

LOCAL SIGNAL DECOMPOSITION BASED
METHODS FOR THE CALCULATION OF
THREE-DIMENSIONAL SCALAR OPTICAL
DIFFRACTION FIELD DUE TO A FIELD GIVEN ON
A CURVED SURFACE

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND

ELECTRONICS ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BİLKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Erdem Şahin

May 2013

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 (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. Haldun M. Özaktas

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. A. 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. Mehmet Tankut Özgen

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

LOCAL SIGNAL DECOMPOSITION BASED METHODS FOR THE CALCULATION OF THREE-DIMENSIONAL SCALAR OPTICAL DIFFRACTION FIELD DUE TO A FIELD GIVEN ON A CURVED SURFACE

Erdem Şahin

Ph.D. in Electrical and Electronics Engineering

Supervisor: Prof. Dr. Levent Onural

May 2013

A three-dimensional scene or object can be optically replicated via the three-dimensional imaging and display method holography. In computer-generated holography, the scalar diffraction field due to a field given on an object (curved surface) is calculated numerically. The source model approaches treat the building elements of the object (such as points or planar polygons) independently to simplify the calculation of diffraction field. However, as a tradeoff, the accuracies of fields calculated by such methods are degraded. On the other hand, field models provide exact field solutions but their computational complexities make their application impractical for meaningful sizes of surfaces. By using the practical setup of the integral imaging, we establish a space-frequency signal decomposition based relation between the ray optics (more specifically the light field representation) and the scalar wave optics. Then, by employing the uncertainty principle inherent to this space-frequency decomposition, we derive

an upper bound for the joint spatial and angular (spectral) resolution of a physically realizable light field representation. We mainly propose two methods for the problem of three-dimensional diffraction field calculation from fields given on curved surfaces. In the first approach, we apply linear space-frequency signal decomposition methods to the two-dimensional field given on the curved surface and decompose it into a sum of local elementary functions. Then, we write the diffraction field as a sum of local beams each of which corresponds to such an elementary function on the curved surface. By this way, we increase the accuracy provided by the source models while keeping the computational complexity at comparable levels. In the second approach, we firstly decompose the three-dimensional field into a sum of local beams, and then, we construct a linear system of equations where we form the system matrix by calculating the field patterns that the three-dimensional beams produce on the curved surface. We find the coefficients of the beams by solving the linear system of equations and thus specify the three-dimensional field. Since we use local beams in three-dimensional field decomposition, we end up with sparse system matrices. Hence, by taking advantage of this sparsity, we achieve considerable reduction in computational complexity and memory requirement compared to other field model approaches that use global signal decompositions. The local Gaussian beams used in both approaches actually correspond to physically realizable light rays. Indeed, the upper joint resolution bound that we derive is obtained by such Gaussian beams.

Keywords: Linear Space-Frequency Signal Decomposition, Gaussian Beam Decomposition, Curved Surfaces, Computer-Generated Holography, Scalar Optical Diffraction, Light Field Representation, Integral Imaging, Ray Optics, Scalar Wave Optics

ÖZET

EĞRİ BİR YÜZEY ÜZERİNDE VERİLEN ALANA KARŞILIK GELEN ÜÇ BOYUTLU SKALAR OPTİK ALANIN HESAPLANMASI İÇİN LOKAL SİNYAL AYRIŞTIRMA TABANLI YÖNTEMLER

Erdem Şahin

Elektrik ve Elektronik Mühendisliği Bölümü Doktora

Tez Yöneticisi: Prof. Dr. Levent Onural

Mayıs 2013

Üç boyutlu sahneler veya nesneler, bir üç boyutlu görüntüleme ve gösterim tekniği olan holografi aracılığıyla optik olarak yeniden üretilebilir. Verilen bir cisim (eğri yüzey) tarafından oluşturulan kırınım deseni, klasik holografının dışında, bilgisayar ortamında sayısal olarak da hesaplanabilir. Kaynak model yaklaşımları, kırınım deseninin hesaplanmasını kolaylaştırmak için nesnenin nokta veya düzlemsel çokgen gibi yapı taşlarını bağımsız bir şekilde ele alırlar. Ancak, bir ödünleşim olarak, bu yöntemler aracılığıyla hesaplanan kırınım desenlerinin doğrulukları azalır. Diğer taraftan, alan modelleri doğru alan çözümleri sunarlar fakat bu modellerin hesap karmaşıklıkları anlamlı boyutlardaki yüzeylere uygulanamamalarına neden olur. Tümlev görüntüleme (*integral imaging*) yönteminin pratik düzenliğini kullanarak, ışın optiği (özel olarak ışık alanı gösterimi) ile skalar dalga optiği arasında uzay-frekans sinyal ayrıştırma tabanlı bir ilişki kuruyoruz. Ardından, bu uzay-frekans ayrıştırma yönteminin doğasında var olan belirsizlik ilkesinden faydalanarak, fiziksel olarak gerçekleştirilebilir ışık alanı gösterimi için bir birleşik uzamsal ve açısal (spektral) üst çözünürlük sınırı buluyoruz. Eğri

yüzeylerden yayılan üç boyutlu kırınım deseninin hesaplanması için esas olarak iki değişik yöntem sunuyoruz. Birinci yaklaşımda, eğri yüzey üzerinde verilen iki boyutlu alana doğrusal uzay-frekans sinyal ayrıştırma yöntemlerini uyguluyor ve bu alanı lokal temel fonksiyonların toplamı olacak şekilde bölüyoruz. Ardından, kırınım desenini her biri eğri yüzey üzerindeki bu temel fonksiyonlardan birine karşılık gelen lokal ışık hüzmelerinin toplamı şeklinde yazıyoruz. Bu sayede, kaynak modelleri tarafından sağlanan alanın doğruluk değerini, hesap karmaşıklığını onlarla karşılaştırılabilir seviyede tutarak, artırıyoruz. İkinci yaklaşımda, ilk olarak üç boyutlu alanı lokal hüzmelerin toplamı olacak şekilde bölüyoruz. Ardından, bir doğrusal denklemler sistemi oluşturuyoruz, öyle ki bu sistemi temsil eden sistem matrisinin kolonlarını üç boyutlu hüzmelerin eğri yüzey üzerinde ürettiği alan desenlerini bularak elde ediyoruz. Hüzmelerin katsayılarını doğrusal denklemler sistemini çözerek buluyoruz. Böylece, üç boyutlu alanı belirlemiş oluyoruz. Üç boyutlu alanı ayrıştırırken lokal hüzmeler kullandığımız için seyrek sistem matrisleri elde ediyoruz. Bundan yararlanarak, global sinyal ayrıştırma yöntemi kullanan alan modellerine göre hesaplama karmaşıklığı ve bellek ihtiyacını önemli derecede azaltıyoruz. İki yaklaşımda da kullanılan lokal Gauss hüzmeleri aslında fiziksel olarak gerçekleştirilebilir ışıklara karşılık gelmektedir. Bulduğumuz birleşik üst çözünürlük sınırı da esasen bu Gauss hüzmeleri tarafından elde edilmektedir.

Anahtar Kelimeler: Doğrusal Uzay-Frekans Sinyal Ayrıştırma, Gauss Hüzme Ayrıştırma, Eğri Yüzeyler, Bilgisayarla Üretilmiş Holografı, Skalar Optik Kırınımı, Işık Alanı Gösterimi, Tümler Görüntüleme, Işın Optiği, Skalar Dalga Optiği

ACKNOWLEDGMENTS

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

I would also like to thank Prof. Dr. Haldun M. Özaktař, Prof. Dr. A. Aydın Alatan, Assoc. Prof. Dr. Mehmet Tankut Özgen 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 acknowledge the financial support of TÜBİTAK (The Scientific and Technological Research Council of Turkey) between 2006-2011.

I also acknowledge the financial support of Bilkent University between 2011-2013.

I would like to thank Erdem Ulusoy (who is like a brother to me) for everything. Finally, I want to express my deepest gratitude to my family for their unconditional and unfailing support. I dedicate this dissertation to them.

Contents

1	Introduction	1
1.2	Holography	4
1.3	The Scalar Wave Optics and Diffraction Theory	6
1.4	Organization of the Dissertation	12
2	The Relation of Light Field Representation to Integral Imaging and Scalar Wave Optics	13
2.1	Light Field Representation	14
2.2	Integral Imaging	17
2.3	Relation of Light Field Representation to Integral Imaging	18
2.4	The Spatio-Angular Resolution Tradeoff in the Light Field Rep- resentation	21
3	Calculation of the 3D Scalar Diffraction Field from the Field Given on a Curved Surface by Using Linear and Local Signal Decomposition Methods	28
3.1	Planar Input Surfaces	30

3.1.1	Continuous Decomposition for Planar Input Surfaces . . .	31
3.1.2	Discrete Decomposition for Planar Input Surfaces	32
3.2	Curved Input Surfaces	36
3.2.1	Continuous Decomposition for Curved Input Surfaces . . .	37
3.2.2	Discrete Decomposition for Curved Input Surfaces	39
3.3	Short Time Fourier Transform	41
3.3.1	STFT for Planar Input Surfaces	41
3.3.2	STFT for Curved Input Surfaces	44
3.4	Wavelet Decomposition	46
3.4.1	Wavelet Decomposition for Planar Input Surfaces	46
3.4.2	Wavelet Decomposition for Curved Input Surfaces	49
4	Calculation of the 3D Scalar Diffraction Field from the Field Given on a Curved Surface via Gaussian Signal Decomposition	54
4.1	Gaussian Signal Decomposition	56
4.1.1	Decomposition for the Planar Input Surface Case	56
4.1.2	Decomposition for the Curved Input Surface Case	61
4.2	Simulation Results	66
5	Calculation of the 3D Scalar Diffraction Field from the Field Given on a Curved Surface via 3D Gaussian Beam Decomposi- tion	81

5.1	Diffraction Field Calculation from the Fields Given on Planar Input Surfaces	83
5.2	Diffraction Field Calculation from the Fields Given on Curved Input Surfaces	85
5.3	Simulation Results	90
6	Conclusions	116

List of Figures

2.1	The relation of solid angle, $d\Omega$, to the differential area, dA_2 . (©2010 The Imaging Science Journal. Reprinted with permission. Published in [40]).	15
2.2	Representation of the ray power density in the discrete case. (©2010 The Imaging Science Journal. Reprinted with permission. Published in [40])	16
2.3	Parameterization of the light power density in integral imaging. (©2010 The Imaging Science Journal. Reprinted with permission. Published in [40])	20
3.1	Local coordinate system on the input curved surface. (©2012 JOSA A. Reprinted with permission. Published in [35])	37
4.1	Mutual couplings between patches on a curved surface (2D cross sections of the surface and beams are shown for the sake of simplicity). (a) Mutual coupling between S_1 and S_3 patches, (b) mutual coupling between S_1 and S_2 patches (in addition to the mutual coupling between S_1 and S_3 patches). (©2012 JOSA A. Reprinted with permission. Published in [35])	64

4.2	The 2D simulation setup which is periodic along the x -axis (the fields used in the simulations repeat themselves between each dashed horizontal lines shown in the figure). (©2012 JOSA A. Reprinted with permission. Published in [35])	67
4.3	The discrete Gaussian synthesis window function $g_d[n] = \exp\left(-\frac{L_s^2}{\sigma^2}n^2\right)$ for $\sigma = 2.20\mu m$ and $L_s = 0.26\mu m$. (©2012 JOSA A. Reprinted with permission. Published in [35])	70
4.4	The discrete analysis window function corresponding to the Gaussian synthesis window function $g_d[n] = \exp\left(-\frac{L_s^2}{\sigma^2}n^2\right)$, for $\sigma = 2.20\mu m$ and $L_s = 0.26\mu m$, for the case that the shift parameters in space and frequency are $M = 16$ and $K = 32$, respectively. (©2012 JOSA A. Reprinted with permission. Published in [35]) .	71
4.5	Periodic Gaussian beams induced by the periodic pattern over the periodic curved line. (©2012 JOSA A. Reprinted with permission. Published in [35])	73
4.6	The real part of the discrete input signal, $u_d[n]$, on the circular arc with a measure of 30 degrees (imaginary part is zero). (©2012 JOSA A. Reprinted with permission. Published in [35])	75
4.7	The magnitude of the 2D diffraction field due to the input signal $\tilde{u}_d[n]$, one period of which is shown in Fig. 4.6, on the circular arc with a measure of 30 degrees. (©2012 JOSA A. Reprinted with permission. Published in [35])	76
4.8	The magnitude of the 2D diffraction field due to a single shifted and modulated Gaussian window function, with parameters $m = 16$ and $k = 1$, on the circular arc with a measure of 30 degrees. (©2012 JOSA A. Reprinted with permission. Published in [35]) .	77

4.9	The real parts of the original, $u_d[n]$, and reconstructed, $u_d^r[n]$, signals on the circular arc with a measure of 30 degrees. (©2012 JOSA A. Reprinted with permission. Published in [35])	80
5.1	The setup of the proposed Gaussian beam decomposition method (a 2D pattern is shown for the sake of simplicity). (©2013 JOSA A. Reprinted with permission. Published in [81])	87
5.2	The periodic 2D simulation setup. (©2013 JOSA A. Reprinted with permission. Published in [81])	94
5.3	The ratios of the complexities of the operations QRPWD, SQR-PWD, QRGBD to the complexity of SQRGBD (note that the result shown for SQRGBD is “1”). (©2013 JOSA A. Reprinted with permission. Published in [81])	96
5.4	The sparsity of the system matrices for the proposed Gaussian beam decomposition method. (©2013 JOSA A. Reprinted with permission. Published in [81])	97
5.5	The absolute value of the system matrix $\mathbf{G}$ which has size 2400×2000 and formed by the application of the proposed Gaussian beam decomposition method to the curved line shown in Fig. 5.2. (©2013 JOSA A. Reprinted with permission. Published in [81])	98
5.6	The real part of the system matrix $\mathbf{G}$ which has size 2400×2000 and formed by the application of the plane wave decomposition method to the curved line shown in Fig. 5.2. (note that the absolute values of all matrix entries are one.) Note the difference with respect to Fig. 5.5 in terms of sparsity. (©2013 JOSA A. Reprinted with permission. Published in [81])	99

5.7	One period of the two-dimensional (x, z) field corresponding to the signal given on the periodic curved line which has loops. . . .	101
5.8	The absolute value of the system matrix $\mathbf{G}$ which has size 1400×986 and formed by the application of the proposed Gaussian beam decomposition where the beam waist positions are chosen inside the loops.	102
5.9	The absolute value of the system matrix $\mathbf{G}$ which has size 1400×986 and formed by the application of the proposed Gaussian beam decomposition where the beam waist positions are chosen on the curved line.	103
5.10	One period of the two-dimensional (x, z) field corresponding to the Gaussian beam decomposition where the beam waist positions are taken inside the loops and all the beam coefficients are taken equal.	104
5.11	The real part of the system matrix $\mathbf{G}$ which has size 1400×1000 and formed by the application of the plane wave decomposition method for the curved line shown in Fig. 5.7. Note that the absolute values of all matrix entries are one.	105
5.12	One period of the 2D periodic curved surface. (©2013 JOSA A. Reprinted with permission. Published in [81])	107
5.13	One period of the 2D periodic object. (Gilles Tran ©2007 www.oyonale.com, used under the Creative Commons Attribution license. Published in [81])	108

5.14	The absolute value of the system matrix $\mathbf{G}$ which has size 11664×9472 and formed by the application of the proposed Gaussian beam decomposition method for the curved surface given in Fig. 5.12. Note that the figure is enhanced for better visualization. ((©2013 JOSA A. Reprinted with permission. Published in [81]))	109
5.15	The real part of the system matrix $\mathbf{G}$ which has size 11664×9475 and formed by the application of the plane wave decomposition method for the curved surface given in Fig. 5.12. Note that the absolute values of all matrix entries are one. Note also that the random noise-like appearance is a consequence of the structure of the surface given in Fig. 5.12. ((©2013 JOSA A. Reprinted with permission. Published in [81]))	110
5.16	The absolute value of the system matrix $\mathbf{G}$ which has size 3973×2809 and formed by the application of the proposed Gaussian beam decomposition method for the object given in Fig. 5.13. ((©2013 JOSA A. Reprinted with permission. Published in [81]))	111
5.17	The real part of the system matrix $\mathbf{G}$ which has size 3973×2819 and formed by the application of the plane wave decomposition method for the object given in Fig. 5.13. Note that the absolute values of all matrix entries are one. ((©2013 JOSA A. Reprinted with permission. Published in [81]))	112

List of Publications

This dissertation is based on the following publications.

[Publication-I] E. Şahin and L. Onural, “A comparative study of light field representation and integral imaging,” *The Imaging Science Journal*, vol. 58, no. 1, pp. 28-31, 2010.

[Publication-II] E. Şahin and L. Onural, “Scalar diffraction field calculation from curved surfaces via Gaussian beam decomposition,” *J. Opt. Soc. Am. A*, vol. 29, no. 7, pp. 1459-1469, 2012.

[Publication-III] E. Şahin and L. Onural, “Calculation of the scalar diffraction field from curved surfaces by decomposing the 3D field into a sum of Gaussian beams,” *J. Opt. Soc. Am. A*, vol. 30, no. 3, pp. 527-536, 2013.

The contributions of the author to the publications which are listed above were as follows: In Publications-I, -II and -III, the author designed and implemented the proposed algorithms and methods, performed the mathematical derivations and prepared the manuscripts. The author conducted and reported the simulations in Publications-II and -III.

Dedicated to my family...

Chapter 1

Introduction

Nowadays, the three-dimensional (3D) media applications become more and more popular as a consequence of successfully produced 3D films and games. In such media applications, usually the stereoscopic viewing and display techniques are used. Although the stereoscopy is actually the simplest 3D visualization technique and were discovered many years ago, it seems to be sufficient to attract people, at least for now. During the capturing stage of stereoscopy, the 3D content is obtained by using two cameras placed side by side [1, 2]. Then in the display stage, the image captured by the right (left) camera is presented to the right (left) eye of the observer [3]. Although the stereoscopic images are displayed on the same screen, each eye of the observer perceives different images by wearing goggles such as polarized glasses or active shutter glasses. The necessity of wearing goggles is one of the main drawbacks of the stereoscopic viewing techniques. In this sense, the autostereoscopic viewing seems to be an attractive alternative to stereoscopy. The autostereoscopy provides 3D viewing without imposing a necessity of wearing goggles [4]. The well known autostereoscopic imaging technique, integral imaging, was discovered by Lippmann at 1908 [5]. Several display devices that provide autostereoscopic views have been produced since then, but nowadays the producers have begun to advertise autostereoscopic

displays more seriously hoping them to be the next generation 3D visualization technique.

Holography provides the richest optical content among other viewing techniques for a given 3D scene. Indeed, the term hologram originates from the Greek words *holos* and *gramma* that mean “whole” and “message”, respectively. Ideally, holography aims to capture a 3D scene by a holographic imaging method and replicate the 3D scene at some other place by an appropriate holographic display device. In classical holography, the optical information related to a 3D scene is captured on a recording device (hologram) by using a reference beam [6, 7]. The interference of the optical wave due to the 3D scene (diffraction field) and the reference beam produce the coded optical information of the 3D scene. This coding is actually an amplitude modulation of the diffraction field by the reference wave as in the communication systems [8]. The 3D scene is replicated by illuminating the hologram with the same reference beam as used in the recording stage.

The recording stage of classical holography makes it unpractical for 3D video applications. With the invention of charged coupled devices (CCD), however, it became possible to record dynamic images. As a consequence of that, the interest of optics community to digital holography gained momentum [9]. The digital holography is different from the classical optical holography in the sense that in digital holography the images of a 3D scene is captured by a CCD camera instead of recording it on a photographic film. A further development in the holography was gained with the inclusion of computers in the process of hologram calculation. In this so-called computer-generated holography (CGH) process, the holograms of desired 3D fields are calculated numerically by computation [10, 11, 12]. On the other hand, the invention of dynamically programmable spatial light modulator (SLM) devices made it also possible to display 3D dynamic images [3]. All these

developments have led the research projects related to end-to-end holographic 3D television systems [3, 13, 14].

As mentioned before, capturing and displaying a 3D content are two main operations in holography. Although usually planar sensor and display devices are used for these operations, several advantages of curved display and sensor devices have been recently reported. For example, spherical geometries are shown to be more useful compared to planar ones to reduce the resolution requirements of the sensor and display devices [15, 16]. In the work of Hahn *et al.* [17], curved array of spatial light modulators (SLMs) is used as a holographic display device to increase the field of view compared to conventional planar array of SLMs. Similar possible advantages of curved display and sensor devices for holography indicate the need for the theoretical analysis of diffraction between curved surfaces. On the other hand, calculation of the diffraction field due to an object is one of the main tasks to be accomplished in computer-generated holography. The challenging part of this task is to provide an approach which is both computationally efficient and satisfactorily accurate, since it is generally a computationally costly process to calculate the 3D diffraction field due to an object. As addressed by these discussions, the methods that we propose in this dissertation provide effective and efficient ways to calculate the 3D diffraction field due to an object or due to a two-dimensional (2D) field given on a curved surface.

In the remaining part of this chapter, we give more detailed information about holography. We also discuss the scalar wave optics and the diffraction theory of light, since we adopt the scalar diffraction theory in the methods that we develop for the problem of diffraction field calculation due to a 2D field given on a curved surface.

1.2 Holography

Holography was invented by D. Gabor in 1947, while he was researching a method to improve the resolution of an electron microscope [18]. The development of holograms started, however, in 1960's with the invention of truly coherent light source laser by the scientist C. Towns, A. Prokhorov and N. Bassov [19]. In 1962, the first laser transmission type hologram of a 3D objet was obtained by E. Leith and J. Upatnieks. The hologram was needed to be illuminated by a laser in order to produce 3D images with realistic depth and clarity. In the same year, N. Denisyuk developed white-light reflection type hologram. It is possible to view such type of holograms under incandescent lighting. Transmission type of white-light holography was invented by S. A. Benton in 1968. White-light transmission type of holograms can be viewed in ordinary white light. Dynamic 3D images were obtained in 1972 with the development of integral hologram method by L. Cross [20]. Integral holography combines white-light transmission holography with the cinematography to produce dynamic 3D images. The first computer-generated hologram was introduced in 1965 by B. R. Brown and A. W. Lohmann [21].

The computer-generated holography makes it possible to produce the holograms of objects (which may be synthetically generated) via numerical methods. That is an important improvement over the classical optical holography. However, as mentioned before, the computational complexity of the diffraction field calculation is the main drawback of CGH. Significant developments have been achieved towards the aim of reducing the computation time since the introduction of first computer-generated hologram [9, 11]. The Fresnel approximation has been a widely used approach to reduce the computation time of digital holograms [8, 22]. In the Fresnel model, the propagation kernel is a 2D separable function. Furthermore, the wave propagation calculation which is modeled by the Fresnel integral can be implemented by the Fourier transform techniques.

The underlying purpose in the development of holographic stereograms [23, 24] is also to reduce the computation time. In the holographic stereogram approach, the hologram plane is partitioned into segments. The 2D complex function to be written inside a segment is determined according to the relative position of the center point of that segment with respect to the object. Thus, the computation time is reduced, since the field is not calculated for each pixel in the segment. Another widely used assumption in CGH is to model the object as a collection of point sources or planar polygons [23, 25, 26, 27, 28, 29, 30]. In these source model approaches, as is the case in other simplifications, there is a tradeoff between the accuracy of the calculated field and the computational complexity. In Chapter 3 and Chapter 4 of this dissertation, we present an approach that increases the accuracy compared to such source model approaches by keeping the computational complexity at comparable levels.

The mathematical derivations related to holography are obtained by using the interference theory [31]. If we denote the reference wave as $\mathbf{U}_r$ and the diffraction field due to the object as $\mathbf{U}_o$, then the intensity of the field captured on the hologram plane (interference pattern) is written as

$$\mathbf{I} = |\mathbf{U}_o + \mathbf{U}_r|^2 \quad . \quad (1.1)$$

The expression given by Eq. 1.1 can be expanded as

$$\mathbf{I} = |\mathbf{U}_o|^2 + |\mathbf{U}_r|^2 + \mathbf{U}_o^* \mathbf{U}_r + \mathbf{U}_o \mathbf{U}_r^* \quad , \quad (1.2)$$

where $*$ is the complex conjugation operator. The interference pattern can be captured by a photographic film (transparency) or by a CCD camera. On the other hand, in the computer-generated holography, the diffraction field $\mathbf{U}_o$ due to the given object is calculated numerically. Afterwards, the interference pattern can be calculated as given by Eq. 1.1. The reconstruction of the object wave is achieved via the diffraction of the same reference wave by the hologram recorded on the transparency. The field just after the transparency is written as

$$\mathbf{I}_r = \mathbf{U}_r |\mathbf{U}_o|^2 + \mathbf{U}_r |\mathbf{U}_r|^2 + \mathbf{U}_r^2 \mathbf{U}_o^* + |\mathbf{U}_r|^2 \mathbf{U}_o \quad . \quad (1.3)$$

The object wave $\mathbf{U}_o$ appears as the fourth term (+1 order term) of Eq. 1.3. The first two terms (zero order terms) represent the modulation of the reference wave by the intensities of the object and reference waves, respectively. The third term (−1 order term) is the modulated conjugate object wave. If these terms are separated in space, then the reconstructed object can be seen clearly. Under the assumption that the object wave is angle-limited, such a separation can be achieved by using an off-axis system where the reference and object waves are offset (there is a small angle between the reference and object waves). The minimum offset angle requirement for an off-axis holographic system can be relieved by taking the intensity of the reference wave much larger than that of the object wave and thus making the first term of Eq. 1.3 negligible [31].

1.3 The Scalar Wave Optics and Diffraction Theory

There are basically four models developed to describe the nature of light. These are ray optics, scalar wave optics (or simply wave optics), electromagnetic optics and quantum optics. The simplest model ray optics defines the light as a collection of rays which travel according to the Fermat's principle [31]. The ray optics can adequately explain many optical phenomena such as reflection and refraction. In the ray optics (geometric optics), it is possible to assign different values to the rays which intersect at a given point and travel towards different directions. This property is not valid according to the diffraction phenomena which is explained by the wave optics. Similarly, the interference phenomenon of light can not be explained by the ray optics. The scalar wave optics encompasses the ray optics such that the ray optics can be seen as the limit of the wave optics when the wavelength of the light gets infinitesimally small [31]. The optical phenomena such as diffraction and interference are explained by the wave nature

of the light. On the other hand, there are situations like the diffraction of light at high angles or the interaction of light with a dielectric media where the scalar wave optics is not adequate. For such cases vector-valued descriptions of the light field are necessary. At this point, the electromagnetic optics provides sufficient descriptions. In electromagnetic optics, the propagation of light is modeled by using two mutually coupled vector waves, an electric field wave and a magnetic field wave. In free space, these fields satisfy the well-known Maxwell's equations. The electromagnetic optics is not sufficient to describe the optical phenomena which are quantum mechanical in nature [31]. Such phenomena can be explained by the quantum optics.

In this dissertation, as mentioned before, we mainly deal with the problem of diffraction field calculation due to a 2D field given on a curved surface that is one of the challenging tasks in holography. In the methods proposed in Chapter 3, Chapter 4 and Chapter 5, we adopt the scalar wave optics and use linear space-frequency signal decompositions. At this point, it is important to note that there is a close relation between the local space-frequency decomposition of a scalar wave field and its ray optics approximation. The construction of a rigorous mathematical relation of the ray optics to the scalar wave optics based on the space-frequency uncertainty is another problem that we deal in this dissertation. The study related to this problem is given in Chapter 2. Here, it is worth noting that the construction of such a relation helped us to consider the approaches proposed in the subsequent chapters from the ray optics point of view.

In scalar wave optics, the light is assumed to propagate in the form of waves. An optical wave is described via a real-valued scalar function $u_t(\mathbf{x}, t)$ of position $\mathbf{x} = [x, y, z]^T \in \mathbb{R}^3$ and time $t \in \mathbb{R}^3$. A given optical wave satisfies the wave equation given by [8, 32],

$$\nabla^2 u_t - \frac{1}{c^2} \frac{\partial^2 u_t}{\partial t^2} = 0 \quad , \quad (1.4)$$

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

As is most often the case in holography, we deal with monochromatic waves in this dissertation. The scalar wave function of a monochromatic wave is written in the form of [31],

$$u_t(\mathbf{x}, t) = a(\mathbf{x}) \cos\{2\pi\nu t + \phi(\mathbf{x})\} \quad , \quad (1.5)$$

where $a(\mathbf{x})$, $\phi(\mathbf{x})$ and ν represent the amplitude, phase and frequency (in cycles/unit time) of the monochromatic wave, respectively. For a purely monochromatic wave, it is possible to omit the time dependence and use the phasor representation of the wave function, since the phasor representation itself adequately describes the light. The phasor representation (complex amplitude) of the wave function is defined as

$$u(\mathbf{x}) = a(\mathbf{x}) \exp\{j\phi(\mathbf{x})\} \quad , \quad (1.6)$$

where $u_t(\mathbf{x}, t) = \mathcal{R}\{u(\mathbf{x}) \exp(j2\pi\nu t)\}$. If we substitute the complex amplitude into the wave equation, we end up with the Helmholtz equation given as

$$\left(\nabla^2 + \frac{4\pi^2}{\lambda^2}\right) u = 0 \quad , \quad (1.7)$$

where λ is the wavelength of the monochromatic wave and $c = \lambda\nu$.

Note that we will use the phasor representation of the wave field from now on. The entire field in 3D space can be specified by defining a 2D field on an infinite extent planar surface. Let the input field is specified on the $z = 0$ plane and defined as $u_0(x, y) = u(x, y, 0)$. According to the scalar diffraction theory, the diffraction field on another plane which is parallel to $z = 0$ plane and placed at a distance z is given by [8, 33, 34],

$$u_z(x, y) = u_0(x, y) ** h_z(x, y) \quad , \quad (1.8)$$

where $**$ denotes the two-dimensional convolution and $h_z(x, y)$ is the free space propagation kernel. The exact field relation (under the scalar wave optics) between such two parallel planes is given by the Rayleigh-Sommerfeld (RS) integral

which is obtained when the RS kernel

$$h_z(x, y) = -\frac{z}{2\pi\sqrt{x^2 + y^2 + z^2}} \left(1 - jk\sqrt{x^2 + y^2 + z^2}\right) \frac{\exp\left(jk\sqrt{x^2 + y^2 + z^2}\right)}{x^2 + y^2 + z^2} \quad (1.9)$$

is substituted into the equation

$$u_z(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(\xi, \eta) h_z(x - \xi, y - \eta) d\xi d\eta \quad . \quad (1.10)$$

Denoting the Fourier transforms of $u_0(x, y)$, $u_z(x, y)$ and $h_z(x, y)$ as $U_0(f_x, f_y)$, $U_z(f_x, f_y)$ and $H_z(f_x, f_y)$, respectively, Eq. 1.10 can be written in the frequency domain as

$$U_z(f_x, f_y) = U_0(f_x, f_y) H_z(f_x, f_y) \quad , \quad (1.11)$$

where, for the RS model, $H_z(f_x, f_y)$ is given as [8, 34],

$$\begin{aligned} H_z(f_x, f_y) &= \mathcal{F}\{h_z(x, y)\} \\ &= \exp\left(j\frac{2\pi}{\lambda}z\sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right) \quad . \end{aligned} \quad (1.12)$$

Here note that the Fourier transform from the (x, y) domain to the (f_x, f_y) domain, of a 2D function $a(x, y)$ is defined as

$$A(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(x, y) \exp\{-j2\pi(f_x x + f_y y)\} dx dy \quad . \quad (1.13)$$

Now, as it can be proved by Eq. 1.11 that the plane waves, which are defined as $\exp(j2\pi \mathbf{f}_x^T \mathbf{x})$ where $\mathbf{x} = [x, y, z]^T$ and $\mathbf{f}_x = [f_x, f_y, f_z]^T$, are actually the eigenfunctions of the scalar diffraction (note that the scalar diffraction is actually a linear shift invariant system). Therefore, the diffraction field at $\mathbf{x}$ that corresponds the 2D field specified on $z = 0$ plane can be also written as

$$u(\mathbf{x}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(f_x, f_y) \exp\{j2\pi(\mathbf{f}_x^T \mathbf{x})\} df_x df_y \quad , \quad (1.14)$$

where $f_x^2 + f_y^2 + f_z^2 = \frac{1}{\lambda^2}$ and λ is the wavelength of the plane waves. The coefficient of each plane wave is found by using the field $u_0(x, y)$ specified on the

reference input plane at $z = 0$ as

$$A(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) \exp\{-j2\pi(f_x x + f_y y)\} dx dy \quad . \quad (1.15)$$

The expression given by Eq. 1.14 is known as the plane wave decomposition. In this dissertation, we take into account only the propagating plane waves whose frequency components along the x and y axes satisfy $f_x^2 + f_y^2 \leq \frac{1}{\lambda^2}$. In other words, we exclude the evanescent waves. The propagating plane waves occupy the 2D circular spatial frequency band which has a radius $\frac{1}{\lambda}$. For the case that the evanescent waves are ignored, the plane wave decomposition is written as

$$u(\mathbf{x}) = \iint_B A(f_x, f_y) \exp(j2\pi \mathbf{f}_x^T \mathbf{x}) df_x df_y \quad , \quad (1.16)$$

where B denotes the 2D frequency band occupied by the propagating plane waves [35]. The propagation kernel and transfer function that correspond to this case are given, respectively, as [8],

$$h_z(x, y) = \frac{z}{j\lambda} \frac{\exp\left(j\frac{2\pi}{\lambda} \sqrt{x^2 + y^2 + z^2}\right)}{x^2 + y^2 + z^2} \quad , \quad (1.17)$$

$$H_z(f_x, f_y) = \begin{cases} \exp\left(j2\pi \frac{z}{\lambda} \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right), & \text{for } \sqrt{f_x^2 + f_y^2} \leq \frac{1}{\lambda} \\ 0, & \text{otherwise} \end{cases} \quad . \quad (1.18)$$

In Eq. 1.16, a monochromatic 3D field is written as a sum of infinite extent plane waves propagating toward different directions defined by the vector $\mathbf{f}_x$. The plane wave decomposition given by Eq. 1.16 together with the analysis equation given by Eq. 1.15, uniquely determines the 3D field corresponding to a 2D field specified on the infinite extent reference plane at $z = 0$, provided that the propagation direction components of the plane waves along the z -axis is always positive (or they are all negative). Note that such a requirement is necessary, since, for example, the two plane waves given by $\exp\{j2\pi(f_x x + f_y y + f_z z)\}$ and $\exp\{j2\pi(f_x x + f_y y - f_z z)\}$ produce the same 2D pattern $\exp\{j2\pi(f_x x + f_y y)\}$

on the $z = 0$ plane and thus they can not be discriminated from this 2D pattern. Therefore, in this dissertation, we also assume that a given 3D field $u(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{C}$ is formed by a superposition of propagating plane waves with $f_z > 0$.

If the plane wave components of the input field which is specified on the $z = 0$ plane make small angles with the z -axis, i.e. $f_x^2 + f_y^2 \ll \frac{1}{\lambda^2}$, then the corresponding wave is called as paraxial wave. For the propagation of paraxial waves, the RS impulse response function can be approximated with a quadratic phase exponential function as

$$h_z(x, y) = \frac{\exp\left(j\frac{2\pi}{\lambda}z\right)}{j\lambda z} \exp\left\{\frac{j\pi}{\lambda z}(x^2 + y^2)\right\} \quad . \quad (1.19)$$

This impulse response function is called as the Fresnel kernel and the diffraction model resulting from application of the Fresnel kernel is known as the Fresnel approximation [31].

The diffraction models such as Rayleigh-Sommerfeld, plane wave decomposition and Fresnel give the field relation (under the scalar diffraction theory) between the fields over two parallel planes. For the case that the input field is specified on a curved surface, direct application of formulations related to these models produce inconsistent 3D field solutions [36]. On the other hand, for such cases, the plane wave decomposition is also employed in a field model approach in the work of Esmer *et al.* [36, 37] and as a result an exact method is achieved. However, there are drawbacks of that approach such as intolerable computational complexity and memory requirement for large size of surfaces. At this point the local signal decomposition method that we propose in Chapter 5 plays a crucial role, in the problem of diffraction field calculation due to a field given on a curved surface, by decreasing both the computation time and memory requirement with respect to the approach introduced in [36], by still providing exact solutions.

1.4 Organization of the Dissertation

The work presented in this dissertation can be categorized into two parts. In the first part, we develop a mathematical relation between the ray optics and the scalar wave optics. More specifically, we firstly relate the light field representation (which is a ray optics based modeling of light) to the integral imaging. Then, by using the integral imaging as a bridge between the ray optics and wave optic, we develop a space-frequency decomposition based relation between the ray optics and the scalar wave optics. As a consequence of the uncertainty principle inherent to this space-frequency decomposition, we also derive an upper bound for the joint spatial and angular (spectral) resolution of a physically realizable light field representation. This part constitutes the second chapter of the dissertation. In the second part (Chapters 3-5), we deal with the problem of the calculation of 3D diffraction field from a 2D field given on a curved surface. In Chapter 3, we develop a generic local signal decomposition framework for the solution of this problem. As a special case of this framework, a Gaussian signal decomposition method is introduced in Chapter 4, where we provide explicit 3D field solutions and related simulations. In the methods proposed in Chapter 3 and Chapter 4, the field given on a 2D curved surface is mainly decomposed into a sum of local elementary signals. Then, the 3D field is written as a sum of local 3D beams each of which corresponds to such an elementary signal. In Chapter 5, we introduce another approach which is based on a 3D Gaussian beam decomposition method. In this method, the problem is formulated as a system of linear equations where the columns of the system matrix are the field patterns that the Gaussian beams produce on the curved surface. Thus, we solve a linear system of equations in order to find the 3D field due to a 2D field given on a curved surface.

