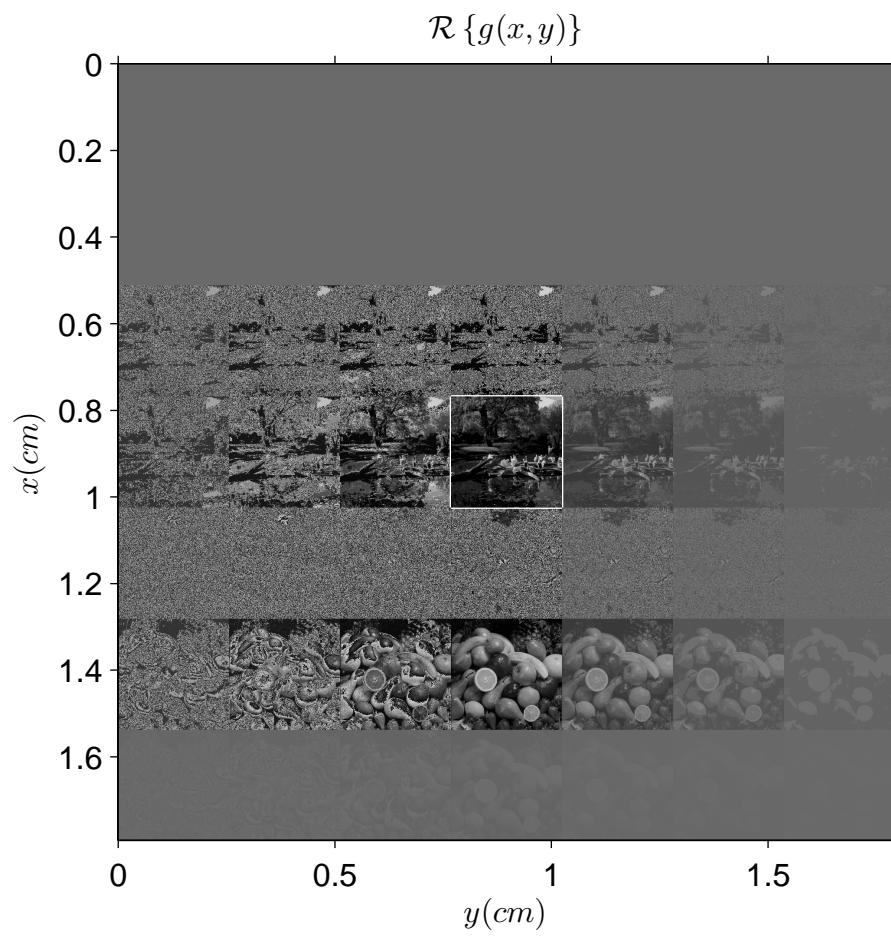


Figure 4.6: Real part of $g(x,y)$. First appeared in [2].

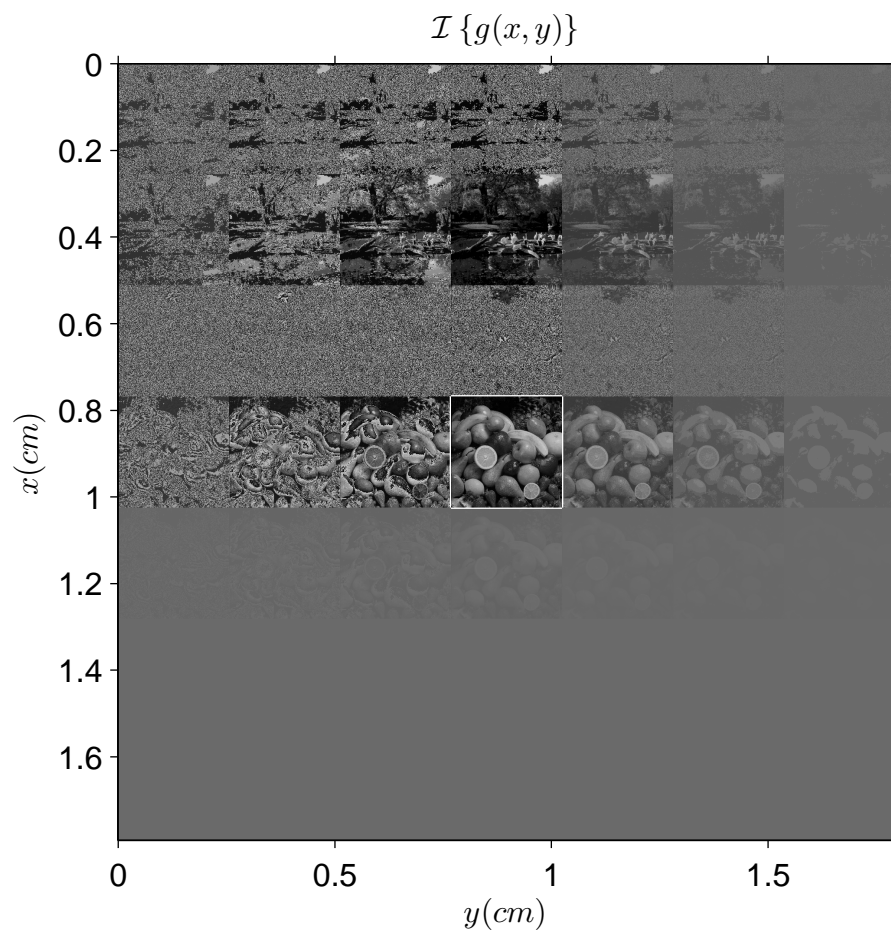


Figure 4.7: Imaginary part of $g(x,y)$. First appeared in [2].

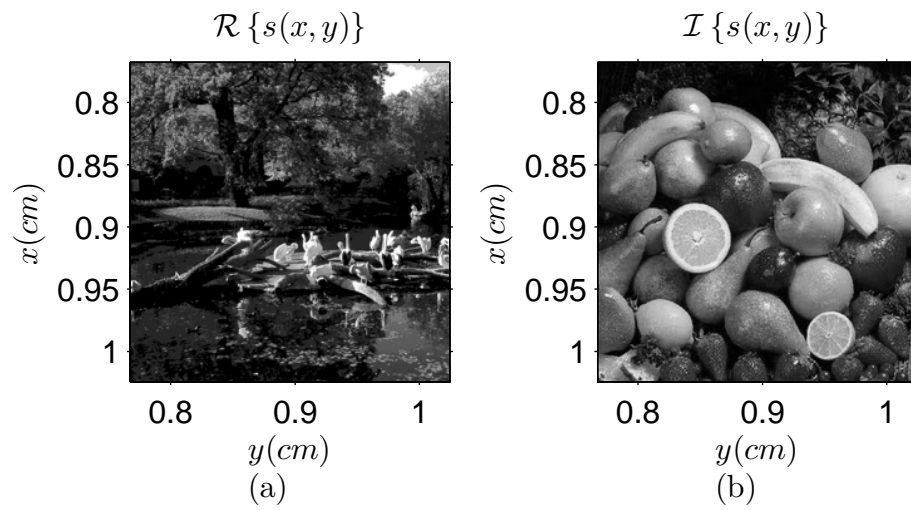


Figure 4.8: (a) Real part of $s(x, y)$. (b) Imaginary part of $s(x, y)$. First appeared in [2].

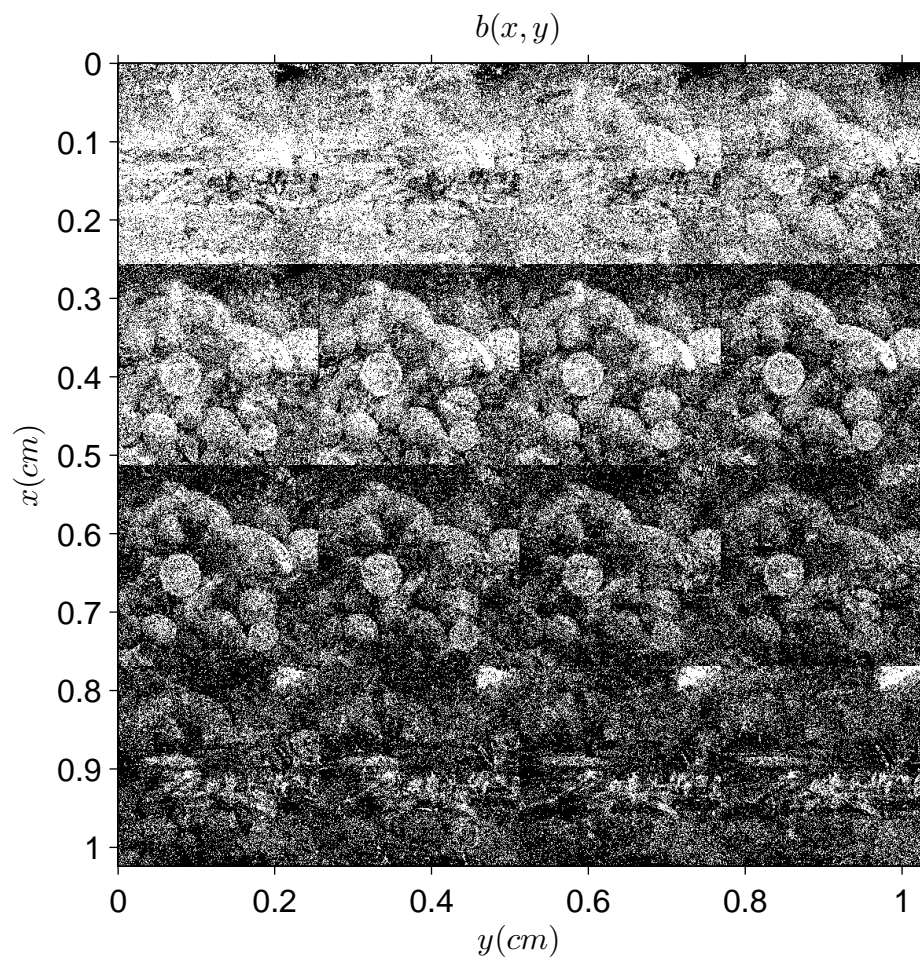


Figure 4.9: Binary SLM pattern for the new weights.

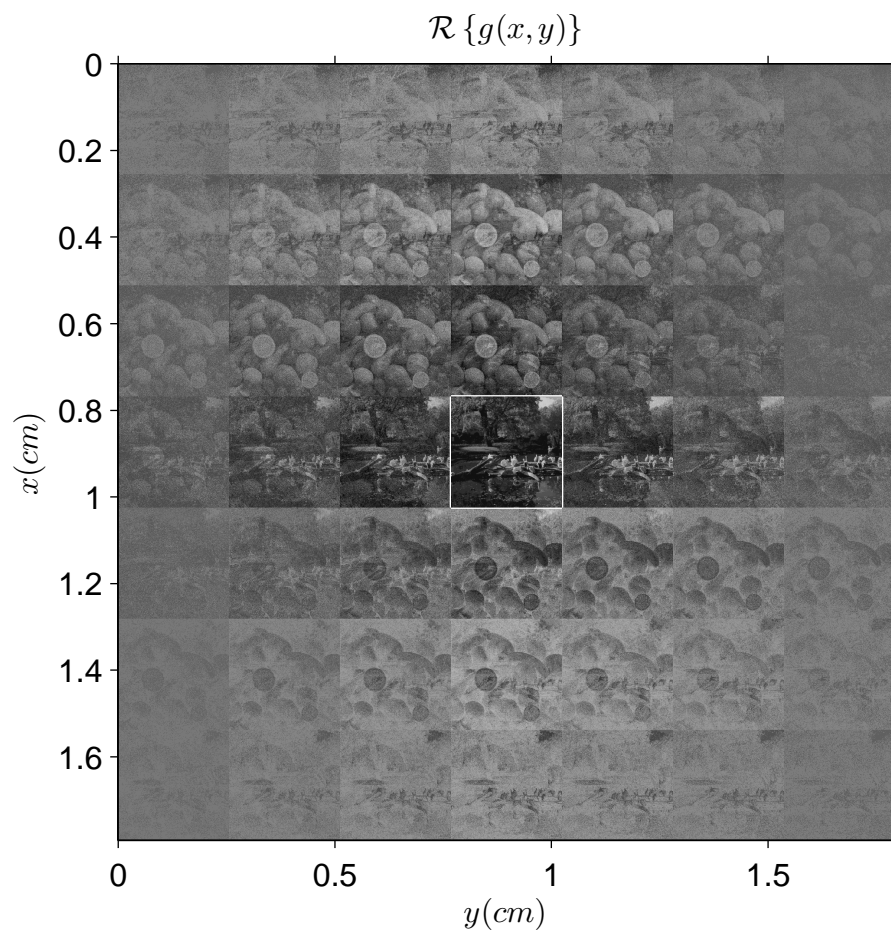


Figure 4.10: Real part of $g(x,y)$.