Chapter 2

The Relation of Light Field Representation to Integral Imaging and Scalar Wave Optics

Light field representation is a three-dimensional image representation method that models the light as a collection of rays [38]. In other words, the ray optics model of light is adopted in the light field representation. The ray optics produces sufficiently accurate results when the light propagates around the objects having much greater dimensions compared to the wavelength of the light. Thus, as mentioned in Chapter 1, it can be viewed as the limit of the scalar wave optics as the wavelength of the light is infinitesimally small [31]. Beyond this observation, we construct a mathematical relation between the ray optics based model light field representation and scalar wave optics in this chapter. The mathematical relation is based on the space-frequency decomposition of the scalar wave field. The 3D optical imaging and display method integral imaging provides a natural way for this decomposition. The physical nature of lenses used in the integral imaging setup together with the diffraction phenomena explains the uncertainty

regarding the spatio-angular resolution tradeoff in the light field representation [39].

In Sec. 2.1 and Sec. 2.2, we introduce the light field representation and integral imaging methods, respectively. Then, we construct an equivalence relation between these two 3D image representation methods in Sec. 2.3. Based on the results obtained in Sec. 2.3, we relate the scalar wave optics to the ray optics (actually the light field representation) and develop an uncertainty relation regarding the spatio-angular resolution tradeoff in the light field representation in Sec. 2.4. Note that most of the content of the work presented in Sec. 2.1, Sec. 2.2 and Sec. 2.3 of this chapter is published in [40].

2.1 Light Field Representation

The appearance of the world at any time from any given point through any given direction can be represented with the so called plenoptic (plenty of optic) function, $P(x, y, z, \theta, \phi, \lambda, t)$ [41]. In other words, the plenoptic function corresponds to the radiance associated with the light ray, at time t , with wavelength λ , that passes from a 3D point (x, y, z) towards the direction (θ, ϕ) . In order to ease the notation, we will omit the variation with respect to t and λ . With further assumption of free space propagation and by keeping the viewing positions outside the convex hull of the scene one can also reduce the dimension of the plenoptic function to four since the radiance associated with the light ray along its path will be constant in this case [38]. The four dimensional plenoptic function is called as light field [38].

The light field is an abstract representation of the optical power flow associated with the light rays. Let us define the infinitesimal power emanating from the differential surface dA_1 (on an arbitrary surface W_1) and reaching to the differential surface dA_2 (on another arbitrary surface W_2) as dP (see Fig. 2.1). We

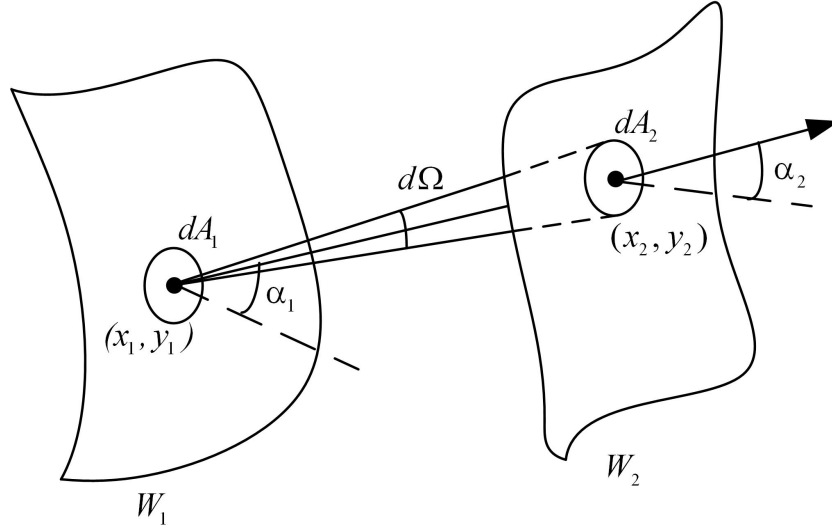


Figure 2.1: The relation of solid angle, $d\Omega$, to the differential area, dA_2 . (©2010 The Imaging Science Journal. Reprinted with permission. Published in [40]).

then associate a power density (for the power flow between W_1 and W_2 surfaces) to the ray crossing the two surfaces at (x_1, y_1) and (x_2, y_2) and define it as

$$L(x_1, y_1, x_2, y_2) = \frac{dP}{dA_1 dA_2} \quad . \quad (2.1)$$

We call $L(x_1, y_1, x_2, y_2)$ as the light field which can also be called ray power density by taking into account physical quantities into consideration [40].

Instead of using the second surface and the differential area dA_2 on it, it is quite common to adopt a solid angle model in the literature [41, 42], and the density in that case is called the radiance. However, we prefer the definition given by Eq. 2.1 for our purposes. Incidentally, it is straightforward to establish the relation between the density given in Eq. 2.1 and the radiance by first noting that the differential solid angle is related to the differential area dA_2 as $d\Omega = dA_2 \cos \alpha_2 / D^2$ where α_2 is the angle between the ray and the outward normal to the W_2 surface at (x_2, y_2) and D is the distance between (x_1, y_1) and (x_2, y_2) (see Fig. 2.1). The radiance associated with the ray crossing the two surfaces at (x_1, y_1) and (x_2, y_2) then becomes

$$\tilde{L}(x_1, y_1, x_2, y_2) = \frac{dP}{dA_1 \cos \alpha_1 d\Omega}, \quad (2.2)$$

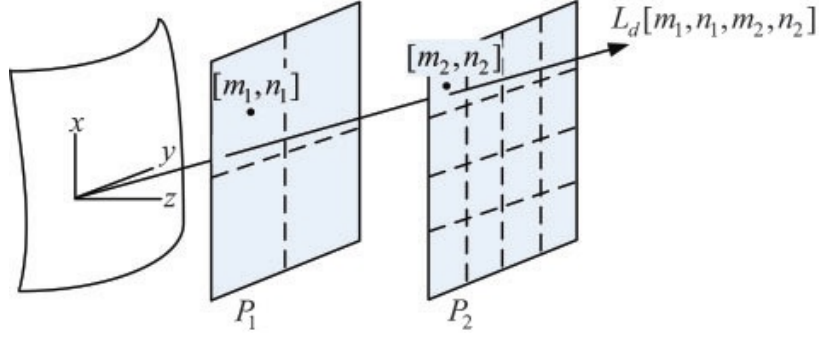


Figure 2.2: Representation of the ray power density in the discrete case. (©2010 The Imaging Science Journal. Reprinted with permission. Published in [40])

where dP is the radiant power emanating from dA_1 and propagating along the cone represented by the solid angle $d\Omega$, $dA_1 \cos \alpha_1$ is the projected differential area on W_1 along the direction of the ray and $d\Omega$ is the solid angle subtended by dA_2 . Therefore, L and $\tilde{L}$ are closely related to each other with a normalization as [40],

$$L(x_1, y_1, x_2, y_2) = \frac{\tilde{L}(x_1, y_1, x_2, y_2)}{D^2} \cos \alpha_1 \cos \alpha_2 \quad . \quad (2.3)$$

Here note that the units of measurement of $\tilde{L}(x_1, y_1, x_2, y_2)$ and $L(x_1, y_1, x_2, y_2)$ are $Wm^{-2}sr^{-1}$ and Wm^{-4} , respectively, where W (Watt) and sr (steradian) are the units of measurement of P and Ω , respectively.

Discretization of $L(x_1, y_1, x_2, y_2)$ is necessary for digital processing. Instead of arbitrary surfaces, we assume the simple two-parallel plane model [38, 43] where the P_1 and P_2 planes are usually defined as the camera and image planes at $z = z_1$ and $z = z_2$, respectively. Let us assume that the index arrays $[m_1, n_1]$ and $[m_2, n_2]$ represent the locations $(m_1 M_1, n_1 N_1, z_1)$ and $(m_2 M_2, n_2 N_2, z_2)$ on the two parallel planes, respectively, where M_1, N_1, M_2 and N_2 are the sampling intervals. Therefore $[m_1, n_1]$ represents the centers of cameras (ideal pinhole camera model), and $[m_2, n_2]$ represents the sample points of the images that are taken by the cameras. We define the discrete power density of the ray crossing the P_1 and P_2 planes, following the definition given by Eq. 2.1, as [40],

$$L_d[m_1, n_1, m_2, n_2] = \frac{P[m_1, n_1, m_2, n_2]}{S_1 S_2} \quad , \quad (2.4)$$

where S_1 and S_2 are the areas of the pixels on the P_1 and P_2 planes, respectively, $P[m_1, n_1, m_2, n_2]$ is the total power emanating from the pixel represented by $[m_1, n_1]$ and reaching to the pixel represented by $[m_2, n_2]$. (See Fig. 2.2) Therefore, L_d represents the power flow between the two pixels $[m_1, n_1]$ and $[m_2, n_2]$ on P_1 and P_2 , respectively. The subscript d denotes that the field is discrete.

2.2 Integral Imaging

Integral imaging is a 3D image capture and display method. The image is captured on a two-dimensional sensor array by a two-dimensional microlens array where the sensor array is placed behind the microlens array in a parallel fashion. Each microlens takes its own image of the 3D scene. The image that is formed behind each microlens is called as an elemental image. If the captured data is displayed properly, autostereoscopic images (allows 3D viewing without wearing eyeglasses) of 3D scenes are provided [5]. The parameterization of integral imaging is similar to the two-plane parameterization of the light field representation where the first plane P_1 is the 2D microlens array plane and the second plane P_2 is the 2D sensor array plane (see Fig. 2.3). The display stage of an integral imaging setup is constructed by placing the microlens array (used in the imaging stage) in front of the two-dimensional display device which displays the elemental images captured by the two-dimensional sensor array. The integral imaging renders a pseudoscopic 3D reconstruction to the observer. There are several ways to convert the pseudoscopic images to orthoscopic ones [44].

The quality of the 3D image that a practical integral imaging setup provides depends on the physical properties of the lenslet and sensor arrays [45]. For example, it is possible to increase the viewing angle by placing the microlenses and sensors on curved surfaces [46]. The aperture size of the microlenses used for imaging is one of the main factors that determines the resolution, the depth

of field and the viewing angle of the 3D images provided by an integral imaging setup. The tradeoffs related to these quality factors are actually the consequences of the spatio-angular resolution tradeoff inherent to integral imaging and they are explained by the physical nature of the lenses and the scalar diffraction phenomena. Indeed, as will be shown in the following sections, this spatio-angular tradeoff makes the integral imaging play a bridging role in the aim of relating the ray optics to the scalar wave optics.

2.3 Relation of Light Field Representation to Integral Imaging

The counterpart of two parallel planes of light field representation in integral imaging is the microlens array plane P_1 and the sensor array plane P_2 . However, the P_2 plane can be easily replaced by its image P_3 which is a hypothetical plane intersecting the 3D object volume (see Fig. 2.3). It is assumed that the captured elemental images are all in focus; by the way, this is necessary for a successful result in integral imaging. We will include both (P_1, P_2) and (P_1, P_3) plane pairs in the discussion.

Let the light field between P_3 and P_1 planes be parameterized via the discrete light field representation $L_d[m_3, n_3, m_1, n_1]$. Let the $[m_1, n_1]$ array represent the locations of the microlenses where the aperture of each microlens corresponds to a pixel on the P_1 plane. Following the definition given by Eq. 2.4, one can find the power density of the light emanating from (x_0, y_0, z_0) on the object surface and crossing the P_3 plane at $[m_3, n_3]$ and reaching the microlens at $[m_1, n_1]$ as

$$\hat{I}_d[m_3, n_3, m_1, n_1] = \frac{P[m_3, n_3, m_1, n_1]}{S_1 S_3} \quad , \quad (2.5)$$

where S_3 is the area of a pixel on P_3 , S_1 is the area of the aperture of the microlens on P_1 and $P[m_3, n_3, m_1, n_1]$ is the total power emanating from the pixel $[m_3, n_3]$

and reaching the aperture of the microlens at $[m_1, n_1]$ [40]. We used the notation $\hat{I}$ to represent the power flow between points on P_3 and P_1 planes; I is reserved for power density between P_1 and P_2 planes; similar notation is adopted for L and $\hat{L}$.

We place P_3 plane at z_3 and adjust the locations of P_1 and P_2 planes such that P_2 and P_3 planes become images of each other due to microlenses. Therefore, according to the lens magnification equation

$$\frac{z_1 - z_3}{z_2 - z_1} = \left(\frac{S_3}{S_2} \right)^{\frac{1}{2}}, \quad (2.6)$$

where z_1 , z_2 and z_3 are again the z -coordinates of the P_1 , P_2 and P_3 planes, respectively, and S_2 is the area of the pixel $[m_2, n_2]$ which is the image of the pixel $[m_3, n_3]$ (see Fig. 2.3).

Since the light power reaching the microlens $[m_1, n_1]$ from an object point (x_0, y_0, z_0) via the pixel $[m_3, n_3]$ on P_3 flows to the point $[m_2, n_2]$ on P_2 unchanged (lossless lenses), where $[m_2, n_2]$ is the image of $[m_3, n_3]$, by using Eq. 2.5, we can write that

$$\begin{aligned} \hat{I}_d[m_3, n_3, m_1, n_1] S_1 S_3 &= P[m_3, n_3, m_1, n_1] \\ &= I_d[m_1, n_1, m_2, n_2] S_1 S_2 \quad . \end{aligned} \quad (2.7)$$

Hence we can write that

$$\hat{I}_d[m_3, n_3, m_1, n_1] = I_d[m_1, n_1, m_2, n_2] \left(\frac{S_2}{S_3} \right), \quad (2.8)$$

and this is consistent, as expected, with lens magnification between the P_2 and P_3 planes [40]. Together with Eq. 2.8, the two equations

$$\begin{aligned} \hat{L}_d[m_3, n_3, m_1, n_1] &= \hat{I}_d[m_3, n_3, m_1, n_1] \quad , \\ L_d[m_1, n_1, m_2, n_2] &= I_d[m_1, n_1, m_2, n_2] \end{aligned} \quad (2.9)$$

establish the desired equivalence either between (P_1, P_3) plane pairs or (P_1, P_2) plane pairs [40].

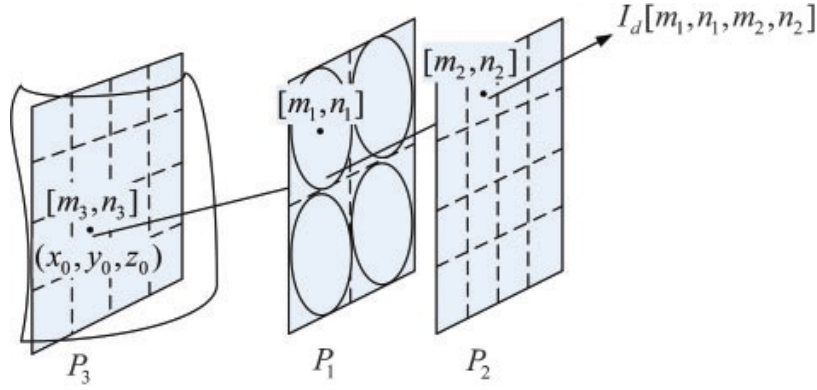


Figure 2.3: Parameterization of the light power density in integral imaging. (©2010 The Imaging Science Journal. Reprinted with permission. Published in [40])

The integral imaging can be represented as a discrete light field either between P_3 and P_1 or between P_1 and P_2 plane pairs, as given by Eq. 2.9, provided that the depth of focus of the microlenses is large enough to focus any point on the 3D object into a pixel on P_2 plane. Please note that this equivalence is valid under the assumption that there is no cross-talk between the elemental images from different microlenses. In other words, we restrict the set of light frustum within a finite propagation angle behind each microlens such that overlaps are prevented [40]. This can simply be achieved by partitioning the sensor array plane to non-overlapping regions such that each partition corresponds to the elemental image of a particular microlens and the leakage from a microlens to the elemental image of any other microlens is prevented [47].

In order to relate the integral imaging to the light field representation in the continuous case, we need infinitely many infinitesimal microlenses on the infinite extent P_1 , and infinitely many infinitesimal sensors on the infinite extent P_2 plane. Let us assume that the microlenses still possess the properties of an ideal lens even when their aperture sizes tend to zero. Even in this case, elimination of the cross-talk is practically impossible, since each elemental image size will also be infinitesimally small. Hence, the integral imaging cannot provide an equivalent representation to the continuous light field representation [40].

In all previous discussions, we assumed that microlenses provide a sufficient angular resolution for our purposes. Physically, a microlens having an infinitesimal aperture size behaves like a point light source and diffracts the incoming ray in an equally weighted manner to all angles. Therefore, our ability to assign an arbitrary propagation angle distribution is lost. This practical issue is a direct consequence of the uncertainty principle which states that we can not achieve arbitrary resolutions in both spatial and spectral (angular) domains of a signal [48]. Thus, the representation of the ray power densities with arbitrary resolutions in both spatial and angular domains is impossible via the light field representation. In Sec. 2.4, we develop a rigorous mathematical relation regarding this spatio-angular resolution tradeoff in the light field representation.

2.4 The Spatio-Angular Resolution Tradeoff in the Light Field Representation

As stated in Sec. 2.3, due to physical constraints we can not assign arbitrary values to the rays, which are defined by a point and a direction, since an infinitesimal point emits the light in an equally weighted manner to all directions. Thus, an equality between the light field representation and the integral imaging is possible, if the microlenses have finite size apertures. Such an equality is developed for the discrete case in Sec. 2.3. In this section, we investigate the integral imaging by using the scalar wave optics [45]. By this way, we are able to see the spatio-angular resolution tradeoff in the light field representation. Note that the link established between the light field representation and integral imaging in Sec. 2.3 leads us towards such an aim.

Firstly, let us write the effect of a thin lens (with focal length f) on a wave field as [8],

$$u_f(x, y) = u(x, y)p_l(x, y) \exp(j\phi_c) \exp \left\{ -\frac{j\pi}{\lambda f}(x^2 + y^2) \right\} \quad , \quad (2.10)$$

where $u(x, y)$ is the wave field incident on the lens plane, $p_l(x, y)$ is the lens aperture function which is defined as 1 inside the lens aperture and 0 outside of it; $u_f(x, y)$ is the wave field just after the lens. Note that $\exp(j\phi_c)$ is a constant phase factor which is determined by the thickness of the lens. We will neglect this constant phase factor in the following discussions. Now let us assume that the lens plane is placed at $z = 0$. Then, the wave field at a distance z can be found as [8],

$$u(x, y, z) = \frac{\exp \left(\frac{j2\pi}{\lambda} z \right)}{j\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(\xi, \eta) p_l(\xi, \eta) \exp \left\{ -\frac{j\pi}{\lambda f}(\xi^2 + \eta^2) \right\} \exp \left\{ \frac{j\pi}{\lambda z} [(x - \xi)^2 + (y - \eta)^2] \right\} d\xi d\eta \quad , \quad (2.11)$$

where $u_0(x, y) = u(x, y, 0)$. Note that here we employ the paraxial approximation to ease the mathematical analysis. If we move the terms independent of ξ and η to out of the integral and rewrite the equation for the back focal plane at $z = f$, we end up with

$$u(x, y, f) = \phi(x, y, f) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(\xi, \eta) p_l(\xi, \eta) \exp \left\{ -\frac{j2\pi}{\lambda f}(x\xi + y\eta) \right\} d\xi d\eta \quad , \quad (2.12)$$

where

$$\phi(x, y, f) = \frac{1}{j\lambda f} \exp \left(\frac{j2\pi}{\lambda} f \right) \exp \left\{ \frac{j\pi}{\lambda f}(x^2 + y^2) \right\} \quad . \quad (2.13)$$

The intensity of the field at (x, y) on the back focal is found as

$$I(x, y) = |u(x, y, f)|^2 = \Phi \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(\xi, \eta) p_l(\xi, \eta) \exp \left\{ -\frac{j2\pi}{\lambda f}(x\xi + y\eta) \right\} d\xi d\eta \right|^2 \quad , \quad (2.14)$$

where

$$\Phi = \phi(x, y, f)\phi^*(x, y, f) = \frac{1}{\lambda^2 f^2} \quad (2.15)$$

is constant on the back focal plane. If we substitute $f_x = \frac{x}{\lambda f}$ and $f_y = \frac{y}{\lambda f}$ into Eq. 2.14 as

$$I(f_x \lambda f, f_y \lambda f) = \Phi \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(\xi, \eta) p_l(\xi, \eta) \exp\{-j2\pi(f_x \xi + f_y \eta)\} d\xi d\eta \right|^2, \quad (2.16)$$

we see that the complex amplitude of the wave field at point (x, y, f) is actually determined by the Fourier component of the wave field $u(x, y, 0)$ (which is incident on the lens plane) at frequencies (f_x, f_y) [8]. (See Eq. 1.13) The 2D Fourier component at frequencies (f_x, f_y) is actually produced by the plane wave $\exp\{j2\pi(f_x x + f_y y + f_z z)\}$ where $f_x^2 + f_y^2 + f_z^2 = \frac{1}{\lambda^2}$. Noting that the propagation direction of the plane wave $\exp\{j2\pi(f_x x + f_y y + f_z z)\}$ is given by the vector $[f_x, f_y, f_z]^T$, we can say that each 2D Fourier component on the $z = 0$ plane actually corresponds to a propagation direction. Therefore, in the following discussions we will use the frequency and angle interchangeably, for the sake of clarity.

Now, in order to extend the formulations to the integral imaging setup let us parameterize the power which is emanated from the microlens at the discrete point $[m_1, n_1]$, on the microlens array plane, and propagates towards the point (x, y) on the back focal plane as $L_o(m_1 M_1, n_1 N_1, f_x, f_y) = I(f_x \lambda f, f_y \lambda f)$. Here, M_1 and N_1 are the distances between the consecutive microlenses along the x and y axes, respectively. Please note that a similar expression to $L_o(m_1 M_1, n_1 N_1, f_x, f_y)$ is defined as the observable light field in [39]. We keep the same definition and therefore use the subscript o to refer to an observable light field. The observability comes from the ability to assign an angular variation to the light emitted from a finite size aperture. By using the observable light field $L_o(m_1 M_1, n_1 N_1, f_x, f_y)$, we can write Eq. 2.16 for the more general case of the

integral imaging setup as

$$L_o(m_1 M_1, n_1 N_1, f_x, f_y) = \Phi \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(\xi, \eta) p_l(\xi - m_1 M_1, \eta - n_1 N_1) \exp\{-j2\pi(f_x \xi + f_y \eta)\} d\xi d\eta \right|^2. \quad (2.17)$$

Thus, $L_o(m_1 M_1, n_1 N_1, f_x, f_y)$ is actually the magnitude squared of the Fourier transform of the windowed scalar wave field (apart from a constant) which is incident on the microlens array plane. In other words, the power which is emanated from a lens on the microlens array plane and propagates towards a point on the back focal plane (can be viewed as the sensor array plane) is given as the magnitude squared of the short time Fourier transform of the wave field incident on the microlens array plane, where the window function is the lens aperture function $p_l(x, y)$. Note that the magnitude squared of short time Fourier transform is called as the spectrogram. The spatio-angular resolution tradeoff in the observable light field is thus a direct consequence of the uncertainty principle inherent to spectrogram or STFT. In order to see that, let us rewrite Eq. 2.17 by using the Fourier transforms $U_0(f_x, f_y)$, $P_l(f_x, f_y)$ of $u_0(x, y)$, $p_l(x, y)$, respectively, as [49],

$$L_o(m_1 M_1, n_1 N_1, f_x, f_y) = \Phi \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_0(\mu, \nu) P_l(f_x - \mu, f_y - \nu) \exp\{j2\pi(m_1 M_1 \mu + n_1 N_1 \nu)\} d\mu d\nu \right|^2. \quad (2.18)$$

The spatial resolution of the light field representation $L_o(m_1 M_1, n_1 N_1, f_x, f_y)$ is determined by the extent of the lens aperture function $p_l(x, y)$ as given by Eq. 2.17. On the other hand, the angular resolution is determined by the extent of the Fourier transform $P_l(f_x, f_y)$ of the lens transfer function as given by Eq. 2.18. According to the uncertainty principle, both $p_l(x, y)$ and $P_l(f_x, f_y)$ can not be made arbitrarily narrow [48]. As a consequence of that, arbitrary resolutions in space

and angle (frequency) of the light field representation $L_o(m_1M_1, n_1N_1, f_x, f_y)$ can not be obtained simultaneously. If we decrease the size of the aperture function, we achieve finer spatial resolution and as a tradeoff the angular (spectral) resolution is degraded, and vice versa. Thus, the spatio-angular resolution tradeoff in the light field representation $L_o(m_1M_1, n_1N_1, f_x, f_y)$ is actually characterized by the aperture function $p_l(x, y)$. Note that we developed the formulations in this section by using the Fourier transform properties of the lenses. Although, we dealt with the intensity of the wave field at the back focal plane, the Fourier transform property of the lenses can be observed in any plane where the source is imaged [8].

The extent of the uncertainty related to the spatial and angular resolutions of the light field representation is given by the uncertainty principle. Before developing a mathematical expression about this uncertainty, let us give some definitions. The effective width Δx of a 2D signal $a(x, y)$ along the x -axis can be defined as [50],

$$\Delta x = \left[\frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_0)^2 |a(x, y)|^2 dx dy}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |a(x, y)|^2 dx dy} \right]^{1/2}, \quad (2.19)$$

where x_0 (the center of $a(x, y)$ along the x -axis) is defined by

$$x_0 = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x |a(x, y)|^2 dx dy}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |a(x, y)|^2 dx dy}. \quad (2.20)$$

The effective width Δy of $a(x, y)$ along the y -axis and effective widths Δf_x , Δf_y of the Fourier transform of $a(x, y)$ along the f_x and f_y axes, respectively, are defined similarly. For a 2D signal $a(x, y)$ which has effective widths Δx , Δy in space and effective widths Δf_x , Δf_y in frequency, as defined by Eq. 2.19, one form of the uncertainty principle states that [39, 50, 51]

$$\Delta x \Delta y \Delta f_x \Delta f_y \geq \frac{1}{16\pi^2}. \quad (2.21)$$

By taking $a(x, y) = p_l(x, y)$ and denoting the effective widths of $p_l(x, y)$ along the x and y axes as Δx and Δy , respectively, and the effective widths of $P_l(f_x, f_y)$ along the f_x and f_y axes as Δf_x and Δf_y , respectively, we can say that as a consequence of the uncertainty principle given by Eq. 2.21, the effective areas of $p_l(x, y)$ and $P_l(f_x, f_y)$ cannot be simultaneously made arbitrarily small. Here, we assume that $\Delta x \Delta y$ and $\Delta f_x \Delta f_y$ give the effective areas occupied by $p_l(x, y)$ and $P_l(f_x, f_y)$, respectively [50]. We define the spatial and angular (spectral) resolutions of the (observable) light field representation $L_o(m_1 M_1, n_1 N_1, f_x, f_y)$ as $1/(\Delta x \Delta y)$ and $1/(\Delta f_x \Delta f_y)$, respectively. Therefore, according to Eq. 2.21, there is a restriction on the joint spatial and angular (spectral) resolution $1/(\Delta x \Delta y \Delta f_x \Delta f_y)$ of $L_o(m_1 M_1, n_1 N_1, f_x, f_y)$ and indeed the joint resolution has the upper bound $16\pi^2$ when the unit of frequencies are in cycles per unit length. We call the light field representations which satisfy this joint resolution requirement as physically realizable light field representations to refer to fields which are valid under the scalar wave optics. Here, it is once more worth noting that the 2D Fourier component of the light field $L_o(m_1 M_1, n_1 N_1, f_x, f_y)$ at frequencies (f_x, f_y) is produced by the plane wave $\exp\{j2\pi(f_x x + f_y y + f_z z)\}$ whose propagation direction is specified by the vector $[f_x, f_y, f_z]^T$. Therefore, we used the frequency and angle interchangeably, for the sake of simplicity. The uncertainty principle derived by using the frequency variables can be actually applied to other scenarios where the light field is defined in a different form. For example, for the observable light field which is defined by using the slope variables $u = x/z$ and $v = y/z$ [39], the upper joint resolution bound $1/(\Delta x \Delta y \Delta u \Delta v)$ becomes $16\pi^2/\lambda^2$ which can be simply found by a coordinate transformation between the frequency and the slope variables, as noted in [39].

In this chapter, we constructed a mathematical relation between the ray optics (light field representation) and scalar wave optics. As a consequence of this relation, we provided an upper bound for the joint spatial and angular (spectral) resolutions of a physically realizable light field representation. In the subsequent

chapters of the dissertation, we adopt the scalar wave optics model of light while developing methods for the calculation of 3D diffraction field due to a 2D field given on a curved surface. The proposed methods provide expressions where the scalar diffraction field is written as a sum of local beams. Indeed, similar to what we do in this chapter, these expressions are achieved by applying space-frequency decomposition techniques to the diffraction field. Thus, there is a close link between the derivations given in this chapter and the methods proposed in subsequent chapters in the sense that the local beam decompositions proposed in the subsequent chapters can be treated as physically realizable light field representations. From this point of view, the implications of this chapter is particularly important for us towards the aim of developing solutions for the problem of scalar diffraction field calculation from a given field on a curved surface by using such physically realizable light field representations.

Chapter 3

Calculation of the 3D Scalar Diffraction Field from the Field Given on a Curved Surface by Using Linear and Local Signal Decomposition Methods

The interest of graphics and optics communities to computer-generated holography has been significant such that several methods have been proposed to generate digital holograms of 3D objects [11, 12]. For the simple case of planar objects, Rayleigh-Sommerfeld formulation and its paraxial approximation (Fresnel model) can be used to find the desired 3D field [8, 52]. For the more general case of 3D objects, having depth, the 3D object is commonly modeled as a collection of self luminous point sources [23, 25, 26, 27] and the 3D diffraction field due to the object is found by superposing the fields due to such point light sources. In some applications the object is assumed to be formed by planar

patches [28, 29, 30]. In this case, the plane wave decomposition method (angular spectrum) is applied to each patch to calculate the 3D diffraction field.

For the problem of finding the 3D field due to a field specified on a curved surface (or an object), the commonly used source model approaches [23, 25, 26, 27, 28, 29, 30] do not produce accurate results, since they ignore the mutual couplings between the sources (point sources or planar patches) used on the curved surface. In other words, if the 3D field is calculated by using such source model approaches, then the subsequently reconstructed field on the same curved surface is found to be inconsistent with the original one. The mutual coupling problem is solved by using a field model approach, based on PWD, in [53, 37] resulting in an exact method to calculate the diffraction field from a given curved surface. However, for large sizes of curved surfaces, this method becomes impractical because of an intolerable increase in the computational complexity and the related memory requirement.

In order to find the 3D scalar diffraction field that corresponds to the 2D field given on a curved surface, we use linear and local signal decompositions on the input surface. Such local signal decompositions can provide more accurate solutions than the source models. The 2D decomposition on the surface corresponds to a local beam decomposition in 3D space. Noting that these local beam decompositions can be treated as physically realizable light field representations, which are discussed in Chapter 2, we can say that the improvement in accuracy is actually a consequence of the spatio-angular tradeoff inherent to such light fields. We use signals (such as Gaussian signal) which have larger spatial supports than the sources used in source model approaches. In turn, the angular spreads of corresponding 3D fields get narrower. As will be discussed in a more detailed fashion in Chapter 4, such a reduction in the angular spread (increase in the angular resolution) is the main factor that makes the proposed local signal decomposition methods more accurate with respect to source model approaches.

Note that in Chapter 4, we present a special case of the framework given in this chapter where we use Gaussian-shaped elementary signals and provide explicit field expressions and related simulations.

In Sec. 3.1, we develop the linear signal decomposition formulation for planar input surfaces. Then, we generalize the approach to the smooth curved surfaces in Sec. 3.2. Under the generic framework developed in Sec. 3.1 and Sec. 3.2, the two linear and local signal decomposition methods short time Fourier transform (STFT) and wavelet decomposition are separately investigated and used for the defined problem in Sec. 3.3 and Sec. 3.4, respectively.

3.1 Planar Input Surfaces

As mentioned in Chapter 1, in this dissertation we deal with 3D monochromatic fields $u(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{C}$ which can be decomposed into a sum of propagating plane waves where the z components of the propagation directions are always positive. Thus,

$$u(\mathbf{x}) = \iint_B A(f_x, f_y) \exp(j2\pi \mathbf{f}_x^T \mathbf{x}) df_x df_y, \quad (3.1)$$

where $\mathbf{x} = [x, y, z]^T \in \mathbb{R}^3$, $\mathbf{f}_x = [f_x, f_y, f_z]^T \in \mathbb{R}^3$; $f_x^2 + f_y^2 + f_z^2 = \frac{1}{\lambda^2}$ and $f_z \geq 0$. In Eq. 3.1, λ is the wavelength of the plane waves and B denotes the 2D circular spatial frequency band $f_x^2 + f_y^2 \leq \frac{1}{\lambda^2}$ which is occupied by the propagating plane waves. We denote the Hilbert space of such 3D fields $u(\mathbf{x})$ as $\mathcal{H}_p$. The coefficient of each such propagating plane wave can be found by using the field values on the reference plane at $z = 0$ as

$$A(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) \exp\{-j2\pi(f_x x + f_y y)\} dx dy, \quad (3.2)$$

where $u_0(x, y) = u(x, y, 0)$. The coefficients given by Eq. 3.2 actually correspond to the two-dimensional Fourier transform of $u_0(x, y)$, where the Fourier transform is defined as in Eq. 1.13. The set of 2D functions $u_0(x, y) : \mathbb{R}^2 \rightarrow \mathbb{C}$ belongs to the

space of bandlimited functions which we define as $\mathcal{B} = \{u_0(x, y) \mid U_0(f_x, f_y) = 0 \text{ for } f_x^2 + f_y^2 > 1/\lambda^2\}$ where $U_0(f_x, f_y)$ is the two-dimensional Fourier transform of $u_0(x, y)$. At this point note that a 2D field given on the infinite extent reference plane at $z = 0$ uniquely determines the corresponding 3D field (provided that the propagation direction components of the waves along the z -axis is always positive). The 2D fields over the planes which are parallel to the reference plane can be found by using the RS diffraction model as described by Eq. 1.10. In this dissertation, what we call as the 3D field is the concatenation of such 2D fields over the parallel planes at different depths [35]. With these definitions, the meaning of a 3D field, given a 2D field on the reference plane at $z = 0$, should be correctly interpreted: The 3D field at a point gives the scalar field on a plane which passes through that point and parallel to the reference plane [35]. $u(\mathbf{x}) \in \mathcal{H}_p$ must be interpreted as such a 3D field.

In this section, we assume that the intersection of a 3D field $u(\mathbf{x}) \in \mathcal{H}_p$ by the plane at $z = 0$ is given as an input. We want to find the 3D field due to $u_0(x, y) \in \mathcal{B}$ by using local beams corresponding to elementary functions on the input surface. Therefore, we decompose $u_0(x, y)$ into a sum of elementary functions on $z = 0$ plane. We continue the discussion by giving the related generic linear and local signal decomposition framework for the continuous and discrete cases.

3.1.1 Continuous Decomposition for Planar Input Surfaces

In the continuous case, the elementary functions $g_{\xi, \eta, \alpha, \beta}(x, y)$ are continuously indexed by $\mathbb{R}^4$. Here, (ξ, η) represents the position of the corresponding elementary function on the planar surface. α and β are the variables related to the

frequency. The coefficients of the elementary functions are found as

$$T_u(\xi, \eta, \alpha, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) g_{\xi\eta\alpha\beta}^*(x, y) dx dy \quad , \quad (3.3)$$

where $u_0(x, y) \in \mathcal{B}$. Using the coefficients given by Eq. 3.3, $u_0(x, y)$ can be written as a sum of elementary functions as

$$u_0(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_u(\xi, \eta, \alpha, \beta) g_{\xi\eta\alpha\beta}(x, y) d\mu(\xi, \eta, \alpha, \beta) \quad , \quad (3.4)$$

provided that the elementary functions satisfy the relation

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_{\xi\eta\alpha\beta}(x, y) g_{\xi\eta\alpha\beta}^*(\tilde{x}, \tilde{y}) d\mu(\xi, \eta, \alpha, \beta) = \delta(x - \tilde{x}, y - \tilde{y}) \quad , \quad (3.5)$$

which is known as the resolution of identity [54]. Here, $d\mu(\xi, \eta, \alpha, \beta)$ is the Haar measure related to the group of elementary functions $g_{\xi\eta\alpha\beta}(x, y)$ obtained by a set of transformations from $g(x, y)$. For example, $d\mu(\xi, \eta, \alpha, \beta) = d\xi d\eta d\alpha d\beta$ for the short time Fourier transform where α and β are the frequency variables.

Now, using the linearity of free space propagation, we can write the 3D field at the observation point $\mathbf{x} = [x, y, z]^T$ due to the input field $u_0(x, y)$ by summing up the field contributions of each elementary function as

$$u(\mathbf{x}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_u(\xi, \eta, \alpha, \beta) \tilde{g}_{\xi\eta\alpha\beta}(\mathbf{x} - \mathbf{r}) d\mu(\xi, \eta, \alpha, \beta) \quad , \quad (3.6)$$

where $\mathbf{r} = [\xi, \eta, 0]^T$ and $\tilde{g}_{\xi\eta\alpha\beta}(\mathbf{x} - \mathbf{r})$ is the 3D field at $\mathbf{x}$ that is due to the elementary function $g_{\xi\eta\alpha\beta}(x, y)$ defined on the input plane at $z = 0$. Thus, the 3D field is written as a sum of local beams each of which corresponds to an elementary function on the input surface.

3.1.2 Discrete Decomposition for Planar Input Surfaces

The continuous decompositions introduced in Sec. 3.1.1 are extremely redundant. It is possible to decompose a given signal into a much sparse set of elementary functions and still obtain a complete synthesis. Thus, a discrete set of

elementary functions may be sufficient. In order to have a decomposition with the desired properties (such as locality in both space and frequency) sometimes overcompleteness of the expansion functions may be a necessity. We may also end up with an overcomplete expansion when we sample continuous decompositions. The frame expansion theory provides a useful framework in order to deal with such cases, since the frames make it possible to work with the analysis and synthesis framework in a controlled manner [54, 55]. Frames provide discrete decompositions which are more general than the orthonormal expansions. Nonetheless, they include orthonormal expansions as a special case. Note that for the discrete decomposition cases we will use the frame expansions. Therefore, we continue this section by giving a more detailed information about the frame theory.

A set of elementary functions $\{g_{mnkl}(x, y); m, n, k, l \in \mathbb{Z}\}$ in a Hilbert space $\mathcal{H}$ constitutes a frame, if there exist some $A, B \in \mathbb{R}$ such that $0 < A \leq B < \infty$ and the frame inequality given as

$$AE_u \leq \sum_m \sum_n \sum_k \sum_l |\langle u_0(x, y), g_{mnkl}(x, y) \rangle|^2 \leq BE_u \quad (3.7)$$

is satisfied for all $u_0(x, y) \in \mathcal{H}$ [54]. Note that E_u is the energy of $u_0(x, y)$ which is given by

$$E_u = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |u_0(x, y)|^2 dx dy \quad . \quad (3.8)$$

By constructing a dual frame $\{w_{mnkl}(x, y)\}$ of $\{g_{mnkl}(x, y)\}$, we are able to write $u_0(x, y)$ as a sum of primal frame elements as [54],

$$u_0(x, y) = \sum_m \sum_n \sum_k \sum_l \langle u_0(x, y), w_{mnkl}(x, y) \rangle g_{mnkl}(x, y) \quad . \quad (3.9)$$

If the frame elements are linearly independent, then the frame is a Riesz basis [56]. In this case, the corresponding dual frame of $\{g_{mnkl}(x, y)\}$ is unique and it is actually the unique biorthogonal sequence of $\{g_{mnkl}(x, y)\}$. However, if the frame elements are not linearly independent, then there exist multiple dual frames

satisfying Eq. 3.9 [56]. In this case, the dual frame which results in an expansion with the set of coefficients having minimum L_2 -norm is mostly preferred. In both cases the dual frame elements are found by computing the inverse of the frame operator where the frame operator is defined as [56],

$$(Fu_0)(x, y) = \sum_m \sum_n \sum_k \sum_l \langle u_0(x, y), g_{mnkl}(x, y) \rangle g_{mnkl}(x, y) \quad . \quad (3.10)$$

If for some reason it is not feasible to find F^{-1} , one can find the coefficients of the frame elements iteratively. The number of necessary iterations, in order to achieve acceptable approximations to $u_0(x, y)$, depends on the ratio of frame bounds, $\frac{B}{A}$. The closer $\frac{B}{A}$ to one, the less number of iterations is needed in order to find a good approximation to $u_0(x, y)$. Note that the maximum of all possible lower bounds and the minimum of all possible upper bounds of the frame inequality (they are called as optimal frame bounds) are given, respectively, as

$$A = \frac{1}{\|F^{-1}\|}, \quad B = \|F\| \quad . \quad (3.11)$$

In the discrete case F is a matrix and hence the optimal frame bounds are given as the minimum and maximum of the eigenvalues of F [55].

Now, let us write $u_0(x, y)$ as

$$u_0(x, y) = \frac{2}{A+B} \sum_m \sum_n \sum_k \sum_l \langle u_0(x, y), g_{mnkl}(x, y) \rangle g_{mnkl}(x, y) + (Ru_0)(x, y) \quad , \quad (3.12)$$

where $R = I - \frac{2}{A+B}F$ [57]. Thus,

$$F^{-1} = \frac{2}{A+B}(I - R)^{-1} = \frac{2}{A+B} \sum_{i=0}^{\infty} R^i \quad . \quad (3.13)$$

Here, the Neumann series expansion of $(I - R)^{-1}$ is used, where R^i denotes i consecutive applications of the operator R . It can be verified by the frame condition that $-\frac{B-A}{B+A}I \leq R \leq \frac{B-A}{B+A}I$. This implies that

$$|R| < \frac{B-A}{B+A} = \frac{\frac{B}{A} - 1}{2 + (\frac{B}{A} - 1)} \quad . \quad (3.14)$$

Since $|R| < \frac{B-A}{B+A} < 1$, the series converge in norm to a limit of $(I - R)^{-1}$ [57]. Note that the closeness of $\frac{B}{A}$ to one determines the convergence rate. If A and B are sufficiently close to each other, i.e. $\frac{B-A}{B+A} \ll 1$, then even the zeroth order approximation produces good results. The exact dual frame elements are given in terms of R as

$$w_{mnkl}(x, y) = (F^{-1}g_{mnkl})(x, y) = \frac{2}{A+B} \sum_{i=0}^{\infty} (R^i g_{mnkl})(x, y) \quad . \quad (3.15)$$

As explained before, by taking into account only finite number of elements in the summation, the dual frame elements can still produce satisfactory results depending on the rate $\frac{B}{A}$. For this aim, $w_{mnkl}^N(x, y)$ is defined as [57],

$$\begin{aligned} w_{mnkl}^N(x, y) &= \frac{2}{A+B} \sum_{i=0}^N (R^i g_{mnkl})(x, y) \\ &= w_{mnkl}(x, y) - \frac{2}{A+B} \sum_{i=N+1}^{\infty} (R^i g_{mnkl})(x, y) \quad , \end{aligned} \quad (3.16)$$

where $w_{mnkl}^N(x, y)$ can be found by iteration as

$$w_{mnkl}^N(x, y) = \frac{2}{A+B} g_{mnkl}(x, y) + (R w_{mnkl}^{N-1})(x, y) \quad . \quad (3.17)$$

We end this section by giving the 3D field expression due to a 2D field $u_0(x, y) \in \mathcal{B}$ given on the planar input surface at $z = 0$. Before doing this, let us define the discrete evaluation points (the positions of the elementary functions) on the input plane as $\mathbf{r} = [\eta_{mnkl}, \xi_{mnkl}, 0]^T$. For example, $\eta_{mnkl} = m\eta_0$ and $\xi_{mnkl} = n\xi_0$ for the short time Fourier transform where η_0 and ξ_0 are the shift steps in η and ξ , respectively. The 3D field at the observation point $\mathbf{x}$ is then written as

$$u(\mathbf{x}) = \sum_m \sum_n \sum_k \sum_l T_u[m, n, k, l] \tilde{g}_{mnkl}(\mathbf{x} - \mathbf{r}) \quad , \quad (3.18)$$

where $T_u[m, n, k, l] = \langle u_0(x, y), w_{mnkl}(x, y) \rangle$ and $\tilde{g}_{mnkl}(\mathbf{x} - \mathbf{r})$ is the 3D field at $\mathbf{x}$ that is due to the elementary function $g_{mnkl}(x, y)$ defined at $(\eta_{mnkl}, \xi_{mnkl})$ on the input plane. Note that if $T_u[m, n, k, l]$ is found by using the approximate

dual frame elements $w_{mnkl}^N(x, y)$ (see Eq. 3.16) then the resulting coefficients of the frame elements, $\langle u_0(x, y), w_{mnkl}^N(x, y) \rangle$, produce an approximate solution to $u(\mathbf{x})$.

3.2 Curved Input Surfaces

In this section, we generalize the formulations developed in Sec. 3.1 to curved surfaces. We consider a continuously differentiable smooth curved surface $S \subset \mathbb{R}^3$. We represent the functions on S as $u(\mathbf{r}) : S \rightarrow \mathbb{C}$ where $\mathbf{r} \in S$. We aim to find the 3D field due to a 2D field $u(\mathbf{r})$ given on the curved surface S by using local beams which correspond to local elementary functions on S . Remember from Sec. 3.1 that the 3D field at a point gives the scalar field on a plane which passes through that point and parallel to the reference plane (which is used to define the entire 3D field as explained in Sec. 3.1). $u(\mathbf{r})$ represents a 2D field resulting from the intersection of such a 3D field $u(\mathbf{x}) \in \mathcal{H}_p$ by the curved surface S . With these definitions, the meaning of a 3D field, given a 2D field on a curved surface S , should be correctly interpreted: The 3D field at a point gives the scalar field on a plane which passes through that point and parallel to the reference plane [35]. Note that the extent of S along the reference plane is assumed to be infinite such that the projection of S onto the reference plane covers the entire plane. Therefore, under the assumption that the propagation direction components of the waves along the z -axis is restricted to be positive, the 3D field is uniquely determined by the 2D field on S [35].

Under the assumption that S is smooth enough within the support of the elementary functions where $u(\mathbf{r})$ is effective, we can represent the local expansion functions on the tangent planes of S . In order to define a local elementary function at $\mathbf{r}$ on S , we define the local coordinate system $(x_{\mathbf{r}}, y_{\mathbf{r}}, z_{\mathbf{r}})$ such that $x_{\mathbf{r}}$ and $y_{\mathbf{r}}$ axes are orthogonal to each other and they are placed on the tangent plane

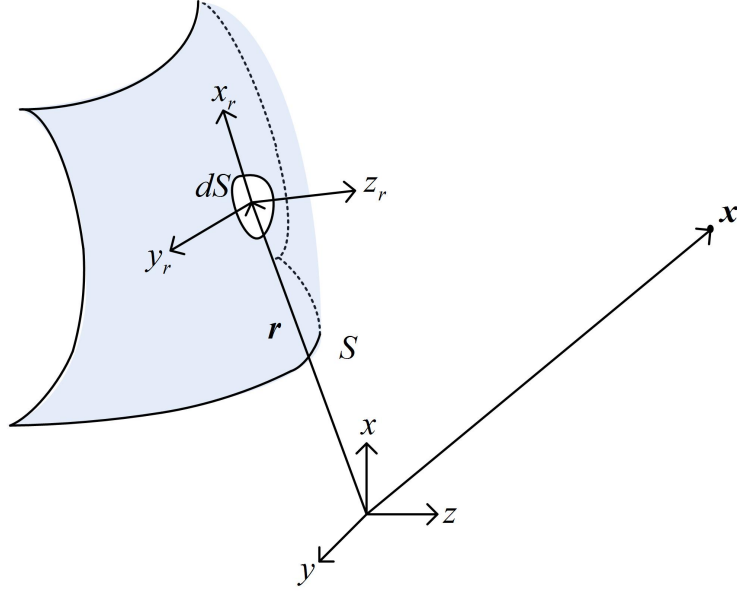


Figure 3.1: Local coordinate system on the input curved surface. (©2012 JOSA A. Reprinted with permission. Published in [35])

$T_{\mathbf{r}}$ of S at $\mathbf{r}$. The $z_{\mathbf{r}}$ -axis is chosen to give a right-handed orthogonal coordinate system (see Fig. 3.1). Please note that S is assumed to be an orientable surface. We represent the elementary function on $T_{\mathbf{r}}$ as $g_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$. Such an approach enables us to do the calculations on the tangent planes of S . For this aim, we firstly define $V_{T_{\mathbf{r}}}$ as the neighborhood of $\mathbf{r}$ on $T_{\mathbf{r}}$ outside of which $g_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ is effectively zero. Then, we define the neighborhood $V_{S_{\mathbf{r}}}$ of $\mathbf{r}$ on S such that its orthogonal projection onto $T_{\mathbf{r}}$ is $V_{T_{\mathbf{r}}}$. We represent the corresponding projected field on $V_{T_{\mathbf{r}}}$ as $u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}})$. Note that the smoothness assumption of S ensures that the projected field $u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ represents $u(\mathbf{r})$ on $V_{S_{\mathbf{r}}}$ with a negligible error.

3.2.1 Continuous Decomposition for Curved Input Surfaces

In the continuous decomposition case, the coefficients of the local expansion function $g_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ can be found by using the projected fields $u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ on

the planes of S as

$$\begin{aligned} T_u(\mathbf{r}, \alpha_{\mathbf{r}}, \beta_{\mathbf{r}}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) g_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}^*(x_{\mathbf{r}}, y_{\mathbf{r}}) dx_{\mathbf{r}} dy_{\mathbf{r}} \\ &\approx \iint_{V_{T_{\mathbf{r}}}} u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) g_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}^*(x_{\mathbf{r}}, y_{\mathbf{r}}) dx_{\mathbf{r}} dy_{\mathbf{r}} \quad . \end{aligned} \quad (3.19)$$

In order to calculate the diffraction field in 3D space due to the field $u(\mathbf{r})$ on S , we need to find the 3D field corresponding to the local functions defined on the tangent planes of S . Before doing this, let us relate the representation of a vector $\mathbf{x}$ in (x, y, z) global coordinate system to its representation $\mathbf{x}_{\mathbf{r}}$ in $(x_{\mathbf{r}}, y_{\mathbf{r}}, z_{\mathbf{r}})$ local coordinate system as

$$\mathbf{x}_{\mathbf{r}} = \mathbf{R}_{\mathbf{r}} \mathbf{x} - \mathbf{r} \quad . \quad (3.20)$$

Here $\mathbf{R}_{\mathbf{r}}$ is the 3×3 rotation matrix. Note that both the local coordinate system $(x_{\mathbf{r}}, y_{\mathbf{r}}, z_{\mathbf{r}})$ and $\mathbf{R}_{\mathbf{r}}$ change with position $\mathbf{r}$ on S . Hence, for a given field $u(\mathbf{r})$ on S we can find the diffraction field at $\mathbf{x}$ by an integration over the fields corresponding to the local elementary functions $g_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ as

$$u(\mathbf{x}) \approx \iint_S \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_u(\mathbf{r}, \alpha_{\mathbf{r}}, \beta_{\mathbf{r}}) \tilde{g}_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(\mathbf{x}_{\mathbf{r}}) d\mu(\mathbf{r}, \alpha_{\mathbf{r}}, \beta_{\mathbf{r}}) \quad , \quad (3.21)$$

where $\mathbf{x}_{\mathbf{r}}$ is the representation of $\mathbf{x}$ with respect to the local coordinate system at $\mathbf{r}$ and $\tilde{g}_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(\mathbf{x}_{\mathbf{r}})$ is the 3D field that corresponds to the local elementary function $g_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ defined on $T_{\mathbf{r}}$. Note that we can find explicit expressions for the 3D field components $\tilde{g}_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(\mathbf{x}_{\mathbf{r}})$, since the corresponding functions $g_{\mathbf{r}\alpha_{\mathbf{r}}\beta_{\mathbf{r}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ are defined on the tangent planes of S .

While developing Eq. 3.21, we treat the patches on the curved surface independently. Here, we use the word patch to refer to a surface element on the curved surface that represents the support of the elementary function written on it. If a beam induced by a patch illuminates another disjoint patch on the curved surface, then such patches can not be treated independently [53]; we call such patches as mutually coupled patches. Independent treatment of mutually

coupled patches produce inconsistent 3D field solutions, since a beam induced by one of the mutually coupled patches alters the field value on the other patch [53]. In this chapter and in Chapter 4, we restrict the set of curved surfaces to a feasible set for which it is possible to keep the mutual couplings at negligible levels, so that, the developed formulations are applicable. Note that a more detailed investigation of mutual coupling problem will be given in Chapter 4.

Although it is not necessary for the development of Eq. 3.21, let us also write the 2D field on S as a sum of elementary functions defined on S . In order to this, we back project the elementary functions, defined on the tangent planes of S , onto S . Thus, we define the elementary function at $\mathbf{r}$ on S in terms of the elementary function defined on T_r as

$$g_{\mathbf{r}\alpha_r\beta_r}(\mathbf{p}) = \begin{cases} g_{\mathbf{r}\alpha_r\beta_r}(x_{\mathbf{r}_p}, y_{\mathbf{r}_p}), & \text{if } \mathbf{p} \in V_{S_r} \\ 0, & \text{else} \end{cases}, \quad (3.22)$$

where $[x_{\mathbf{r}_p}, y_{\mathbf{r}_p}, z_{\mathbf{r}_p}]^T = \mathbf{R}_r \mathbf{p} - \mathbf{r}$. Hence, the 2D field on the curved surface S can be approximated as a sum of elementary functions on it as,

$$u(\mathbf{r}) \approx \iint_S \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_u(\mathbf{p}, \alpha_{\mathbf{p}}, \beta_{\mathbf{p}}) g_{\mathbf{p}\alpha_{\mathbf{p}}\beta_{\mathbf{p}}}(\mathbf{r}) d\mu(\mathbf{p}, \alpha_{\mathbf{p}}, \beta_{\mathbf{p}}) \quad . \quad (3.23)$$

3.2.2 Discrete Decomposition for Curved Input Surfaces

By choosing the discrete set of evaluation points $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n\}$ on S on a sufficiently dense grid and shift steps in $\alpha_{\mathbf{r}_i}$ and $\beta_{\mathbf{r}_i}$ sufficiently small, it is still possible to decompose the field on S into a discrete set of elementary functions on the tangent planes of S . Assuming that the set of elementary functions $\{g_{\mathbf{r}_i kl}(\mathbf{r}); \mathbf{r}_i \in S, k, l \in \mathbb{Z}\}$ on S (see Eq. 3.22) constitutes a frame, the coefficient of the

elementary function $g_{\mathbf{r}_i kl}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i})$ defined on $T_{\mathbf{r}_i}$ can be found as

$$\begin{aligned} T_u[\mathbf{r}_i, k, l] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{\mathbf{r}_i}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) w_{\mathbf{r}_i kl}^*(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) dx_{\mathbf{r}_i} dy_{\mathbf{r}_i} \\ &\approx \iint_{V_{T_{\mathbf{r}_i}}} u_{\mathbf{r}_i}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) w_{\mathbf{r}_i kl}^*(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) dx_{\mathbf{r}_i} dy_{\mathbf{r}_i} \quad , \end{aligned} \quad (3.24)$$

where $\{w_{\mathbf{r}_i kl}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i})\}$ is the set of projected dual elementary functions on the tangent planes of S at the discrete evaluation points.

In the discrete case, the 3D field at the observation point $\mathbf{x}$ can be found by summing up the 3D field components corresponding to the elementary functions defined on the tangent planes of S as

$$u(\mathbf{x}) \approx \sum_{i=1}^n \sum_k \sum_l T_u[\mathbf{r}_i, k, l] \tilde{g}_{\mathbf{r}_i kl}(\mathbf{x}_{\mathbf{r}_i}) \quad (3.25)$$

and $u(\mathbf{r})$ can be approximated using the elementary functions, defined by Eq. 3.22, on S as

$$u(\mathbf{r}) \approx \sum_{i=1}^n \sum_k \sum_l T_u[\mathbf{r}_i, k, l] g_{\mathbf{r}_i kl}(\mathbf{r}) \quad . \quad (3.26)$$

Note that in the discrete case, $V_{T_{\mathbf{r}}}$ is used to represent the support of the corresponding projected dual elementary function on $T_{\mathbf{r}}$. If the neighborhood $V_{S_{\mathbf{r}}}$ of $\mathbf{r}$ on S is defined such that its orthogonal projection onto $T_{\mathbf{r}}$ is $V_{T_{\mathbf{r}}}$, then for the applicability of Eq. 3.24, the dual elementary functions should be sufficiently narrow such that the surface can be treated as locally planar within all $V_{S_{\mathbf{r}_i}}$ where $u(\mathbf{r})$ is effective.

Up to this point we gave the 3D field expressions due to a 2D field specified on a planar or curved input surface by using continuous or discrete decompositions. Here it is worth noting that for all cases, once we have the entire 3D field expression due to the continuous field specified on the input surface, we can find the 2D field on another surface by intersecting the 3D field. By this way, we construct a field relation between two curved surfaces. Such a field relation can

be also constructed for the special cases, that are explained in Sec. 3.3, Sec. 3.4 and Chapter 4, of the generic framework given up to this point.

3.3 Short Time Fourier Transform

In this section and in Sec. 3.4, we present two linear and local signal decomposition methods, namely the short time Fourier transform and wavelet decomposition, respectively. We give the formulations for both planar and smooth curved input surfaces. We develop 3D field expressions due to 2D fields given on such surfaces. The 3D field is written as a sum of local beams each of which corresponds to an elementary function on the input surface. The short time Fourier transform and wavelet decomposition are the special cases of the generic linear and local signal decomposition framework, given in Sec. 3.1 and Sec. 3.2, in the sense of construction of the set of elementary functions to be used in the decomposition. Therefore, we use the generic framework presented up to this section in this chapter, while developing the 3D field expressions. Note that in Chapter 4, we go further to develop explicit 3D field expressions using Gaussian signal decomposition method which is actually a special case of the short time Fourier decomposition.

3.3.1 STFT for Planar Input Surfaces

The elementary functions, continuously indexed by $\mathbb{R}^4$, used in STFT are obtained as $g_{\xi,\eta,f_x,f_y}(x,y) = g(x - \xi, y - \eta) \exp\{j2\pi(f_x x + f_y y)\}$, where f_x and f_y are the frequency variables. Thus, the field $u_0(x,y) \in \mathcal{B}$ given on the input plane at $z = 0$ is written as a sum of shifted and modulated versions of $g(x,y)$ as [58],

$$u_0(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} STFT_u(\xi, \eta, f_x, f_y) g_{\xi,\eta,f_x,f_y}(x,y) d\xi d\eta df_x df_y \quad , \quad (3.27)$$

where

$$STFT_u(\xi, \eta, f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) g_{\xi\eta f_x f_y}^*(x, y) dx dy \quad . \quad (3.28)$$

Note that in order to satisfy Eq. 3.5, the elementary functions should satisfy the relation

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g_{\xi\eta f_x f_y}(x, y)|^2 d\xi d\eta df_x df_y = 1 \quad . \quad (3.29)$$

The 3D field at the observation point $\mathbf{x}$ due to $u_0(x, y)$ can thus be found as

$$u(\mathbf{x}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} STFT_u(\xi, \eta, f_x, f_y) \tilde{g}_{\xi\eta f_x f_y}(\mathbf{x} - \mathbf{r}) d\xi d\eta df_x df_y \quad , \quad (3.30)$$

where $\mathbf{r} = [\xi, \eta, 0]^T$ and $\tilde{g}_{\xi\eta f_x f_y}(\mathbf{x} - \mathbf{r})$ is the 3D field at $\mathbf{x}$ corresponding to the elementary function $g_{\xi\eta f_x f_y}(x, y)$ defined on the plane at $z = 0$. Note that elementary functions with different modulation frequencies correspond to beams with different propagation directions. Thus, the 3D field is written as a sum of local beams which are induced from different positions on $z = 0$ plane and have different propagation directions.

In the discrete case, $u_0(x, y)$ is decomposed into a sum of discrete number of elementary functions which are obtained as $g_{mnkl}(x, y) = g(x - mX, y - nY) \exp\{j2\pi(kF_x x + lF_y y)\}$, where X, Y are the shift steps in space and F_x, F_y are the shift space in frequency. Note that such a uniform sampling of the space frequency domain is a natural choice for the STFT. As explained in Sec. 3.1.2, if $\{g_{mnkl}(x, y); m, n, k, l \in \mathbb{Z}\}$ constitutes a frame, a numerically stable reconstruction is possible. In order that $\{g_{mnkl}(x, y); m, n, k, l \in \mathbb{Z}\}$ constitutes a frame, the relations $XF_x \leq 1$ and $YF_y \leq 1$ should be satisfied [57, 59]. Nevertheless, at the critical density, $XF_x = 1$ and $YF_y = 1$, it is not possible to have a frame with elementary functions that are localized in both space and frequency. This is explained by the Balian Low theorem [60]. Such a behavior makes it meaningless to work at the critical density, since otherwise we loose the localization in either

space or frequency. Under the assumption that $XF_x < 1$ and $YF_y < 1$, the discrete version of Eq. 3.27 can be written using the frame expansion given by Eq. 3.9 as

$$u_0(x, y) = \sum_m \sum_n \sum_k \sum_l \langle u_0(x, y), w_{mnkl}(x, y) \rangle g_{mnkl}(x, y) \quad , \quad (3.31)$$

where $\{w_{mnkl}(x, y); m, n, k, l \in \mathbb{Z}\}$ constitutes the dual frame. Note that Daubechies [61] gives the frame bound estimates and dual functions for various step sizes satisfying the relations $XF_x < 1$ and $YF_y < 1$. Although these results are developed for 1D signals they can be applied to 2D signals, defined on a planar surface, for separable elementary functions. Once the dual elementary function $w(x, y)$ of $g(x, y)$ is found, the elementary functions constituting the dual frame are simply obtained as $w_{mnkl}(x, y) = w(x - mX, y - nY) \exp\{j2\pi(kF_x x + lF_y y)\}$.

In the discrete case, the 3D field at the observation point $\mathbf{x}$ due to $u_0(x, y)$ can be written as

$$u(\mathbf{x}) = \sum_m \sum_n \sum_k \sum_l STFT_u[m, n, k, l] \tilde{g}_{mnkl}(\mathbf{x} - \mathbf{r}) \quad , \quad (3.32)$$

where $\mathbf{r} = [mX, nY, 0]^T$, $STFT_u[m, n, k, l] = \langle u_0(x, y), w_{mnkl}(x, y) \rangle$ and $\tilde{g}_{mnkl}(\mathbf{x} - \mathbf{r})$ is the 3D field at $\mathbf{x}$ corresponding to the elementary function $g_{mnkl}(x, y)$ defined on the plane at $z = 0$.

For the case that $g(x, y)$ is a Gaussian elementary signal, the decomposition of a 2D signal as a sum of such elementary signals was suggested by Gabor [62]. Bastiaans has several studies [63, 64] that aim to calculate the Gabor's expansion coefficients and expand an optical signals into a set of Gaussian beams. However, in these works an explicit 3D Gaussian beam decomposition expression is not given. Note that in Chapter 4, we will give explicit 3D field expressions using Gaussian beams.

3.3.2 STFT for Curved Input Surfaces

Under the assumption that S is smooth enough within the support of the elementary functions where $u(\mathbf{r})$ is effective, it is possible to represent the elementary functions on the tangent plane $T_{\mathbf{r}}$ of S at $\mathbf{r}$ as $g_{\mathbf{r}f_{x_{\mathbf{r}}}f_{y_{\mathbf{r}}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$. Then, the coefficient of $g_{\mathbf{r}f_{x_{\mathbf{r}}}f_{y_{\mathbf{r}}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ can be found as

$$\begin{aligned} STFT_u(\mathbf{r}, f_{x_{\mathbf{r}}}, f_{y_{\mathbf{r}}}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) g_{\mathbf{r}f_{x_{\mathbf{r}}}f_{y_{\mathbf{r}}}}^*(x_{\mathbf{r}}, y_{\mathbf{r}}) dx_{\mathbf{r}} dy_{\mathbf{r}} \\ &\approx \iint_{V_{T_{\mathbf{r}}}} u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) g_{\mathbf{r}f_{x_{\mathbf{r}}}f_{y_{\mathbf{r}}}}^*(x_{\mathbf{r}}, y_{\mathbf{r}}) dx_{\mathbf{r}} dy_{\mathbf{r}} \quad , \quad (3.33) \end{aligned}$$

where $u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ is the projected local field as defined in Sec. 3.2. By using the back projected elementary function expression given by Eq. 3.22, the field on S can be approximated as

$$u(\mathbf{r}) \approx \iint_S \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} STFT_u(\mathbf{p}, f_{x_{\mathbf{p}}}, f_{y_{\mathbf{p}}}) g_{\mathbf{p}f_{x_{\mathbf{p}}}f_{y_{\mathbf{p}}}}(\mathbf{r}) dS df_{x_{\mathbf{p}}} df_{y_{\mathbf{p}}} \quad . \quad (3.34)$$

In order to calculate the diffraction field in 3D space due to the field $u(\mathbf{r})$ on S , the local functions $g_{\mathbf{r}f_{x_{\mathbf{r}}}f_{y_{\mathbf{r}}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ defined on the tangent planes of S can be used. We represent the 3D field at $\mathbf{x}$ due to $g_{\mathbf{r}f_{x_{\mathbf{r}}}f_{y_{\mathbf{r}}}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ as $\tilde{g}_{\mathbf{r}f_{x_{\mathbf{r}}}f_{y_{\mathbf{r}}}}(\mathbf{x}_{\mathbf{r}})$, where $\mathbf{x}_{\mathbf{r}}$ is the representation of $\mathbf{x}$ with respect to the local coordinate system at $\mathbf{r}$ (see Eq. 3.20). By using the defined local 3D beam expression, we can find the diffraction field at $\mathbf{x}$, given the field $u(\mathbf{r})$ on S , via an integration over the local beams as

$$u(\mathbf{x}) \approx \iint_S \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} STFT_u(\mathbf{r}, f_{x_{\mathbf{r}}}, f_{y_{\mathbf{r}}}) \tilde{g}_{\mathbf{r}f_{x_{\mathbf{r}}}f_{y_{\mathbf{r}}}}(\mathbf{x}_{\mathbf{r}}) dS df_{x_{\mathbf{r}}} df_{y_{\mathbf{r}}} \quad . \quad (3.35)$$

In the discrete case, if the set of evaluation points $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n\}$ on the curved surface S are taken on a sufficiently dense regular grid, then the dual frame elements can be calculated as if the problem is stated on a planar surface. However, it may not possible to place a regular grid on the entire curved surface.

For the case of irregular sampling of the elementary function positions, the frame conditions and frame bound estimates are derived in [65, 66, 67], for 1D signals. Although it is not straightforward to generalize these results to 2D setting, here we assume (for the completeness of the discussion) that the set of elementary functions $\{g_{\mathbf{r}_i kl}(\mathbf{r}); \mathbf{r}_i \in S, k, l \in \mathbb{Z}\}$ on S constitutes a frame and the dual frame elements are available as $\{w_{\mathbf{r}_i kl}(\mathbf{r}); \mathbf{r}_i \in S, k, l \in \mathbb{Z}\}$. Then, we can decompose $u(\mathbf{r})$ into a sum of elementary functions on the tangent planes of S by finding the coefficients of the elementary functions defined on $T_{\mathbf{r}_i}$ as

$$\begin{aligned} STFT_u[\mathbf{r}_i, k, l] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{\mathbf{r}_i}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) w_{\mathbf{r}_i kl}^*(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) dx_{\mathbf{r}_i} dy_{\mathbf{r}_i} \\ &\approx \iint_{V_{T_{\mathbf{r}_i}}} u_{\mathbf{r}_i}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) w_{\mathbf{r}_i kl}^*(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) dx_{\mathbf{r}_i} dy_{\mathbf{r}_i} \quad , \end{aligned} \quad (3.36)$$

where $\{w_{\mathbf{r}_i kl}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i})\}$ is the set of projected dual elementary functions at the discrete evaluation points $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n\}$.

By using the elementary functions on S (see Eq. 3.22) $u(\mathbf{r})$ can be approximated as

$$u(\mathbf{r}) \approx \sum_{i=1}^n \sum_k \sum_l STFT_u[\mathbf{r}_i, k, l] g_{\mathbf{r}_i kl}(\mathbf{r}) \quad . \quad (3.37)$$

In the discrete case, the 3D field at the observation point $\mathbf{x}$ is written as a sum of 3D field components corresponding to the elementary functions $g_{\mathbf{r}_i kl}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i})$ defined on the tangent planes of S ,

$$u(\mathbf{x}) \approx \sum_{i=1}^n \sum_k \sum_l STFT_u[\mathbf{r}_i, k, l] \tilde{g}_{\mathbf{r}_i kl}(\mathbf{x}_{\mathbf{r}_i}) \quad , \quad (3.38)$$

where $\tilde{g}_{\mathbf{r}_i kl}(\mathbf{x}_{\mathbf{r}_i})$ is the 3D field expression corresponding to the elementary function $g_{\mathbf{r}_i kl}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i})$ defined on $T_{\mathbf{r}_i}$. Note that for the applicability of Eq. 3.38 with the coefficients obtained by Eq. 3.36, the dual elementary functions should be sufficiently narrow such that the surface can be treated as locally planar within all $V_{S_{\mathbf{r}_i}}$ where $u(\mathbf{r})$ is effective.

A special case of short time Fourier decomposition, for the problem of diffraction field calculation from a given field on a curved surface, is investigated in Chapter 4. The elementary functions that we choose in that chapter are modulated and shifted versions of a Gaussian window function. Over the framework given in this chapter, we develop explicit 3D field expressions by introducing an approximate Gaussian beam expression and conduct related simulations in Chapter 4. The simulations regarding the application of short time Fourier decomposition in the problem of diffraction field calculation from a given field on a curved surface can be found in Sec. 4.2.

3.4 Wavelet Decomposition

3.4.1 Wavelet Decomposition for Planar Input Surfaces

The elementary functions of the wavelet decomposition are given as $g_{\xi\eta\alpha\theta}(x, y) = |\alpha|^{-1}g\left(\frac{1}{\alpha}\mathbf{r}_\theta^{-1}[x - \xi, y - \eta]^T\right)$ where $\alpha \in \mathbb{R}^+$ denotes the scale, θ is the rotation angle and $\mathbf{r}_\theta$ is the 2×2 (unitary) rotation matrix of angle θ such that $\theta \in [0, 2\pi)$. $[\xi, \eta] \in \mathbb{R}^2$ denotes the position of the corresponding elementary function. Thus, we can write the field $u_0(x, y) \in \mathcal{B}$ on the input plane at $z = 0$ as a sum of shifted, scaled and rotated versions of $g(x, y)$ (mother wavelet) [68],

$$u_0(x, y) = \frac{1}{C_g} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{2\pi} W_u(\xi, \eta, \alpha, \theta) g_{\xi\eta\alpha\theta}(x, y) d\xi d\eta \frac{d\alpha}{\alpha^3} d\theta \quad , \quad (3.39)$$

where

$$W_u(\xi, \eta, \alpha, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) g_{\xi\eta\alpha\theta}^*(x, y) dx dy \quad . \quad (3.40)$$

If the admissibility condition of the wavelet decomposition given as

$$C_g \triangleq 4\pi^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{|G(f_x, f_y)|^2}{f_x^2 + f_y^2} df_x df_y < \infty \quad , \quad (3.41)$$

is satisfied, then Eq. 3.39 produce the correct $u_0(x, y)$. For an integrable $g(x, y)$, Eq. 3.41 implies that $g(x, y)$ has a zero mean. In other words $G(0, 0) = 0$, where $G(f_x, f_y)$ is the 2D Fourier transform of $g(x, y)$. Note that the admissibility condition is sufficient for the existence of the synthesis given by Eq. 3.39 with the analysis coefficients given by Eq. 3.40. However, it is not necessary. As given in [69], it is still possible to obtain a synthesis equation for the wavelet used in [69] which does not satisfy the admissibility condition given by Eq. 3.41.

The elementary functions of the wavelet decomposition are given as $g_{\xi\alpha}(x) = |\alpha|^{-1/2}g\left(\frac{x-\xi}{\alpha}\right)$ in the 1D case. Thus, the elementary functions are the shifted and scaled versions of $g(x, y)$. While passing to the 2D case, instead of adding a second scale variable the rotated versions of $g(x, y)$ are used [68]. Note that a rotation by $\theta = \pi$ together with a positive scaling result in a negative scaling. Thus in order to represent $u_0(x, y) : \mathbb{R}^2 \rightarrow \mathbb{C}$, it is sufficient to take into account only the positive scales of the mother wavelet function.

By using the wavelet decomposition given by Eq. 3.39, we can write the 3D field at the observation point $\mathbf{x}$ due to $u_0(x, y)$ as

$$u(\mathbf{x}) = \frac{1}{C_g} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{2\pi} W_u(\xi, \eta, \alpha, \theta) \tilde{g}_{\xi\eta\alpha\theta}(\mathbf{x} - \mathbf{r}) d\xi d\eta \frac{d\alpha}{\alpha^3} d\theta \quad , \quad (3.42)$$

where $\mathbf{r} = [\xi, \eta, 0]^T$ and $\tilde{g}_{\xi\eta\alpha\theta}(\mathbf{x} - \mathbf{r})$ is the 3D field at $\mathbf{x}$ corresponding to the elementary function $g_{\xi\eta\alpha\theta}(x, y)$ defined on the plane at $z = 0$. Note that in the wavelet decomposition, the beams having different propagation directions are obtained by using differently scaled versions of the mother wavelet defined on the input surface.

The discrete set of elementary functions are obtained by taking the positions, scale and rotation variables at the discrete values $\xi = mXA_x^k$, $\eta = nXA_x^k$, $\alpha = A_x^k$ and $\theta = l\Theta$ where X and Θ are the shift steps in space and angle; $A_x > 1$ is a real constant determining the discrete scales. Thus, $g_{mnkl}(x, y) = A_x^{-k}g\left(\frac{1}{A_x^k}\mathbf{r}l\Theta^{-1}[x - mXA_x^k, y - nXA_x^k]^T\right)$; $m, n, k, l \in \mathbb{Z}$ and the discrete version

of Eq. 3.27 is written as a frame expansion,

$$u_0(x, y) = \sum_m \sum_n \sum_k \sum_l \langle u_0(x, y), w_{mnkl}(x, y) \rangle g_{mnkl}(x, y) \quad , \quad (3.43)$$

where $\{w_{mnkl}(x, y); m, n, k, l \in \mathbb{Z}\}$ constitutes the dual frame. The frame conditions for the wavelet decomposition, regarding the choices of A_x and X , are given by Daubechies [57] for the 1D case and generalized to 2D case in [68, 70]. Note that the discrete set of wavelet functions are constructed such that at coarser scales (corresponding to lower frequency band) the spectral resolution is finer at the expense of degradation in its spatial resolution and at finer scales (corresponding to higher frequency band) the spatial resolution is finer at the expense of degradation in its spectral resolution. Such a paving strategy of the space-frequency domain makes it possible to construct orthonormal wavelet decompositions which have localized elementary functions in both space and frequency. Remember that this is not possible for the STFT. Orthonormal wavelet decomposition is constructed in [71] where separable 2D wavelets at dyadic scales (2^k) are used. In that work of Mallat, orthonormal bases are constructed by a systematic procedure based on the multiresolution signal approximation framework. Various other orthonormal wavelets (such as Daubechies, Meyer and Battle-Lemarié) can be analyzed and constructed in this framework. The rotated versions of the wavelets are not included in this scheme. Instead of that, the 2D wavelets are constructed by separable products of the 1D scaling and wavelet functions [71]. Note that the dual elementary functions are equal to the primal wavelet functions in such orthonormal decompositions.