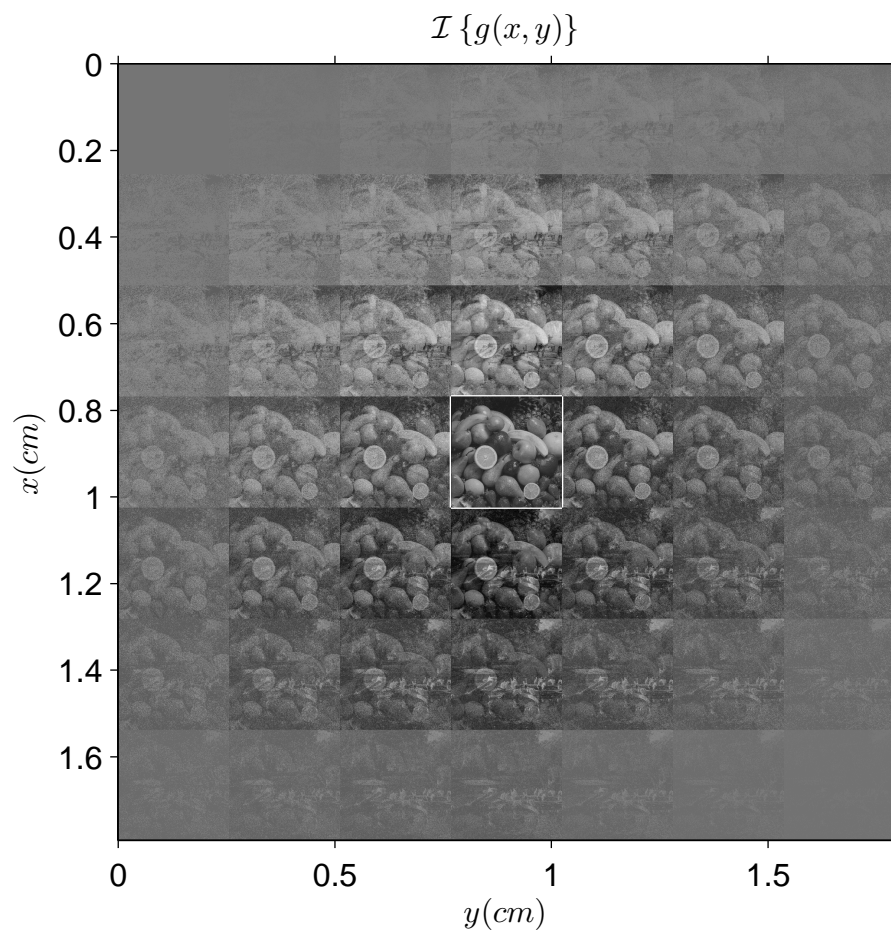


Figure 4.11: Imaginary part of $g(x,y)$.

4.2.2 Imposing a Bandwidth Limitation

In theory, the LSI system proposed in the previous subsection enables us to effectively create a PQ -bit $M \times N$ SLM out of a $PM \times QN$ binary SLM. It is well known that $4f$ systems can be used to optically implement LSI systems, and we will do so in the next subsection. However, before proceeding, we will analyze the effects of imposing a bandwidth restriction to the LSI system we used in the previous subsection. The reason is, the frequency response $H(\nu_x, \nu_y)$ of that LSI system occupies the entire frequency spectrum, whereas a $4f$ setup which consists of finite-sized lenses and optical masks can only support a finite bandwidth. In this respect, let us consider a new LSI system with impulse response $h_L(x, y)$ such that

$$h_L(x, y) = h(x, y) * * h_B(x, y) \quad (4.13)$$

where $h_B(x, y)$ denotes the impulse response of an ideal low-pass filter with bandwidths B_x and B_y such that

$$h_B(x, y) = B_x B_y \text{sinc}(x B_x) \text{sinc}(y B_y) \quad (4.14)$$

with $\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$. Hence, the new LSI system is the bandlimited version of the original LSI system. Note that if $H_L(\nu_x, \nu_y)$ and $H_B(\nu_x, \nu_y)$ respectively denote the Fourier transforms of $h_L(x, y)$ and $h_B(x, y)$, we have $H_B(\nu_x, \nu_y) = \text{rect}\left(\frac{\nu_x}{B_x}\right) \text{rect}\left(\frac{\nu_y}{B_y}\right)$ and $H_L(\nu_x, \nu_y) = H(\nu_x, \nu_y) H_B(\nu_x, \nu_y)$. Let $g_L(x, y)$ denote the output when the binary SLM is processed by the new LSI system, so that $g_L(x, y) = b(x, y) * * h_L(x, y)$. We can also write

$$g_L(x, y) = g(x, y) * * h_B(x, y) \quad (4.15)$$

which indicates that the output of the new LSI system is a blurred version of the output of the original LSI system. Recall that when the original LSI system was used, the full-complex SLM (represented by $s(x, y)$) was selected out of $g(x, y)$ with a simple windowing operation in space. Suppose we apply the same window to $g_L(x, y)$ and denote the resulting output with $s_L(x, y)$ such that $s_L(x, y) =$

$g_L(x, y) \text{rect} \left(\frac{x - (P-0.5)M\Delta_x}{M\Delta_x} \right) \text{rect} \left(\frac{y - (Q-0.5)N\Delta_y}{N\Delta_y} \right)$. Assuming that the blurring is not too strong (that is, B_x and B_y are sufficiently large, or $h_B(x, y)$ is sufficiently narrow) so that leakages due to infinite tails of $h_B(x, y)$ can be ignored, we can write:

$$s_L(x, y) \approx s(x, y) * h_B(x, y). \quad (4.16)$$

Hence, when the new LSI system is used, we approximately obtain a blurred version of the full-complex SLM pattern represented by $s(x, y)$. Since the free-space propagation is also a LSI system, the light field produced by the SLM at any distance will also experience the same blurring. Obviously, we do not want to lose any important information present in the generated light field due to this blurring effect. So there is a limit to the degree of blurring we can tolerate. As we saw in Sec. 2.3, the light field produced by a pixellated SLM consists of diffraction orders, which are shifted, modulated and dispersed versions of each other, so they essentially carry the same information. The order which has the lowest frequency content is called the central order. The blurring will not cause any information loss as long as the central order remains unaffected from it. This is the case if B_x and B_y are greater than the bandwidths of the central order, which are given as $\frac{1}{\Delta_x}$ and $\frac{1}{\Delta_y}$ in our case. Hence, the bandlimited LSI system can be confidently used instead of the original LSI system if:

$$B_x > \frac{1}{\Delta_x} \text{ and } B_y > \frac{1}{\Delta_y}. \quad (4.17)$$

Indeed, if the above conditions are met near the limit (i.e., $B_x \approx \frac{1}{\Delta_x}$ and $B_y \approx \frac{1}{\Delta_y}$), central order of the light field produced by the full-complex SLM is preserved (with little distortion) while higher orders are almost eliminated. This result is actually explicitly desired in 3D displays where presence of higher orders is disturbing. As for an illustration, assume $P = Q = 4$, $M = N = 256$, and consider the 1024×1024 binary SLM depicted in Fig. 4.12 ($\Delta_x = \Delta_y = 10\mu m$, the SLM has rectangular pixels). Suppose the weighting coefficients are as given in Eq. 4.11. If we processed this binary SLM with the original LSI system (no band limitation), we would obtain the full-complex SLM depicted in Fig. 4.13a,

that would produce the diffraction field depicted in Fig. 4.13b at a distance of 50cm . Note that since the SLM in Fig. 4.13a has rectangular pixels, the diffraction field in Fig. 4.13b consists of diffraction orders (the bright guitar at the center is the central diffraction order while its replicas are the higher diffraction orders). Next, Fig. 4.13c depicts the blurred version of the full-complex SLM depicted in Fig. 4.13a that we obtain when we process the binary SLM with the new LSI system with bandwidths given by $B_x = \frac{1}{\Delta_x}$ and $B_y = \frac{1}{\Delta_y}$. Fig. 4.13d displays the new diffraction field. As explained above, the central order is almost unaffected from the blurring, while the higher orders are almost eliminated. (In these figures, we computed the diffraction fields taking $h_z(x, y)$ as in Eq. 2.11, and $\lambda = 632.9\text{nm}$.)

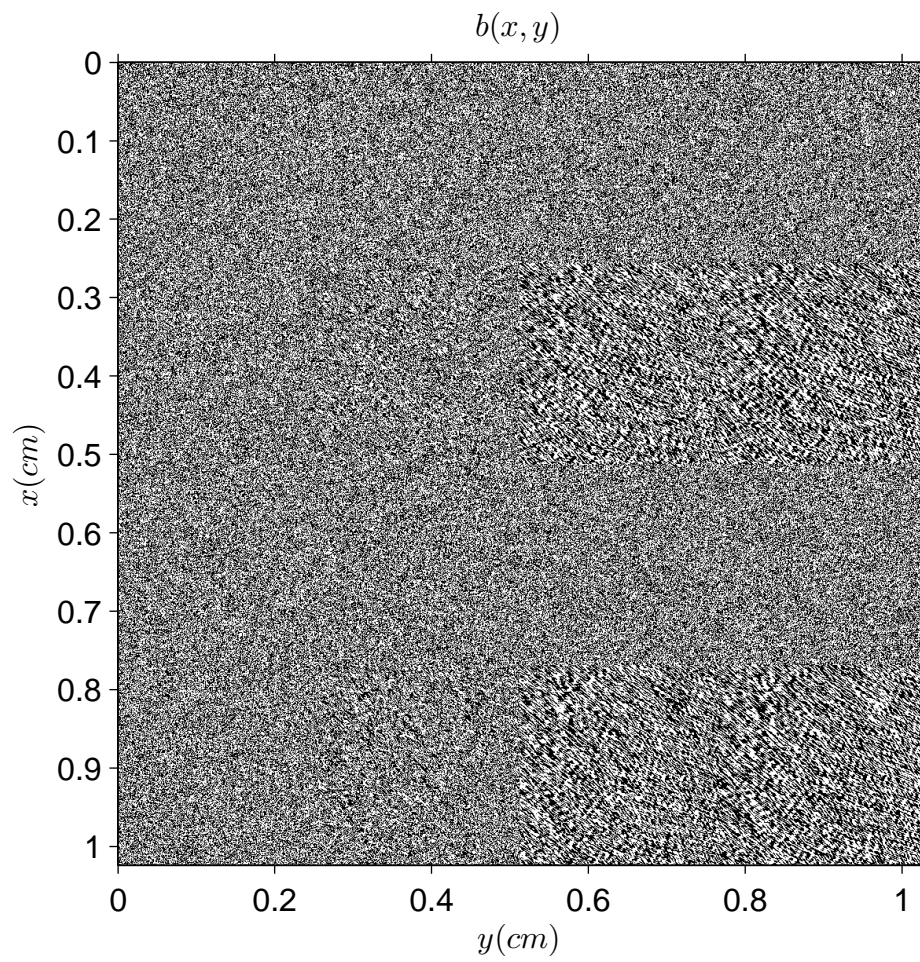


Figure 4.12: Binary SLM Pattern. First appeared in [2].

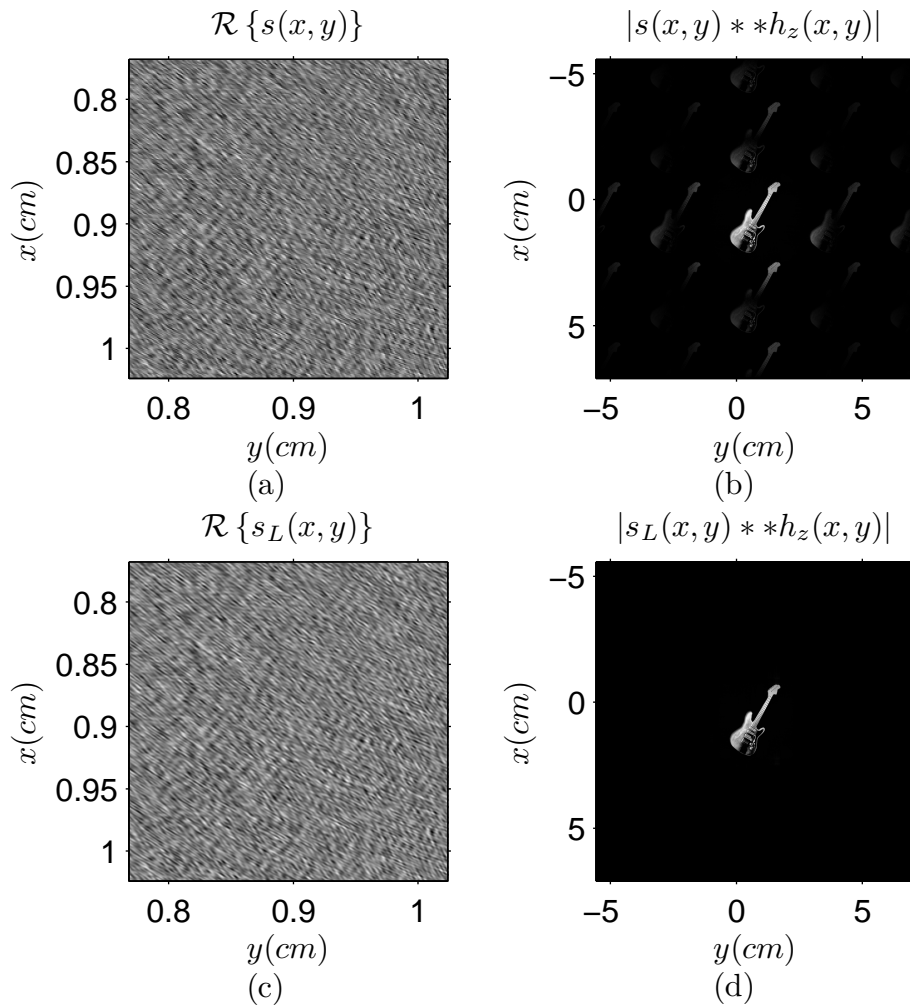


Figure 4.13: (a) Full-complex SLM pattern obtained by processing the binary SLM pattern in Fig. 4.12 with the LSI system described by $h(x, y)$. (b) Resulting diffraction field at $50cm$. (c) Full-complex SLM pattern obtained with the LSI system described by $h_B(x, y)$. (d) Resulting diffraction field at $50cm$. First appeared in [2].

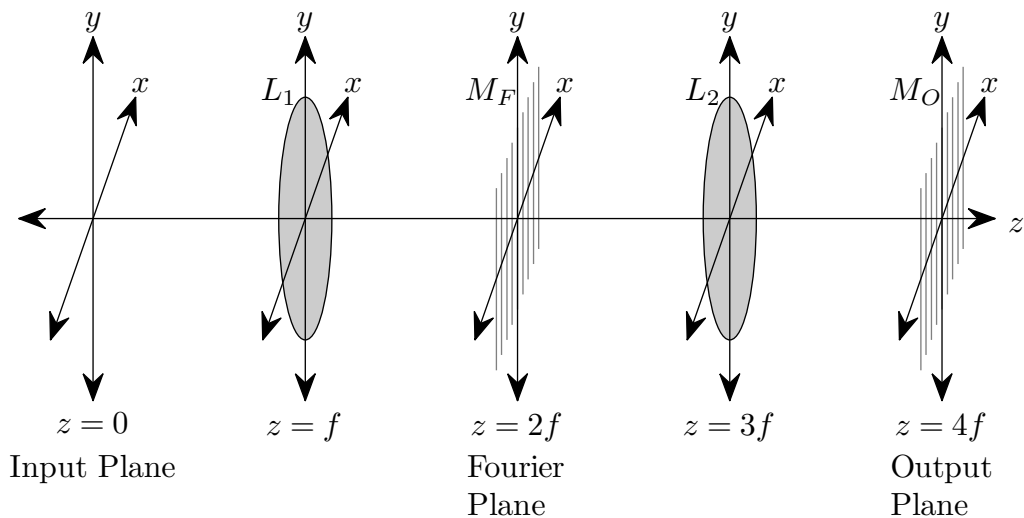


Figure 4.14: $4f$ setup. L_1 and L_2 denote positive thin lenses of focal length f . M_F and M_O respectively denote the Fourier and output plane masks. First appeared in [2].

4.2.3 Implementation with a $4f$ System

Now, the ground for optical implementation is established. For a single wavelength, the LSI system described by Eq. 4.13 can be optically implemented using a $4f$ system. Consider the system depicted in Fig. 4.14. As seen, two positive thin lenses (denoted by L_1 and L_2) with focal lengths f ($f > 0$) are placed at $z = f$ and $z = 3f$ planes. If the illumination wavelength is λ , the complex transmittances of these lenses are given by

$$t_l(x, y) = \exp \left\{ -\frac{j\pi}{\lambda f} (x^2 + y^2) \right\}. \quad (4.18)$$

At the Fourier plane ($z = 2f$ plane), an optical thin mask (that we name the Fourier plane mask) denoted by M_F is placed. Let $m_F(x, y)$ denote the complex transmittance of this mask. At the output plane ($z = 4f$ plane), another thin mask (that we name the output plane mask) denoted by M_O is placed. Let $m_O(x, y)$ denote the complex transmittance of this mask. We assume that both of these masks are passive components, implying that their magnitude transmission at any point should be less than or equal to unity. We simply assume that

$\max \{|m_F(x, y)|\} = 1$ and $\max \{|m_O(x, y)|\} = 1$. Let $u_0(x, y)$ denote the light field over the input plane ($z = 0$ plane). As explained in [104], according to the Fresnel scalar diffraction theory, the light field just before the output plane mask is given as:

$$u_{4f^-}(x, y) = \frac{e^{\frac{j8\pi f}{\lambda}}}{(j\lambda f)^2} u_0(-x, -y) ** M_F \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \quad (4.19)$$

where $M_F(\nu_x, \nu_y)$ denotes the Fourier transform of $m_F(x, y)$. We see that if we take $u_0(x, y) = j^2 e^{-\frac{j8\pi f}{\lambda}} b(-x, -y)$ (which corresponds to placing the 180° rotated version of the binary SLM pattern to the third quadrant of the input plane and illuminating it with a normally incident plane wave of complex amplitude $j^2 e^{-\frac{j8\pi f}{\lambda}}$) and if we have $M_F \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) = \frac{1}{\eta} (\lambda f)^2 h_L(x, y)$ (where $\frac{1}{\eta}$ is included to satisfy the passive mask condition), we can get $u_{4f^-}(x, y) = \frac{1}{\eta} g_L(x, y)$. Therefore, we should have:

$$\begin{aligned} m_F(x, y) &= \frac{1}{\eta} H_L \left(-\frac{x}{\lambda f}, -\frac{y}{\lambda f} \right) \\ &= \frac{1}{\eta} H_B \left(-\frac{x}{\lambda f}, -\frac{y}{\lambda f} \right) H \left(-\frac{x}{\lambda f}, -\frac{y}{\lambda f} \right) \\ &= \frac{1}{\eta} \text{rect} \left(\frac{x}{W_x} \right) \text{rect} \left(\frac{y}{W_y} \right) \\ &\quad \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \bar{w}[p, q] e^{\frac{j2\pi}{\lambda f} \{xpM\Delta_x + yqN\Delta_y\}} \end{aligned} \quad (4.20)$$

where $W_x = B_x \lambda f$, $W_y = B_y \lambda f$. It is easy to see that $m_F(x, y)$ corresponds to the complex transmittance of a periodic grating that is windowed in space, where the grating periods are $\frac{\lambda f}{M\Delta_x}$ and $\frac{\lambda f}{N\Delta_y}$ and the window widths are W_x and W_y . Because of Eq. 4.17, we should have $W_x > \frac{\lambda f}{\Delta_x}$ and $W_y > \frac{\lambda f}{\Delta_y}$, so at least $M \times N$ periods of the grating should be preserved after windowing. For the weights given in Eq. 4.11, $m_F(x, y)$ is illustrated in Fig. 4.15a for 5 periods in each dimension. (In this figure, $P = Q = 4$, $M = N = 256$, $\Delta_x = \Delta_y = 10\mu m$, $f = 10cm$, $\lambda = 632.9nm$.) Recall that the $\frac{1}{\eta}$ factor is included in Eq. 4.20 to ensure that $\max \{|m_F(x, y)|\} = 1$. The value of η is given as

$$\eta = \max \left\{ \left| \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \bar{w}[p, q] e^{j2\pi \{x'p+y'q\}} \right| \right\} \quad (4.21)$$

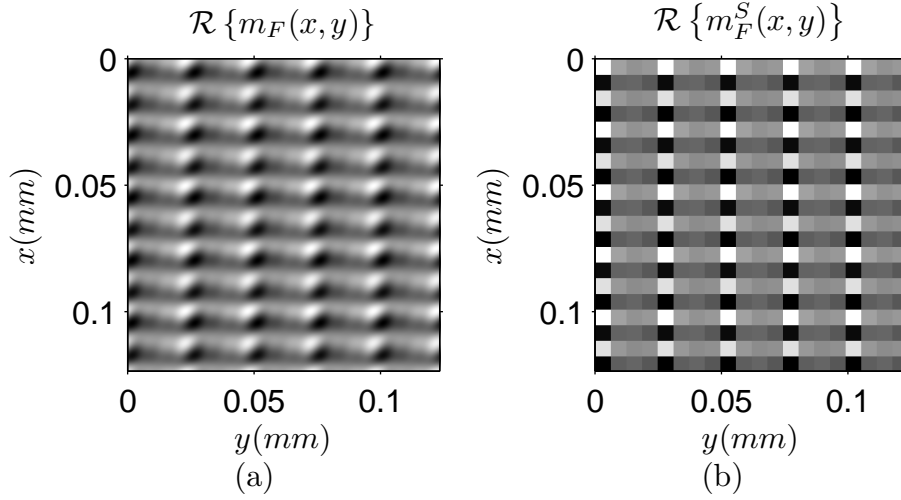


Figure 4.15: (a) Fourier plane mask for the weights given in Eq. 4.11. (b) Pixelated Fourier plane mask that should be used for the weights given in Eq. 4.11. Both masks are displayed for 5 periods in each dimension, and only real parts are shown. First appeared in [2].

for $x', y' \in \mathcal{R}$.

Finally, to select $\frac{1}{\eta}s_L(x,y)$ out of $u_{4f-}(x,y) = \frac{1}{\eta}g_L(x,y)$, we can use the following simple output plane mask

$$m_O(x,y) = \text{rect}\left(\frac{x - (P - 0.5)M\Delta_x}{M\Delta_x}\right) \text{rect}\left(\frac{y - (Q - 0.5)N\Delta_y}{N\Delta_y}\right). \quad (4.22)$$

At the end, we get $u_{4f+}(x,y) = u_{4f-}(x,y)m_O(x,y) \approx \frac{1}{\eta}s_L(x,y)$ as desired.

4.2.4 Discussion About the $4f$ Setup

Firstly, we should remind that the proposed $4f$ setup is analyzed using Fresnel scalar diffraction theory which is accurate under paraxial cases, i.e. the light rays traveling throughout the system must be confined to the vicinity of the optical axis and they should have small angles. Hence, the physical optical setup must be prepared accordingly. The binary SLM size should not be too large, and the focal length of the positive lenses should not be too small. Usually, these are

already a straightforward consequence of typical component sizes in an optical lab environment. We also assumed during the analysis that the overall bandwidth of the system is mainly restricted by the Fourier plane mask. This means, the lens apertures should not be too small, so that they do not cause a further restriction on the bandwidth. Under these conditions, Eq. 4.18 — Eq. 4.22 will provide a fairly accurate description of behavior of the physical setup.

Secondly, we should note that other optical implementations are also possible for the generic method proposed in Sec. 4.1. Assuming that we start with a single $PM \times QN$ binary SLM, the critical issue is that, the binary SLM should be divided into PQ sub-SLMs of size $M \times N$ and a *properly weighted superposition* of these sub-SLMs must be formed optically in a coherent manner. That superposition can be effectively formed using other optical components such as beam splitters or prisms. But in such options, each sub-SLM must be illuminated with a plane wave whose complex amplitude is equal to the corresponding weighting coefficient of that sub-SLM. Hence, a nonuniform illumination must be used for the binary SLM. In addition, since there are many sub-SLMs, we would need many beam splitters or prisms, whose physical dimensions must be suitable for placing side by side. All these factors complicate the implementation. But the presented $4f$ system only requires the lenses and the Fourier and output plane masks. A common plane wave illumination is sufficient for the entire binary SLM. Then, the $4f$ system automatically handles the mentioned properly weighted superposition of the sub-SLMs. Moreover, while creating the full-complex SLM, adjusting the widths of the Fourier plane mask, we can get rid of the diffraction orders of the SLM output. Hence, while not being the only possible option, we believe that the $4f$ system is a convenient option.