In the discrete case, the 3D field at the observation point $\mathbf{x}$ due to $u_0(x, y)$ is written as

$$u(\mathbf{x}) = \sum_m \sum_n \sum_k \sum_l W_u[m, n, k, l] \tilde{g}_{mnkl}(\mathbf{x} - \mathbf{r}) \quad , \quad (3.44)$$

where $W_u[m, n, k, l] = \langle u_0(x, y), w_{mnkl}(x, y) \rangle$ and $\tilde{g}_{mnkl}(\mathbf{x} - \mathbf{r})$ is the 3D field at $\mathbf{x}$ that is due to the elementary function $g_{mnkl}(x, y)$ on the plane at $z = 0$. Note

that the elementary functions at different scales correspond to beams having different propagation directions. Unlike the STFT, in the wavelet decomposition the angular extent of the beams are not the same. Indeed, as the scale gets finer the angular extents of the corresponding beams become larger.

The Fresnelet bases which are developed in [72] can be viewed as a special case of Eq. 3.44. In that development, the B-spline wavelets are used to construct bases for the input field defined on a planar surface and the Fresnel model is assumed while calculating the 3D fields corresponding to such basis functions. Thus, the Fresnel transform of each wavelet function $g_{mnkl}(x, y)$ defined on the input plane is called a Fresnelet and the field on a planar surface, parallel to the input plane, is written in the so called Fresnelet basis. As noted in [72], such an approach enables the reconstruction of the input field from the output field at various user-defined scales.

3.4.2 Wavelet Decomposition for Curved Input Surfaces

In order to write the field on a curved surface S as a sum of wavelets defined on the tangent planes of S , the wavelets used in the decomposition should be sufficiently local such that S can be treated as locally planar within the supports of the wavelet functions where $u(\mathbf{r})$ is effective. We define the neighborhood V_{S_r} of $\mathbf{r}$ on S such that S can be treated as locally planar within V_{S_r} for all $\mathbf{r} \in S$. Under the locality assumption of the wavelets, we can represent the elementary functions on the tangent plane T_r of S at $\mathbf{r}$ as $g_{r\alpha\theta}(x_r, y_r) = |\alpha|^{-1} g_r \left(\frac{1}{\alpha} \mathbf{r}_\theta^{-1} [x_r, y_r]^T \right)$. Here the 2D rotation is applied about the local z_r -axis. We can find the coefficients of the wavelets as

$$\begin{aligned} W_u(\mathbf{r}, \alpha, \theta) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_r(x_r, y_r) g_{r\alpha\theta}^*(x_r, y_r) dx_r dy_r \\ &\approx \iint_{V_{T_r}} u_r(x_r, y_r) g_{r\alpha\theta}^*(x_r, y_r) dx_r dy_r \quad , \end{aligned} \quad (3.45)$$

where $u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ is the projected local field onto $V_{T_{\mathbf{r}}}$ which is defined as the orthogonal projection of $V_{S_{\mathbf{r}}}$ onto $T_{\mathbf{r}}$. We denote the coarsest scale used in the decomposition as α_0 . Note that this coarsest scale is determined by the smoothness of the surface. $V_{T_{\mathbf{r}}}$ actually represents the neighborhood outside of which the wavelet at scale α_0 is effectively zero. The signal components at scales coarser than α_0 (larger than α_0 or smaller than $-\alpha_0$) can be represented by using lowpass scaling functions $s_{\mathbf{r}\alpha_0\theta}(x_{\mathbf{r}}, y_{\mathbf{r}}) = |\alpha_0|^{-1} s_{\mathbf{r}} \left(\frac{1}{\alpha_0} \mathbf{r} \boldsymbol{\theta}^{-1} [x_{\mathbf{r}}, y_{\mathbf{r}}]^T \right)$ at scale α_0 [73]. Note that the scaling functions are also assumed to be sufficiently narrow such that the surface can be treated as locally planar within all $V_{T_{\mathbf{r}}}$ where $u(\mathbf{r})$ is effective. If the scaling function $s_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}})$ is chosen such that

$$\int_0^{2\pi} |S_{\mathbf{r}\alpha_0\theta}(f_{x_{\mathbf{r}}}, f_{y_{\mathbf{r}}})|^2 d\theta = \int_0^{2\pi} \int_{\alpha_0}^{\infty} |G_{\mathbf{r}\alpha\theta}(f_{x_{\mathbf{r}}}, f_{y_{\mathbf{r}}})|^2 \frac{d\alpha}{\alpha^3} d\theta \quad , \quad (3.46)$$

then we can write $u(\mathbf{r})$, by using the back projected wavelet and scaling functions onto S , as

$$\begin{aligned} u(\mathbf{r}) \approx & \frac{1}{C_g} \iint_S \int_0^{\alpha_0} \int_0^{2\pi} W_u(\mathbf{p}, \alpha, \theta) g_{\mathbf{p}\alpha\theta}(\mathbf{r}) dS \frac{d\alpha}{\alpha^3} d\theta \\ & + \frac{1}{C_g \alpha_0^2} \iint_S \int_0^{2\pi} A_u(\mathbf{p}, \alpha_0, \theta) s_{\mathbf{p}\alpha_0\theta}(\mathbf{r}) dS d\theta \quad , \end{aligned} \quad (3.47)$$

where the coefficients of $s_{\mathbf{r}\alpha_0\theta}(x_{\mathbf{r}}, y_{\mathbf{r}})$ are found by

$$\begin{aligned} A_u(\mathbf{r}, \alpha_0, \theta) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) s_{\mathbf{r}\alpha_0\theta}^*(x_{\mathbf{r}}, y_{\mathbf{r}}) dx_{\mathbf{r}} dy_{\mathbf{r}} \\ &\approx \iint_{V_{T_{\mathbf{r}}}} u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) s_{\mathbf{r}\alpha_0\theta}^*(x_{\mathbf{r}}, y_{\mathbf{r}}) dx_{\mathbf{r}} dy_{\mathbf{r}} \quad . \end{aligned} \quad (3.48)$$

Note that $S_{\mathbf{r}\alpha_0\theta}(f_{x_{\mathbf{r}}}, f_{y_{\mathbf{r}}})$ and $G_{\mathbf{r}\alpha\theta}(f_{x_{\mathbf{r}}}, f_{y_{\mathbf{r}}})$ are the 2D Fourier transforms of the scaling and wavelet functions $s_{\mathbf{r}\alpha_0\theta}(x_{\mathbf{r}}, y_{\mathbf{r}})$, $g_{\mathbf{r}\alpha\theta}(x_{\mathbf{r}}, y_{\mathbf{r}})$, respectively. The complex phases of the scaling functions can be chosen arbitrarily. A more restrictive condition

$$|S_{\mathbf{r}\alpha_0}(f_{x_{\mathbf{r}}}, f_{y_{\mathbf{r}}})|^2 = \int_{\alpha_0}^{\infty} |G_{\mathbf{r}\alpha}(f_{x_{\mathbf{r}}}, f_{y_{\mathbf{r}}})|^2 \frac{d\alpha}{\alpha^3} \quad (3.49)$$

that results in the same expansion given by Eq. 3.47, can also be imposed on $s_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}})$.

We represent the 3D field at $\mathbf{x}$ due to the wavelet $g_{\mathbf{r}\alpha\theta}(x_{\mathbf{r}}, y_{\mathbf{r}})$ defined on the tangent plane $T_{\mathbf{r}}$ as $\tilde{g}_{\mathbf{r}\alpha\theta}(\mathbf{x}_{\mathbf{r}})$ where $\mathbf{x}_{\mathbf{r}}$ is the representation of $\mathbf{x}$ with respect to the local coordinate system at $\mathbf{r}$ on S . Similarly, we represent the 3D field at $\mathbf{x}$ due to the scaling function $s_{\mathbf{r}\alpha_0\theta}(x_{\mathbf{r}}, y_{\mathbf{r}})$ as $\tilde{s}_{\mathbf{r}\alpha_0\theta}(\mathbf{x}_{\mathbf{r}})$. Thus, for a given field $u(\mathbf{r})$ on S we can find the diffraction field at $\mathbf{x}$ by an integration over the fields corresponding to wavelets at scales $0 \leq \alpha \leq \alpha_0$ and scaling functions at scale α_0 as

$$\begin{aligned} u(\mathbf{x}) \approx & \frac{1}{C_g} \iint_S \int_0^{\alpha_0} \int_0^{2\pi} W_u(\mathbf{r}, \alpha, \theta) \tilde{g}_{\mathbf{r}\alpha\theta}(\mathbf{x}_{\mathbf{r}}) dS \frac{d\alpha}{\alpha^3} d\theta \\ & + \frac{1}{C_g \alpha_0^2} \iint_S \int_0^{2\pi} A_u(\mathbf{r}, \alpha_0, \theta) \tilde{s}_{\mathbf{r}\alpha_0\theta}(\mathbf{x}_{\mathbf{r}}) dS d\theta \quad . \end{aligned} \quad (3.50)$$

The discrete set of wavelet functions explained for the planar input surface case are obtained by translating and scaling a particular function which is called as the mother wavelet. These wavelets are known as the first generation wavelets [74]. For the case of curved surfaces, such an approach can not be applicable, unless the discrete positions of the wavelet functions are not taken on a regular grid placed over the curved surface. In order to overcome such a situation, second generation wavelets [75] are well suited, since they can be used for irregular sampling schemes. Note that for the application of discrete STFT to curved surfaces, where the discrete positions of elementary functions are taken on an irregular grid, we did not talk about an explicit construction scheme for the primal and dual elementary functions. The reason is that, as far as we know, an extension taking into account the irregular sampling cases does not exist for the STFT case. Here for the wavelet case, we give some explanation about the construction of discrete decomposition for irregular sampling schemes based on the second generation wavelets.

Let us assume that $u(\mathbf{r})$ is specified on a sufficiently fine mesh which we call as the representation of the field at scale $k = 0$. We can then construct a wavelet decomposition by hierarchically representing $u(\mathbf{r})$ on coarser meshes at scales $k \geq 1$. We denote the indices of nodes of the mesh at scale k as $M(k)$ and the nodes which are added to $M(k)$ to obtain the finer mesh $M(k-1)$ as $N(k)$. Note that the subdivision points $N(k)$ can be related to $M(k)$ by a recursive model [74]. In the simplest case, they are taken as the mid-points of $M(k)$. By using the lifting scheme introduced in [75], it is possible to construct second generation wavelets and their corresponding scaling functions for irregularly sampled domains. Note that the resulting scaling functions $s_{km}(\mathbf{r})$ and wavelet functions $g_{kn}(\mathbf{r})$ are located at nodes with indices $m \in M(k)$ and $n \in N(k)$, respectively. Under these definitions, we can write the approximated version of $u(\mathbf{r})$ at scale $k = k_0$ as

$$u_{k_0}(\mathbf{r}) = \sum_{m \in M(k_0)} \langle u(\mathbf{r}), h_{k_0 m}(\mathbf{r}) \rangle s_{k_0 m}(\mathbf{r}) \quad , \quad (3.51)$$

where $h_{km}(\mathbf{r})$ is the dual (biorthogonal) function of $s_{km}(\mathbf{r})$. By adding the wavelet functions hierarchically at scales $k \leq k_0$, we can improve the approximation and end up with the original signal defined on the finest mesh at scale zero. Thus,

$$\begin{aligned} u(\mathbf{r}) &= \sum_{m \in M(k_0)} \langle u(\mathbf{r}), h_{k_0 m}(\mathbf{r}) \rangle s_{k_0 m}(\mathbf{r}) \\ &+ \sum_{k=1}^{k_0} \sum_{n \in N(k)} \langle u(\mathbf{r}), w_{kn}(\mathbf{r}) \rangle g_{kn}(\mathbf{r}) \quad , \end{aligned} \quad (3.52)$$

where $w_{kn}(\mathbf{r})$ is the dual function of $g_{kn}(\mathbf{r})$ [76]. Note that the wavelet and scaling functions are defined on the curved surface in the explained scheme. In order to use such a decomposition for the calculation of diffraction field from a given field on a curved surface, the scaling and wavelet functions used in Eq. 3.52 should be sufficiently narrow such that the surface can be treated as locally planar within their supports. In other words, the scale k_0 which we start the decomposition should be sufficiently small. Under these assumptions, we can

project the scaling and wavelet functions on the tangent planes of S and then calculate the 3D field by summing up their contributions as

$$\begin{aligned}
u(\mathbf{x}) = & \sum_{m \in M(k_0)} \langle u(\mathbf{r}), h_{k_0 m}(\mathbf{r}) \rangle \tilde{s}_{k_0 m}(\mathbf{x} - \mathbf{r}) \\
& + \sum_{k=1}^{k_0} \sum_{n \in N(k)} \langle u(\mathbf{r}), w_{kn}(\mathbf{r}) \rangle \tilde{g}_{kn}(\mathbf{x} - \mathbf{r}) \quad , \quad (3.53)
\end{aligned}$$

where $\tilde{s}_{km}(\mathbf{x} - \mathbf{r})$ and $\tilde{g}_{km}(\mathbf{x} - \mathbf{r})$ are the 3D fields at $\mathbf{x}$ corresponding to the projected versions of $s_{km}(\mathbf{r})$ and $g_{km}(\mathbf{r})$, respectively.

In this chapter, we presented a generic approach that provides a unified framework, for the problem of 3D diffraction field calculation from 2D fields given on planar or smooth curved surfaces, by using linear and local signal decomposition methods. Under this framework, the choice of a specific decomposition depends on the desired properties of the proposed solution. In the subsequent chapter, we introduce a Gaussian signal decomposition method which is actually a specific case of the short time Fourier decomposition given in Sec. 3.3. The nice properties of Gaussian beams, that will be discussed in Chapter 4, makes such a decomposition attractive. On the other hand, the wavelet decomposition can be suitable, for example, for the cases where the diffraction field is desired to be calculated at various user-defined scales. In fact, as mentioned before, the wavelet decomposition is used for such an aim in [72] for planar input surfaces.

Chapter 4

Calculation of the 3D Scalar Diffraction Field from the Field Given on a Curved Surface via Gaussian Signal Decomposition

In this chapter, we introduce a Gaussian signal decomposition method for the calculation of the 3D diffraction field from a given 2D field on a curved surface. The Gaussian signal decomposition is a linear and local signal decomposition method and it is actually a special case of the short time Fourier decomposition presented in Sec. 3.3 where the elementary functions are the shifted and modulated versions of a Gaussian window function. Therefore, we develop the formulations in this chapter under the framework given in Sec. 3.2 and Sec. 3.3. In Chapter 3, we did not give an explicit 3D field expression for a 2D field specified on a curved surface. In this chapter, we develop explicit 3D field expressions for planar and curved input surfaces by using the proposed approximate Gaussian beam expression. In the proposed Gaussian signal decomposition method, we firstly decompose a given field on a 2D curved surface into a sum of properly

shifted and modulated Gaussian-shaped elementary signals. Then we write the 3D diffraction field as a sum of Gaussian beams each of which corresponding to a modulated Gaussian window function on the curved surface. The Gaussian beams are propagated according to the derived approximate expression which is based on the Rayleigh-Sommerfeld diffraction model. The assumptions related to the smoothness of the curved surface and mutual couplings that was stated in Sec. 3.2 are still valid in this chapter: We assume that the given curved surface is smooth enough such that the Gaussian window functions on it can be treated as written on planar patches and we restrict the set of curved surfaces to a feasible set for which it is possible to keep the mutual couplings at negligible levels. For the surfaces that satisfy this assumption, we show in Sec. 4.2 that the proposed method produces quite accurate 3D field solutions. This is an important development over the commonly used source model approaches [23, 25, 26, 27, 28, 29, 30], since they do not produce accurate results for the problem of finding the 3D field due to a field specified on a curved surface (or an object). The reason is that significant mutual couplings are seen in these models. Therefore, if the 3D field is calculated by using such source model approaches, then the subsequently reconstructed field on the same curved surface is found to be inconsistent with the original one.

The field model approach proposed in [37, 53] solves the mutual coupling problem and provides an exact method to calculate the diffraction field from a given curved surface. However, for large sizes of curved surfaces, the method introduced in [37] becomes impractical because of an intolerable increase in the computational complexity and the related memory requirement.

The local Gaussian signal decomposition method that we propose in this chapter provides more accurate solutions than the source models without being as computational highly demanding as the plane wave decomposition based method [53]. For the case of a planar input surface, the decomposition of a 2D signal as

a sum of Gaussian elementary signals was suggested by Gabor [62]. After that, several studies, that aim to find Gabor's expansion coefficients and expand an optical signal into a set of Gaussian beams, related to this decomposition were introduced by Bastiaans [63, 64]. In those works, the input surface is restricted to be planar and an explicit 3D Gaussian beam decomposition expression is not given. Here in this chapter, we study the more general case of curved input surfaces. Furthermore, as mentioned before, for the propagation of 3D Gaussian beams, we develop an approximate explicit expression which is based on the RS diffraction model. Such an expression enables us to use an optically valid model for the propagation of nonparaxial Gaussian beams. Note that most of the content of the work presented in this chapter is published in [35].

4.1 Gaussian Signal Decomposition

Similar to what we do in Chapter 3, we start with developing formulations for the calculation of the diffraction field due to a field specified on a planar input surface. This will relieve us from dealing with the structure of the surface. We will take into account the curvature of the input surface in Sec. 4.1.2 and extend the formulations developed for planar input surfaces.

4.1.1 Decomposition for the Planar Input Surface Case

We define a local Gaussian beam in 3D space as the beam that corresponds to a shifted and modulated version of a Gaussian window function on the given input surface. Hence, we start with writing the field $u_0(x, y) = u(x, y, 0) \in \mathcal{B}$ specified on a planar surface at $z = 0$ in the form of a continuous short time

Fourier decomposition as [58],

$$u_0(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(\xi, \eta, f_x, f_y) g(x - \xi, y - \eta) \exp\{j2\pi(f_x x + f_y y)\} d\xi d\eta df_x df_y \quad , \quad (4.1)$$

where $g(x, y)$ is a unit energy Gaussian window function. The corresponding analysis equation of Eq. 4.1 is given as

$$a(\xi, \eta, f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) g^*(x - \xi, y - \eta) \exp\{-j2\pi(f_x x + f_y y)\} dx dy \quad . \quad (4.2)$$

Note that the definitions of $u(x, y, z)$ and $u_0(x, y)$ that are given in Sec. 3.1.1 are also valid in this chapter. Therefore, here we do not redefine the signal spaces that we deal with. Eq. 4.1 can be viewed as a decomposition of the input signal $u_0(x, y)$ into a sum of modulated and shifted versions of the Gaussian window function $g(x, y)$. The discrete version of Eq. 4.1 is written as

$$u_0(x, y) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g(x - mX, y - nY) \exp\{j2\pi(kF_x x + lF_y y)\} \quad , \quad (4.3)$$

where X, Y are the shift steps in space and F_x, F_y are the shift steps in frequency. In the discrete case, the corresponding analysis window function is not same as the synthesis window function. One can find an analysis window function $w(x, y)$ corresponding to the Gaussian synthesis window function $g(x, y)$ such that the signal $u_0(x, y)$ is obtained exactly with the coefficients a_{mnkl} as [63, 77],

$$a_{mnkl} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) w^*(x - mX, y - nY) \exp\{-j2\pi(kF_x x + lF_y y)\} dx dy \quad . \quad (4.4)$$

Here note that the modulated and shifted versions of the synthesis and analysis window functions correspond to the primal and dual elementary functions, respectively, as explained in Sec. 3.

Each modulated and shifted version of a Gaussian-shaped elementary signal occupies the smallest possible area in the space-frequency domain. This is a

particularly desirable property for our aims in this chapter. In [62], the decomposition is restricted to the case of critical sampling ($XF_x = 1, YF_y = 1$) of the space-frequency domain. However, for the choice of Gaussian-shaped signal as the synthesis window function, it is shown in [78] that oversampling ($XF_x < 1, YF_y < 1$) should be preferred, for a numerically stable reconstruction.

The Gaussian signal decomposition on the input surface corresponds to the Gaussian beam decomposition in the 3D space. Let us define $g_h(\mathbf{x} - \mathbf{r}, f_x, f_y)$ as the 3D field expression of the Gaussian beam at the observation point $\mathbf{x} = [x, y, z]^T$ due to the modulated Gaussian window function, at $\mathbf{r} = [\xi, \eta, 0]^T$ on the $z = 0$ plane, with the modulation frequencies f_x and f_y along the x and y axes, respectively. The 3D field at $\mathbf{x}$ due to the field $u_0(x, y)$ specified on $z = 0$ plane is then written as an integration over Gaussian beams by using the decomposition given by Eq. 4.1,

$$u(\mathbf{x}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(\xi, \eta, f_x, f_y) g_h(\mathbf{x} - \mathbf{r}, f_x, f_y) d\xi d\eta df_x df_y \quad . \quad (4.5)$$

Now, let us find an explicit expression for $g_h(\mathbf{x} - \mathbf{r}, f_x, f_y)$. We know that the beams having higher propagation angles with respect to the direction of the surface normal correspond to the modulated Gaussian window functions with higher modulation frequencies on the input plane. Hence, the range of propagation directions of the beams with respect to the direction of the surface normal depends on the frequency content of the signal being analyzed. Because of this reason, as the bandwidth of the signal increases the propagating beams are no longer paraxial. Although the accurate RS model can be successfully used for such nonparaxial cases, it is difficult to obtain an explicit expression for Gaussian beams under the RS model (to our knowledge there is no explicit expression of the Gaussian beam obtained under the RS model, in the literature). Therefore, we propose an approximate formula which is still based on the RS model. For a local field $u_0(x, y)$ given on the $z = 0$ plane, we find the corresponding diffraction field over a planar surface, located at a distance z and parallel to the input plane,

as [35],

$$u_z(x, y) \approx h_z(x, y)U_0 \left(\frac{x}{\lambda\sqrt{x^2 + y^2 + z^2}}, \frac{y}{\lambda\sqrt{x^2 + y^2 + z^2}} \right) . \quad (4.6)$$

This expression produces quite accurate results even for nonparaxial cases as long as the observation distance z is sufficiently large and the given input field $u_0(x, y)$ has a sufficiently narrow support around $(x, y) = (0, 0)$. In Eq. 4.6,

$$h_z(x, y) = \frac{z}{j\lambda} \frac{\exp \left(j \frac{2\pi}{\lambda} \sqrt{x^2 + y^2 + z^2} \right)}{x^2 + y^2 + z^2} \quad (4.7)$$

is the 2D RS kernel; $\frac{x}{\lambda\sqrt{x^2 + y^2 + z^2}}$ and $\frac{y}{\lambda\sqrt{x^2 + y^2 + z^2}}$ are the instantaneous frequencies of the RS kernel at (x, y) along x and y directions, respectively, and $U_0(f_x, f_y)$ is the Fourier transform of the given input field $u_0(x, y)$. The expression given by Eq. 4.6 is applicable for a sufficiently large observation distance z , since in this case the RS kernel can be successfully approximated locally as a single frequency complex exponential. For example, the diffraction field at $z = 0.1mm$ due to a Gaussian window function on $z = 0$ plane having a parameter of $10^{-5}m$ (typical case in our experiments) is calculated by the approximate expression with a normalized error in the order of 10^{-2} . However, for $z > 0.1m$ the normalized error is found to be in the order of 10^{-8} . Note that we define the normalized error, in approximating a signal $f(x, y)$ as $\tilde{f}(x, y)$, as

$$1 - \frac{|\langle f(x, y), \tilde{f}(x, y) \rangle|^2}{\langle f(x, y), f(x, y) \rangle \langle \tilde{f}(x, y), \tilde{f}(x, y) \rangle} , \quad (4.8)$$

where $\langle f_1(x, y), f_2(x, y) \rangle$ is the inner product of $f_1(x, y)$ and $f_2(x, y)$ [35]. We will use the defined normalized error as the error measure in the simulations.

We can use Eq. 4.6 to calculate the diffraction field corresponding to a Gaussian-shaped elementary signal given on the $z = 0$ plane, since the Gaussian signals that we use are sufficiently local. In order to do that, we need the Fourier domain expressions of the modulated and shifted Gaussian window functions. The Fourier transform of a Gaussian window function $g(x, y) = c \exp \left(\frac{x^2 + y^2}{\sigma^2} \right)$ is given as [79], $\mathcal{F}\{g(x, y)\} = c\sigma^2\pi \exp \{ -\pi^2\sigma^2(f_x^2 + f_y^2) \}$, where $\mathcal{F}$ denotes

the Fourier transform from the (x, y) domain to (f_x, f_y) domain as defined by Eq. 1.13 and c is a constant making $g(x, y)$ a unit energy function. The Fourier transform of a shifted Gaussian window function is then written as $\mathcal{F}\{g(x-\xi, y-\eta)\} = c\sigma^2\pi \exp\{-\pi^2\sigma^2(f_x^2 + f_y^2)\} \exp\{-j2\pi(f_x\xi + f_y\eta)\}$. Finally, the Fourier transform of a shifted and modulated Gaussian window function is obtained as

$$\begin{aligned} \mathcal{F}\{g(x-\xi, y-\eta)e^{j2\pi\mu x}e^{j2\pi\nu y}\} &= c\sigma^2\pi e^{-\pi^2\sigma^2\{(f_x-\mu)^2+(f_y-\nu)^2\}} \\ &\quad e^{-j2\pi\{(f_x-\mu)\xi+(f_y-\nu)\eta\}} \quad . \end{aligned} \quad (4.9)$$

Hence, $\tilde{g}(\mathbf{x} - \mathbf{r}, f_x, f_y)$ is found by substituting Eq. 4.9 into Eq. 4.6 as [35],

$$\begin{aligned} \tilde{g}(\mathbf{x} - \mathbf{r}, f_x, f_y) &\approx h_z(x - \xi, y - \eta) \\ &\quad c\sigma^2\pi e^{-\pi^2\sigma^2\left\{\left(\frac{x-\xi}{\lambda\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}-f_x\right)^2+\left(\frac{y-\eta}{\lambda\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}-f_y\right)^2\right\}} \\ &\quad e^{-j2\pi\left\{\left(\frac{x-\xi}{\lambda\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}-f_x\right)\xi+\left(\frac{y-\eta}{\lambda\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}-f_y\right)\eta\right\}} \quad , \end{aligned} \quad (4.10)$$

The 3D field that corresponds to the input field $u_0(x, y)$ defined on $z = 0$ plane is then found by substituting $\tilde{g}(\mathbf{x} - \mathbf{r}, f_x, f_y)$ given by Eq. 4.10 into Eq. 4.5 as [35],

$$\begin{aligned} u(\mathbf{x}) &\approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(\xi, \eta, f_x, f_y) h_z(x - \xi, y - \eta) \\ &\quad c\sigma^2\pi e^{-\pi^2\sigma^2\left\{\left(\frac{x-\xi}{\lambda\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}-f_x\right)^2+\left(\frac{y-\eta}{\lambda\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}-f_y\right)^2\right\}} \\ &\quad e^{-j2\pi\left\{\left(\frac{x-\xi}{\lambda\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}-f_x\right)\xi+\left(\frac{y-\eta}{\lambda\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}-f_y\right)\eta\right\}} d\xi d\eta df_x df_y \quad . \end{aligned} \quad (4.11)$$

At this point we should note that, as mentioned before, the accuracy of the 3D Gaussian beam expressions used in Eq. 4.11 is better for a narrower Gaussian window function. Therefore, in order to calculate the 3D field with an acceptable accuracy by using Eq. 4.11, a sufficiently narrow Gaussian window function should be chosen.

The discrete version of Eq. 4.11 is obtained by sampling the space-frequency domain, at $\xi = mX, \eta = nY$ and $f_x = kF_x, f_y = lF_y$, as

$$u(\mathbf{x}) \approx \sum_m \sum_n \sum_k \sum_l a_{mnkl} h_z(x - mX, y - nY) \\ c\sigma^2\pi e^{-\pi^2\sigma^2\left\{\left(\frac{x-mX}{\lambda\sqrt{(x-mX)^2+(y-nY)^2+z^2}} - kF_x\right)^2 + \left(\frac{y-nY}{\lambda\sqrt{(x-mX)^2+(y-nY)^2+z^2}} - lF_y\right)^2\right\}} \\ e^{-j2\pi\left\{\left(\frac{x-mX}{\lambda\sqrt{(x-mX)^2+(y-nY)^2+z^2}} - kF_x\right)mX + \left(\frac{y-nY}{\lambda\sqrt{(x-mX)^2+(y-nY)^2+z^2}} - lF_y\right)nY\right\}}, \quad (4.12)$$

where the discrete analysis coefficients a_{mnkl} are found by using Eq. 4.4 [35].

4.1.2 Decomposition for the Curved Input Surface Case

The smoothness assumption of S , stated in Sec. 3.2, enables us to represent a field $u(\mathbf{r})$ given on S locally on the tangent planes of S . Therefore, in order to decompose $u(\mathbf{r})$ as a sum of modulated Gaussian window functions on the tangent planes of S , we first define the Gaussian window function at $\mathbf{r}$ on the tangent plane $T_{\mathbf{r}}$ as

$$g_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) = c \exp\left(-\frac{x_{\mathbf{r}}^2 + y_{\mathbf{r}}^2}{\sigma^2}\right). \quad (4.13)$$

The modulated Gaussian window function with modulation frequencies of f_{x_r} and f_{y_r} is then given as

$$g_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) \exp\{j2\pi(f_{x_r}x_{\mathbf{r}} + f_{y_r}y_{\mathbf{r}})\} = c \exp\left(-\frac{x_{\mathbf{r}}^2 + y_{\mathbf{r}}^2}{\sigma^2}\right) \exp\{j2\pi(f_{x_r}x_{\mathbf{r}} + f_{y_r}y_{\mathbf{r}})\}. \quad (4.14)$$

We find the coefficients of the modulated Gaussian window functions at $\mathbf{r}$ by the analysis equation, defined on the tangent plane $T_{\mathbf{r}}$, as [35],

$$a(\mathbf{r}, f_{x_r}, f_{y_r}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) g_{\mathbf{r}}^*(x_{\mathbf{r}}, y_{\mathbf{r}}) \exp\{-j2\pi(f_{x_r}x_{\mathbf{r}} + f_{y_r}y_{\mathbf{r}})\} dx_{\mathbf{r}} dy_{\mathbf{r}} \\ \approx \int \int_{V_{T_{\mathbf{r}}}} u_{\mathbf{r}}(x_{\mathbf{r}}, y_{\mathbf{r}}) g_{\mathbf{r}}^*(x_{\mathbf{r}}, y_{\mathbf{r}}) \exp\{-j2\pi(f_{x_r}x_{\mathbf{r}} + f_{y_r}y_{\mathbf{r}})\} dx_{\mathbf{r}} dy_{\mathbf{r}}, \quad (4.15)$$

where the definitions of the neighborhood V_{T_r} of $\mathbf{r}$ on T_r and the projected field $u_r(x_r, y_r)$ are same as the definitions given in Sec. 3.2.

In order to calculate the diffraction field in 3D space due to the field $u(\mathbf{r})$ on S , we need to propagate the Gaussian beams corresponding to the modulated Gaussian window functions on S . We gave the expression for the propagation of a Gaussian beam corresponding to a shifted and modulated Gaussian window function on a planar surface in Eq. 4.10. Now, using the definition given in Sec. 4.1.1, we define $\tilde{g}(\mathbf{x}, f_x, f_y)$ as the 3D field at the observation point $\mathbf{x} = [x, y, z]^T$ on a plane parallel to $z = 0$ plane, due to a modulated Gaussian window function at $(x, y) = (0, 0)$ on $z = 0$ plane, with the modulation frequencies f_x and f_y along the x and y axes, respectively. Remember that (x, y, z) is the global coordinate system as shown in Fig. 3.1. By using the expression derived in Sec. 4.1.1 and given by Eq. 4.10, $\tilde{g}(\mathbf{x}, f_x, f_y)$ is written as [35],

$$\tilde{g}(\mathbf{x}, f_x, f_y) \approx h_z(x, y) c \sigma^2 \pi \exp \left\{ -\pi^2 \sigma^2 \left[\left(\frac{x}{\lambda |\mathbf{x}|} - f_x \right)^2 + \left(\frac{y}{\lambda |\mathbf{x}|} - f_y \right)^2 \right] \right\} , \quad (4.16)$$

where $|\mathbf{x}| = \sqrt{x^2 + y^2 + z^2}$.

We use the expression given by Eq. 4.16, with proper coordinate transformations, for the propagation of the Gaussian beams induced by the curved surface S . Hence, for a given field $u(\mathbf{r})$ on S we find the diffraction field at $\mathbf{x}$ by an integration over all Gaussian beams as

$$\begin{aligned} u(\mathbf{x}) &\approx \iiint_S \int a(\mathbf{r}, f_{x_r}, f_{y_r}) \tilde{g}(\mathbf{x}_r, f_{x_r}, f_{y_r}) df_{x_r} df_{y_r} dS \\ &\approx \iiint_S \int a(\mathbf{r}, f_{x_r}, f_{y_r}) h_{z_r}(x_r, y_r) \\ &\quad c \sigma^2 \pi \exp \left\{ -\pi^2 \sigma^2 \left[\left(\frac{x_r}{\lambda |\mathbf{x}_r|} - f_{x_r} \right)^2 + \left(\frac{y_r}{\lambda |\mathbf{x}_r|} - f_{y_r} \right)^2 \right] \right\} df_{x_r} df_{y_r} dS , \end{aligned} \quad (4.17)$$

where $\mathbf{x}_r = \mathbf{R}_r \mathbf{x} - \mathbf{r}$, as defined by Eq. 3.20 [35].

The continuous signal decomposition given by Eq. 4.17 is extremely redundant. By choosing the discrete set of evaluation points on S and shift steps in frequency properly, we can still decompose the field on S into a discrete set of modulated Gaussian window functions on the tangent planes of S . That is, it is possible to write $u(\mathbf{x})$ as a sum of Gaussian beams that correspond to Gaussian window functions at discrete positions $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n\}$ on S and having discrete modulation frequencies as multiples of F_x and F_y , as [35],

$$\begin{aligned}
u(\mathbf{x}) &\approx \sum_{i=1}^n \sum_k \sum_l a_{\mathbf{r}_i kl} \tilde{g}(\mathbf{x}_{\mathbf{r}_i}, kF_x, lF_y) \\
&\approx \sum_{i=1}^n \sum_k \sum_l a_{\mathbf{r}_i kl} h_{z_{\mathbf{r}_i}}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) \\
&\quad c\sigma^2 \pi \exp \left\{ -\pi^2 \sigma^2 \left[\left(\frac{x_{\mathbf{r}_i}}{\lambda |\mathbf{x}_{\mathbf{r}_i}|} - kF_x \right)^2 + \left(\frac{y_{\mathbf{r}_i}}{\lambda |\mathbf{x}_{\mathbf{r}_i}|} - lF_y \right)^2 \right] \right\} \quad ,
\end{aligned} \tag{4.18}$$

where

$$\begin{aligned}
a_{\mathbf{r}_i kl} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{\mathbf{r}_i}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) w_{\mathbf{r}_i kl}^*(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) dx_{\mathbf{r}_i} dy_{\mathbf{r}_i} \\
&\approx \iint_{V_{T_{\mathbf{r}_i}}} u_{\mathbf{r}_i}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) w_{\mathbf{r}_i kl}^*(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}) dx_{\mathbf{r}_i} dy_{\mathbf{r}_i} \quad ,
\end{aligned} \tag{4.19}$$

and $\{w_{\mathbf{r}_i kl}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}); \mathbf{r}_i \in S, k, l \in \mathbb{Z}\}$ is the set of projected dual elementary functions as described in Sec. 3.3.2. If the discrete evaluation points are taken on a sufficiently dense regular grid, then, similar to the planar input surface case, a set of dual elementary functions can be found on S as explained in [80]. When it is not possible to place such a regular grid on the entire surface construction of a dual elementary function set may not be that simple. Here, however, let us assume that such a construction is available and the corresponding set of projected dual elementary functions is given as $\{w_{\mathbf{r}_i kl}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i}); \mathbf{r}_i \in S, k, l \in \mathbb{Z}\}$. Then, we can decompose $u(\mathbf{r})$ into a sum of modulated Gaussian window functions on the tangent planes of S and find their coefficients by using Eq. 4.19. Note that in the discrete case, $V_{T_{\mathbf{r}_i}}$ is used to represent the support of the analysis window function $w_{\mathbf{r}_i kl}(x_{\mathbf{r}_i}, y_{\mathbf{r}_i})$ on $T_{\mathbf{r}_i}$. For the applicability of Eq. 4.19, the analysis window

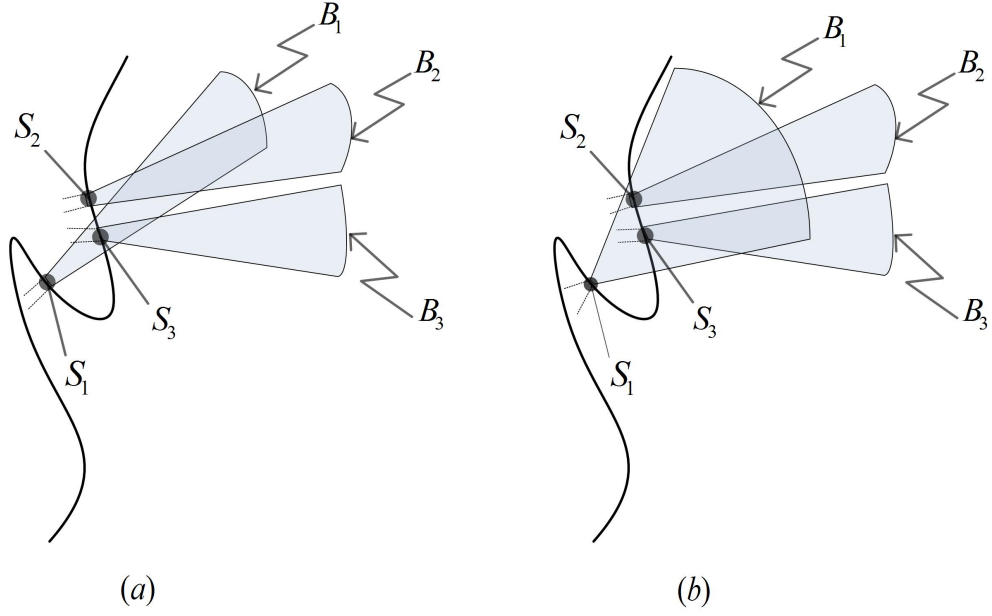


Figure 4.1: Mutual couplings between patches on a curved surface (2D cross sections of the surface and beams are shown for the sake of simplicity). (a) Mutual coupling between S_1 and S_3 patches, (b) mutual coupling between S_1 and S_2 patches (in addition to the mutual coupling between S_1 and S_3 patches). (©2012 JOSA A. Reprinted with permission. Published in [35])

function should be sufficiently narrow such that within its support the surface can be treated as locally planar. By a sufficiently dense sampling in space and frequency, it is possible to have an analysis window function which is as narrow as the synthesis window function [77].