Finally, let us discuss the main drawbacks of the proposed $4f$ system. One of the significant drawbacks is that precise alignment is required between the optical components. For instance, if other components are perfectly placed but

the Fourier plane mask is slightly off-positioned on the transverse plane, the sub-SLMs will be superposed with weights that are different than intended, and this will result in a malfunctioning of the system. However, we believe that easy test procedures can be developed to achieve the required precision in alignment in an optical environment.

Another drawback might be due to the light efficiency of the $4f$ system. In practice, the input power (that is, the power used to illuminate the binary SLM) will be partly lost as the light passes through the binary SLM, the lenses, and the masks; so that only a fraction of the input power will be delivered to the full-complex SLM and to the observation region. Actually, if we ignore the losses due to the binary SLM, the lenses, and the finite aperture size of the Fourier plane mask; and if we assume that the binary SLM pixels are independently distributed and for each pixel the values -1 or 1 are equally likely, a straightforward analysis yields that on the average, the fraction of the input power delivered to the full-complex SLM (i.e., the light efficiency of the system) is given by

$$L_{\text{eff}} = \frac{1}{PQ} \frac{1}{\eta^2} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} |\bar{w}[p, q]|^2 \quad (4.23)$$

where η is as given in Eq. 4.21. Hence, L_{eff} depends on the selection for $\bar{w}[p, q]$. It can be shown that L_{eff} varies between $\frac{1}{(PQ)^2}$ and $\frac{1}{PQ}$. For instance, for the $P = Q = 4$ case, L_{eff} changes between 0.39% and 6.25%, and for the weights given in Eq. 4.11, it is about 1.08%. For some applications, these efficiencies might be low. But as we pointed in Sec. 4.1, for holographic purposes, even 4-bit quantization is usually sufficient [82, 154], so taking $P = 2$ and $Q = 2$, L_{eff} can be made to vary between 12.5% and 25%. We believe that at least for 3D display purposes, this efficiency is sufficient, comparable to that of other schemes based on binary SLMs, and can be tolerated to enjoy the benefits of having a full-complex SLM.

4.3 Pixellated and Quantized Fourier Plane Masks

The Fourier plane mask, denoted by $m_F(x, y)$ and given in Eq. 4.20, is the key component of the proposed $4f$ system. This mask should be physically produced and placed in the Fourier plane. The problem is, the mask given in Eq. 4.20 is a continuous function of space coordinates taking on continuously varying gray values, so it is hard to physically produce. In this section, we will consider the usage of pixellated and quantized Fourier plane masks, since such masks are easier to produce in practice.

Actually, the mask given in Eq. 4.20 is a continuous function of space coordinates because $H(\nu_x, \nu_y)$ given in Eq. 4.7 is a continuous function of ν_x and ν_y . Recall that $H(\nu_x, \nu_y)$ denotes the frequency response of the LSI system discussed in Sec. 4.2.1. Now, suppose instead of that system, we use another LSI system whose frequency response $H_S(\nu_x, \nu_y)$ is defined as:

$$H_S(\nu_x, \nu_y) = \text{rect}(\nu_x PM\Delta_x) \text{rect}(\nu_y QN\Delta_y) ** \left\{ H(\nu_x, \nu_y) \sum_{r=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \delta\left(\nu_x - \frac{r}{PM\Delta_x}, \nu_y - \frac{t}{QN\Delta_y}\right) \right\} \quad (4.24)$$

As seen, $H_S(\nu_x, \nu_y)$ is obtained by sampling $H(\nu_x, \nu_y)$ and then applying zero order interpolation on the resulting discrete signal. Recall that $H(\nu_x, \nu_y)$ is periodic with $\frac{1}{M\Delta_x}$ and $\frac{1}{N\Delta_y}$. Since it is sampled with sampling periods $\frac{1}{PM\Delta_x}$ and $\frac{1}{QN\Delta_y}$, $H_S(\nu_x, \nu_y)$ is also periodic with $\frac{1}{M\Delta_x}$ and $\frac{1}{N\Delta_y}$. Due to zero order hold interpolation, $H_S(\nu_x, \nu_y)$ has a piecewise constant structure. Actually, this is the main reason for us to consider the new LSI system instead of the original LSI system; because if we manage to create the full-complex SLM using the new LSI system, the new Fourier plane mask given as

$$m_F^S(x, y) = \frac{1}{\eta} H_B\left(-\frac{x}{\lambda f}, -\frac{y}{\lambda f}\right) H_S\left(-\frac{x}{\lambda f}, -\frac{y}{\lambda f}\right) \quad (4.25)$$

will become a pixellated mask. However, we should first show that we can also create the full-complex SLM using the new LSI system.

Taking the inverse Fourier transform of $H_S(\nu_x, \nu_y)$, we see that the impulse response of the new LSI system is given as

$$h_S(x, y) = \text{sinc}\left(\frac{x}{PM\Delta_x}\right) \text{sinc}\left(\frac{y}{QN\Delta_y}\right) \times \sum_{r=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} h(x - rPM\Delta_x, y - tQN\Delta_y) \quad (4.26)$$

where $h(x, y)$ denotes the impulse response of the original LSI system (see Eq. 4.6). As seen, sampling of $H(\nu_x, \nu_y)$ causes a periodic replication of $h(x, y)$ in space where the replicas are spaced by $PM\Delta_x$ and $QN\Delta_y$; and zero order hold interpolation creates the sinc roll-off factor. Similar to $h(x, y)$, $h_S(x, y)$ consists of impulses that are spaced by $M\Delta_x$ and $N\Delta_y$, but unlike $h(x, y)$, the number of impulses in $h_S(x, y)$ is infinite. Similar to what we did in Sec. 4.2.1, let $g_S(x, y) = b(x, y) ** h_S(x, y)$ and $s_S(x, y) = g_S(x, y) \text{rect}\left(\frac{x - (P-0.5)M\Delta_x}{M\Delta_x}\right) \text{rect}\left(\frac{y - (Q-0.5)N\Delta_y}{N\Delta_y}\right)$. It is easy to show that

$$s_S(x, y) = \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \bar{w}'[p, q] b_{p,q}(x - x_0, y - y_0) \quad (4.27)$$

where $x_0 = (P - 1)M\Delta_x$, $y_0 = (Q - 1)N\Delta_y$, and

$$\bar{w}'[p, q] = \bar{w}[p, q] \text{sinc}\left(\frac{p}{P}\right) \text{sinc}\left(\frac{q}{Q}\right) \quad (4.28)$$

for $0 \leq p \leq P - 1$ and $0 \leq q \leq Q - 1$. We see upon comparison of Eq. 4.27 with Eq. 4.10 that, when we use the new LSI system, the only change is, when forming $s_S(x, y)$, sub-SLMs are weighted by $\bar{w}'[p, q]$ instead of $\bar{w}[p, q]$. The main reason for this change is the zero order hold interpolation that is used when obtaining $H_S(\nu_x, \nu_y)$ from $H(\nu_x, \nu_y)$. However, this change does not create any problem. In particular, now we should specify $\bar{w}'[p, q]$ rather than specifying $\bar{w}[p, q]$. After specifying $\bar{w}'[p, q]$, we should find $\bar{w}[p, q]$ according to Eq. 4.28, and then we should design the new LSI system according to Eq. 4.7 and Eq. 4.24. When this is done, the new LSI system will produce the same output as the original

LSI system (i.e., $s_S(x, y) = s(x, y)$); implying that instead of the old Fourier plane mask given in Eq. 4.20, we can use the mask given in Eq. 4.25. This new mask, which has a pixellated structure, is also periodic in its spatial support with periods $\frac{\lambda f}{M\Delta_x}$ and $\frac{\lambda f}{N\Delta_y}$. The pixel widths of the new mask are given by $\frac{\lambda f}{PM\Delta_x}$ and $\frac{\lambda f}{QN\Delta_y}$. Therefore, in each period of the mask there are $P \times Q$ pixels. Note that since the mask widths must be greater than $\frac{\lambda f}{\Delta_x}$ and $\frac{\lambda f}{\Delta_y}$ by Eq. 4.17, the new mask should have at least $PM \times QN$ pixels.

If the physical production process only dictates that the Fourier plane mask should be pixellated, but no quantization on pixel values is required, given $\bar{w}'[p, q]$, we only need to compute $\bar{w}[p, q]$, $H(\nu_x, \nu_y)$ and $H_S(\nu_x, \nu_y)$ as explained above, and prepare the pixellated Fourier plane mask $m_F^S(x, y)$ according to Eq. 4.25. For instance, for the $P = Q = 4$ case, if we want $\bar{w}'[p, q]$ to be equal to the weights given in Eq. 4.11, the pixellated mask shown in Fig. 4.15b should be used. The mask in Fig. 4.15b produces the same output (possibly up to a constant amplitude factor) with the mask in Fig. 4.15a.

In practice, usually, there is also a quantization constraint on the pixel values of the Fourier plane mask. In such cases, the correct approach is to take the Fourier plane mask as given and determine the implied $\bar{w}'[p, q]$. Given $m_F^S(x, y)$; using simple Fourier transform relations, it can be shown that the implied $\bar{w}'[p, q]$ becomes:

$$\bar{w}'[p, q] = \frac{1}{16} \text{sinc}\left(\frac{p}{P}\right) \text{sinc}\left(\frac{q}{Q}\right) \sum_{r=0}^{P-1} \sum_{t=0}^{Q-1} m_F^S\left(\frac{r\lambda f}{PM\Delta_x}, \frac{t\lambda f}{QN\Delta_y}\right) e^{j2\pi\left(\frac{pr}{P} + \frac{qt}{Q}\right)} \quad (4.29)$$

for $0 \leq p \leq P - 1$ and $0 \leq q \leq Q - 1$. Note that in the above equation, $m_F^S\left(\frac{r\lambda f}{PM\Delta_x}, \frac{t\lambda f}{QN\Delta_y}\right)$ denotes the pixel values of the Fourier plane mask. In the presence of a quantization constraint, Fourier plane masks should be designed according to the above equation. In particular, we should first assign the values of the pixels of the mask taking into account the quantization constraint on them. Then, we should compute the implied weights according to the above equation,

and then we should compute the complex values achievable by a pixel of the new SLM based on these weights. If these values are few in number or have poor coverage of the complex plane, we should re-design the mask.

Assuming $P = Q = 4$, we will go through a number of examples and show that even Fourier plane masks with quite limited pixel values can lead to $\bar{w}'[p, q]$'s that generate a large number of complex values for the pixels of the new SLM. Fig. 4.16a illustrates a mask whose pixels are equal to ± 1 or $\pm j$. Hence, there are only four levels available for a pixel of the mask. Fig. 4.16b illustrates the complex numbers available for a pixel of the new SLM when we use this mask. There are 2^{16} different complex numbers on this figure. Another 4-level example is illustrated in Fig. 4.16c and Fig. 4.16d. The number of achievable complex numbers is again 2^{16} . (In Fig. 4.16a and Fig. 4.16c; white, light gray, dark gray and black pixels respectively have values $1, j, -j$ and -1 .) Therefore, even using the simple masks illustrated in Fig. 4.16, we can produce the new full-complex SLM without any loss in the information content. Moreover, the achievable complex numbers have a good coverage of the complex plane. Even simpler masks can be used if we accept a slight degradation in this coverage. Fig. 4.17a and Fig. 4.17c illustrate two masks whose pixels are equal to ± 1 or 0 . Hence, there are only three levels available for a pixel of the mask. Fig. 4.17b and Fig. 4.17d illustrate the resulting complex numbers that can be achieved. There are again 2^{16} different complex numbers in both figures, but their coverage of the complex plane is slightly worse than the 4-level examples. (In Fig. 4.17a and Fig. 4.17c; white, gray and black pixels respectively have values $1, 0$, and -1 .) Yet even simpler masks can be used if we accept to achieve a reduced number of complex numbers (that is, if we tolerate some loss in the information content). Fig. 4.18a and Fig. 4.18c illustrate two binary masks whose pixels are equal to ± 1 . Fig. 4.18b and Fig. 4.18d illustrate the resulting complex numbers that can be achieved. This time, there are only 2^{15} different complex numbers in both figures (implying that the full-complex SLM is 15-bit, so 1-bit

of information is lost per pixel), which is lower than 2^{16} , but which is still high, and the coverage of the complex plane is acceptable. (Fig. 4.18a and Fig. 4.18c; white and black pixels respectively have values 1 and -1 .) The 1-bit per pixel loss in the information content may be tolerated for the convenience of using binary masks which are quite easy to physically produce. These examples show that as the quantization constraint on the Fourier plane mask gets harsher, the number of available complex values for a pixel of the SLM can decrease and the coverage of the complex plane can get worse. However, since the number of available levels is still large, given a typical desired full-complex SLM pattern, the quantization error will be still quite low (though it may increase slightly) and no noticeable degradation in final reconstruction quality will take place.