The accuracy of Eq. 4.17 mainly depends on three factors [35]. The first factor is the adaption of an approximate 3D Gaussian beam expression (see Eq. 4.10). As mentioned in Sec. 4.1.1, the accuracy of the approximate 3D Gaussian beam expression $\tilde{g}(\mathbf{x}, f_x, f_y)$ is better for a narrower Gaussian window function. The second factor is the application of the developed Gaussian beam expression to curved surfaces even though it is developed for a planar surface. A narrower Gaussian window function is also desired to decrease the error resulting from the application of such an expression to a curved surface, since for a narrower Gaussian window function the locally planar assumption of the given curved surface becomes more reasonable. Finally, the last factor is the independent treatment

of mutually coupled patches on the curved surface. If a beam induced by a patch illuminates another disjoint patch on the curved surface, then independent treatment of such mutually coupled patches produce inconsistent 3D field solutions, since a beam induced by one of the mutually coupled patches alters the field value on the other patch [53]. Since a narrower Gaussian window function results in Gaussian beams having larger angular spreads, the likelihood that such a wider Gaussian beam illuminates another patch on the surface increases. Therefore, as opposed to first two factors given above, a narrower Gaussian window function produces worse results in terms of mutual couplings. Note that the tradeoff regarding the spatial and angular resolutions of a Gaussian beam is a consequence of the uncertainty principle that we discuss in Chapter 2.

In order to show the effect of the width of the Gaussian window function on the mutual couplings, two scenarios are illustrated in Fig. 4.1. S_1 and S_2 patches shown in Fig. 4.1(a) are not mutually coupled, since the beam induced by S_1 does not illuminate S_2 and vice versa. In Fig. 4.1(b) we decrease the width of the Gaussian window function on S_1 patch. Therefore, the angular spread of B_1 increases, and S_2 is now illuminated by also B_1 . Thus, S_1 and S_2 patches become mutually coupled. Note that as we continue to decrease the width of the Gaussian window functions on the curved surface, the angular spreads of the corresponding beams become larger and larger. As a result, the mutual couplings increase.

In addition to the locality of the Gaussian window functions, mutual couplings also depend on the structure of the input curved surface, as well. As an example, mutual coupling between S_1 and S_3 patches on a curved surface is shown in Fig. 4.1(a) and Fig. 4.1(b). Because of the structure of the curved surface, the mutual coupling between these two patches can not be reduced to negligible levels even if we change the width of the Gaussian window function on S_1 . In this study, we restrict the set of curved surfaces to a feasible set for which it is possible to keep

the mutual couplings at negligible levels, so that, the developed formulations are applicable.

Gaussian beams having large angular spread are not desirable because of associated computational reasons. Such beams illuminate a larger volume of space compared to narrower beams, and therefore, a higher number of them become effective (superpose) at a given observation point; this brings an extra computational burden in the computation of Eq. 4.17.

All these discussions lead us to choose a Gaussian window function having a sufficiently small width such that the accuracy of the derived formulations are satisfactory in field calculations, but nevertheless, the computational complexity is still in tolerable limits and the mutual couplings are negligible [35].

4.2 Simulation Results

The simulations are presented for the 2D (x, z) -space, for the sake of simplicity. Note that the z variable still represents the depth. We aim to find the 2D diffraction field due to a field specified on a curved line (instead of a curved surface in 3D space) by using the proposed Gaussian signal decomposition method.

In order to write the signal space as a sum of finite number of plane waves (whose propagation direction components along the y -axis is zero), we assume a periodicity along the x -axis. The 2D cross sections of such plane waves by the (x, z) -plane are used in (x, z) -space. The periodicity, together with the discretization of the functions, enable us to calculate the coefficients of the Gaussian beams by using DFT (see Eq. 4.21).

An example simulation setup is presented in Fig. 4.2. As shown in the figure, the curved line is periodic along the x -axis with period X_p . The 2D (x, z) field is also periodic along the x -axis with the same period. Therefore, the signal on

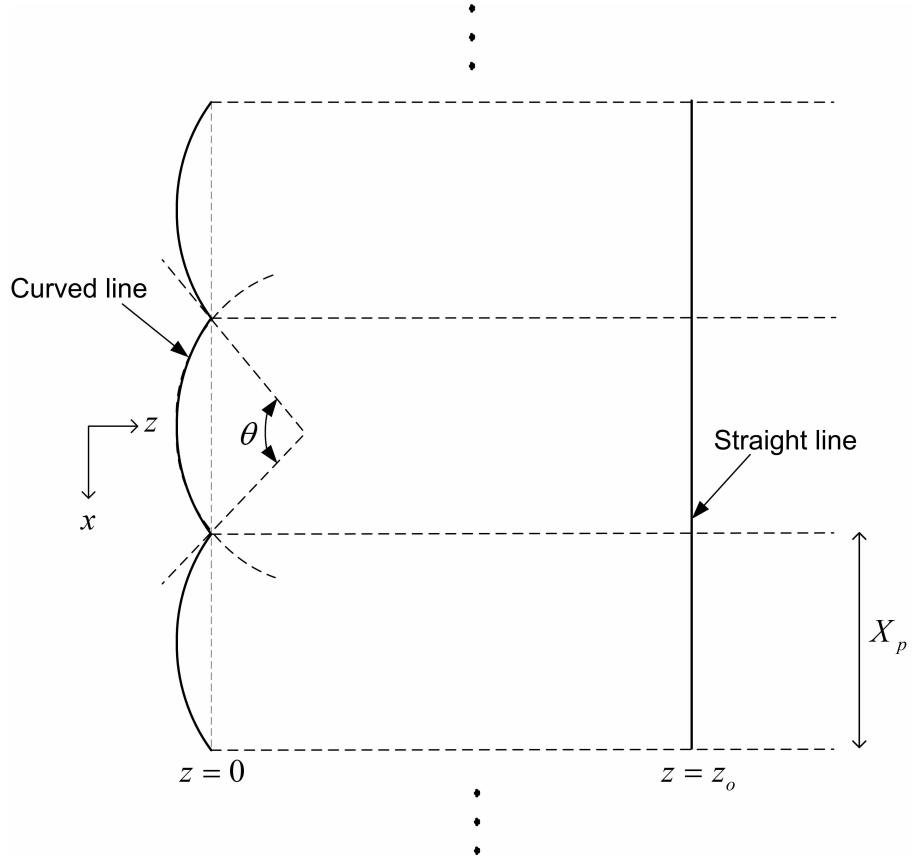


Figure 4.2: The 2D simulation setup which is periodic along the x -axis (the fields used in the simulations repeat themselves between each dashed horizontal lines shown in the figure). (©2012 JOSA A. Reprinted with permission. Published in [35])

the periodic curved line, $\tilde{u}(\mathbf{r}(l))$, that is obtained by intersecting such a periodic 2D field, is also periodic. Note that we represent the signals on the curved line, with respect to the 2D (x, z) global coordinate system, by using the arc-length parameterization $\mathbf{r}(l)$. Also note that we use the tilde sign to denote periodic signals on the curved line and we represent one period of such signals without the tilde sign.

The discrete and periodic input signal $\tilde{u}_d[n]$ is obtained by uniformly sampling the continuous signal $\tilde{u}(\mathbf{r}(l))$ with the sampling step L_s , i.e. $\tilde{u}_d[n] = \tilde{u}(\mathbf{r}(L_s n))$. N samples per period are obtained after discretization. The finite length Gaussian synthesis window function $g(\mathbf{r}(l))$ is sampled with the same sampling step and periodically extended over the periodic curved line to have the discrete and periodic synthesis window function $\tilde{g}_d[n]$. The total number of Gaussian beams to be included in the diffraction field calculation depends on the choice of discrete shift steps in the position and modulation frequency of the Gaussian synthesis window function. The shift steps in n and the normalized frequency domain k are denoted as M *samples* and $1/K$ *cycles/sample*, respectively. The corresponding total number of Gaussian beams used in the decomposition is therefore equal to $\frac{NK}{M}$. The ratio $\frac{K}{M}$ also affects the shape of the analysis window function corresponding to $\tilde{g}_d[n]$. For the oversampled case, $K/M > 1$, the expansion coefficients are dependent, since the set of modulated and shifted Gaussian windows is overcomplete. Therefore, the analysis window function corresponding to the synthesis window function is not unique. In the simulations we choose M and K such that $K/M = 2$, which corresponds to oversampling by a factor of two. For this case, we find the analysis window function $\tilde{w}_d[n]$ corresponding to the specified Gaussian synthesis window function by using the relation which gives the minimum L_2 -norm solution, developed in [77]. The parameter of the Gaussian synthesis window function, σ , is chosen as $2.20\mu m$. One periods of the resulting discrete synthesis window function, $g_d[n] = \exp\left(-\frac{L_s^2}{\sigma^2}n^2\right)$ for $L_s = 0.26\mu m$, and

the corresponding discrete analysis window function, $w_d[n]$, are shown in Fig. 4.3 and Fig. 4.4, respectively, for $n = -64, -63, \dots, 63$.

Note that we presented the formulations in Sec. 4.1.2 by using the tangent planes of the curved surface, for the 3D case. However, in the 2D simulations we do the calculations on the curved line with the signals defined above (instead of the tangent lines of the curved line), for the sake of clarity. The curved lines, that we work with, are almost locally linear within the supports of the analysis and synthesis window functions. Therefore such an approach is appropriate to simulate the developed formulations in 2D (x, z) -space. Regardless of a dimension reduction, we omit the constant multipliers and simply assume them as unity during the computations. For a given periodic input signal $\tilde{u}_d[n]$, the coefficients of the modulated Gaussian window functions on the curved line are found by the discrete analysis equation as

$$\begin{aligned}
 a_{mk} &= \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} \tilde{u}_d[n] \tilde{w}_d^*[n - mM] \exp\left(-j2\pi \frac{k}{K}\right) \quad , \\
 m &\in \left\{ \frac{-N}{2M}, \frac{-N}{2M} + 1, \dots, \frac{N}{2M} - 1 \right\} \quad , \\
 k &\in \left\{ \frac{-K}{2}, \frac{-K}{2} + 1, \dots, \frac{K}{2} - 1 \right\} \quad , \tag{4.20}
 \end{aligned}$$

where N, M and K are positive integer valued (N and K are an even integer) simulation parameters that are defined above. Eq. 4.20 can be rewritten as

$$\begin{aligned}
 a_{mk} &= \sum_{n_1=-\frac{K}{2}}^{\frac{K}{2}-1} \sum_{n_2=-\frac{N}{2K}}^{\frac{N}{2K}-1} \tilde{u}_d[n_1 + n_2K] \tilde{w}_d^*[n_1 + n_2K - mM] \exp\left(-j2\pi \frac{k}{K} n_1\right) \quad , \\
 m &\in \left\{ \frac{-N}{2M}, \frac{-N}{2M} + 1, \dots, \frac{N}{2M} - 1 \right\} \quad , \\
 k &\in \left\{ \frac{-K}{2}, \frac{-K}{2} + 1, \dots, \frac{K}{2} - 1 \right\} \quad . \tag{4.21}
 \end{aligned}$$

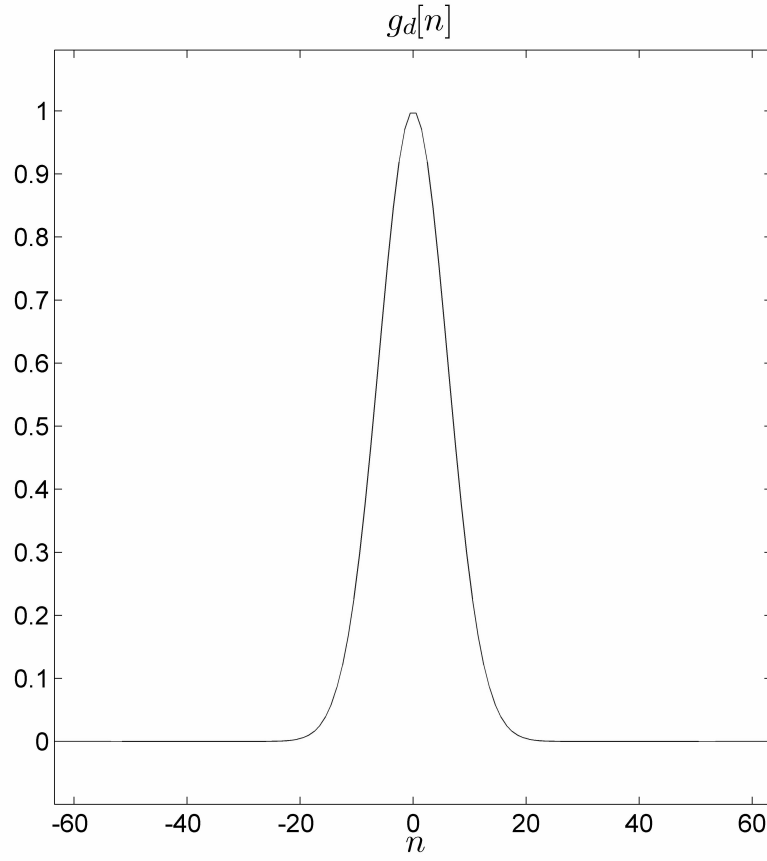


Figure 4.3: The discrete Gaussian synthesis window function $g_d[n] = \exp\left(-\frac{L_s^2}{\sigma^2}n^2\right)$ for $\sigma = 2.20\mu m$ and $L_s = 0.26\mu m$. (©2012 JOSA A. Reprinted with permission. Published in [35])

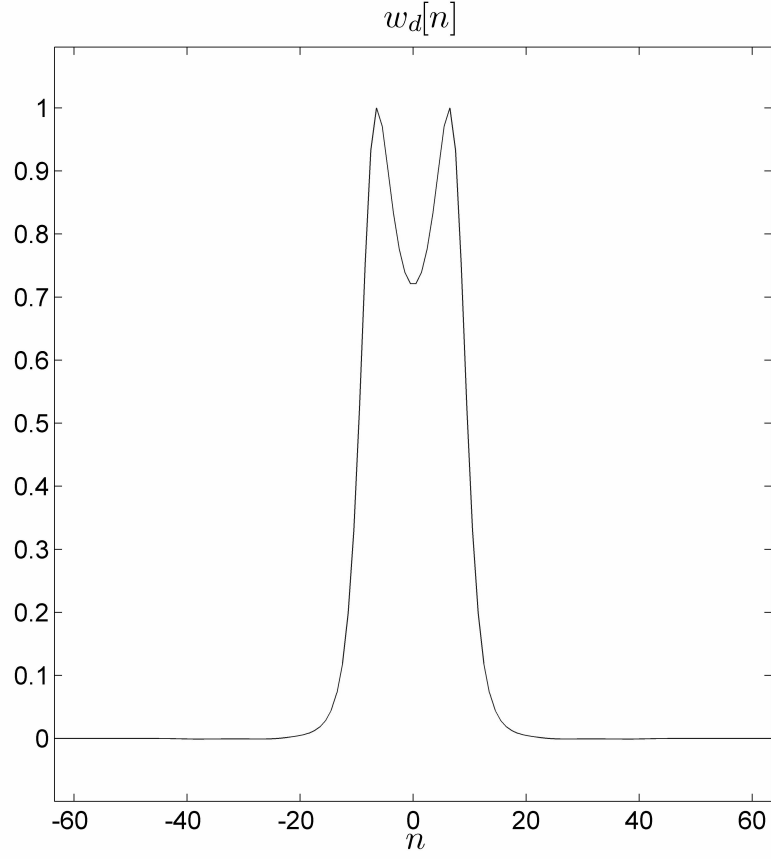


Figure 4.4: The discrete analysis window function corresponding to the Gaussian synthesis window function $g_d[n] = \exp\left(-\frac{L_s^2}{\sigma^2}n^2\right)$, for $\sigma = 2.20\mu m$ and $L_s = 0.26\mu m$, for the case that the shift parameters in space and frequency are $M = 16$ and $K = 32$, respectively. (©2012 JOSA A. Reprinted with permission. Published in [35])

For a fixed m we can rewrite Eq. 4.21 as

$$a_{mk} = \sum_{n_1=0}^{K-1} \tilde{u}_{w_d} \left[n_1 - \frac{K}{2} \right] \exp \left(-j2\pi \frac{k}{K} n_1 \right) ,$$

$$k \in \left\{ \frac{-K}{2}, \frac{-K}{2} + 1, \dots, \frac{K}{2} - 1 \right\} , \quad (4.22)$$

where

$$\tilde{u}_{w_d}[n_1] = \sum_{n_2=\frac{-N}{2K}}^{\frac{N}{2K}-1} \tilde{u}_d[n_1 + n_2 K] \tilde{w}_d^*[n_1 + n_2 K - mM] \exp(j\pi K) .$$

Thus, for each fixed m , by computing a K -point DFT of the signal $\tilde{u}_{w_d} \left[n_1 - \frac{K}{2} \right]$ (which is periodic with period K), we firstly obtain the analysis coefficients for $k \in \{0, 1, \dots, K-1\}$. Then, the coefficients for $k \in \left\{ \frac{-K}{2}, \frac{-K}{2} + 1, \dots, \frac{K}{2} - 1 \right\}$ are simply found by using the periodicity of a_{mk} , i.e. $a_{mk} = a_{m(k+K)}$. Please note that even though some of the indicated (m, k) pairs correspond to Gaussian beams having propagation directions with negative components along the z -axis, the coefficients of such beams are zero. The reason is that we restrict the 2D field to a superposition of propagating waves whose propagation direction components along the z -axis are positive.

As a consequence of the periodicity together with the discretization of the signals on the periodic curved line, for each (m, k) pair, Eq. 4.20 gives the coefficient of the corresponding parallel Gaussian beams induced by the curved line (see Fig. 4.5). Since the Gaussian beams are local, only a few of these parallel duplicates are effective at a given observation point. We label the windows that are taken into account, for the (m, k) pair, as the set of integers $\{b_{mk}, b_{mk} + 1, \dots, b_{mk} + c\}$. Here, the window label b_{mk} refers to the region of the 2D space with $b_{mk}X_p - \frac{X_p}{2} \leq x < b_{mk}X_p + \frac{X_p}{2}$, taking the center window $\left(\frac{-X_p}{2} \leq x < \frac{X_p}{2} \right)$ as a reference. For example, while calculating the field at (x_o, z_o) due to the parallel beams shown in Fig. 4.5, only the windows with labels 1 and 2 (that induce B_1 and B_2 beams, respectively) are taken into account.

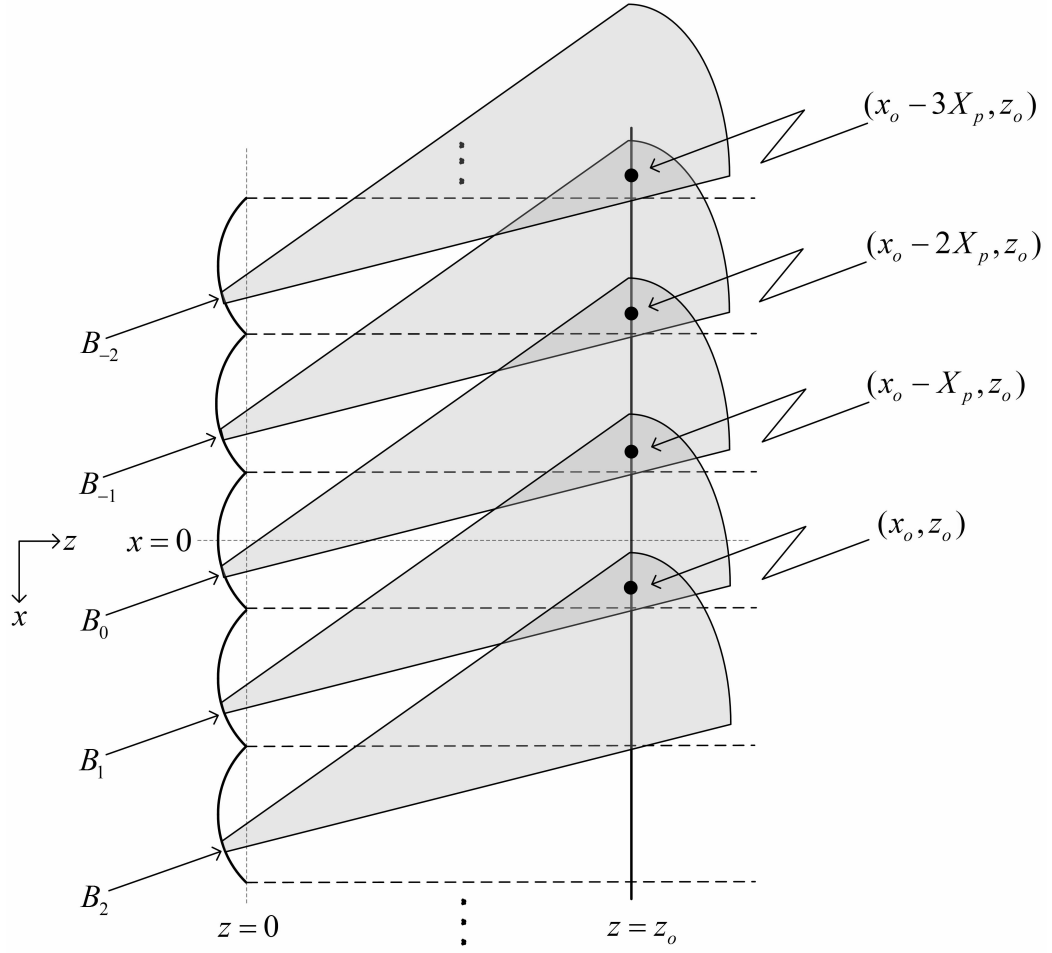


Figure 4.5: Periodic Gaussian beams induced by the periodic pattern over the periodic curved line. (©2012 JOSA A. Reprinted with permission. Published in [35])

Under the above definitions, we find the field at (x, z) due to the periodic input curved line by using the following equation as

$$u(x, z) \approx \sum_{m=\frac{-N}{2M}-1}^{\frac{N}{2M}-1} \sum_{k=\frac{-K}{2}-1}^{\frac{K}{2}-1} \sum_{b \in S_w} a_{mk} \tilde{g} \left(\mathbf{R}_{\mathbf{r}(mML_s)} [x - bX_p, z]^T - \mathbf{r}(mML_s), \frac{k}{KL_s} \right) \quad ,$$

$$S_w = \{b_{mk}, b_{mk} + 1, \dots, b_{mk} + c\} \quad ,$$
(4.23)

where $\mathbf{R}_{\mathbf{r}(mML_s)}$ is the 2×2 rotation matrix that relates the 2D (x, z) global coordinate system and the local coordinate system (one axis is along the tangent line and the other axis is along the normal line of the curved line) at $\mathbf{r}(mML_s)$.

A circular arc with a measure of $\theta = 30$ degrees is used as one period of the input curved line (see Fig 4.2). The period X_p along the x -axis is chosen as $0.26mm$. We uniformly sample the signal on the curved line with a sampling step $L_s = 0.51\lambda$ resulting in $N = 1024$ samples in one period of the signal. Note that λ , which is taken as $500nm$, is the wavelength of the monochromatic light. We take $M = 16$ and $K = 32$ resulting in 2048 different parallel Gaussian beams. Thus, the oversampling factor is two.

As an illustration of the decomposition procedure on the curved line, three discrete sinc functions, shifted to different positions, are written on the circular arcs. One period, $u_d[n]$, of this signal is shown in Fig. 4.6. One period of the diffraction field due to the input signal $\tilde{u}_d[n]$ is given in Fig. 4.7. This field is found by superposing various Gaussian beam components that correspond to modulated Gaussian window functions on the circular segments with different m and k parameters. An example Gaussian beam component, that corresponds to the modulated Gaussian window function with parameters $m = 16$ and $k = 1$, is presented in Fig. 4.8. Note that Fig. 4.7 and Fig. 4.8 are obtained by sampling the $u(x, z)$ field with a sampling step $X_s = 0.5\lambda$ along both axes. Also note that the beams shown in Fig. 4.7 and Fig. 4.8 are propagating from $z = -\infty$ to $z = +\infty$ and their beam waists are taken on the curved line.

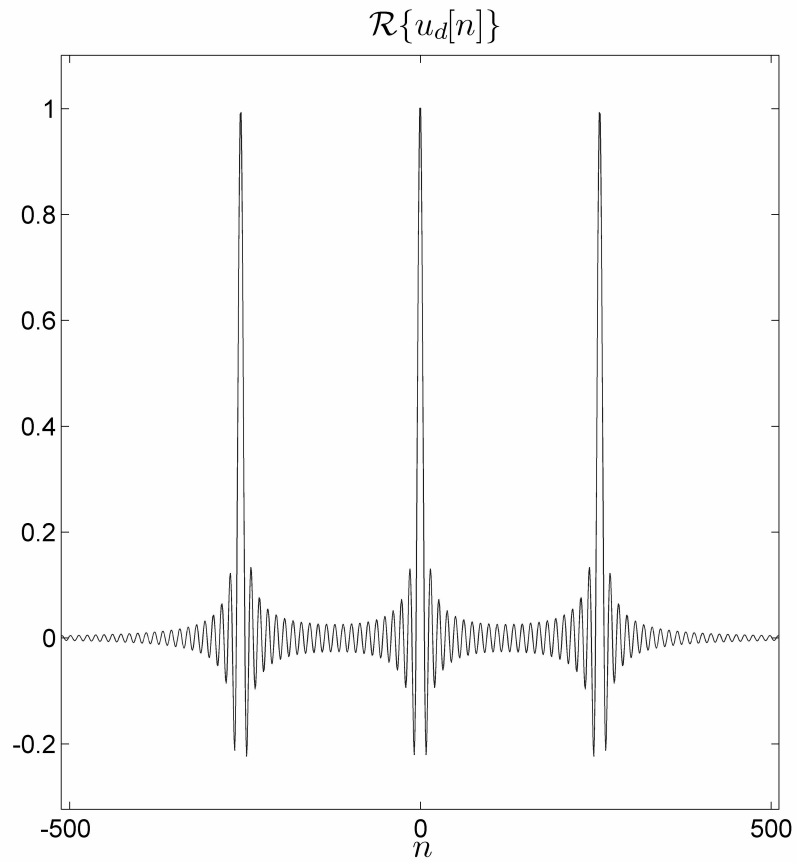


Figure 4.6: The real part of the discrete input signal, $u_d[n]$, on the circular arc with a measure of 30 degrees (imaginary part is zero). (©2012 JOSA A. Reprinted with permission. Published in [35])

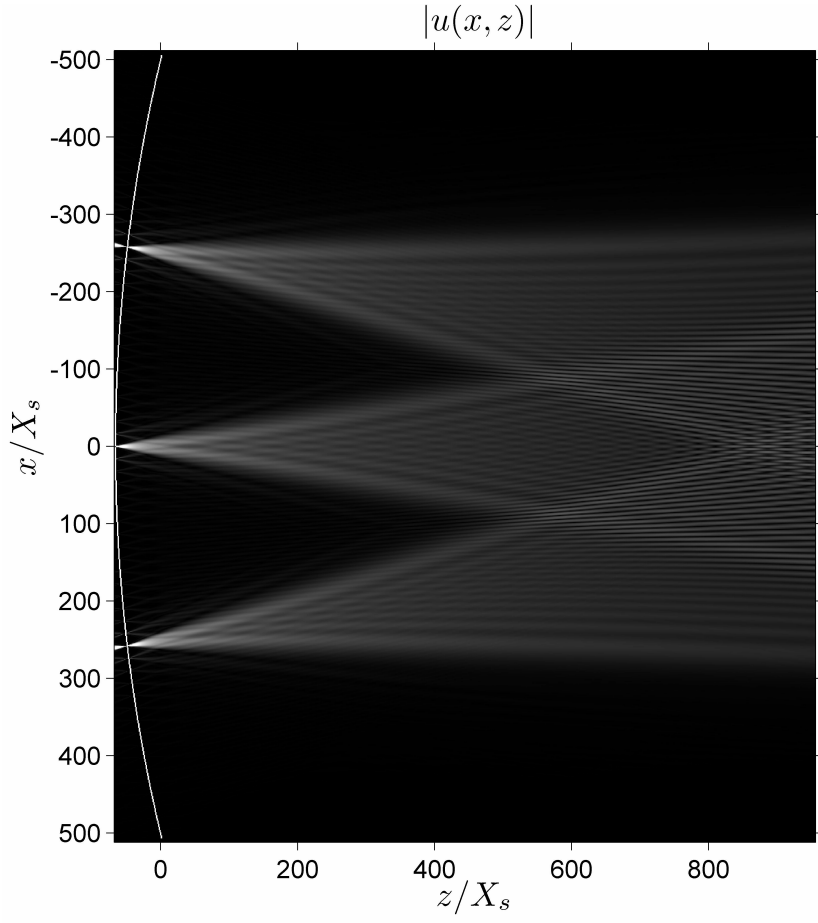


Figure 4.7: The magnitude of the 2D diffraction field due to the input signal $\tilde{u}_d[n]$, one period of which is shown in Fig. 4.6, on the circular arc with a measure of 30 degrees. (©2012 JOSA A. Reprinted with permission. Published in [35])

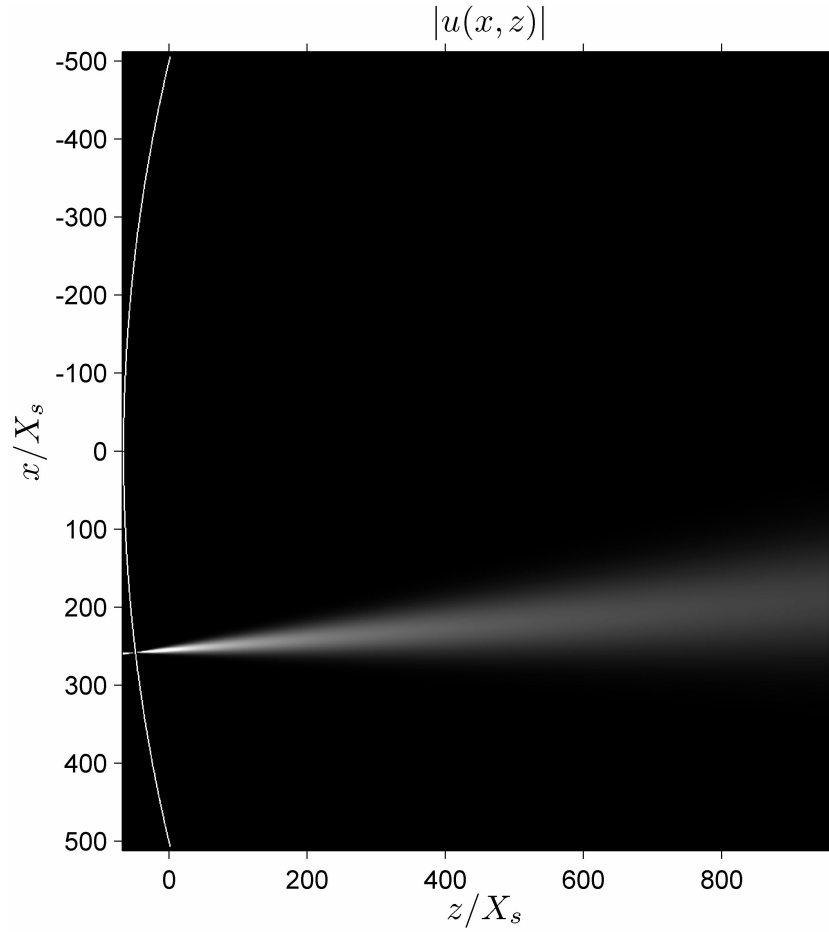


Figure 4.8: The magnitude of the 2D diffraction field due to a single shifted and modulated Gaussian window function, with parameters $m = 16$ and $k = 1$, on the circular arc with a measure of 30 degrees. (©2012 JOSA A. Reprinted with permission. Published in [35])

It is clearly seen in Fig. 4.8 that the mutual couplings are kept at negligible levels. If point sources were used on the curved line instead of Gaussian window functions, the angular spreads of the resulting 2D (x, z) fields would be much larger than the angular spreads of Gaussian beams and thus the mutual couplings would be unavoidable.

For a more general case, we start with specifying a 2D (x, z) field by randomly choosing the coefficients of the finite number of plane waves whose frequency components along the x -axis occupy the $\left[\frac{-0.70}{\lambda}, \frac{0.70}{\lambda}\right]$ *cycles/m* frequency band. Then we intersect this test field by the curved line. Since the developed formulations are not applicable for the non-smooth regions of the curved line, we window the resulting field so that the field values around the edges of the curved line (non-smooth regions) are set to zero (see Fig. 4.9). We firstly write the field on the curved line as a sum of modulated and shifted Gaussian window functions and then find the coefficients of each Gaussian elementary signal according to Eq. 4.21. Afterwards, we calculate the samples of the diffraction field on the observation line at $z = 0.1m$ via Eq. 4.23. Note that this distance is sufficiently large, and therefore the developed Gaussian beam expression, given by Eq. 4.10, yields accurate field calculations. In order to test the proposed approach, we firstly use the plane wave decomposition method, explained in Sec. 1.3 and 3.1, to find the 2D continuous field due to the discrete and periodic signal computed on the observation line. The periodic continuous field due to the discrete and periodic signal $u(nX_s, z_o)$, observed at $z = z_o$, is computed by the 2D PWD as

$$u(x, z) = \frac{1}{N} \sum_{k=\frac{-N}{2}}^{\frac{N}{2}-1} A_k \exp \left\{ j2\pi \left(\frac{k}{X_s N} x + \sqrt{\frac{1}{\lambda^2} - \frac{k^2}{X_s^2 N^2}} (z - z_o) \right) \right\} \quad , \quad (4.24)$$

where

$$\begin{aligned} A_k &= \sum_{n=\frac{-N}{2}}^{\frac{N}{2}-1} u(nX_s, z_o) \exp \left(-j2\pi \frac{kn}{N} \right) \quad , \\ k &\in \left\{ \frac{-N}{2}, \frac{-N}{2} + 1, \dots, \frac{N}{2} - 1 \right\} \quad , \end{aligned} \quad (4.25)$$

where, as mentioned before, X_s is the sampling step along the x -axis. Note that because of the discretization and the periodicity along the x -axis, we find only the coefficients of 1024 (total number of samples in one period of $u(x, z_o)$) plane waves by using 1024-point DFT of the sampled field at $z = z_o$. In order to find the reconstruction error, we resample the continuous field obtained by Eq. 4.24 at the discrete points on the curved line and compare the resulting discrete reconstructed signal with the original discrete test signal written on the curved line. One periods of the real parts of the original and reconstructed signals are shown in Fig. 4.9 (the imaginary parts have similar plots). We are able to reconstruct the field on the periodic circular segments with a normalized error (see Eq. 4.8) in the order of 10^{-4} . This shows that the developed formulations work well in computing the diffraction field in (x, z) -space for a field specified on the given periodic circular arcs. Similar results are obtained for many other curved lines that satisfy the smoothness assumption and for which the mutual couplings between the disjoint patches on it are negligible for a specified Gaussian window function. Here, once more, it is worth noting that for the circular arcs that we use, the mutual couplings between the disjoint patches inducing the Gaussian beams are at negligible levels.

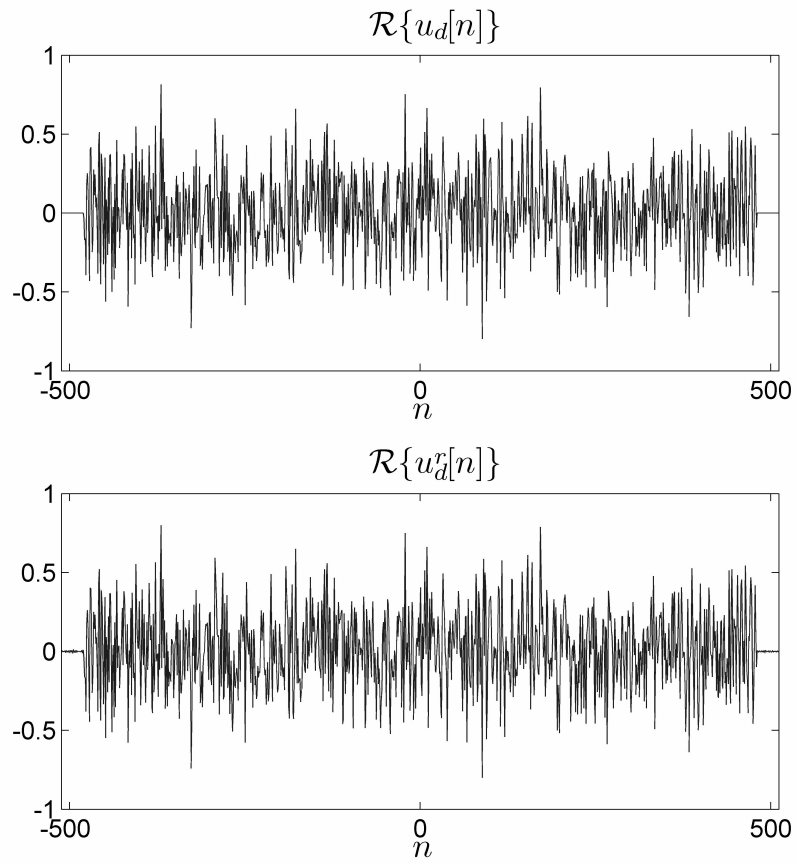


Figure 4.9: The real parts of the original, $u_d[n]$, and reconstructed, $u_d^r[n]$, signals on the circular arc with a measure of 30 degrees. (©2012 JOSA A. Reprinted with permission. Published in [35])

Chapter 5

Calculation of the 3D Scalar Diffraction Field from the Field Given on a Curved Surface via 3D Gaussian Beam Decomposition

In this chapter, we present another local signal decomposition method for the calculation of the 3D scalar diffraction field due to a 2D field specified on a curved surface. As pointed out before, the commonly used source model approaches [23, 25, 26, 27, 28, 29, 30] do not produce accurate results for this problem, since they ignore the mutual couplings between the field samples on the given curved surface. The local Gaussian signal decomposition that we introduced in Chapter 4 (and its general framework given in Chapter 3), reduces the mutual coupling effects with respect to such source models considerably. However, there are some underlying constraints and assumptions regarding its application for

the mentioned problem. For example, for some surfaces this method still produces inconsistent 3D field solutions. The reason is that the disjoint patches which represent the supports of the modulated Gaussian window functions on the surface are treated independently. Thus, possible mutual couplings between such patches are ignored. As a result of that, the method proposed in Chapter 4 is applicable to the surfaces such that the mutual couplings between the disjoint patches on the surface are negligible. In addition to that, a smoothness constraint is imposed on the curved surface for the applicability of related formulations. At this point it is worth noting, however, that the approach presented in Chapter 3 and Chapter 4 lead us to understand the problematic issues related to the diffraction field calculation from 2D fields given on curved surfaces and thus develop the second local signal decomposition method that we introduce in this chapter.