To sum up, Eq. 4.27 and Eq. 4.28 indicate that instead of any continuous Fourier plane mask, we can design and use an equivalent pixellated mask and get the same output. And the examples through Fig. 4.16—Fig. 4.18 indicate that even in the case of a severe quantization constraint, it is possible to design Fourier plane masks such that the complex values that are available for a pixel of the full-complex SLM are large in number and have a good coverage of the complex plane. Therefore, pixellation and quantization of the Fourier plane mask do not cause any noticeable degradation in the system performance in terms of reconstruction quality. However, in the case of a pixellated mask, the light efficiency will be slightly decreased relative to the continuous mask case. This is because of the fact that the pixellated mask will cause the emergence of higher order waves which divert some of the input power. These waves travel in high angles and are blocked at the output plane mask stage, causing a smaller portion of the input power to be delivered to the full-complex SLM and thus to the observation region. Roughly, on the average, the efficiency will be decreased by about 20% at this stage relative to the continuous mask case. This decrease can be minimized if phase only Fourier plane masks such as the ones shown in Fig. 4.16 and Fig. 4.18 are used. We assume that this additional loss can be

tolerated for the convenience of using pixellated and quantized Fourier plane masks.

4.4 Conclusion

In this chapter, we first proposed a generic method for effectively creating full-complex SLMs out of binary SLMs. The method relies on forming a properly weighted superposition of binary SLMs. We showed that in this manner, information-wise, binary SLMs are utilized in the most efficient manner. Then, we proposed a $4f$ system as a possible optical implementation of our generic method. In addition to forming the full-complex SLM, this $4f$ setup also enables us to get rid of the disturbing higher diffraction orders of the SLM output. We showed that the parameters and components of the system can easily be customized for different production technologies. One main drawback of the system is the precise alignment requirement, but we believe that easy to apply optical test procedures can be designed to satisfy it. Another drawback may be due to light efficiency, but we assume that in 3D display applications, the levels are tolerable. Compared to previous approaches, the most important feature of our approach is that we tried to use the full potential of the binary SLMs when creating the full-complex SLMs. Actually, our generic method can be tailored to create full-complex SLMs out of multilevel amplitude-only or phase-only SLMs. In this case, less complicated optical systems can be used for optical implementation. However, we believe that the robust behavior of binary SLMs justify our choice for selecting them to create the full complex SLM. We believe that for commercially available binary SLMs, the proposed $4f$ system can be implemented within a small volume. Therefore, multiple replicas of the $4f$ system can be conveniently placed side by side to create full-complex SLM arrays to be used in applications such as 3D displays.

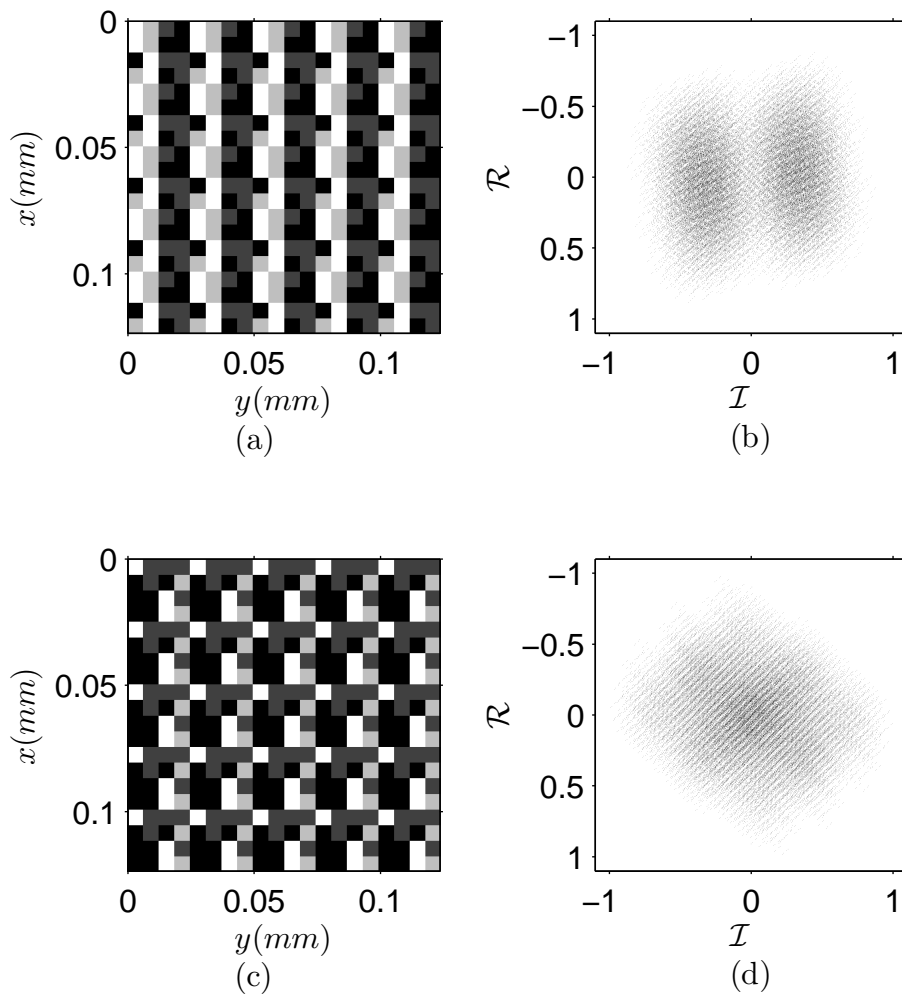


Figure 4.16: (a) A 4-level Fourier plane mask. (b) Achievable complex numbers. (c) Another 4-level mask. (d) Achievable complex numbers. First appeared in [2].

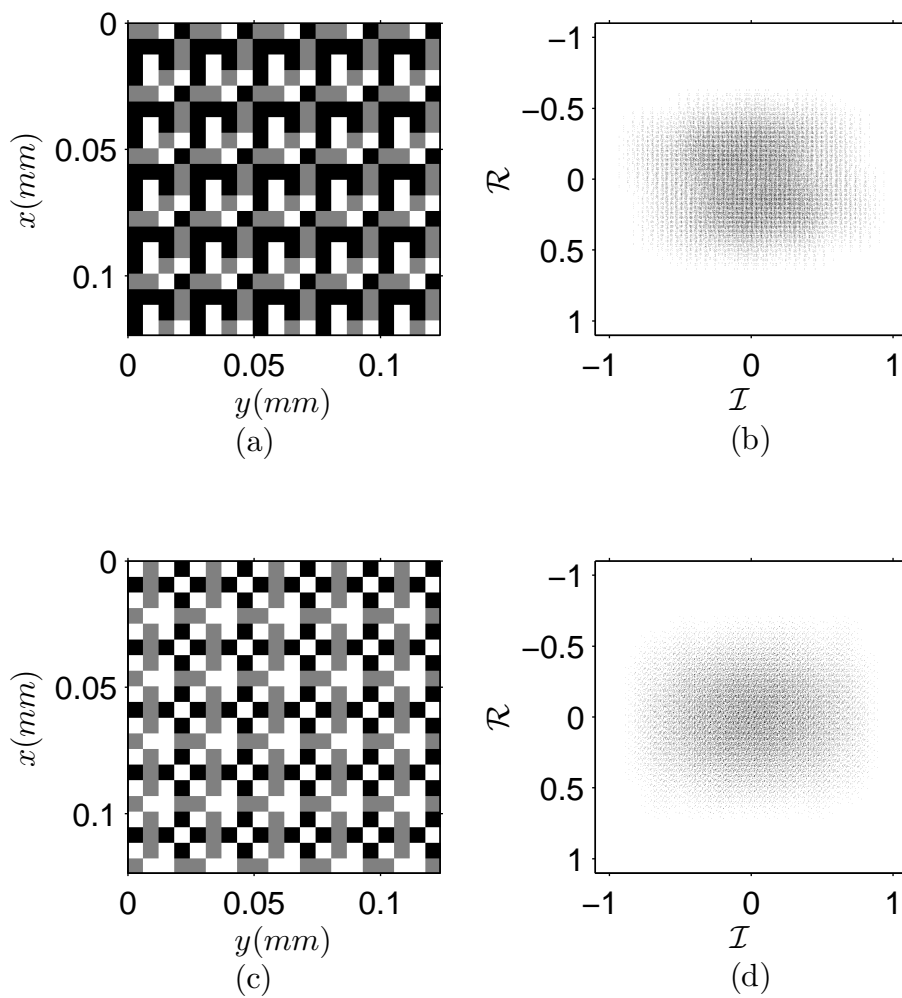


Figure 4.17: (a) A 3-level Fourier plane mask. (b) Achievable complex numbers. (c) Another 3-level mask. (d) Achievable complex numbers. First appeared in [2].

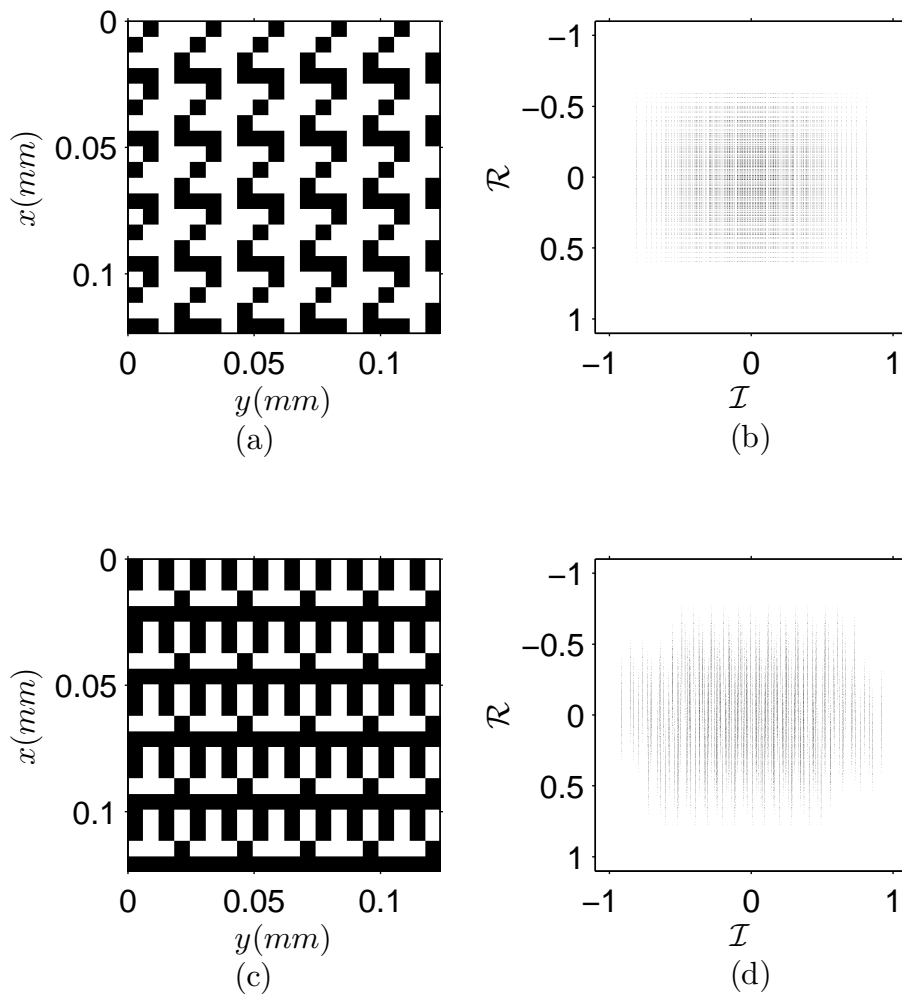


Figure 4.18: (a) A binary Fourier plane mask. (b) Achievable complex numbers. (c) Another binary mask. (d) Achievable complex numbers. First appeared in [2].