As mentioned in Chapter 4, the mutual coupling problem is solved in [37, 53] by using a field model approach, based on the plane wave decomposition, that results in an exact method to calculate the diffraction field from a field given on a curved surface. However, for large sizes of curved surfaces, this method becomes impractical because of an intolerable increase in the computational complexity and the related memory requirement.

Here in this chapter, we reformulate the local Gaussian beam decomposition method introduced in Chapter 4, in order to achieve the exact field solutions. The 3D Gaussian beam decomposition method that we propose in this chapter does not impose any restriction on the curved surface and still produces consistent 3D field solutions. We write the 3D field as a sum of 3D beams whose waist positions are determined by the given curved surface. The directions of the beams are fixed independent of the surface. We find the coefficients of the beams by constructing the linear system of equations where the system matrix is formed by calculating the field patterns that the Gaussian beams produce on the curved surface. In

this sense, the formulation of the proposed method is similar to the field model approach introduced in [53]. However, the locality of the signals used in our proposed decomposition results in important advantages over the approaches that use global signal decomposition methods (such as the one presented in [53]) in terms of computational complexity and memory allocation. In Sec. 5.1, we firstly give the Gaussian beam decomposition for planar input surfaces. Then, we introduce the proposed Gaussian beam decomposition for curved input surfaces in Sec. 5.2. We test our approach and present the simulation results in Sec. 5.3. Note that most of the content of the work presented in this chapter is published in [81].

5.1 Diffraction Field Calculation from the Fields Given on Planar Input Surfaces

In Sec. 4.1.1, both the continuous and discrete versions of the Gaussian signal decomposition were presented for the case of a planar input surface. In this chapter, we deal with the discrete version which is given by Eq. 4.3. Therefore, for an input field $u_0(x, y) = u(x, y, 0) \in \mathcal{B}$ specified on a planar surface at $z = 0$, here let us rewrite the discrete version of the Gaussian signal decomposition as

$$u_0(x, y) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g(x - mX, y - nY) \exp\{j2\pi(kF_x x + lF_y y)\} \quad , \quad (5.1)$$

where X, Y are the shift steps in space and F_x, F_y are the shift steps in frequency and

$$a_{mnkl} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) w^*(x - mX, y - nY) \exp\{-j2\pi(kF_x x + lF_y y)\} dx dy \quad . \quad (5.2)$$

The 3D field at an observation point $\mathbf{x} = [x, y, z]^T$ due to the field $u_0(x, y)$ specified on $z = 0$ plane can be found by summing up the contributions of the

Gaussian beams that correspond to the modulated Gaussian window functions on the input plane. Thus,

$$u(\mathbf{x}) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{x} - \mathbf{p}_{mn}) \quad . \quad (5.3)$$

Here, $g_{mnkl}(\mathbf{x} - \mathbf{p}_{mn})$ is defined as the 3D field expression of the Gaussian beam at the observation point $\mathbf{x} = [x, y, z]^T$. Note that in Chapter 3 and Chapter 4, we reserved g for the elementary functions defined on the input surface and used $\tilde{g}$ for the corresponding 3D beams. However, in this section we prefer not to use a tilde sign for 3D Gaussian beams, since we do not directly deal with the local elementary functions (more specifically Gaussian window function) defined on the input surface. The Gaussian beam defined as $g_{mnkl}(\mathbf{x} - \mathbf{p}_{mn})$ corresponds to the modulated Gaussian window function which is placed at $\mathbf{p}_{mn} = [mX, nY, 0]^T$ on the $z = 0$ plane and has the modulation frequencies kF_x and lF_y along the x and y axes, respectively. Note that for the case of uniform sampling of the space-frequency domain (as described above), there exists an analysis window function $w(x, y)$ corresponding to the Gaussian window function $g(x, y)$ that gives the expansion coefficients by the application of Eq. 5.2 [63, 77]. Therefore, one does not need not to solve a linear system of equations to find the coefficients of the Gaussian window functions or Gaussian beams.

In this chapter, we also use the plane wave decomposition to compare and show the efficiency of the proposed approach. Therefore, let us remember the plane wave decomposition which was presented in 3.1. The plane wave decomposition for a monochromatic scalar 3D field $u(\mathbf{x}) \in \mathcal{H}_p : \mathbb{R}^3 \rightarrow \mathbb{C}$ is given as [8],

$$u(\mathbf{x}) = \iint_B A(f_x, f_y) \exp(j2\pi \mathbf{f}_x^T \mathbf{x}) df_x df_y \quad , \quad (5.4)$$

where $\mathbf{x} = [x, y, z]^T \in \mathbb{R}^3$, $\mathbf{f}_x = [f_x, f_y, f_z]^T \in \mathbb{R}^3$; $f_x^2 + f_y^2 + f_z^2 = \frac{1}{\lambda^2}$, and λ is the wavelength of the monochromatic light. Here, remember that we exclude the evanescent waves and take into account only the propagating plane waves whose frequency components along the x and y axes satisfy $f_x^2 + f_y^2 \leq \frac{1}{\lambda^2}$. This

2D circular spatial frequency band occupied by the propagating plane waves is denoted as B . In this chapter, we also keep to adopt the assumption that the z component of the propagation direction is always positive. The coefficient of each propagating plane wave used in Eq. 5.3 can be found using the field $u_0(x, y)$ given on the $z = 0$ plane as

$$A(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) \exp\{-j2\pi(f_x x + f_y y)\} dx dy \quad . \quad (5.5)$$

5.2 Diffraction Field Calculation from the Fields Given on Curved Input Surfaces

Let us consider a curved surface $S \subset \mathbb{R}^3$ on which we represent the points by the vector $\mathbf{r}$. In Chapter 3 and in Chapter 4, we introduced our first approach for the problem of diffraction field calculation due to a field $u(\mathbf{r})$ given on the curved surface S . These methods are actually based on the linear and local signal decomposition of the 2D field given on the curved input surface. In other words, the 3D local beam decompositions are achieved as a consequence of the linear and local signal decompositions of the 2D fields given on curved input surfaces. In this chapter, we approach the same problem from a different point of view. We decompose the 3D scalar wave field into a sum of 3D Gaussian beams without dealing with the decomposition of the 2D field given on the curved input surface. This relieves us from the smoothness constraint and the mutual coupling problem which are actually consequences of the 2D signal decomposition framework presented in Chapter 3 and Chapter 4.

Now, consider the 3D plane wave decomposition and Gaussian beam decomposition given by Eq. 5.4 and Eq. 5.3, respectively. We can actually find the diffraction field due to a field $u(\mathbf{r})$ given on the curved surface S by using these decompositions. In order to that, we need to construct the system of linear

equations [81],

$$u(\mathbf{r}) = \iint_B A(f_x, f_y) \exp(j2\pi \mathbf{f}_x^T \mathbf{r}) df_x df_y, \quad (5.6)$$

$$u(\mathbf{r}) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{r} - \mathbf{p}_{mn}) \quad (5.7)$$

that correspond to Eq. 5.4 and Eq. 5.3, respectively. Given the field $u(\mathbf{r})$ on S , the coefficients of the expansion elements can be found by solving these equations. Note that the decomposition given by Eq. 5.6 is, indeed, used in the work of Esmer *et al.* [36, 37] for the problem of diffraction field calculation due to a 2D field given on an object. Since the 3D field $u(\mathbf{x})$ is assumed to be formed by a superposition of propagating plane waves which occupy the frequency band B , the coefficients of the Gaussian window functions whose frequency supports are outside this band are effectively zero. Thus, such Gaussian window functions need not be taken into account in Eq. 5.7. Once the coefficients of the expansion functions given by Eq. 5.6 and Eq. 5.7 are found, the 3D field at an arbitrary point $\mathbf{x}$ can be calculated by using Eq. 5.4 or Eq. 5.3, respectively.

The sparsity of the system matrix related to Eq. 5.7 is one of the main factors affecting the computational complexity in the solution of Eq. 5.7. The main benefit of the sparse matrices is that the computational complexity related to the linear system of equations solver is reduced if the solver can take advantage of the percentage and distribution of the zero elements of the system matrix [82, 83]. In addition to this, the memory requirement, related to the storage of the system matrix, during the solution of the linear system of equations can be reduced [82, 83, 84]. In Eq. 5.3 and Eq. 5.7, the waist positions of the Gaussian beams are all chosen on a reference plane. However, since the Gaussian beams spread (their widths increase) as they propagate beyond the waist position, to use the decomposition given by Eq. 5.3 and the corresponding linear system of equations given by Eq. 5.7 is not the best choice for curved input surfaces.

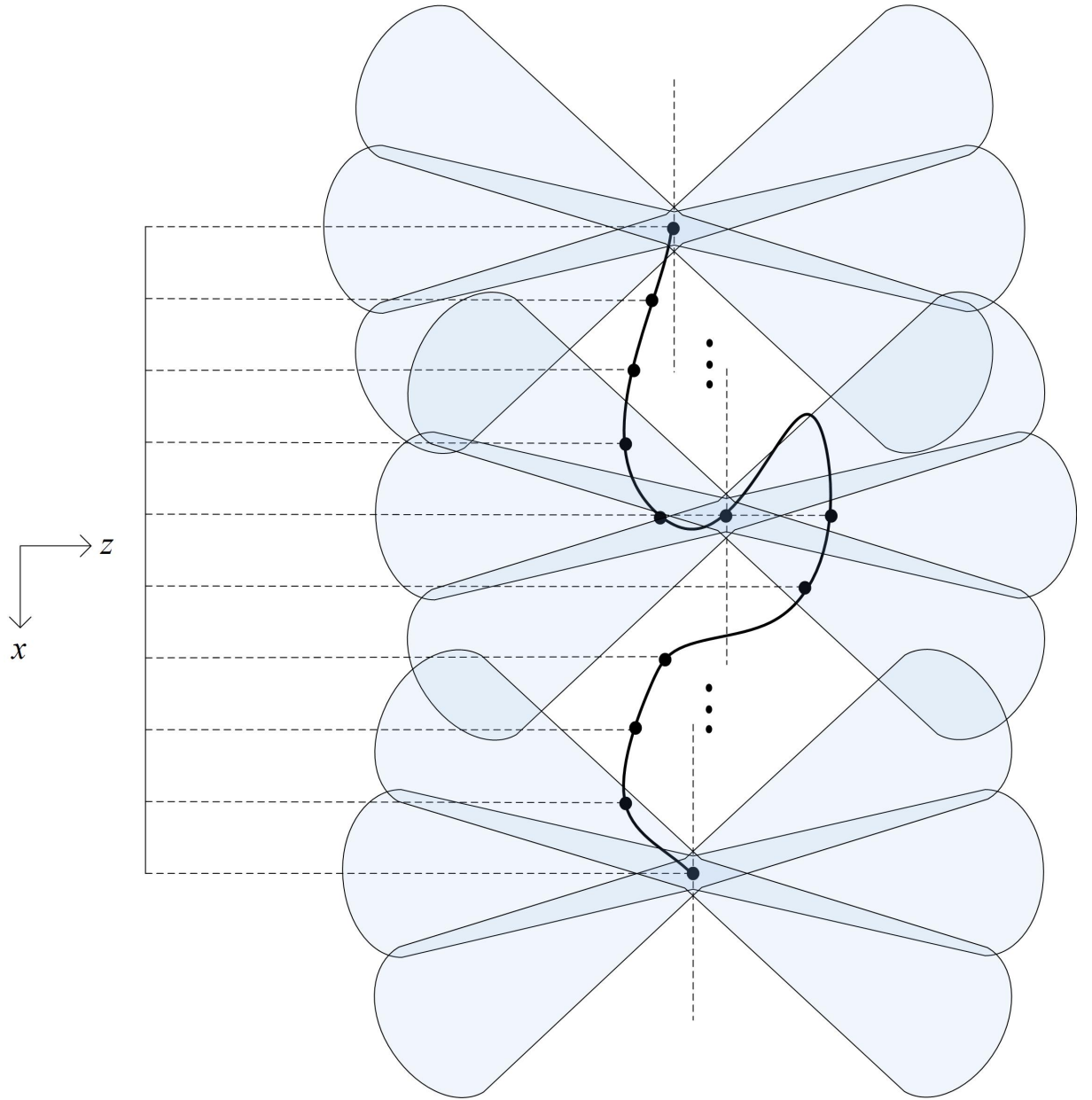


Figure 5.1: The setup of the proposed Gaussian beam decomposition method (a 2D pattern is shown for the sake of simplicity). (©2013 JOSA A. Reprinted with permission. Published in [81])

In this chapter, as a more reasonable approach, we choose the waist positions of the Gaussian beams on the curved surface (see Fig. 5.1). By doing so, the patterns resulting from the intersection of Gaussian beams by the given curved surface become more local than the case where Eq. 5.3 is used. Hence, noting that the patterns of the Gaussian beams on the curved surface actually represent the columns of the system matrix, more sparse system matrices are obtained. Note that each modulated and shifted version of a Gaussian-shaped elementary signal has the smallest possible support in the space-frequency domain. Thus, a Gaussian signal achieves the upper joint spatial and angular resolution bound derived in Sec. 2.4. This is a particularly desirable property for our aims in this chapter, since it is another factor affecting the sparsity of the system matrix related to Eq. 5.7.

In the proposed method, we define the modulated Gaussian window functions on the hypothetical planes which are parallel to $z = 0$ plane and pass through the discrete positions $\mathbf{s}_{mn} = [mX, nY, \zeta_{mn}]^T$ on S (see Fig. 5.1) [81]. Here, ζ_{mn} represents the depth of the point on S for the given discrete position $[mX, nY]^T$ on $z = \zeta_{mn}$ plane. The modulated Gaussian window functions defined on the planes at $z = \zeta_{mn}$ correspond to Gaussian beams in 3D space where $\mathbf{s}_{mn}$ represents the waist positions of these beams. Note that the discrete beam waist positions are obtained by sampling of S using a regular grid on the transversal plane. For some surfaces, after applying such a sampling, there may be multiple discrete points that have the same transversal position but are at different depths (see Fig. 5.1). For such a case, the depth value of the beam waist position should be chosen such that the system matrix is as sparse as possible. Indeed, this position need not be on the curved surface as in the case of example shown in Fig. 5.7. The details regarding to this example will be given in Sec. 5.3. By using the proposed idea, we find the field at an observation point $\mathbf{x} = [x, y, z]^T$ due to the field $u(\mathbf{r})$ specified on the curved surface S , by summing up the contributions of

the Gaussian beams, as

$$u(\mathbf{x}) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{x} - \mathbf{s}_{mn}) \quad , \quad (5.8)$$

where $g_{mnkl}(\mathbf{x} - \mathbf{s}_{mn})$ represents the 3D field expression of the Gaussian beam at the observation point $\mathbf{x} = [x, y, z]^T$ due to the modulated Gaussian window function which is defined at the point $\mathbf{s}_{mn}$ on $z = \zeta_{mn}$ plane and has the modulation frequencies kF_x and lF_y along the x and y axes, respectively [81]. Note that in the proposed method, the transversal waist positions and the discrete set of modulation frequencies are fixed independent of the surface (see Fig. 5.1). Eq. 5.8 actually defines a 3D signal space for $u(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{C}$. In order that such a signal space covers the signal space which is defined in Sec. 5.1 and denoted as $\mathcal{H}_p$, the discrete shift steps X, Y in the positions and F_x, F_y in the modulation frequencies of the Gaussian window functions should be chosen sufficiently small. The number of Gaussian beams to be included in Eq. 5.8 depends also on the frequency band $f_x^2 + f_y^2 \leq \frac{1}{\lambda^2}$ of the signal space defined in Sec. 5.1. Indeed, the Gaussian beams corresponding to Gaussian window functions whose frequency supports are outside the region defined by $f_x^2 + f_y^2 \leq \frac{1}{\lambda^2}$ need not be taken into account. Similar to the case where the positions of the Gaussian window functions are taken on a reference plane (see Eq. 5.7), we find the coefficients of the Gaussian beams by solving the system of linear equations given as [81],

$$u(\mathbf{r}) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{r} - \mathbf{s}_{mn}) \quad . \quad (5.9)$$

We calculate the field at a given point $\mathbf{x}$ by

$$u(\mathbf{x}) = \sum_m \sum_n \sum_k \sum_l \hat{a}_{mnkl} g_{mnkl}(\mathbf{x} - \mathbf{s}_{mn}) \quad , \quad (5.10)$$

where the estimated coefficients $\hat{a}_{mnkl}$ denote the least squares solution of Eq. 5.9 defined by

$$\hat{a}_{mnkl} = \underset{a_{mnkl}}{\operatorname{argmin}} \left\{ \iint_S |u(\mathbf{r}) - \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{r} - \mathbf{s}_{mn})|^2 dS \right\} \quad . \quad (5.11)$$

Here, once more note that by using the proposed Gaussian beam decomposition method explained above, we end up with systems of linear equations, as given by Eq. 5.9, where the system matrices are sparse. This sparsity is the fact that we take advantage of and achieve considerable improvements in terms of computational complexity regarding the solution of the linear system of equations, for the problem of scalar diffraction field calculation from curved surfaces [81]. The related improvements are reported in the following section.

5.3 Simulation Results

We show the applicability of the developed formulations for both 1D curved lines and 2D curved surfaces. Note that the z variable represents the depth in both cases. For the 2D (x, z) -space case, Eq. 5.8 and Eq. 5.9 correspond to the equations,

$$u(\mathbf{x}) = \sum_m \sum_k a_{mk} g_{mk}(\mathbf{x} - \mathbf{s}_m) \quad , \quad (5.12)$$

$$u(\mathbf{r}) = \sum_m \sum_k a_{mk} g_{mk}(\mathbf{r} - \mathbf{s}_m) \quad , \quad (5.13)$$

where $\mathbf{x} = [x, z]^T$ and $\mathbf{s}_m = [mX, \zeta_m]^T$. Note that here $\mathbf{r}$ vector is used to represent the points on the given curved line.

Assuming that we deal with the fields that are formed by a superposition of propagating plane waves with $f_z > 0$ (see Eq.(5.4)), we can uniquely determine such fields given the intersection of these fields by anyone of the surfaces having infinite extents along the x and y axes. For the 2D (x, z) space case, we use the 2D cross sections of such plane waves by the (x, z) -plane. Because of numerical concerns, we deal with the fields that are periodic along the x -axis and we intersect such fields by the curved lines that are also periodic with the same

period along the x -axis. By this way, with a proper discretization, we restrict the signal space so that it is spanned by finite number of plane waves. We denote the number of plane waves constructing the signal space as N . In order to span such a signal space, we write each Gaussian beam as a sum of appropriately weighted plane waves (which are in the signal space) and obtain the periodic Gaussian beam expression

$$g_{mk}(\mathbf{x} - \mathbf{s}_m) = \frac{1}{N} \sum_{l=-\frac{N}{2}}^{\frac{N}{2}-1} \hat{G}_{mk} \left(\frac{l}{X_p} \right) \exp \left\{ j2\pi \left[\frac{l}{X_p} x + \sqrt{\frac{1}{\lambda^2} - \frac{l^2}{X_p^2}} (z - \zeta_m) \right] \right\} , \quad (5.14)$$

where X_p is the period of the curved line along the x -axis. The Fourier transform $\hat{G}_{mk}(f_x)$ of the shifted and modulated version of the Gaussian window function $g(x) = c \exp(x^2/\sigma^2)$ is found as

$$\begin{aligned} \hat{G}_{mk}(f_x) &= \mathcal{F}\{g(x - mX) \exp(j2\pi k F_x x)\} \\ &= \sqrt{c\pi}\sigma \exp\{-\pi^2\sigma^2 (f_x - kF_x)^2\} \\ &\quad \exp\{-j2\pi(f_x - kF_x)mX\} . \end{aligned} \quad (5.15)$$

Note that since the spectra of the Gaussian window functions are local, only a small percentage of plane waves that are in the signal space are sufficient to represent the corresponding Gaussian beams.

We uniformly discretize the given periodic curved line with sampling step L_s and have T samples per period. Denoting the resulting discrete points taken on the curved line as $\mathbf{r}_n = \mathbf{r}(L_s n)$, where $\mathbf{r}(l)$ is the arc-length parameterization of the curved line, we construct the linear system of equations by using Eq. 5.13 for each discrete point $\mathbf{r}_n$ as

$$u(\mathbf{r}_n) = \sum_{m=-\frac{M}{2}}^{\frac{M}{2}-1} \sum_{k=-\frac{K}{2}}^{\frac{K}{2}-1} a_{mk} g_{mk}(\mathbf{r}_n - \mathbf{s}_m) , \quad (5.16)$$

where M is the total number of shift positions in one period and K is the total number of modulation frequencies. Note that the discrete beam waist positions

$\mathbf{s}_m$ are obtained by using a uniform sampling grid on the x -axis with a sampling step X . Thus, $\mathbf{s}_m = [mX, \zeta_m]^T$.

Using the matrix notation, we write Eq. 5.16 in a more compact form as

$$\mathbf{U} = \mathbf{G}\mathbf{a} \quad , \quad (5.17)$$

where $\mathbf{U}$ is the vector which represents one period of the field given on the curved surface, G is the matrix whose columns represent one period of the field patterns that different parallel Gaussian beams produce on the curved line and $\mathbf{a}$ is the vector of coefficients of the parallel Gaussian beams to be found. Here, $\mathbf{U}$ and $\mathbf{a}$ are vectors that have lengths T and N , respectively; $\mathbf{G}$ is a $T \times N$ matrix. Note that the total number of parallel Gaussian beams used in the decomposition is equal to N (the degree of freedom of the signal space), i.e. $MK = N$. We form the system matrix $\mathbf{G}$ as

$$\mathbf{G} = [g_{11}(\mathbf{r} - \mathbf{s}_{11}) \mid g_{12}(\mathbf{r} - \mathbf{s}_{12}) \mid \dots \mid g_{MK}(\mathbf{r} - \mathbf{s}_{MK})] \quad (5.18)$$

and the coefficients vector $\mathbf{a}$ as

$$\mathbf{a} = [a_{11}, a_{12}, \dots, a_{MK}]^T \quad . \quad (5.19)$$

In order to obtain the exact field, the number of samples given on the curved line must be equal to or larger than the number of Gaussian beams used in the decomposition, i.e. $T \geq N$. Otherwise, multiple solutions exist. Denoting the least squares solution of the resulting overdetermined system of linear equations, given by Eq. 5.17, as

$$\begin{aligned} \hat{\mathbf{a}} &= [\hat{a}_{11}, \hat{a}_{12}, \dots, \hat{a}_{MK}]^T \\ &= \underset{\mathbf{a}}{\operatorname{argmin}} \left\{ \sum_{n=\frac{-N}{2}}^{\frac{N}{2}-1} |u(\mathbf{r}_n) - \sum_m \sum_k a_{mk} g_{mk}(\mathbf{r}_n - \mathbf{s}_m)|^2 \right\} \quad , \quad (5.20) \end{aligned}$$

the field at an observation point $\mathbf{x}$ is found as

$$u(\mathbf{x}) = \sum_{m=\frac{-M}{2}}^{\frac{M}{2}-1} \sum_{k=\frac{-K}{2}}^{\frac{K}{2}-1} \hat{a}_{mk} g_{mk}(\mathbf{x} - \mathbf{s}_m) \quad . \quad (5.21)$$

Note that as long as the observation distance z is sufficiently large and the waist of the Gaussian beams is sufficiently small, the Gaussian beam expression introduced in [35] can be used in Eq. 5.21. The mentioned expression which is based on the Rayleigh-Sommerfeld diffraction model gives the scalar 3D field corresponding to a modulated Gaussian window function quite accurately even for nonparaxial cases [35, 85].

Once we have the entire 2D field expression given by Eq. 5.21 due to the continuous field specified on the input curved line, we can find the continuous field on another curved line, the points on which are represented by $\mathbf{p} \in \mathbb{R}^2$, as

$$u(\mathbf{p}) = \sum_{m=-\frac{M}{2}}^{\frac{M}{2}-1} \sum_{k=-\frac{K}{2}}^{\frac{K}{2}-1} \hat{a}_{mk} g_{mk}(\mathbf{p} - \mathbf{s}_m) \quad . \quad (5.22)$$

In order to have a one-to-one continuous field relation between two different curved lines, the columns of the system matrix $\mathbf{G}$ must be linearly independent. Note that this requirement already implies that $T \geq N$.

In order to test the accuracy and efficiency of the developed formulations, we start with a 2D periodic and continuous (x, z) field which is specified by randomly choosing the coefficients of the periodic Gaussian beams from the uniform distribution on the open interval $(0, 1)$. By using Eq. 5.21, we calculate the resulting field on the $z = 0$ line to obtain our test signal. Similarly, we intersect the 2D test field by the curved line at the discrete points $\mathbf{r}_n$ to obtain the $\mathbf{U}$ vector. Using Eq. 5.14, we find the field samples that each Gaussian beam produces at the discrete points $\mathbf{r}_n$ on the curved line and form the $\mathbf{G}$ matrix. We also construct another overdetermined system of linear equations where the columns of the $\mathbf{G}$ matrix are the field patterns that the plane waves produce on the curved line. We solve these two overdetermined systems of linear equations to find the sets of Gaussian beam and plane wave coefficients. By using the resulting coefficients, we calculate the samples of the diffraction field on the reference line at $z = 0$ via the Gaussian beam decomposition given by Eq. 5.21

and the plane wave decomposition given as

$$u(x, z) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \hat{A}_k \exp \left\{ j2\pi \left(\frac{k}{X_p} x + \sqrt{\frac{1}{\lambda^2} - \frac{k^2}{X_p^2}} z \right) \right\} , \quad (5.23)$$

where $\left\{ \hat{A}_k; k \in \left\{ \frac{-N}{2}, \frac{-N}{2} + 1, \dots, \frac{N}{2} - 1 \right\} \right\}$ is the set of estimated plane wave coefficients. The accuracies of both methods are verified by comparing the reconstructed signals with the original test signal defined on the $z = 0$ line.

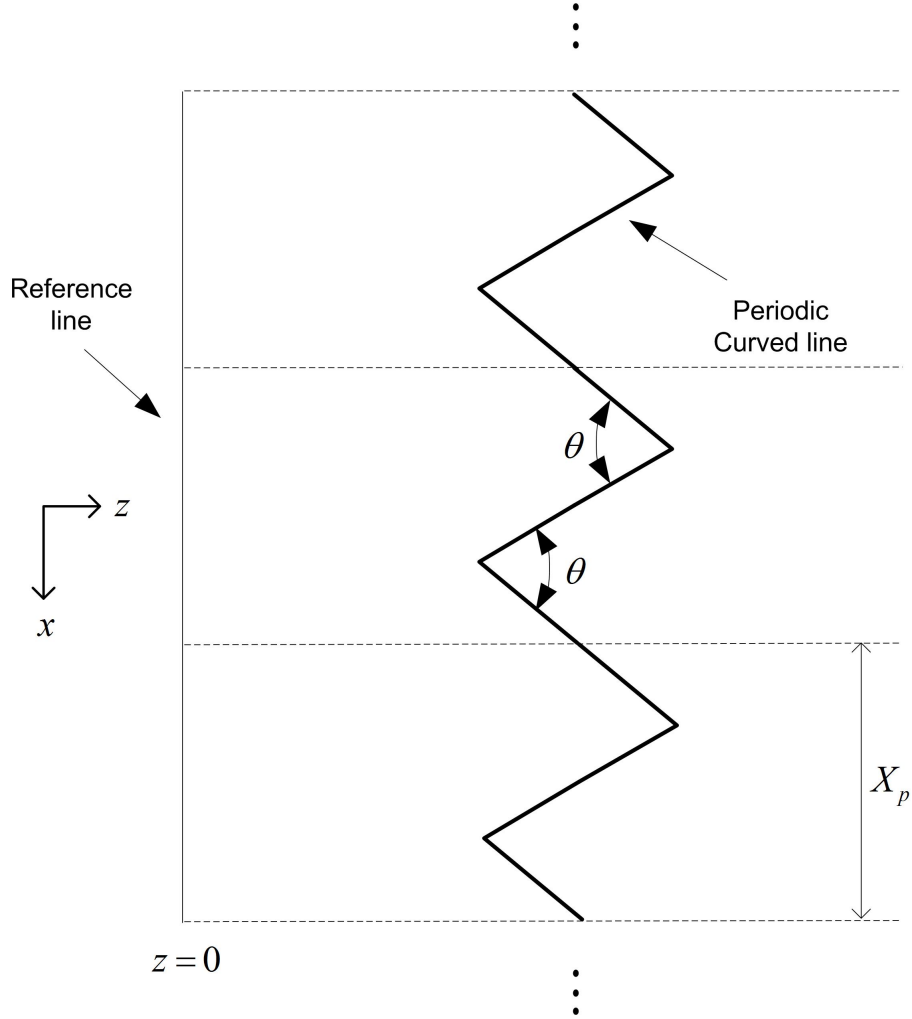


Figure 5.2: The periodic 2D simulation setup. (©2013 JOSA A. Reprinted with permission. Published in [81])

The coefficients of the Gaussian beams and plane waves used in the Gaussian beam decomposition and plane wave decomposition, respectively, are found via the QR factorization method [86]. In order to test the efficiency of the proposed method, we compare the computation times necessary to solve the two systems of

linear equations corresponding to these decompositions. We apply the usual QR factorization routine of MATLAB as well as the suitesparseQR (SPQR) package [83] which provides MATLAB routines to implement multifrontal sparse QR factorization. Note that most of the elements of the $\mathbf{G}$ matrix in the proposed method are practically negligible. Therefore, while using SPQR, we simply apply a thresholding to the $\mathbf{G}$ matrix to obtain a sparse matrix and take advantage of the large number of zero elements of such a sparse matrix.

The simulation setup is seen in Fig. 5.2. In order to see the effect of the size of the input manifold, we change the period X_p along the x -axis and do the simulations for each case. While doing this, we preserve the structure given in Fig. 5.2. Therefore, as the period of the curved line is increased, the number of samples T taken on the curved line and the number of parallel Gaussian beams N are also increased. We take $T = 1.2N$ and keep the sampling step L_s constant at 0.43λ where λ is the wavelength of the monochromatic light and taken as $500nm$. Note that by taking $T > N$, we ensure that the resulting systems of linear equations are overdetermined. θ is chosen as 30 degrees. The computational complexities of the four operations explained above are given in Fig. 5.3 with respect to N . The operations mentioned in Fig. 5.2 are: 1) QR factorization is applied to the PWD formulation (QRPWD), 2) Multifrontal sparse QR factorization is applied to the PWD formulation (SQRPWD), 3) QR factorization is applied to the proposed Gaussian beam decomposition formulation (QRGBD) and 4) Multifrontal sparse QR factorization is applied to the proposed Gaussian beam decomposition formulation (SQRGBD). Note that the computational complexity of SQRGBD is taken as one unit time as a reference for all T . The computational complexities of other operations represent the relative times with respect to SQRGBD. For the case that the proposed Gaussian beam decomposition method is used, the sparsity of the $\mathbf{G}$ matrix is shown in Fig. 5.4 for the same set of N . Here, the sparsity is defined as the ratio of the number of zero elements of a matrix to its total number of elements.

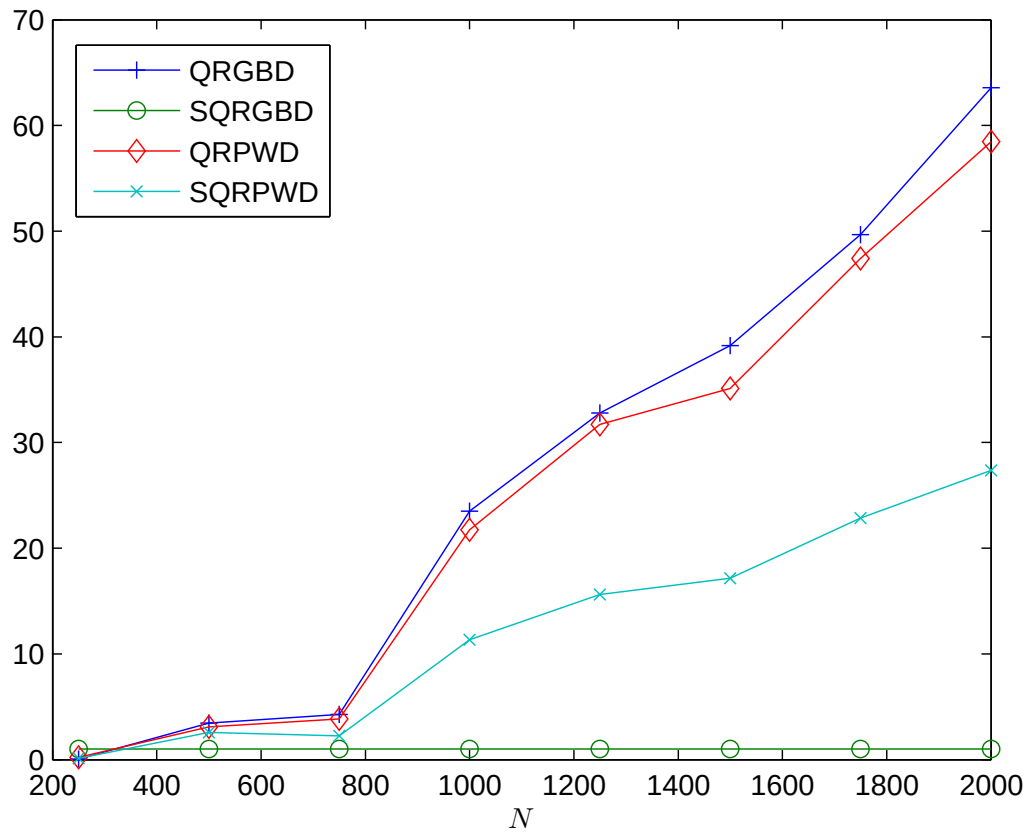


Figure 5.3: The ratios of the complexities of the operations QRPWD, SQRPWD, QRGBD to the complexity of SQRGBD (note that the result shown for SQRGBD is “1”). (©2013 JOSA A. Reprinted with permission. Published in [81])

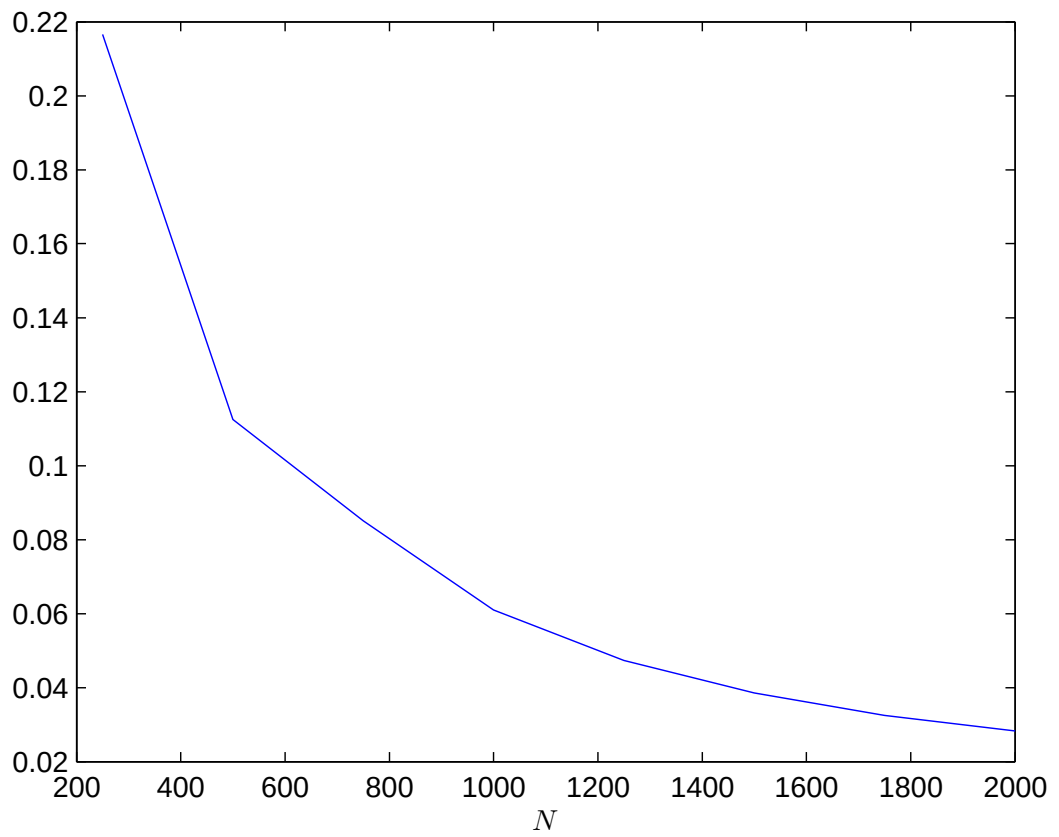


Figure 5.4: The sparsity of the system matrices for the proposed Gaussian beam decomposition method. (©2013 JOSA A. Reprinted with permission. Published in [81])

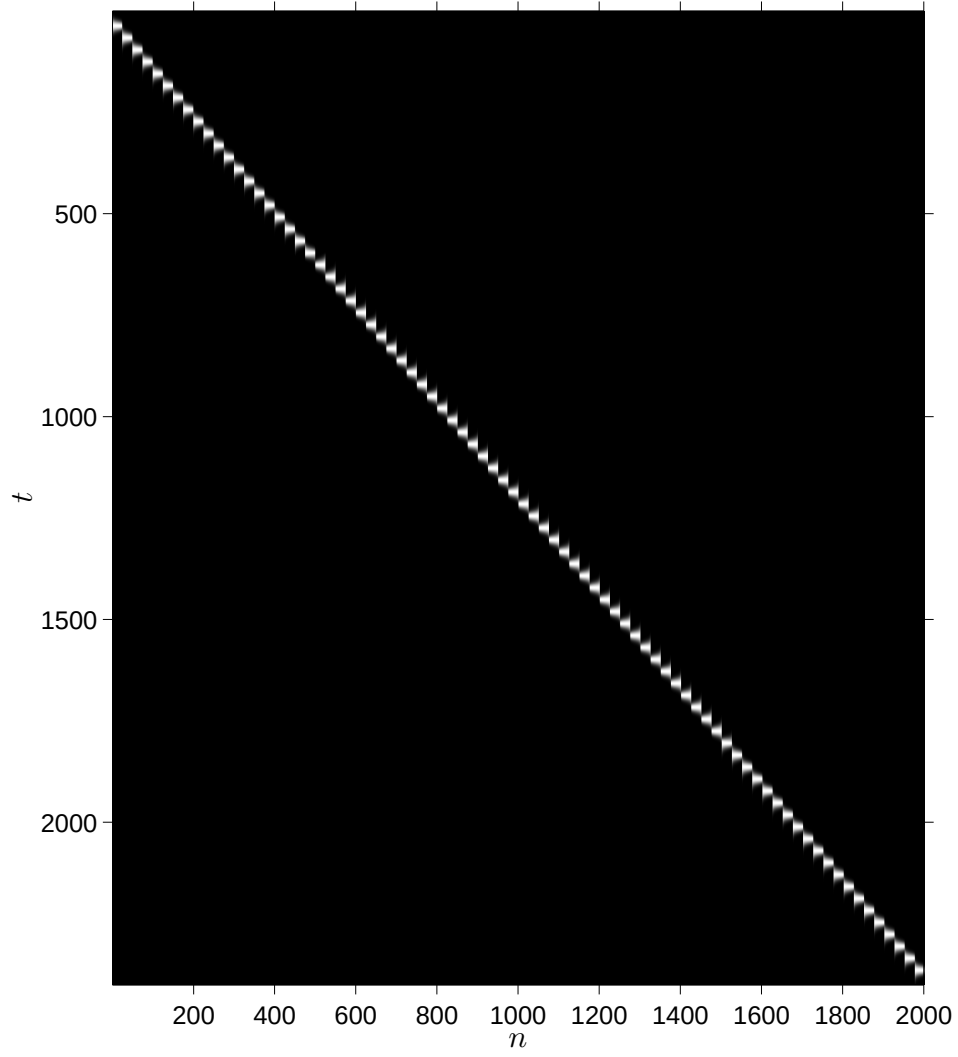


Figure 5.5: The absolute value of the system matrix $\mathbf{G}$ which has size 2400×2000 and formed by the application of the proposed Gaussian beam decomposition method to the curved line shown in Fig. 5.2. (©2013 JOSA A. Reprinted with permission. Published in [81])

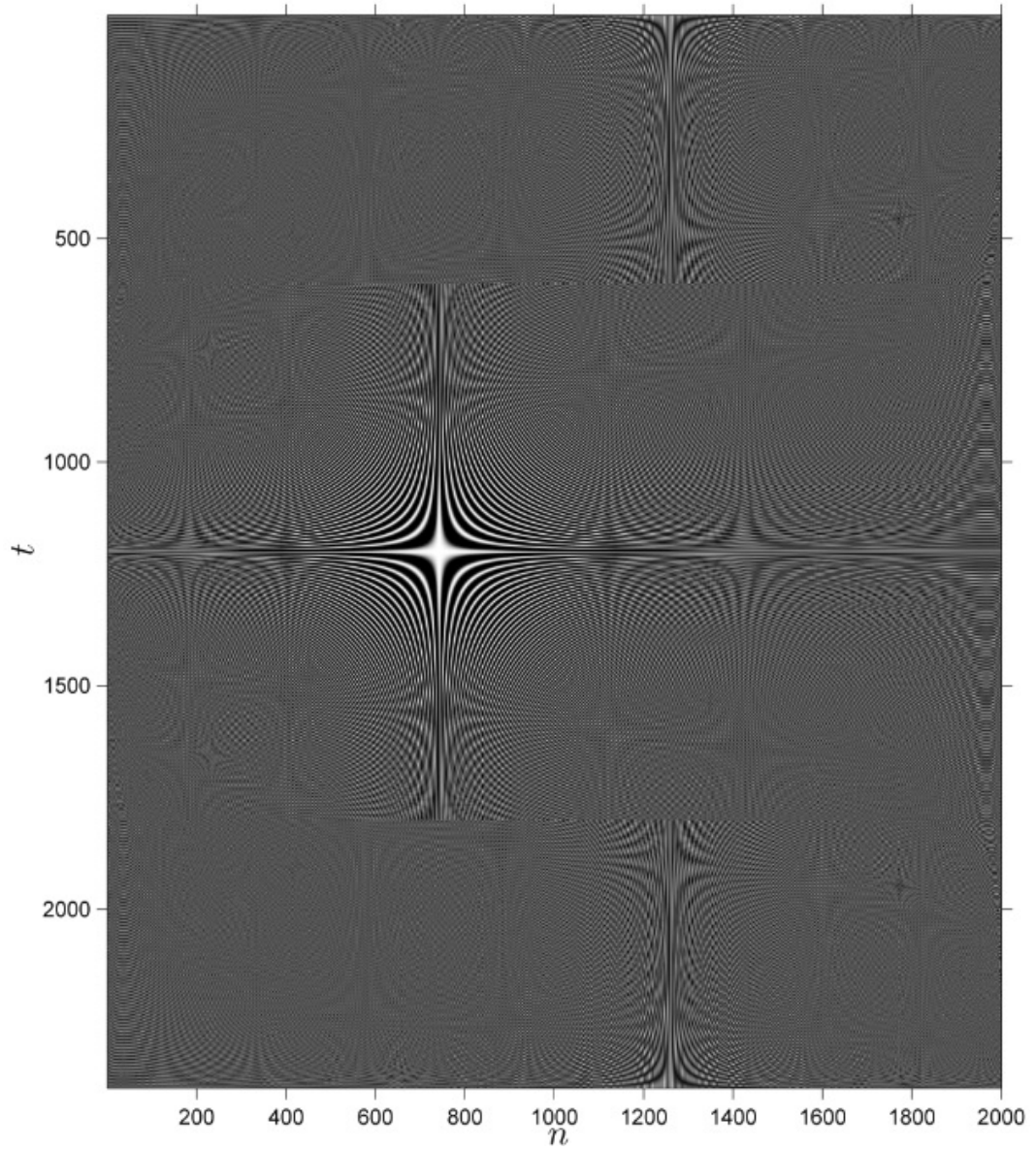


Figure 5.6: The real part of the system matrix $\mathbf{G}$ which has size 2400×2000 and formed by the application of the plane wave decomposition method to the curved line shown in Fig. 5.2. (note that the absolute values of all matrix entries are one.) Note the difference with respect to Fig. 5.5 in terms of sparsity. (©2013 JOSA A. Reprinted with permission. Published in [81])

Examples of the system matrices for the proposed Gaussian beam decomposition and the plane wave decomposition methods are shown in Fig. 5.5 and Fig. 5.6, respectively, for $T = 2400$. Note that we use t for indexing over the sample points of the curved line (or curved surface) and n is used for indexing over the profiles which are produced on the curved line (or curved surface) by different expansion functions. Thus, $g(t, n)$ represents the entry of the system matrix $\mathbf{G}$ that has row number t and column number n . The sparsity of the $\mathbf{G}$ matrix corresponding to the proposed method is clearly seen in Fig. 5.5. All the figures mentioned above justify that as the system matrix corresponding to proposed method gets more sparse, the resulting computational saving becomes more significant. Such a computational advantage makes the proposed method more critical for meaningful sizes of surfaces.

In order to show the applicability and efficiency of the proposed method for curved lines which have loops, we conduct another simulation. We choose the curved line which has two loops as shown in Fig. 5.7. In this simulation, we firstly specify the test signal on the $z = 0$ line as a rectangular window function. Then, by using the plane wave decomposition given by Eq. 4.24, we find the 1D field on the curved line. From this given 1D field, we achieve the entire 2D field as shown in Fig. 5.7 by using the proposed Gaussian beam decomposition method. As can be seen in Fig. 5.7, the intersection of calculated 2D field by the $z = 0$ line gives the original test signal which is a rectangular window function. There are two depth points corresponding to each horizontal beam waist position on the loops. The system matrices for the cases where the beam waist positions are taken inside the loops and on the curved line are shown in Fig. 5.8 and Fig. 5.9, respectively. Note that the choice of beam waist positions inside the loops results in a slightly sparser system matrix than the case where the beam waist positions are chosen on the curved line. For the case that the beam waist positions are chosen inside the loops, the beam waist positions can be clearly seen in Fig. 5.10. Here, we take the beam coefficients equal for illustration purposes.

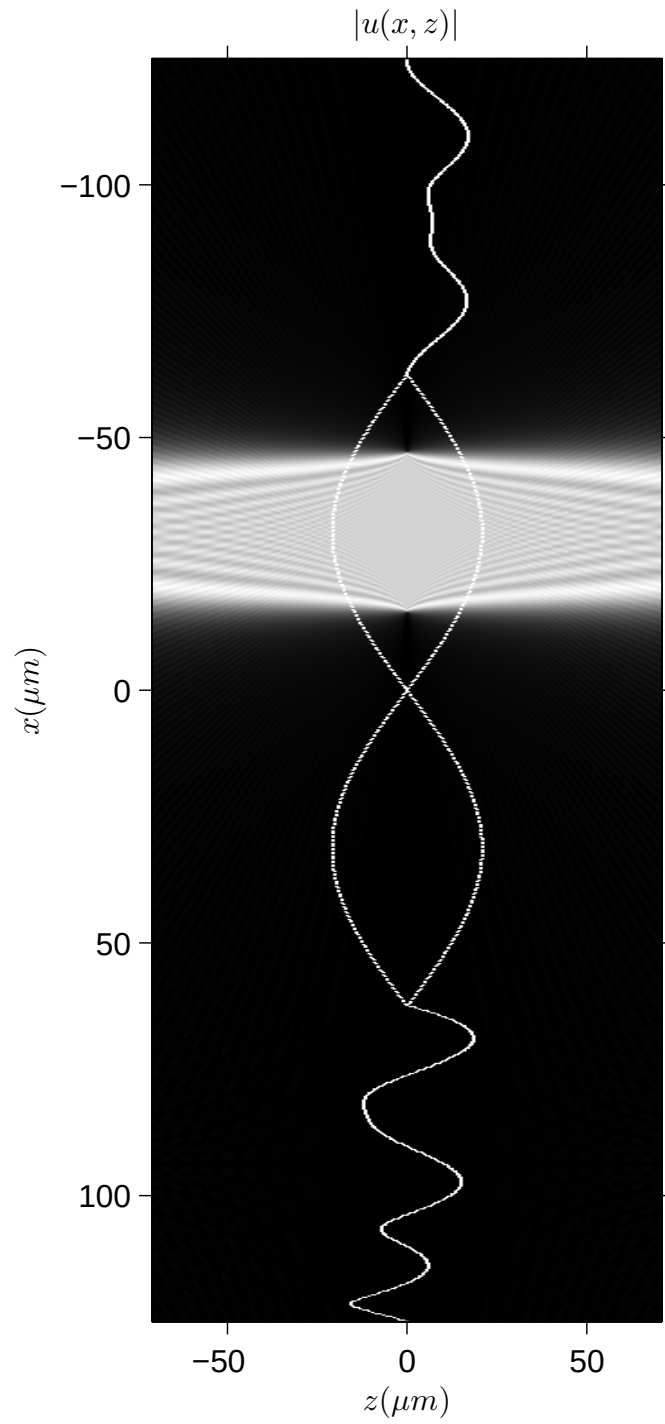


Figure 5.7: One period of the two-dimensional (x, z) field corresponding to the signal given on the periodic curved line which has loops.

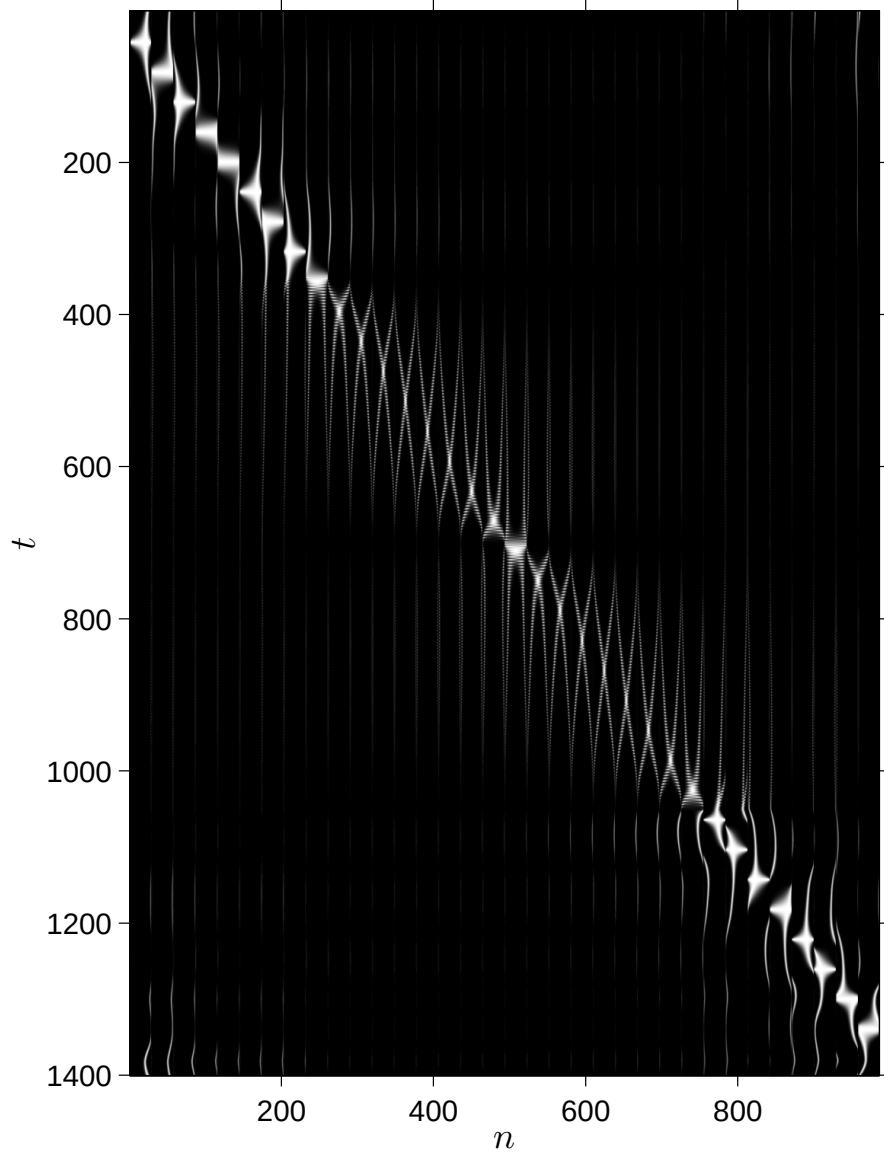


Figure 5.8: The absolute value of the system matrix $\mathbf{G}$ which has size 1400×986 and formed by the application of the proposed Gaussian beam decomposition where the beam waist positions are chosen inside the loops.

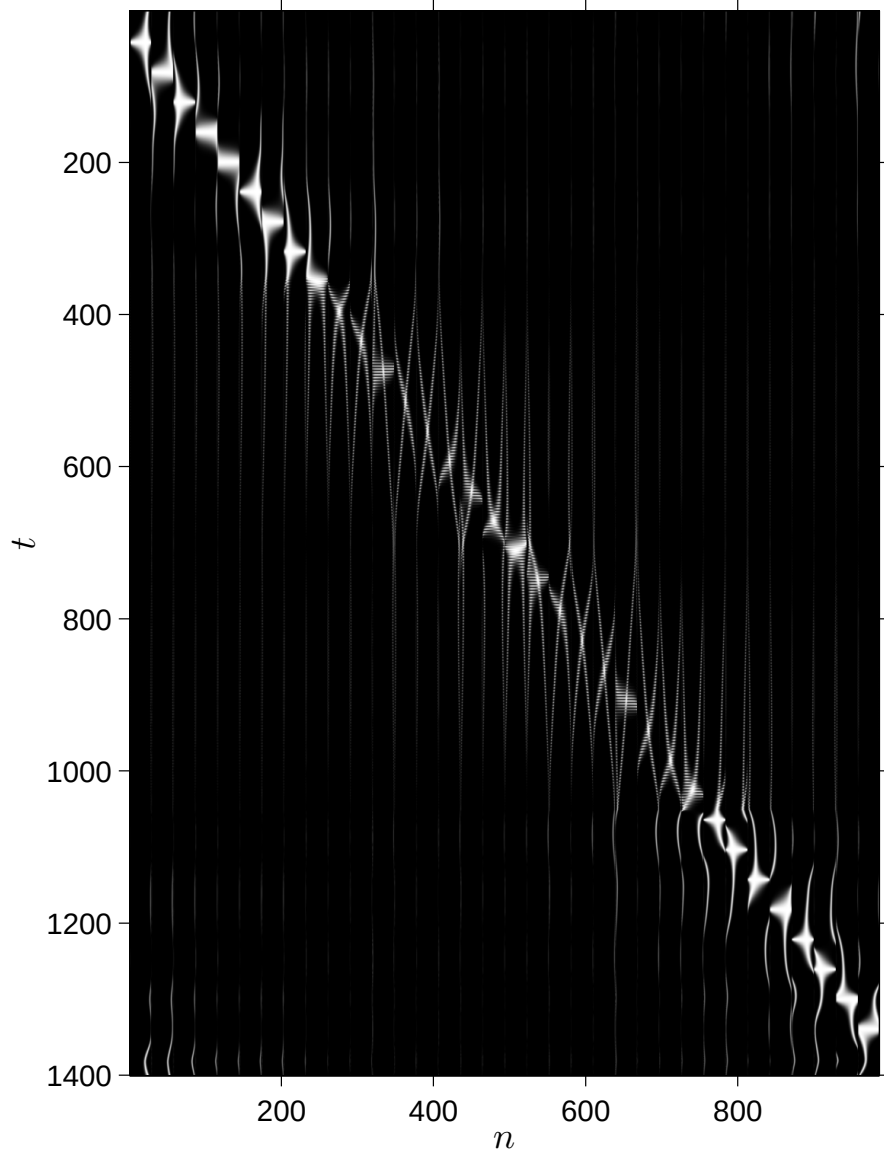


Figure 5.9: The absolute value of the system matrix $\mathbf{G}$ which has size 1400×986 and formed by the application of the proposed Gaussian beam decomposition where the beam waist positions are chosen on the curved line.

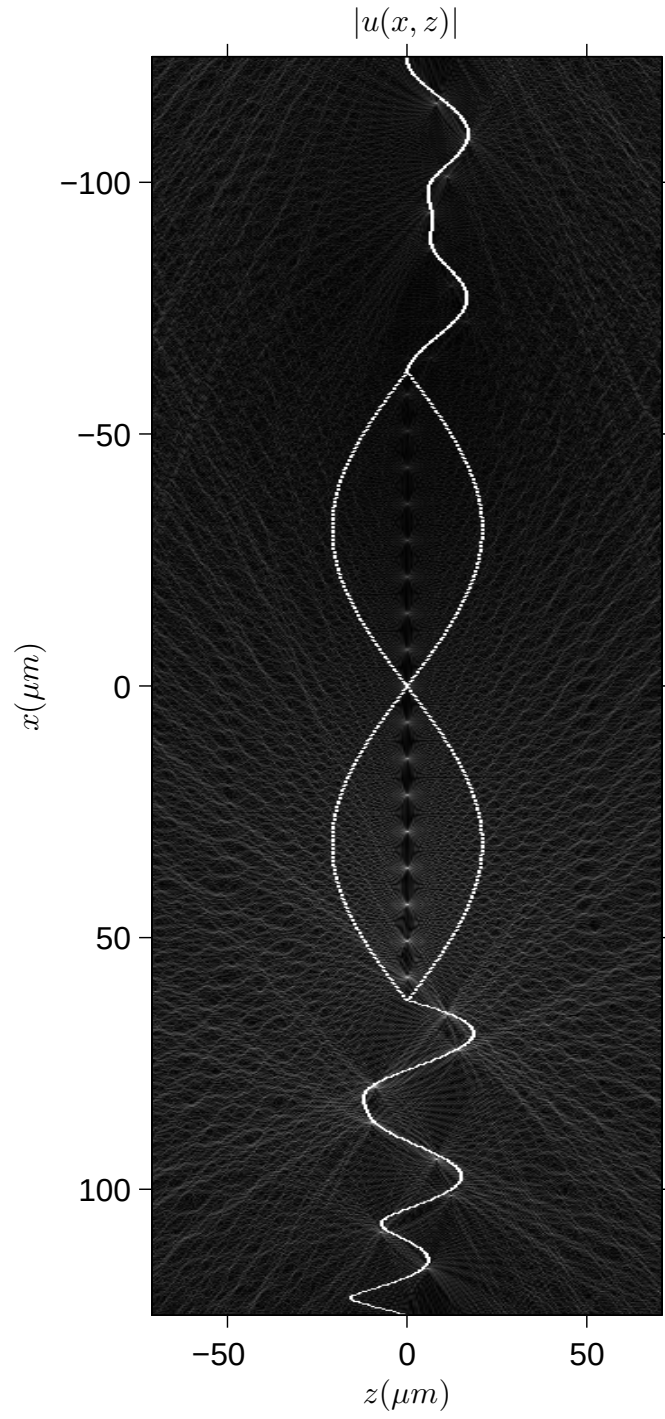


Figure 5.10: One period of the two-dimensional (x, z) field corresponding to the Gaussian beam decomposition where the beam waist positions are taken inside the loops and all the beam coefficients are taken equal.

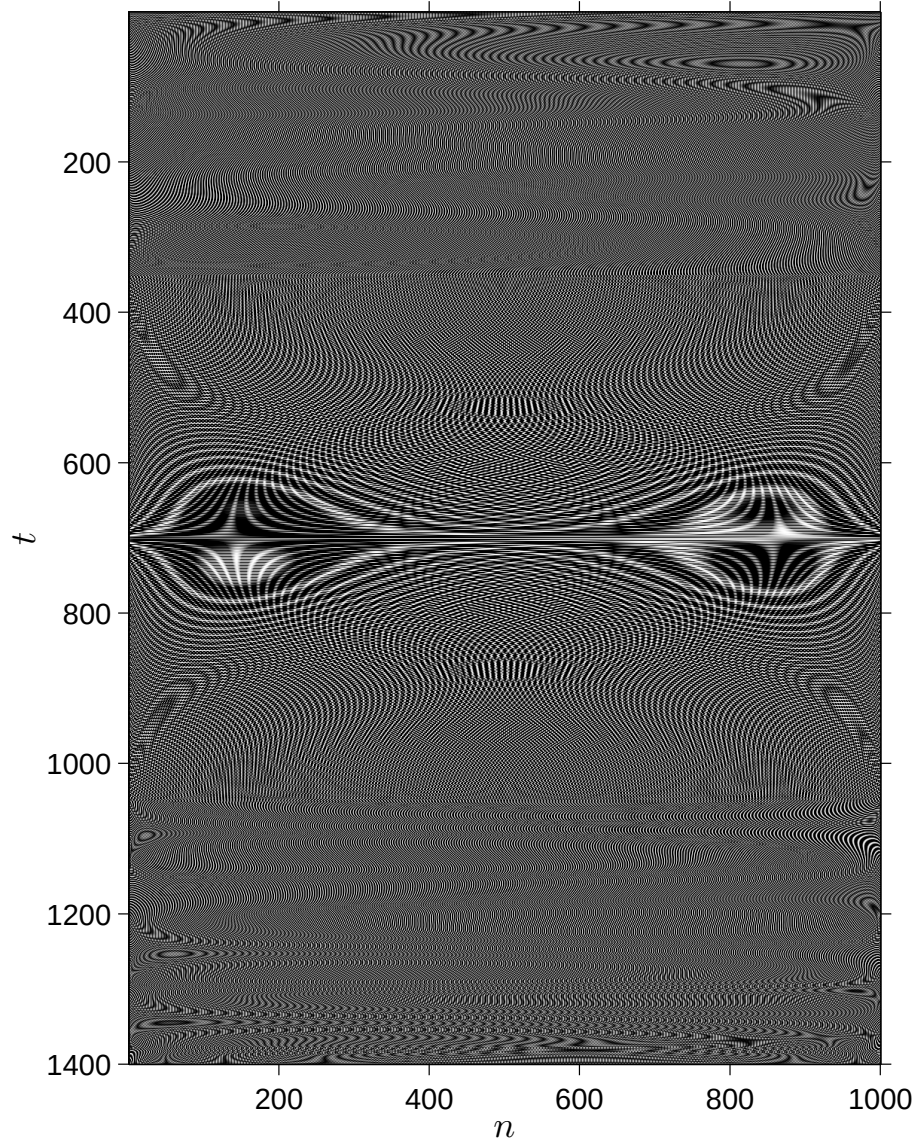


Figure 5.11: The real part of the system matrix $\mathbf{G}$ which has size 1400×1000 and formed by the application of the plane wave decomposition method for the curved line shown in Fig. 5.7. Note that the absolute values of all matrix entries are one.

The system matrix which results in after the application of the plane wave decomposition method for the curved line shown in Fig 5.7 is given in Fig. 5.11. When the sparse QR factorization is applied to the system matrices given in Fig. 5.8 and Fig. 5.11, it is observed that about five times of computational gain is achieved as a consequence of the sparsity of the system matrix corresponding to the proposed method.

As a proof of concept, we also test the developed formulations for 2D curved surfaces. We assume that both the 3D field and the curved surface are periodic along the x and y axes with the same period. The periodicity assumption is due to computational reasons. Note that with a proper indexing over the elements of the samples of the given field on the curved surface, we end up with a 1D array $\mathbf{U}$. Using the same indexing, we obtain 1D arrays of field pattern samples that the expansion functions produce on the curved surface. These arrays are actually columns of the system matrix $\mathbf{G}$. Thus, they form the $\mathbf{G}$ matrix. Therefore, the problem formulation given by Eq. 5.17 is also valid for curved surfaces.

We use a periodic curved surface, which we obtain by lowpass filtering a periodic 2D discrete function whose spectral coefficients are randomly chosen from the uniform distribution on the open interval $(0,1)$, and a synthetically generated periodic object in 3D simulations. One periods of the curved surface and object are shown in Fig. 5.12 and Fig. 5.13, respectively.

For the curved surface given by Fig. 5.12, the period along both the x and y axes is taken as 54.5λ . We discretize the curved surface by using a uniform transversal sampling grid where the sampling step is taken as 0.5λ along both the x and y axes. One period of the resulting 2D discrete signal is of size 108×108 . The total number of shift positions and the total number of modulation frequencies of the Gaussian window functions along the x and y axes are chosen to be same. The total number of plane waves and Gaussian beams used in the decompositions are 9475 and 9472, respectively. Thus, noting that the total

number of samples on one period of the curved surface is 11664, the system matrices are overdetermined.

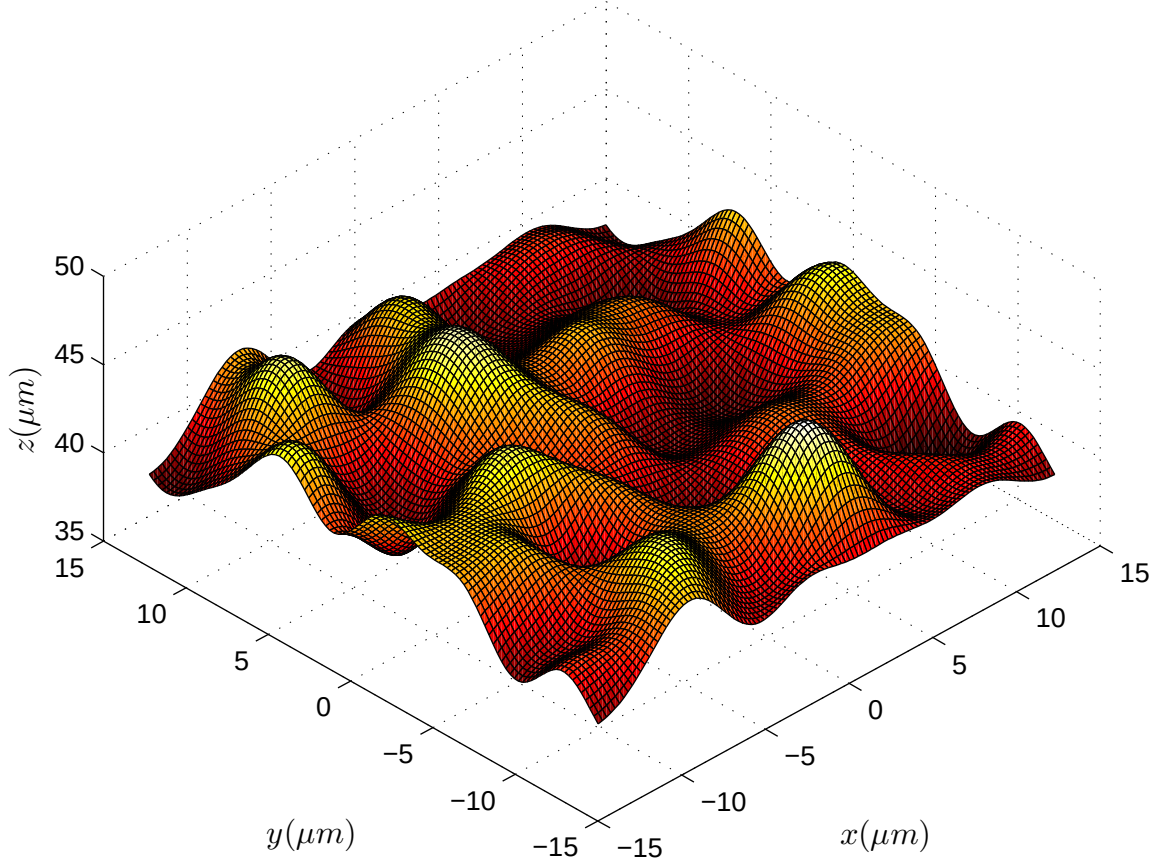


Figure 5.12: One period of the 2D periodic curved surface. (©2013 JOSA A. Reprinted with permission. Published in [81])

The period along both the x and y axes is taken as 30λ for the object given by Fig. 5.13. The object is composed of 3973 vertices. In this simulation, the total number of plane waves and Gaussian beams used in the decompositions are 2819 and 2809, respectively. Therefore, the system matrices for both the Gaussian beam and the plane wave decomposition methods are overdetermined.

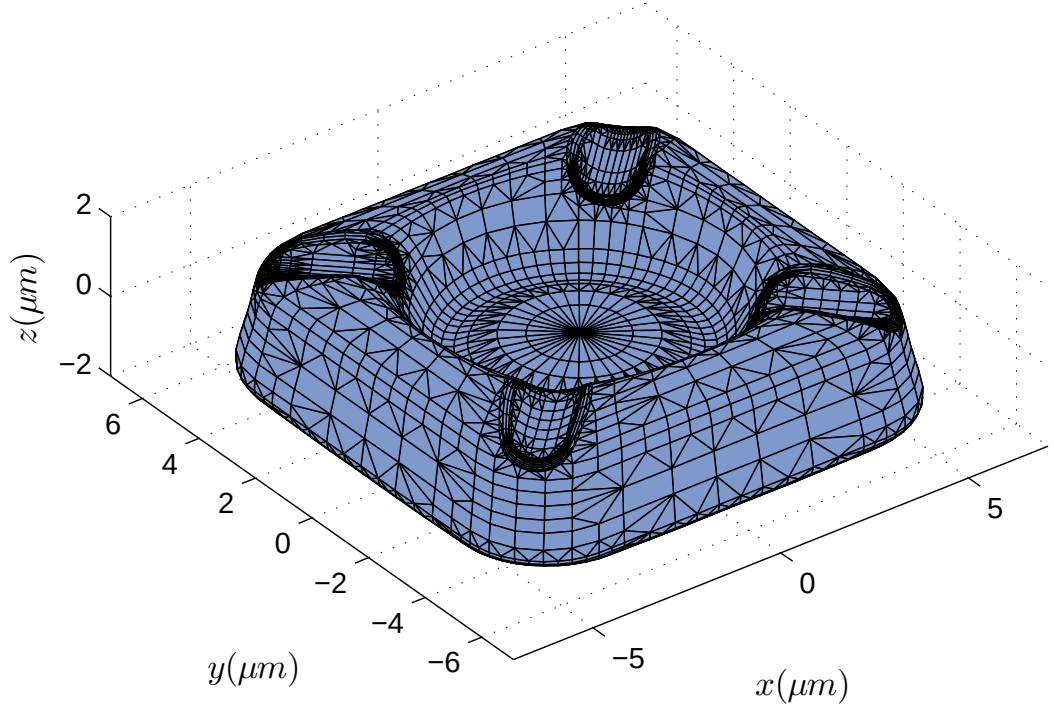


Figure 5.13: One period of the 2D periodic object. (Gilles Tran ©2007 www.oyonale.com, used under the Creative Commons Attribution license. Published in [81])

By randomly choosing the coefficients of the periodic Gaussian beams (from the uniform distribution on the open interval $(0, 1)$), we define the 3D field and form the corresponding system of linear equations. We repeat the proposed procedure several times for randomly chosen different 3D field examples. Examples of the system matrices related to the proposed Gaussian beam decomposition and plane wave decomposition methods are shown in Fig. 5.14 and Fig. 5.15 for the curved surface given by Fig. 5.12. The system matrices for the object given by Eq. 5.13 are shown in Fig. 5.16 and Fig. 5.17. The sparsities of the system matrices corresponding to the Gaussian beam decomposition method are clearly seen in Fig. 5.14 and Fig. 5.16.