Chapter 5

SUMMARY AND CONCLUSIONS

5.1 Summary

In this thesis, we study the signal processing problems related to the holographic 3D video displays constructed using binary SLMs. In Chapter 1, we explore several approaches to realize a 3D video display, and highlight the holographic approach as the most promising alternative despite the fact that many crucial problems still remain to be solved. In particular, we identify the imperfections (such as amplitude coupled phase modulation or vice versa) and non-robust behavior (such as drastic changes in pixel values with wavelength) of currently existing multilevel SLMs as troublesome factors that hinder the quality of holographic displays. We suggest that most of the related problems can be avoided by using the more robust binary SLMs, if only the associated signal processing problems (“determination of binary SLM patterns that synthesize desired light fields” and “achieving full-complex modulation using binary SLMs”) are solved

to a satisfactory extent. We undertake these problems. As a review of the background material, in Chapter 2, we discuss the basics of scalar wave optics theory and scalar diffraction theory. We also include in that chapter an analysis of the light field generated by a finite-size SLM. This analysis is particularly important for understanding the restrictions that the SLM structure imposes on the light fields that are wished to be generated.

In Chapter 3, we undertake the problem of determination of binary SLM patterns that synthesize desired light fields. We first concentrate on the determination of the spatial domain over which it makes sense to specify the desired light field in the case of a binary SLM. For this, we examine the effects of applying a low-pass filter to the SLM pattern written on a finite size SLM, and find out that the light field generated by the SLM approximately remains unchanged within a sub-region of the central order region. We choose this region as the synthesis region. We show that if the desired light field is confined within the synthesis region, the ideal gray level complex-valued SLM pattern generating it becomes a low-pass pattern. We show that this complex-valued pattern can be successfully halftoned into a binary SLM pattern if the pixels of the binary SLM are partitioned into two groups, such that the first group is used for halftoning the real part and the second group is used for halftoning the imaginary part. To optically implement the mentioned partition, we first propose placing a thin mask after the SLM. Then we show that if we remove the mask and instead use oblique illumination at a specific incidence angle, we can still achieve the desired partition. Finally, we show that proper adjustment of the incidence angle is not critical from a visual viewpoint. With these solutions, we effectively extend the theory of halftoning for classical real-valued images to complex-valued holograms. Our simulations illustrate that the proposed technique can be used to generate planar as well as volumetric light field distributions.

In Chapter 4, we undertake the problem of achieving full-complex modulation using binary SLMs. Here, compared to Chapter 3, we take a different approach and focus on effectively creating a full complex SLM out of binary SLMs. For this purpose, we first propose a generic method which is based on the well known concepts of bit plane representation and decomposition for ordinary gray level images. In particular, we effectively create a full-complex SLM by forming a properly weighted superposition of a number of binary SLMs, where superposition weights are allowed to be complex-valued as well. The method actually involves a trade off between pixel count and dynamic range. We show that with our method, information-wise, binary SLMs are utilized in the most efficient manner that is possible. As a possible optical implementation, we propose a $4f$ system. In addition to forming the full-complex SLM, this $4f$ setup also enables us to get rid of the disturbing side beams generated by the SLM. We show that the parameters and components of the system can easily be customized for different production technologies, and even in the case of severe restrictions (such as a binary Fourier plane mask), the values available for a pixel of the new SLM may have a quite satisfactory coverage of the complex plane.

5.2 Conclusions

When we decided to undertake this thesis, we made the observation that the theory of halftoning for ordinary gray scale images has been excelled to such an extent that binary printers have been the paradigm in the printing industry for quite many decades. Based on this observation, we decided that if the theory of halftoning for ordinary gray scale images is extended to complex-valued holograms with a correct approach (in which the optics and signal processing related aspects of the problem are treated in a coherent manner), it should be possible to achieve quite satisfactory holographic reconstructions with binary SLMs

in a simple manner (without using complicated optical setups or making intensive computations) as well. We make the mentioned extension in Chapter 3 of this thesis. As our simulation results indicate, with our approach, it is possible to synthesize high quality light fields by using binary SLMs just as multilevel SLMs. The achieved quality is comparable to that provided by a lower resolution full complex SLM; resembling the image quality offered by a binary printer. Therefore, there is no reason to think any more that because of their low dynamic range, binary SLMs are intrinsically inadequate for holographic display purposes. Indeed, it is possible to make holographic displays more efficient and flexible by using the more robust binary SLMs.

Most of the methods proposed until now on the problem studied in Chapter 3 assume that the mathematical relation between the binary SLM pattern and the desired light field is given by a Fourier transform, and solve the problem accordingly [122, 124, 126, 127, 128, 129]. To be successful, these methods require the desired light field to be specified on a plane that lies either in the far field or on the Fourier plane of a $2f$ setup. On the other hand, in our formulation, we assume from the very beginning that the desired light field lies in the non-far field range and allow it to be specified within a volume as well. And as our simulation results indicate, with our method, we can successfully synthesize light fields specified within a volume that lies in the non-far-field range merely after free-space propagation from the SLM plane. In this sense, our method requires optical setups that are much less complicated. Moreover, it gives users more flexibility in the specification of desired light fields. Therefore, especially within the context of holographic displays, our solutions are much ready and suitable to be used.

Another main improvement provided by the solution we develop in Chapter 3 becomes apparent when we consider its algorithmic details. Recall that in

our solution, given a properly specified desired field, we first compute the associated ideal complex-valued gray level pattern, and then solve two decoupled real-valued halftoning problems to find the required binary SLM pattern. In this way, we decouple the hologram related computations and halftoning related computations. This aspect makes our approach much more modular compared to many of the algorithms proposed before. And this modularity has quite important advantages. For instance, the designer has the freedom to choose any of the existing algorithms for hologram computation. As long as the hologram is computed with sufficient accuracy, the details of the algorithm are unimportant. Similarly, as long as the desired accuracy is achieved, halftoning related computations can be carried out using any of the existing algorithms. (Indeed, it can be noticed that in Chapter 3 we do not propose any particular algorithm either for hologram computation or for classical halftoning. We only identify the places in which such algorithms are needed, and deliberately leave the selection of the related algorithm to the designer.) A straightforward consequence is that, the overall accuracy and computational complexity of our approach is determined by the individual accuracies and computational complexities of the algorithms used for hologram computation and halftoning. In this respect, the accuracy and computational performance of our solution will directly benefit from any possible improvements in the methods for hologram computation and halftoning. It is hard to say that the algorithms developed for binary SLM pattern computation before our study have these nice properties. Most of these algorithms are iterative algorithms where during a typical iteration, the output field produced by some current binary SLM pattern is computed, and then that binary SLM pattern is updated according to the error between the output and desired fields [121, 130, 131, 132, 133]. Therefore, in these algorithms, free space propagation and halftoning related computations are strongly mingled. This mingled structure greatly increases the computational complexity of those iterative algorithms. Moreover, since those algorithms are also not modular, they lose the

chance to benefit from the improvements in the algorithms for halftoning or hologram computation. Actually, examination of the related literature will reveal that many researchers complain about the poor computational performance of existing algorithms [134, 135, 136, 137]. This is understood to be one of the reasons that binary SLMs are considered in holographic display designs relatively less frequently compared to multilevel SLMs. In this respect, we find out that besides increasing the reconstruction quality significantly, our solution also provides important improvements in computation performance and modularity, paving the way for more frequent employment of binary SLMs in holographic display designs.

In the beginning of Chapter 4, we explain that most of the schemes proposed until now for achieving full complex modulation by using restricted type multilevel SLMs suffer from the imperfections and non-robust behavior of those multilevel SLMs [145, 146, 147, 148]. We indicate that in the proposed schemes, especially the distribution of the values achievable by a pixel of the new SLM on the complex plane is poor; and that the proposed schemes operated in the desired fashion only within a limited wavelength range [149, 150, 151, 152]. Then we propose our scheme which uses binary SLMs. Though our scheme still has some drawbacks such as poor light efficiency or precise alignment requirements, we show through our simulations that even in the presence of severe practical constraints on system components, the coverage of the complex plane is much more satisfactory than that achieved by previous schemes. Meanwhile, the scheme can be tailored to operate at any wavelength within a wide range. Noting that the drawbacks of our scheme are also shared to a certain extent by previous schemes, we can say that our solution turns out to be a strong option for designers, especially to those who wish enjoy the robust nature of binary SLMs. Actually, we believe that our design also has the potential to be quite beneficial and instructive to SLM designers as well. For instance, rather than trying to design a physical SLM which provides full complex modulation in a direct

manner, these designers can focus on the optical implementation of the $4f$ system that we propose in Chapter 4. We believe that possibly with some minor modifications, this $4f$ system can be implemented within a small volume. If the required alignment precision is also achieved, it is possible to effectively obtain a full-complex SLM in an indirect manner, where that full complex SLM will most probably have much superior qualifications compared to any existing multilevel SLM (better coverage of complex plane, eliminated diffraction orders, etc.). Needless to say, holographic displays will enormously benefit from all these possible developments.

To sum up, we believe that in this thesis, we provide systematic solutions to the signal processing problems that arise within the context of holographic displays constructed using binary SLMs and that were thought to be challenging. The solutions we present are quite ready to be directly applied to these problems. Therefore, we expect that with the increased usage of the more robust binary SLMs, holographic displays will be more robust, efficient and flexible. Actually, in our belief, another consequence of this study is that, researchers working on SLM design (in fields such as MEMS, MOEMS, nanotechnology etc.) can now concentrate on developing robust binary SLMs with improved properties (such as finer pixels, higher light throughput, larger pixel count, smaller response time etc.) rather than trying to improve the dynamic range of currently existing non-robust multilevel SLMs. Our signal processing based methods can easily compensate for the low dynamic range of binary SLMs, making it possible for us to enjoy their many advantages.

Bibliography

- [1] E. Ulusoy, L. Onural, and H. M. Ozaktas, “Synthesis of three-dimensional light fields with binary spatial light modulators,” *Journal of the Optical Society of America A*, vol. 28, pp. 1211–1223, 2011.
- [2] E. Ulusoy, L. Onural, and H. M. Ozaktas, “Full-complex amplitude modulation with binary spatial light modulators,” *Journal of the Optical Society of America A*, vol. 28, pp. 2310–2321, 2011.
- [3] E. Ulusoy, L. Onural, and H. M. Ozaktas, “Signal processing for three-dimensional holographic television displays that use binary spatial light modulators,” in *Proceedings of IEEE Conference on Signal Processing and Communications Applications Conference*, pp. 41–44, 2010.
- [4] T. Okoshi, *Three Dimensional Imaging Techniques*. New York Academic Press, 1976.
- [5] S. A. Benton, *Three Dimensional Imaging*. SPIE, 1977.
- [6] J. Ebbeni and A. Monfils, *Three Dimensional Imaging*. SPIE, 1983.
- [7] M. W. Vannier, J. Marsh, and P. R. Biondetti, *Three Dimensional Imaging*. Pergamon Press, 1988.
- [8] G. T. Herman, *Three Dimensional Imaging*. Wiley, 2000.
- [9] B. Javidi, F. Okano, and J. Y. Son, *Three Dimensional Imaging, Visualization, and Display*. Springer, 2009.

- [10] O. Schreer, P. Kauff, and T. Sikora, *3D Videocommunication: Algorithms, Concepts, and Real-time Systems in Human Centred Communication*. Wiley, 2005.
- [11] L. Onural, *3D Video Technologies: An Overview of Research Trends*. SPIE Press, 2011.
- [12] M. Mrak, M. Grgic, and M. Kunt, *High-Quality Visual Experience: Creation, Processing and Interactivity of High-Resolution and High-Dimensional Video Signals*. Springer, 2010.
- [13] R. Ronfard and G. Taubin, *Image and Geometry Processing for 3D Cinematography*. Springer, 2010.
- [14] B. Javidi and F. Okano, *Three-Dimensional Video and Display: Devices and Systems*. SPIE, 2001.
- [15] B. Mendiburu, *3D TV and 3D Cinema: Tools and Processes for Creative Stereoscopy*. Elsevier Science, 2011.
- [16] E. B. Goldstein, *Sensation and Perception*. Wadsworth Cengage Learning, 2009.
- [17] I. P. Howard, *Perceiving in Depth, Volume 1: Basic Mechanisms*. Oxford University Press, USA, 2012.
- [18] I. P. Howard and B. J. Rogers, *Perceiving in Depth, Volume 2: Stereoscopic Vision*. Oxford University Press, USA, 2012.
- [19] N. A. Valius, *Stereoscopy*. Focal P., 1966.
- [20] J. Merritt and S. S. Fisher, *Stereoscopic Displays and Applications*. SPIE, 1990.
- [21] I. Sexton and P. Surman, "Stereoscopic and autostereoscopic display systems: An in-depth review of past, present and future technologies," *IEEE Signal Processing Magazine*, vol. 16, pp. 85–99, 1999.