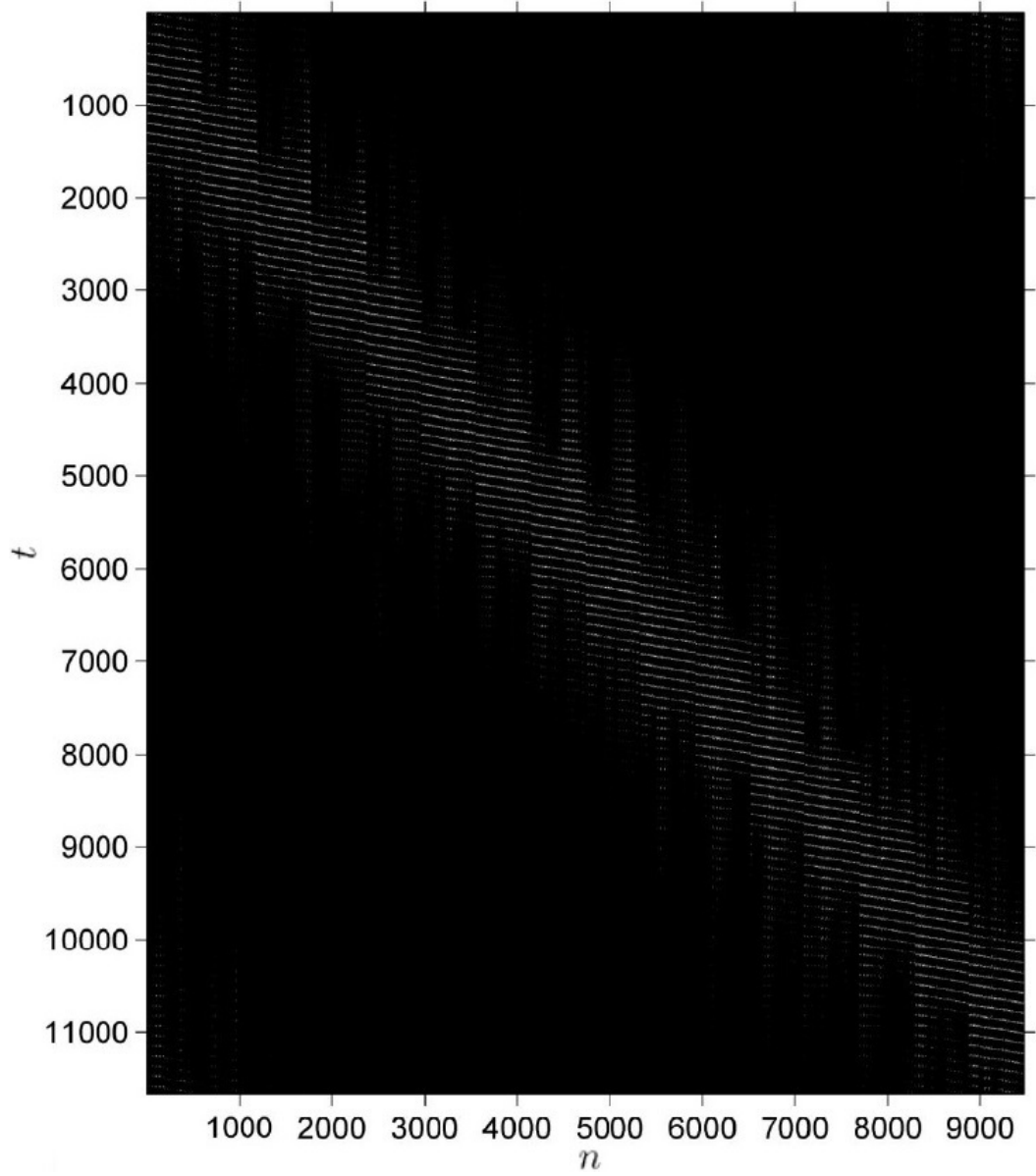


Figure 5.14: The absolute value of the system matrix $\mathbf{G}$ which has size 11664×9472 and formed by the application of the proposed Gaussian beam decomposition method for the curved surface given in Fig. 5.12. Note that the figure is enhanced for better visualization. (©2013 JOSA A. Reprinted with permission. Published in [81])

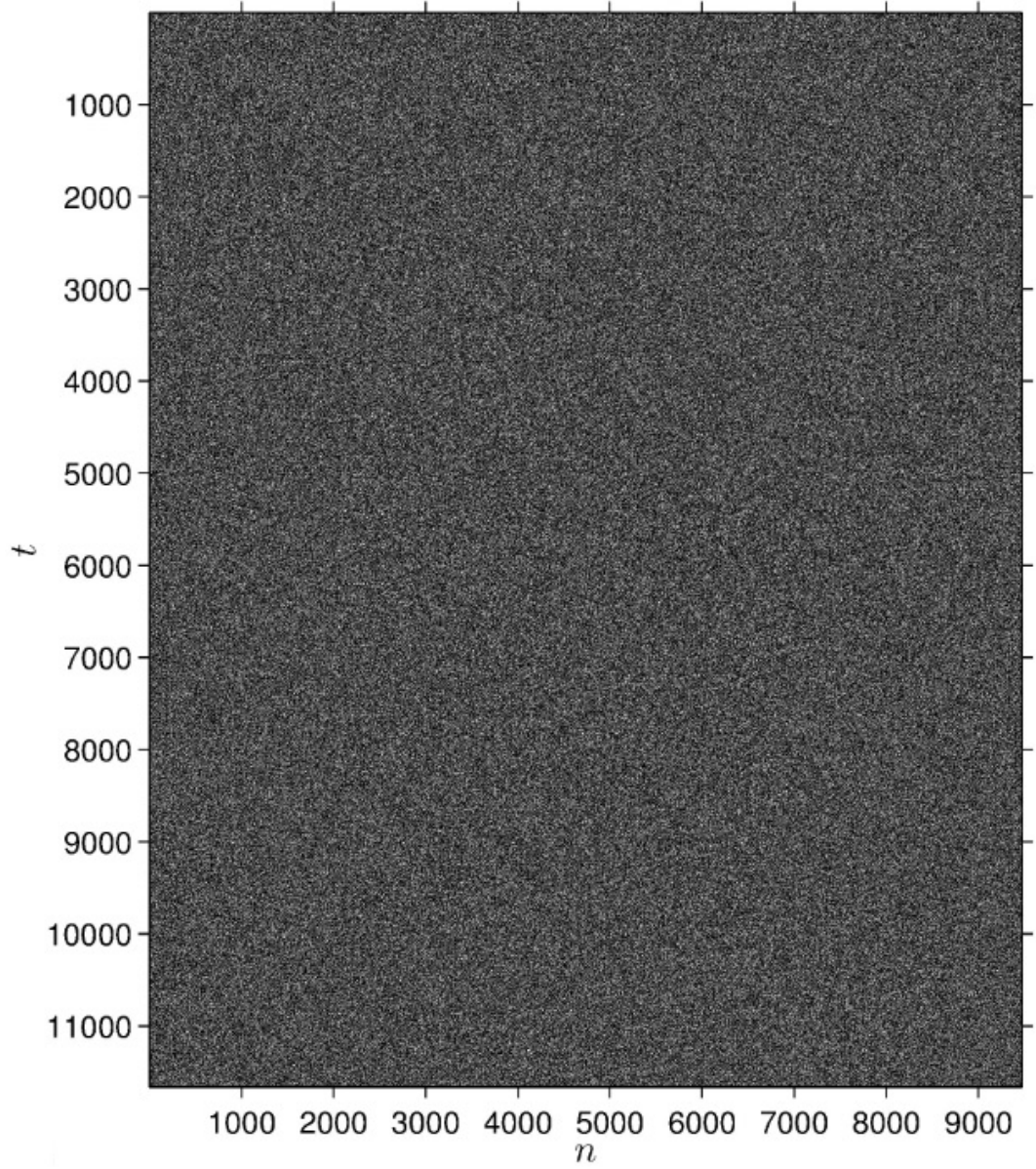


Figure 5.15: The real part of the system matrix $\mathbf{G}$ which has size 11664×9475 and formed by the application of the plane wave decomposition method for the curved surface given in Fig. 5.12. Note that the absolute values of all matrix entries are one. Note also that the random noise-like appearance is a consequence of the structure of the surface given in Fig. 5.12. (©2013 JOSA A. Reprinted with permission. Published in [81])

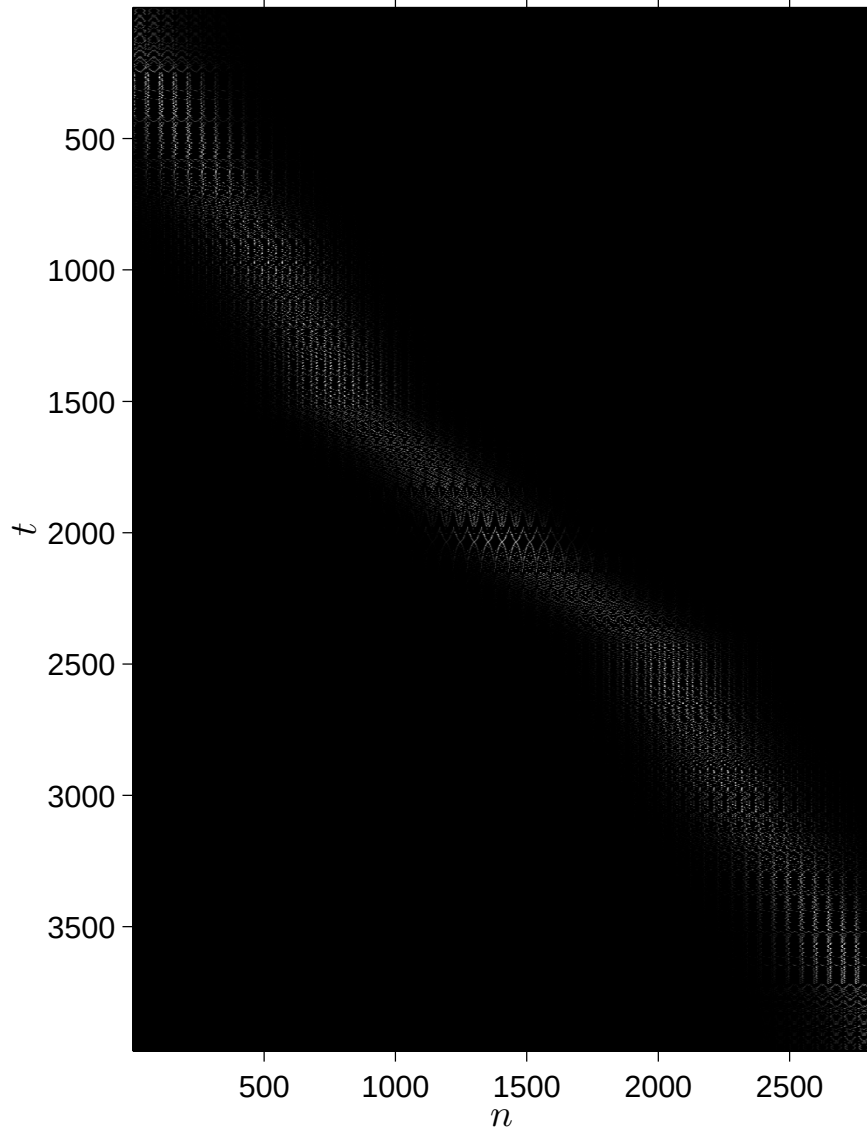


Figure 5.16: The absolute value of the system matrix $\mathbf{G}$ which has size 3973×2809 and formed by the application of the proposed Gaussian beam decomposition method for the object given in Fig. 5.13. (©2013 JOSA A. Reprinted with permission. Published in [81])

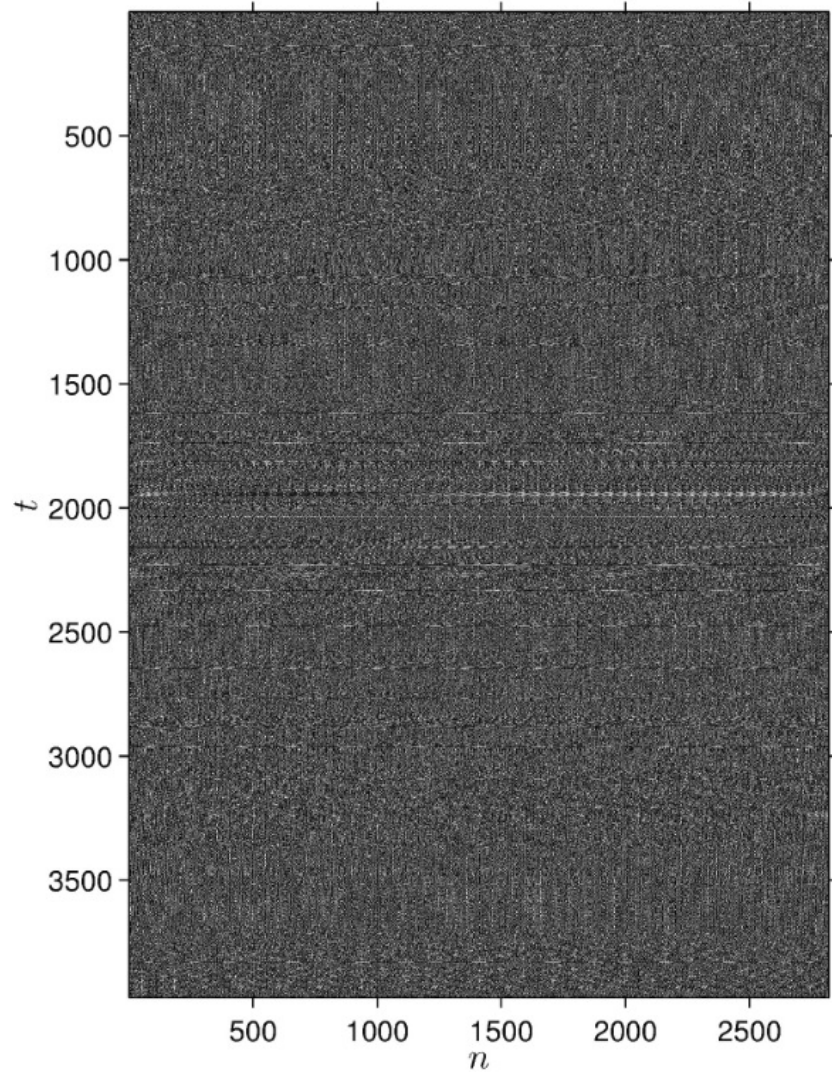


Figure 5.17: The real part of the system matrix $\mathbf{G}$ which has size 3973×2819 and formed by the application of the plane wave decomposition method for the object given in Fig. 5.13. Note that the absolute values of all matrix entries are one. ((©2013 JOSA A. Reprinted with permission. Published in [81]))

We solve the linear systems of equations related to the proposed Gaussian beam decomposition and plane wave decomposition methods by applying the usual QR decomposition and multifrontal sparse QR factorization to the system matrices. For the curved surface given by Fig. 5.12, we report that the computational gain due to the usage of sparse QR factorization (with respect to the usual QR decomposition) is about two for the plane wave decomposition method. However, this ratio is about 20 for the proposed method. When the sparse QR factorization is used, the time necessary to solve the linear system of equations related to the plane wave decomposition method is about 10 times more than the time necessary to solve the linear system of equations related to the proposed Gaussian beam decomposition method.

For the object given by Fig. 5.13, if the sparse QR factorization is applied to the system matrix corresponding to the plane wave decomposition method, the computation time is reduced by a factor of about two with respect to the usual QR decomposition. However, this reduction is by a factor of about 10 for the proposed method. When the sparse QR factorization is used, the time necessary to solve the linear system of equations corresponding to the plane wave decomposition method is about five times more than the time necessary to solve the linear system of equations corresponding to the proposed Gaussian beam decomposition method. Thus, these simulation results show that the sparsity of the system matrices related to the proposed method is advantageous even for the small scale examples that we demonstrate above.

The sparsity of the system matrix $\mathbf{G}$ is also beneficial in terms of memory. If the system matrix is sparse, the memory requirement related to the storage of the system matrix during the solution of the linear system of equations can be relieved by using a sparse storage scheme. Thus, the sparsity of $\mathbf{G}$ corresponding to the proposed method results in a reduction in the memory requirement compared to cases where the system matrix is a full matrix as in the case of

plane wave decomposition method. In the proposed method, before applying the sparse QR factorization to the system matrix, thresholding is used and then the system matrix is stored using the sparse storage scheme provided by MATLAB [84]. Once we have the matrix stored using the sparse storage scheme, we do not need the original matrix. Therefore, noting that the memory size of the original matrix is considerably higher than the memory size of the matrix stored using the sparse storage scheme, a memory gain is achieved. Even for the small scale examples that we explain above, the memory requirements of the system matrices corresponding to the proposed Gaussian beam decomposition method can be reduced by considerable amounts. The reductions are by factors of about 50 and 10 for the curved surface and object given by Fig. 5.12 and Fig. 5.13, respectively.

In this section, we dealt with various examples and showed the efficiency of the proposed method in each case. The reported simulation results lead us to state that we can expect considerable amount of gains in computational complexity and memory for most of the curved surfaces. Here, it is worth noting that the structure of the curved surface actually has an effect on the amount of the gain. For example, the intersection of Gaussian beams by the curved surface at multiple locations (as in the case of loop example shown in Fig. 5.7) reduces the sparsity of the system matrix and thus the computational gain decreases. The size of the curved surface is another factor that affects the computational and memory gains. The system matrices for the proposed method generally have a banded matrix structure (see Fig. 5.5, Fig. 5.14 and Fig. 5.16). The width of that band is determined mainly by the waist width of the Gaussian beams. As the size of the given curved surface increases, the size of the system matrix also increases. However, the width of the Gaussian beam waists need not be increased proportionally. Therefore, the computational and memory gains of the proposed method related to the solution of linear system of equations increases as well, since for larger size of surfaces the system matrices become

more sparse (the percentage of total number of nonzero elements of the matrix become less). Although there is no exact explicit expression for the dependence of the computational complexity of the sparse QR factorization to sparsity, as the above explanations imply, we can note that the computation time and memory usage in the process of sparse QR factorization is roughly $O(|\mathbf{G}|)$ (here $|\mathbf{G}|$ denotes the total number of nonzero entries of $\mathbf{G}$) [83]. Indeed, this dependence can be observed in Fig. 5.3 and Fig. 5.4 by comparing the computation times of the operations SQRGBD and QRGBD with respect to different sparsities at different N values (especially for $N \geq 1000$).

Chapter 6

Conclusions

In Chapter 2, we introduce a comparative study of the light field representation and integral imaging to relate the ray optics to the scalar wave optics. The integral imaging is shown to be equivalent to the discrete light field representation provided that the light rays are restricted within a finite propagation angle so that there is no cross-talk between the elemental images from different microlenses. The apertures of the microlenses and sensors correspond to pixels on the two parallel planes which are used in the parameterization of the light field representation. In the continuous case, it is shown that an equivalence relation between the light field representation and integral imaging can not be established, since the elimination of the cross-talk between the elemental images becomes practically impossible. Furthermore, in this case, an infinitesimal microlens behaves like a point light source and thus diffracts the incoming ray by equally distributing the incoming power in all directions. Thus, it is impossible to get arbitrary resolutions in both space and angle. These facts are the direct consequences of the uncertainty principle. At that point, we incorporate the uncertainty principle into the formulation by the analysis of integral imaging under the scalar wave optics. In this analysis, the linear space-frequency decomposition of the scalar wave field is the key point in establishing a relation between the

scalar wave optics and (observable) light field representation. More specifically, the light field representation is written as the magnitude squared of the windowed (short time) Fourier transform of the scalar wave field. This established link leads us to derive the upper joint spatial and angular (spectral) resolution bound, $16\pi^2$, for physically realizable light field representations which in turn helps us to understand the limits of practical implementation of light fields.

In Chapter 4, we present a local Gaussian signal decomposition method for the problem of diffraction field calculation from a field given on a curved surface. The proposed signal decomposition method provides a model with better accuracy compared to the source model approaches [23, 25, 26, 27, 28, 29, 30] for the specified problem. The reason is that the angular spread of a local Gaussian beam is considerably smaller than the angular spread of a 3D field induced by a point source (or a planar polygon). Thus, the mutual couplings between the point sources on a curved surface are significantly higher than the mutual couplings between the patches that induce local Gaussian beams.

The field model introduced in [53] provides an exact solution for the problem of scalar diffraction field calculation from a field given on a curved surface. That is an important development over the source model approaches [23, 25, 26, 27, 28, 29, 30]. However, the computational complexity of the exact field model given in [53] increases significantly for large sizes of surfaces, since the defined problem is desired to be solved by an inverse problem approach. Thus, in the discrete case a large matrix inversion should be performed; this is a computationally highly demanding operation for a meaningful size of surface. Moreover, since the intersections of infinite extent plane waves by the given curved surface are needed to be stored as a part of the algorithm proposed in [53], a memory allocation problem occurs for large sizes of surfaces. On the other hand, there is no such a requirement for the Gaussian signal decomposition method that we present in Chapter 4. Furthermore, our proposed method is suitable for parallel

programming to reduce the computation time considerably, since we find the 3D field as a sum of Gaussian beams each of which can be calculated separately, in a parallel fashion. That is, although our method does not produce exact field solutions as the model given in [53], it provides quite accurate results and applicable to even large size surfaces.

The choice of the width of the Gaussian window function used in the decomposition given in Chapter 4 is based on a tradeoff. The developed approximate Gaussian beam expression gives better results as the width of the Gaussian window function is reduced. Such a narrower Gaussian window function is also desired in order to satisfy the smoothness constraint of a given surface for the applicability of the developed formulations. On the other hand, Gaussian beams corresponding to narrower Gaussian window functions have larger angular spreads. Therefore, the likelihood that the patches on the curved surface (inducing such beams) become mutually coupled increases. Moreover, since higher number of such beams should be taken into account (compared to the case where Gaussian beams having smaller angular spreads are used) while calculating the diffraction field at a given observation point, the computational burden also increases. Thus, the width of the Gaussian window function should be chosen sufficiently small such that the derived formulations produce satisfactory results in field calculations, but the mutual couplings are still negligible and the computational complexity is still in tolerable limits. For some surfaces such a choice may not exist. The method proposed in Chapter 4 produces quite accurate results with a tolerable complexity for surfaces for which it is possible to find such a width for the Gaussian window function.

In Chapter 5, we present another local signal decomposition approach for the problem of diffraction field calculation from a field given on a curved surface. In this approach, a given 3D field is decomposed into a sum of 3D Gaussian beams. The problem is formulated as a system of linear equations where the columns

of the system matrix are the field patterns that the Gaussian beams produce on the curved surface. In the sense of problem formulation, the proposed method is similar to the method introduced in [53]. However, our proposed method provides considerable amount of reduction in computational complexity with respect to the methods that use global signal decompositions (such as the method introduced in [53] where infinite extent plane waves are used). This reduction is a consequence of the locality of the expansion functions that we use in our proposed method. The locality of the Gaussian beams results in sparse system matrices. Therefore, we take advantage of this sparsity while solving the linear system of equations and reduce the computation time.

In order to show the efficiency of the proposed method, the systems of linear equations, related to the proposed Gaussian beam decomposition and the plane wave decomposition [53], are solved by using the usual QR factorization method [86] and the multifrontal sparse QR factorization method [83]. For both 1D curved lines and 2D curved surfaces, these methods are applied several times to the system matrices having various sizes. It is shown that the sparsities of the system matrices related to the proposed method can be taken advantage of by a sparse solver in terms of both the computational complexity and the memory requirement. For example, for the curved surface shown in Fig. 5.12, the usual and sparse QR decomposition methods are applied several times for different 3D fields to the system matrices shown in Fig. 5.14 and Fig. 5.15. Even for this small size curved surface which is represented by 108×108 samples, the computational gain of the proposed method with respect to the plane wave decomposition method is reported to be about 10 times, when the sparse solver [83] is used. By using the sparse storage scheme provided by MATLAB [84], the reduction in the storage space that is required to store the sparse system matrices shown Fig. 5.14 and 5.16 are reported to be by factors of about 50 and 10, respectively. Even for the small scale examples that we deal, the achieved computational and memory gains are satisfactory. Noting that as the size of

the curved surface increases the corresponding system matrices for the proposed method become more sparse, we expect the related computational and memory gains to be much more for large sizes of curved surfaces. Indeed, as can be observed in Fig. 5.3, the computation time (and memory usage) in the process of sparse QR factorization is roughly $O(|\mathbf{G}|)$ where $|\mathbf{G}|$ denotes the total number of nonzero entries of the system matrix $\mathbf{G}$.

The multifrontal sparse QR decomposition method [83] performs the matrix factorization by applying a sequence of operations to the smaller submatrices which are called as frontal matrices. Therefore, such an approach provides even more improvements in the computational complexity, if the parallelism among the frontal matrices is exploited by parallel processing.

For the problem of diffraction field calculation from a field given on a curved surface, although the first approach that we introduce in Chapter 4 provides quite accurate solutions, it imposes some restrictions on the curved surface such as smoothness. On the other hand, the second method that we propose in Chapter 5 does not impose any restriction on the curved surface and still provides exact scalar field solutions. Nevertheless, there is a necessity of a matrix factorization in the second proposed method. This makes it computationally costly with respect to the first method, even if the matrices are sparse. Therefore, for curved surfaces which obey the restrictions imposed for the applicability of the first proposed method, the computational advantage of the first approach may be preferred to the accuracy of the second method.

As described in Sec. 2.4, the ray optics is related to the scalar wave optics by writing the light field representation as the magnitude squared of the windowed Fourier transform of the scalar wave field. Thus, the Gaussian beams that we use in Chapter 4 and Chapter 5 for the decomposition of scalar diffraction field can actually be treated as physically realizable light rays. This point of view links Chapter 2 to the subsequent chapters of the dissertation. The upper bound

regarding the joint spatial and angular resolution of physically realizable light field representations, that is derived in Sec. 2.4, is achieved by such Gaussian beams, since the corresponding Gaussian window functions are the set of functions which are optimally concentrated in both space and frequency. At this point it is worth noting that the choice of Gaussian beams in the proposed decompositions is particularly important, since the optimal concentration (in both space and frequency) property of the Gaussian signal makes it highly desirable for our aims in this dissertation.

The local signal decomposition methods that we introduce in Chapter 4 and Chapter 5 can be also applicable with other types of linear (and local) signal decompositions. Indeed, the generic approach presented in Chapter 3, provides a unified framework for the approach given in Chapter 4. In addition to the short time Fourier transform, the application of the wavelet decomposition is also explained in Chapter 3. The wavelet decomposition can be attractive for the cases where the diffraction field is desired to be calculated at various user-defined scales. In fact, for planar input surfaces the wavelet decomposition is used for such an aim in [72].

Bibliography

- [1] N. A. Valius, *Stereoscopy*. Focal P., 1966.
- [2] A. J. Woods, *Stereoscopic Displays and Virtual Reality Systems XII*. SPIE, 2005.
- [3] E. Ulusoy, *Signal Processing Based Solutions for Holographic Displays that Use Binary Spatial Light Modulators*. PhD thesis, Bilkent University, Turkey, 2012.
- [4] M. Halle, “Autostereoscopic displays and computer graphics,” *Computer Graphics, ACM SIGGRAPH*, vol. 31, no. 2, pp. 58–62, 1997.
- [5] G. Lippmann, “Epreuves reversibles donnant la sensation du relief,” *J. Phys.*, vol. 7, pp. 821–825, 1908.
- [6] S. A. Benton and V. M. Bove, *Holographic Imaging*. Wiley-Interscience, 2008.
- [7] G. K. Ackermann, J. Eichler, and J. Eichler, *Holography: A Practical Approach*. John Wiley and Sons, 2008.
- [8] J. W. Goodman, *Introduction to Fourier Optics*. McGraw-Hill, 2nd ed., 1996.
- [9] T. C. Poon, *Digital Holography and Three-Dimensional Display: Principles and Applications*. Springer, 2006.

- [10] G. Tricoles, “Computer-generated holograms: an historical review,” *Appl. Opt.*, vol. 26, no. 20, pp. 4351–4360, 1987.
- [11] W. J. Dallas, “Computer-generated holograms,” in *Digital Holography and Three-Dimensional Display* (T. C. Poon, ed.), pp. 1–49, Springer, 2006.
- [12] C. Slinger, C. Cameron, and M. Stanley, “Computer-generated holography as a generic display technology,” *Computer*, vol. 38, no. 8, pp. 46 – 53, 2005.
- [13] L. Onural, *3D Video Technologies: An Overview of Research Trends*. SPIE Press, 2011.
- [14] L. Onural, “Television in 3D: What are the prospects?,” *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1143–1145, 2007.
- [15] L. Onural, F. Yaraş, and H. Kang, “Digital holographic three-dimensional video displays,” *Proc. IEEE*, vol. 99, no. 4, pp. 576 –589, 2011.
- [16] N. Sato, H. Aritake, M. Kato, M. Ishimoto, and M. Nakashima, “Stereoscopic display apparatus.” US Patent. 5594559, Jan 1997.
- [17] J. Hahn, H. Kim, Y. Lim, G. Park, and B. Lee, “Wide viewing angle dynamic holographic stereogram with a curved array of spatial light modulators,” *Opt. Express*, vol. 16, no. 16, pp. 12372–12386, 2008.
- [18] J. Kasper and S. Feller, *The Complete Book of Holograms: How They Work and How to Make Them*. Dover Publ., 1985.
- [19] G. Saxby, *Practical Holography, Third Edition*. Taylor & Francis, 2010.
- [20] S. Johnston, *Holographic Visions: a History of New Science*. Oxford University Press, 2006.
- [21] B. R. Brown and A. W. Lohmann, “Complex spatial filtering with binary masks,” *Appl. Opt.*, vol. 5, no. 6, pp. 967–969, 1966.

- [22] L. Yaroslavsky, *Digital Holography and Digital Image Processing: Principles, Methods, Algorithms*. Kluwer Academic, 2004.
- [23] T. Yatagai, “Stereoscopic approach to 3-D display using computer-generated holograms,” *Appl. Opt.*, vol. 15, no. 11, pp. 2722–2729, 1976.
- [24] H. Kang, *Quality Improvements of the Coherent Holographic Stereogram for Natural 3D Display and Its Applications*. PhD thesis, Nihon University, Japan, 2008.
- [25] J. Waters, “Holographic image synthesis utilizing theoretical methods,” *Appl. Phys. Letters*, vol. 9, pp. 405–407, 1966.
- [26] K. Matsushima and M. Takai, “Recurrence formulas for fast creation of synthetic three-dimensional holograms,” *Appl. Optics*, vol. 39, no. 35, pp. 6587–6594, 2000.
- [27] M. Lucente, “Optimization of hologram computation for real-time display,” in *Practical Holography VI*, S. A. Benton, ed., *Proc. SPIE 1667*, pp. 32–43, 1992.
- [28] K. Matsushima, “Computer-generated holograms for three-dimensional surface objects with shade and texture,” *Appl. Optics*, vol. 44, no. 22, pp. 4607–4614, 2005.
- [29] M. Janda, I. Hanák, and L. Onural, “Hologram synthesis for photorealistic reconstruction,” *J. Opt. Soc. Am. A*, vol. 25, no. 12, pp. 3083–3096, 2008.
- [30] L. Ahrenberg, *Methods for transform, analysis and rendering of complete light representations*. PhD thesis, Max-Planck-Institut für Informatik, Germany, 2007.
- [31] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. Wiley Series in Pure and Applied Optics, John Wiley & Sons, 2007.

- [32] G. F. Wheeler and W. P. Crummett, “The vibrating string controversy,” *American Journal of Physics*, vol. 55, no. 1, pp. 33–37, 1987.
- [33] A. Sommerfeld, “Mathematische theorie der diffraction.,” *Mathematische Annalen*, vol. 47, pp. 317–374, 1896.
- [34] A. Sommerfeld, *Mathematical Theory of Diffraction*. Birkhäuser, 2004. Translated from German by R. J. Nagem.
- [35] E. Şahin and L. Onural, “Scalar diffraction field calculation from curved surfaces via Gaussian beam decomposition,” *J. Opt. Soc. Am. A*, vol. 29, no. 7, pp. 1459–1469, 2012.
- [36] G. B. Esmer, *Calculation of Scalar Optical Diffraction Field from Its Distributed Samples over the Space*. PhD thesis, Bilkent University, Turkey, 2010.
- [37] G. B. Esmer, L. Onural, and H. M. Ozaktas, “Exact diffraction calculation from fields specified over arbitrary curved surfaces,” *Opt. Communications*, vol. 284, no. 24, pp. 5537 – 5548, 2011.
- [38] M. Levoy and P. Hanrahan, “Light field rendering,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH ’96, (New York, NY, USA), pp. 31–42, ACM, 1996.
- [39] Z. Zhang and M. Levoy, “Wigner distributions and how they relate to the light field,” in *IEEE International Conference on Computational Photography (ICCP)*, 2009.
- [40] E. Şahin and L. Onural, “A comparative study of light field representation and integral imaging,” *Imaging Science Journal*, vol. 58, no. 1, pp. 28–31, 2010.

- [41] E. H. Adelson and J. R. Bergen, “The plenoptic function and the elements of early vision,” in *Computational Models of Visual Processing* (M. S. Landy and A. J. Movshon, eds.), pp. 3–20, Cambridge, MA: MIT Press, 1991.
- [42] A. Gershun, “The light field,” *Journal of Mathematics and Physics*, vol. 18, pp. 51–151, 1939.
- [43] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH ’96, (New York, NY, USA), pp. 43–54, ACM, 1996.
- [44] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, “Formation of real, orthoscopic integral images by smart pixel mapping,” *Opt. Express*, vol. 13, no. 23, pp. 9175–9180, 2005.
- [45] A. O. Yöntem and L. Onural, “Integral imaging using phase-only lcos spatial light modulators as Fresnel lenslet arrays,” *J. Opt. Soc. Am. A*, vol. 28, no. 11, pp. 2359–2375, 2011.
- [46] Y. Kim, J.-H. Park, S.-W. Min, S. Jung, H. Choi, and B. Lee, “Wide-viewing-angle integral three-dimensional imaging system by curving a screen and a lens array,” *Appl. Opt.*, vol. 44, no. 4, pp. 546–552, 2005.
- [47] J. Arai, F. Okano, H. Hoshino, and I. Yuyama, “Gradient-index lens-array method based on real-time integral photography for three-dimensional images,” *Appl. Opt.*, vol. 37, no. 11, pp. 2034–2045, 1998.
- [48] F. Hlawatsch and G. F. Boudreaux-Bartels, “Linear and quadratic time-frequency signal representations,” *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.
- [49] L. Cohen, “Time-frequency distributions-a review,” *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.

- [50] J. G. Daugman, “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters,” *J. Opt. Soc. Am. A*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [51] N. A. Ahuja and N. K. Bose, “Spatiotemporal-bandwidth product of m-dimensional signals,” *Signal Processing Letters, IEEE*, vol. 12, no. 2, pp. 123 – 125, 2005.
- [52] J. D. Gaskill, “The propagation and diffraction of optical wave fields,” in *Linear Systems, Fourier Transforms, and Optics*, Wiley, 1978.
- [53] G. B. Esmer, *Calculation of Scalar Optical Diffraction Field from Its Distributed Samples over the Space*. PhD thesis, Bilkent University, Turkey, 2010.
- [54] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. Wavelet Analysis and Its Applications, Academic Press, 1999.
- [55] O. Christensen, *Frames and Bases: An Introductory Course*. Applied and Numerical Harmonic Analysis, Birkhäuser, 2008.
- [56] O. Christensen, “Frames, Riesz bases, and discrete Gabor/wavelet expansions,” *Bull. Amer. Math. Soc*, pp. 273–291, 2001.
- [57] I. Daubechies, *Ten Lectures on Wavelets*. Cbms-Nsf Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992.
- [58] P. Flandrin, *Time-frequency/time-scale analysis*. Academic Press, 1999.
- [59] M. J. Bastiaans, “Samling theorem for the complex spectrogram, and Gabor’s expansion of a signal in Gaussian elementary signals.,” *Optical Engineering*, vol. 20, no. 4, pp. 594–598, 1981.

- [60] R. Balian, “Un principe d’incertitude fort en théorie du signal ou en mécanique quantique,” *C. R. Acad. Sci. Paris*, vol. 292, no. 20, pp. 1357–1362, 1981.
- [61] I. Daubechies, “The wavelet transform, time-frequency localization and signal analysis,” *Information Theory, IEEE Trans.*, vol. 36, no. 5, pp. 961–1005, 1990.
- [62] D. Gabor, “Theory of communication,” *J. Inst. Elec. Eng.*, vol. 93, no. 3, pp. 429–457, 1946.
- [63] M. J. Bastiaans, “Gabor’s signal expansion and the Zak transform,” *Appl. Opt.*, vol. 33, no. 23, pp. 5241–5255, 1994.
- [64] M. J. Bastiaans, “Expansion of an optical signal into a discrete set of Gaussian beams,” *Optik*, vol. 57, no. 1, pp. 95–102, 1980.
- [65] K. Gröchenig, “Irregular sampling of wavelet and short-time Fourier transforms,” *Constructive Approximation*, vol. 9, no. 2-3, pp. 283–297, 1993.
- [66] P. G. Casazza and O. Christensen, “Gabor frames over irregular lattices,” *Advances in Computational Mathematics*, vol. 18, no. 2-4, pp. 329–344, 2003.
- [67] W. Sun and X. Zhou, “Irregular wavelet/gabor frames,” *Applied and Computational Harmonic Analysis*, vol. 13, no. 1, pp. 63 – 76, 2002.
- [68] J. Antoine, *Two-Dimensional Wavelets and Their Relatives*. Cambridge University Press, 2004.
- [69] L. Onural, M. Kocatepe, and H. M. Ozaktas, “A class of wavelet kernels associated with wave propagation,” in *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, vol. iii, pp. III/9 –III/12, 1994.
- [70] T. S. Lee, “Image representation using 2D Gabor wavelets,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, 1996.

- [71] S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [72] M. Liebling, T. Blu, and M. Unser, “Fresnelets: new multiresolution wavelet bases for digital holography,” *IEEE Trans. Image Process*, vol. 12, no. 1, pp. 29–43, 2003.
- [73] S. Mallat, *A Wavelet Tour of Signal Processing*. Wavelet Analysis and Its Applications Series, Academic Press, 1999.
- [74] I. Daubechies, I. Guskov, P. Schröder, and W. Sweldens, “Wavelets on irregular point sets,” *Phil. Trans. R. Soc. Lond. A*, vol. 357, no. 1760, pp. 2397–2413, 1999.
- [75] W. Sweldens, “The lifting scheme: a construction of second generation wavelets,” *SIAM J. Math. Anal.*, vol. 29, no. 2, pp. 511–546, 1998.
- [76] K. Amaratunga and J. E. Castrillon-Candas, “Surface wavelets: a multiresolution signal processing tool for 3D computational modelling,” *International Journal for Numerical Methods in Engineering*, vol. 52, no. 3, pp. 239–271, 2001.
- [77] M. J. Bastiaans, “Oversampling in Gabor’s signal expansion by an integer factor,” in *Time-Frequency and Time-Scale Analysis, 1994., Proceedings of the IEEE-SP International Symposium on*, pp. 280–283, 1994.
- [78] A. Janssen, “Gabor representation of generalized functions,” *J. Math. Anal. Appl.*, vol. 83, no. 2, pp. 377–394, 1981.
- [79] K. Tang, *Mathematical Methods for Engineers and Scientists: Fourier analysis, partial differential equations and variational models*. Springer, 2007.
- [80] M. J. Bastiaans and A. J. van Leest, “Gabor’s signal expansion and the Gabor transform based on a non-orthogonal sampling geometry,” in *Signal*

- Processing and its Applications, Sixth International, Symposium on. 2001*, pp. 162–163, 2001.
- [81] E. Şahin and L. Onural, “Calculation of the scalar diffraction field from curved surfaces by decomposing the three-dimensional field into a sum of Gaussian beams,” *J. Opt. Soc. Am. A*, vol. 30, no. 3, pp. 527–536, 2013.
 - [82] I. S. Duff, “A survey of sparse matrix research,” *Proceedings of the IEEE*, vol. 65, no. 4, pp. 500 – 535, 1977.
 - [83] T. A. Davis, “Algorithm 915, SuiteSparseQR: Multifrontal multithreaded rank-revealing sparse qr factorization,” *ACM Trans. Math. Softw.*, vol. 38, no. 1, pp. 8:1–8:22, 2011.
 - [84] J. R. Gilbert, C. Moler, and R. Schreiber, “Sparse matrices in MATLAB: Design and implementation,” *SIAM J. Matrix Anal. Appl.*, vol. 13, no. 1, pp. 333–356, 1992.
 - [85] E. Ulusoy, L. Onural, and H. M. Ozaktas, “Synthesis of three-dimensional light fields with binary spatial light modulators,” *J. Opt. Soc. Am. A*, vol. 28, no. 6, pp. 1211–1223, 2011.
 - [86] G. Stewart, *Matrix Algorithms*. Society for Industrial and Applied Mathematics, 1998.