- [22] A. J. Woods, *Stereoscopic Displays and Virtual Reality Systems XII*. SPIE, 2005.
- [23] F. P. Miller, A. F. Vandome, and J. McBrewster, *Stereoscopy: Head-mounted Display, LCD Shutter Glasses, Anaglyph Image, Autostereogram, Pulfrich Effect, Lenticular Printing, Stereopsis, Photogrammetry, Stereogram, Computer Vision, Depth Perception, Binocular Vision, CIE 1931 Color Space*. Alphascript Publishing, 2009.
- [24] J. E. Melzer and K. W. Moffitt, *Head-Mounted Displays: designing for the user*. McGraw-Hill, 1997.
- [25] M. Halle, “Autostereoscopic displays and computer graphics,” *Computer Graphics, ACM SIGGRAPH*, vol. 31, pp. 58–62, 1997.
- [26] N. A. Dodgson, “Autostereoscopic 3D displays,” *Computer*, vol. 38, pp. 31–36, 2005.
- [27] K. Talmi and J. Liu, “Eye and gaze tracking for visually controlled interactive stereoscopic displays,” *Signal Processing: Image Communication*, vol. 14, pp. 799–810, 1999.
- [28] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, pp. 519–528, 2006.
- [29] F. L. Kooi and A. Toet, “Visual comfort of binocular and 3D displays,” *Displays*, vol. 25, pp. 99–108, 2004.
- [30] M. Lambooi, W. IJsselsteijn, and M. Fortuin, “Visual discomfort and visual fatigue of stereoscopic displays: A review,” *Journal of Imaging Science and Technology*, vol. 53, pp. 1–14, 2009.

- [31] L. Onural, “Television in 3D: What are the prospects?,” *Proceedings of the IEEE*, vol. 95, pp. 1143–1145, 2007.
- [32] R. Zone, *Stereoscopic Cinema and the Origins of 3D film, 1838-1952*. The University Press of Kentucky, 2007.
- [33] D. Gabor, “A new microscopic principle,” *Nature*, vol. 161, pp. 777–778, 1948.
- [34] D. Gabor, “Holography, 1948-1971,” *Proceedings of the IEEE*, vol. 60, pp. 655–668, 1972.
- [35] R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*. Bell Telephone Laboratories, 1983.
- [36] P. Hariharan, *Basics of Holography*. Cambridge University Press, 2002.
- [37] C. J. Kuo and M. H. Tsai, *Three-Dimensional Holographic Imaging*. Wiley, 2002.
- [38] G. Saxby, *Practical Holography*. Institute of Physics Publication, 2004.
- [39] G. K. Ackermann, J. Eichler, and J. Eichler, *Holography: A Practical Approach*. John Wiley and Sons, 2008.
- [40] S. A. Benton and V. M. Bove, *Holographic Imaging*. Wiley-Interscience, 2008.
- [41] H. M. Smith and R. A. Bartolini, *Holographic Recording Materials*. Springer-Verlag, 1977.
- [42] H. J. Caulfield and E. N. Leith, *The Art and Science of Holography: a Tribute to Emmett Leith and Yuri Denisyuk*. SPIE, 2004.
- [43] S. Johnston, *Holographic Visions: a History of New Science*. Oxford University Press, 2006.

- [44] J. E. Ludman, H. J. Caulfield, and J. Riccobono, *Holography for the New Millennium*. Springer, 2002.
- [45] L. Onural, “Research trends in holographic 3DTV displays,” in *LEOS Annual Meeting Conference Proceedings, 2009.*, pp. 6–7, 2009.
- [46] L. Onural, “An overview of research in 3DTV,” in *Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, p. 3, 2007.
- [47] L. Onural, “Holographic 3DTV research within the european 3DTV project,” in *Digital Holography and Three-Dimensional Imaging*, p. DWA1, Optical Society of America, 2008.
- [48] L. Onural, F. Yaras, and H. Kang, “Current research activities on holographic video displays,” in *SPIE Conference Series*, vol. 7690, 2010.
- [49] T. C. Poon, *Digital Holography and Three-Dimensional Display: Principles and Applications*. Springer, 2006.
- [50] U. Schnars and W. Jueptner, *Digital Holography: Digital Hologram Recording, Numerical Reconstruction, and Related Techniques*. Springer, 2005.
- [51] L. Onural, A. Gotchev, H. M. Ozaktas, and E. Stoykova, “A survey of signal processing problems and tools in holographic three-dimensional television,” *IEEE Trans on Circuits and Systems for Video Technology*, vol. 17, pp. 1631–1646, 2007.
- [52] L. Onural and H. M. Ozaktas, “Signal processing issues in diffraction and holographic 3DTV,” *Sign Pro: Image Comm*, vol. 22, pp. 169–177, 2007.
- [53] M. Lucente and T. A. Galyean, “Rendering interactive holographic images,” in *Proc. of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 387–394, 1995.

- [54] M. Lucente, “Interactive three-dimensional holographic displays: seeing the future in depth,” *SIGGRAPH Comput. Graph.*, vol. 31, pp. 63–67, 1997.
- [55] Y. Frauel, T. J. Naughton, O. Matoba, E. Tajahuerce, and B. Javidi, “Three-dimensional imaging and processing using computational holographic imaging,” *Proc. IEEE*, vol. 94, pp. 636–653, 2006.
- [56] G. B. Esmer, V. Uzunov, L. Onural, H. M. Ozaktas, and A. Gotchev, “Diffraction field computation from arbitrarily distributed data points in space,” *Signal Processing: Image Communication*, vol. 22, pp. 178–187, 2007.
- [57] M. Janda, I. Hanak, and L. Onural, “Hologram synthesis for photorealistic reconstruction,” *J. Opt. Soc. Am. A*, vol. 25, pp. 3083–3096, 2008.
- [58] G. B. Esmer, L. Onural, and H. M. Ozaktas, “Exact diffraction calculation from fields specified over arbitrary curved surfaces,” *Optics Communications*, vol. 284, pp. 5537–5548, 2011.
- [59] S. A. Benton, *Selected Papers on Three-Dimensional Displays*. SPIE Optical Engineering Press, 2000.
- [60] H. M. Ozaktas and L. Onural, *Three-Dimensional Television: Capture, Transmission, Display*. Springer, 2008.
- [61] J. W. Goodman, *Introduction to Fourier Optics*. McGraw-Hill, 1996.
- [62] L. Yaroslavsky, *Digital Holography and Digital Image Processing: Principles, Methods, Algorithms*. Kluwer Academic, 2004.
- [63] I. Yamaguchi and T. Zhang, “Phase-shifting digital holography,” *Optics Letters*, vol. 22, pp. 1268–1270, 1997.
- [64] G. Tricoles, “Computer-generated holograms: an historical review,” *Appl. Opt.*, vol. 26, pp. 4351–4360, 1987.

- [65] O. Bryngdahl and F. Wyrowski, “Digital holography - computer-generated holograms,” in *Progress in Optics* (E. Wolf, ed.), vol. 28, pp. 1 – 86, Elsevier, 1990.
- [66] J. Y. Chen, *Computer Generated Holography*. University of Massachusetts. Lowell, 1991.
- [67] S. H. Lee, *Selected Papers on Computer Generated Holograms and Diffractive Optics*. SPIE Optical Engineering Press, 1992.
- [68] D. Abookasis and J. Rosen, “Computer-generated holograms of three-dimensional objects synthesized from their multiple angular viewpoints,” *J. Opt. Soc. Am. A*, vol. 20, pp. 1537–1545, 2003.
- [69] C. Slinger, C. Cameron, and M. Stanley, “Computer-generated holography as a generic display technology,” *Computer*, vol. 38, pp. 46–53, 2005.
- [70] W. J. Dallas, “Computer-generated holograms,” in *Digital Holography and Three-Dimensional Display* (T. C. Poon, ed.), pp. 1–49, Springer US, 2006.
- [71] B. R. Brown and A. W. Lohmann, “Complex spatial filtering with binary masks,” *Appl. Opt.*, vol. 5, pp. 967–969, 1966.
- [72] A. W. Lohmann and D. P. Paris, “Binary Fraunhofer holograms, generated by computer,” *Appl. Opt.*, vol. 6, pp. 1739–1748, 1967.
- [73] J. P. Waters, “Three-dimensional Fourier-transform method for synthesizing binary holograms,” *J. Opt. Soc. Am.*, vol. 58, pp. 1284–1288, 1968.
- [74] R. A. Gabel and B. Liu, “Minimization of reconstruction errors with computer-generated binary holograms,” *Appl. Opt.*, vol. 9, pp. 1180–1191, 1970.
- [75] A. R. Sass, “Binary intensity holograms,” *J. Opt. Soc. Am.*, vol. 61, pp. 910–915, 1971.

- [76] P. L. Ransom and R. F. Henton, “Analysis of a computer-generated binary-phase hologram,” *Appl. Opt.*, vol. 13, pp. 2765–2767, 1974.
- [77] W. Lee, “Binary synthetic holograms,” *Appl. Opt.*, vol. 13, pp. 1677–1682, 1974.
- [78] R. A. Gabel, “Reconstruction errors in computer-generated binary holograms: a comparative study,” *Appl. Opt.*, vol. 14, pp. 2252–2255, 1975.
- [79] W. Lee, “Binary computer-generated holograms,” *Appl. Opt.*, vol. 18, pp. 3661–3669, 1979.
- [80] M. Kovachev, R. Ilieva, P. Benzie, G. B. Esmer, L. Onural, J. Watson, and T. Reyhan, “Holographic 3DTV displays using spatial light modulators,” in *Three-Dimensional Television: Capture, Transmission, Display* (H. M. Ozaktas and L. Onural, eds.), pp. 529–555, Springer, 2008.
- [81] F. Yaras, H. Kang, and L. Onural, “State of the art in holographic displays: A survey,” *Journal of Display Technology*, vol. 6, pp. 443–454, 2010.
- [82] L. Onural, F. Yaras, and H. Kang, “Digital holographic three-dimensional video displays.” Accepted for publication in *Proc. IEEE*.
- [83] F. Yaras, H. Kang, and L. Onural, “Circularly configured multi-SLM holographic display system,” in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, pp. 1–4, 2011.
- [84] J. A. Neff, R. A. Athale, and S. H. Lee, “Two-dimensional spatial light modulators: a tutorial,” *Proc. IEEE*, vol. 78, pp. 826–855, 1990.
- [85] U. Efron, *Spatial Light Modulator Technology: Materials, Devices, and Applications*. Marcel Dekker, 1994.
- [86] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. von Kopylow, “A survey of 3DTV displays: Techniques

and technologies,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1647–1658, 2007.

- [87] R. W. Cohn, “Pseudorandom encoding of complex-valued functions onto amplitude-coupled phase modulators,” *J. Opt. Soc. Am. A*, vol. 15, pp. 868–883, 1998.
- [88] R. W. Cohn, “Analyzing the encoding range of amplitude-phase coupled spatial light modulators,” *Optical Engineering*, vol. 38, pp. 361–367, 1999.
- [89] Y. Yang, H. Stark, D. Gurkan, C. L. Lawson, and R. W. Cohn, “High-diffraction-efficiency pseudorandom encoding,” *J. Opt. Soc. Am. A*, vol. 17, pp. 285–293, 2000.
- [90] C. Stolz, L. Bigue, and P. Ambs, “Implementation of high-resolution diffractive optical elements on coupled phase and amplitude spatial light modulators,” *Appl. Opt.*, vol. 40, pp. 6415–6424, 2001.
- [91] V. Arrizon, “Optimum on-axis computer-generated hologram encoded into low-resolution phase-modulation devices,” *Optics Letters*, vol. 28, pp. 2521–2523, 2003.
- [92] J. A. Davis, K. O. Valadez, and D. M. Cottrell, “Encoding amplitude and phase information onto a binary phase-only spatial light modulator,” *Appl. Opt.*, vol. 42, pp. 2003–2008, 2003.
- [93] V. Arrizon, G. Mendez, and D. S. de La-Llave, “Accurate encoding of arbitrary complex fields with amplitude-only liquid crystal spatial light modulators,” *Optics Express*, vol. 13, pp. 7914–7927, 2005.
- [94] V. Arrizon, U. Ruiz, R. Carrada, and L. A. Gonzalez, “Pixelated phase computer holograms for the accurate encoding of scalar complex fields,” *J. Opt. Soc. Am. A*, vol. 24, pp. 3500–3507, 2007.

- [95] T. Kreis, P. Aswendt, and R. Hoffing, “Hologram reconstruction using a digital micromirror device,” *Opt. Eng.*, vol. 40, pp. 926–933, 2001.
- [96] B. Javidi and F. Okano, *Three-Dimensional Television, Video and Display Technology*. Springer, 2002.
- [97] D. Dudley, W. Duncan, and J. Slaughter, “Emerging Digital Micromirror Device (DMD) applications,” 2003. White Paper, Texas Instruments.
- [98] F. P. Miller, A. F. Vandome, and J. McBrewster, *Digital Micromirror Device*. VDM Publishing House Ltd., 2010.
- [99] M. E. Motamedi, *MOEMS: Micro-Opto-Electro-Mechanical Systems*. SPIE Press, 2005.
- [100] E. Ulusoy, G. Esmer, H. M. Ozaktas, L. Onural, A. Gotchev, and V. Uzunov, “Signal processing problems and algorithms in display side of 3DTV,” in *Proceedings of 2006 IEEE Conference on Image Processing*, pp. 2985–2988, 2006.
- [101] E. Ulusoy, V. Uzunov, L. Onural, H. M. Ozaktas, and A. Gotchev, “Three-dimensional monochromatic light field synthesis with a deflectable mirror array device,” in *Proceedings of SPIE Volume 6187*, 2006.
- [102] E. Ulusoy, L. Onural, and H. M. Ozaktas, “Analysis of the complex light field generated by a deflectable mirror array device,” in *Proceedings of SPIE Volume 6252*, 2006.
- [103] E. Ulusoy, L. Onural, H. M. Ozaktas, V. Uzunov, and A. Gotchev, “Three-dimensional complex scalar light field synthesis with a deflectable mirror array device,” in *Proceedings of 2nd Workshop on Immersive Communication and Broadcast Systems, ICOB 2005, Berlin, Germany*, 2005.
- [104] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics, 2nd edition*. Wiley, 2007.

- [105] G. C. Sherman, “Application of the convolution theorem to Rayleigh’s integral formulas,” *J. Opt. Soc. Am.*, vol. 57, pp. 546–547, 1967.
- [106] E. Lalor, “Conditions for the validity of the angular spectrum of plane waves,” *J. Opt. Soc. Am.*, vol. 58, pp. 1235–1237, 1968.
- [107] L. Onural, “Exact analysis of the effects of sampling of the scalar diffraction field,” *J. Opt. Soc. Am. A*, vol. 24, pp. 359–367, 2007.
- [108] L. Onural, “Sampling of the diffraction field,” *Appl. Opt.*, vol. 39, pp. 5929–5935, 2000.
- [109] S. B. Tucker, J. O. Castaneda, and W. T. Cathey, “Matrix description of near-field diffraction and the fractional Fourier transform,” *J. Opt. Soc. Am. A*, vol. 16, pp. 316–322, 1999.
- [110] H. M. Ozaktas, S. O. Arik, and T. Coskun, “Fundamental structure of Fresnel diffraction: natural sampling grid and the fractional Fourier transform,” *Optics Letters*, vol. 36, pp. 2524–2526, 2011.
- [111] H. M. Ozaktas, S. O. Arik, and T. Coskun, “Fundamental structure of Fresnel diffraction: longitudinal uniformity with respect to fractional Fourier order,” *Optics Letters*, vol. 37, pp. 103–105, 2012.
- [112] F. Gori, “Fresnel transform and sampling theorem,” *Opt. Comm.*, vol. 39, pp. 293–297, 1981.
- [113] F. Wyrowski, “Diffractive optical elements: iterative calculation of quantized, blazed phase structures,” *J. Opt. Soc. Am. A*, vol. 7, pp. 961–969, 1990.
- [114] J. N. Mait, “Understanding diffractive optic design in the scalar domain,” *J. Opt. Soc. Am. A*, vol. 12, pp. 2145–2158, 1995.

- [115] Y. H. Wu and P. Chavel, "Cell-oriented on-axis computer-generated holograms for use in the Fresnel diffraction mode," *Appl. Opt.*, vol. 23, pp. 228–238, 1984.
- [116] M. Li, A. Larsson, N. Eriksson, and M. Hagberg, "Continuous-level phase-only computer-generated hologram realized by dislocated binary gratings," *Opt. Lett.*, vol. 21, pp. 1516–1518, 1996.
- [117] R. Hauck and O. Bryngdahl, "Computer-generated holograms with pulse-density modulation," *J. Opt. Soc. Am. A*, vol. 1, pp. 5–10, 1984.
- [118] O. K. Ersoy, J. Zhuang, and J. Brede, "Iterative interlacing approach for synthesis of computer-generated holograms," *Appl. Opt.*, vol. 31, pp. 6894–6901, 1992.
- [119] P. Thorston, F. Wyrowski, and O. Bryngdahl, "Importance of initial distribution for iterative calculation of quantized diffractive elements," *Journal of Modern Optics*, vol. 40, pp. 591–600, 1993.
- [120] C. Wu, C. Chen, and M. A. Fiddy, "Iterative procedure for improved computer-generated hologram reconstruction," *Appl. Opt.*, vol. 32, pp. 5135–5140, 1993.
- [121] N. Yoshikawa and T. Yatagai, "Phase optimization of a kinoform by simulated annealing," *Appl. Opt.*, vol. 33, pp. 863–868, 1994.
- [122] E. Zhang, S. Noehte, C. H. Dietrich, and R. Manner, "Gradual and random binarization of gray-scale holograms," *Appl. Opt.*, vol. 34, pp. 5987–5995, 1995.
- [123] L. Legiard, P. Refregier, and P. Ambs, "Multicriteria optimality for iterative encoding of computer-generated holograms," *Appl. Opt.*, vol. 36, pp. 7444–7449, 1997.

- [124] H. H. Suh, "Color-image generation by use of binary phase holograms," *Opt. Lett.*, vol. 24, pp. 661–663, 1999.
- [125] R. Eschbach, "Comparison of error diffusion methods for computer-generated holograms," *Appl. Opt.*, vol. 30, pp. 3702–3710, 1991.
- [126] A. Kirk, K. Powell, and T. Hall, "A generalization of the error diffusion method for binary computer-generated hologram design," *Opt. Comm.*, vol. 92, pp. 12–18, 1992.
- [127] R. Eschbach and Z. Fan, "Complex valued error diffusion for off-axis computer-generated holograms," *Appl. Opt.*, vol. 32, pp. 3130–3136, 1993.
- [128] F. Fetthauer, S. Weissbach, and O. Bryngdahl, "Equivalence of error diffusion and minimal average error algorithms," *Opt. Comm.*, vol. 113, pp. 365–370, 1995.
- [129] K. Heggarty and R. Chevallier, "Signal window minimum average error algorithm for computer-generated holograms," *J. Opt. Soc. Am. A*, vol. 15, pp. 625–635, 1998.
- [130] M. A. Seldowitz, J. P. Allebach, and D. W. Sweeney, "Synthesis of digital holograms by direct binary search," *Applied Optics*, vol. 26, pp. 2788–2798, 1987.
- [131] B. K. Jennison and J. P. Allebach, "Efficient design of direct-binary-search computer-generated holograms," *J. Opt. Soc. Am. A*, vol. 8, pp. 652–660, 1991.
- [132] J. Zhuang and O. K. Ersoy, "Fast decimation-in-frequency direct binary search algorithms for synthesis of computer-generated holograms," *J. Opt. Soc. Am. A*, vol. 11, pp. 135–143, 1995.

- [133] J. Zhuang and O. K. Ersoy, “Optimal decimation-in-frequency iterative interlacing technique for synthesis of computer-generated holograms,” *J. Opt. Soc. Am. A*, vol. 12, pp. 1460–1468, 1995.
- [134] B. B. Chhetri, S. Yang, and T. Shimomura, “Stochastic approach in the efficient design of the direct-binary-search algorithm for hologram synthesis,” *Appl. Opt.*, vol. 39, pp. 5956–5964, 2000.
- [135] F. Wyrowski, “Iterative quantization of digital amplitude holograms,” *Appl. Opt.*, vol. 28, pp. 3864–3870, 1989.
- [136] L. Bigue and P. Ambs, “Optimal multicriteria approach to the iterative Fourier transform algorithm,” *Appl. Opt.*, vol. 40, pp. 5886–5893, 2001.
- [137] S. H. Tao and X. Yuan, “Practical implementation of the phase-quantization technique in an iterative Fourier-transform algorithm,” *Appl. Opt.*, vol. 43, pp. 2089–2092, 2004.
- [138] J. P. Allebach, “Representation-related errors in binary digital holograms: a unified analysis,” *Appl. Opt.*, vol. 20, pp. 290–299, 1981.
- [139] B. K. Jennison and J. P. Allebach, “Analysis of the leakage from computer-generated holograms synthesized by direct binary search,” *J. Opt. Soc. Am. A*, vol. 6, pp. 234–243, 1989.
- [140] F. Wyrowski, “Diffraction efficiency of analog and quantized digital amplitude holograms: analysis and manipulation,” *J. Opt. Soc. Am. A*, vol. 7, pp. 383–393, 1990.
- [141] Y. Chang, P. Zhou, and J. H. Burge, “Analysis of phase sensitivity for binary computer-generated holograms,” *Appl. Opt.*, vol. 45, pp. 4223–4234, 2006.

- [142] C. Maurer, A. Schwaighofer, A. Jesacher, S. Bernet, and M. Ritsch-Marte, “Suppression of undesired diffraction orders of binary phase holograms,” *Appl. Opt.*, vol. 47, pp. 3994–3998, 2008.
- [143] R. Piestun, B. Spektor, and J. Shamir, “On-axis binary-amplitude computer-generated holograms,” *Opt. Comm.*, vol. 136, pp. 85–92, 1997.
- [144] R. Ulichney, *Digital Halftoning*. MIT Press, 1987.
- [145] R. Tudela, I. Labastida, E. Martin-Badosa, S. Vallmitjana, I. Juvells, and A. Carnicer, “A simple method for displaying fresnel holograms on liquid crystal panels,” *Opt. Comm.*, vol. 214, pp. 107–114, 2002.
- [146] R. Tudela, E. Martin-Badosa, I. Labastida, S. Vallmitjana, I. Juvells, and A. Carnicer, “Wavefront reconstruction by adding modulation capabilities of two liquid crystal devices,” *Opt. Eng.*, vol. 43, pp. 2650–2657, 2004.
- [147] D. A. Gregory, J. C. Kirsch, and E. C. Tam, “Full complex modulation using liquid-crystal televisions,” *Appl. Opt.*, vol. 31, pp. 163–165, 1992.
- [148] R. D. Juday, “Full-complex modulation with two one-parameter spatial light modulators.” United States Patent 5,416,618, 1995.
- [149] L. G. Neto, D. Roberge, and Y. Sheng, “Full-range, continuous, complex modulation by the use of two coupled-mode liquid-crystal televisions,” *Appl. Opt.*, vol. 35, pp. 4567–4576, 1996.
- [150] P. M. Birch, R. Young, D. Budgett, and C. Chatwin, “Two-pixel computer-generated hologram with a zero-twist nematic liquid-crystal spatial light modulator,” *Optics Letters*, vol. 25, pp. 1013–1015, 2000.
- [151] V. Arrizon, “Complex modulation with a twisted-nematic liquid-crystal spatial light modulator: double-pixel approach,” *Optics Letters*, vol. 28, pp. 1359–1361, 2003.

- [152] E. G. van Putten, I. M. Vellekoop, and A. P. Mosk, “Spatial amplitude and phase modulation using commercial twisted nematic LCDs,” *Appl. Opt.*, vol. 47, pp. 2076–2081, 2008.
- [153] R. C. Gonzales and R. E. Woods, *Digital Image Processing, 2nd edition*. Prentice Hall, 2002.
- [154] T. J. Naughton, Y. Frauel, B. Javidi, and E. Tajahuerce, “Compression of digital holograms for three-dimensional object reconstruction and recognition,” *Appl. Opt.*, vol. 41, pp. 4124–4132, 2002.