UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Optical Phase Space Measurements and Applications to 3D Imaging and Light Scattering

Permalink

https://escholarship.org/uc/item/5kk7w66s

Author

Liu, Hsiou-Yuan

Publication Date

2018

Peer reviewed|Thesis/dissertation

Optical Phase Space Measurements and Applications to 3D Imaging and Light Scattering

by

Hsiou-Yuan Liu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Laura Waller, Chair
Professor Austin Roorda
Professor Yi-Ren Ng

Spring 2018

Optical Phase Space Measurements and Applications to 3D Imaging and Light Scattering

Abstract

Optical Phase Space Measurements and Applications to 3D Imaging and Light Scattering

by

Hsiou-Yuan Liu

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Laura Waller, Chair

4D phase space descriptions of light capture wave-optical and angular information, enabling digitally refocusing, 3D reconstructions and aberration removal. The wave-optical theory includes diffraction and interference effects, making phase space applicable to scales near the wavelength of light (e.g. in microscopy); however, at the cost of making phase space functions more complicated than their ray optics counterparts (light fields). In this thesis, we aim at bridging the gap between the abstract high-dimensional phase space and actual experiments upon which the reconstruction of unknown objects relies. We achieve the aim by 1) providing practical methods of measuring phase-space functions with good resolution in all 4D and 2) developing phase-space theories that we use to computationally mitigate scattering in experimental situations. We extend phase-space measurement schemes from from lenslet arrays to a scanning-based coded aperture method in order to improve information throughput. Theory and experiment for designed coded apertures is proposed that can efficiently capture the entire 4D phase space. Next, we develop a phase-space theory for imaging through scattering and apply it to experimentally imaging point sources through scattering and tracking neural activity in a scattering environment, such as mouse brain tissue. The method utilizes the dimension mismatch between 3D object and 4D phase-space measurements, along with a sparsity prior, to ensure robustness and allow 3D localization of point sources relatively deep into scattering tissue. We develop theory and verify the mathematical phase-space scattering operator, then study how light interacts with scatterers and propose a fast wave-equation solver. This method uses an accelerated gradient descent solver and expands the solution to the wave equation as a series of the gradient solver updates. The method outperforms the first Born approximation and the Rytov approximation in predicting the scattered field as well as in reconstructing the scatterer distributions.

To people who supported me and to obstacles

Contents

List of Figures

List of Tables

Acknowledgments

During the past five years of graduate school, a great deal of my time and efforts have been put into making this thesis complete. Studying at Berkeley has enriched me intellectually and made me a more independently thinking person. But more importantly, living in Berkeley and interacting with people taught me the lessons of life that I found of great value. I am truly thankful for the people surrounding me, who have been kind and supportive to me, who have instructed me, and who have posed emotional or professional challenges to me. Without them I would not be who I am.

The first person I want to thank is Laura for her great advice. She has given me a great freedom to work on my research and has been very supportive in that regard as well as when I encounter other difficulties. She knew my affinity for math and theory, helped me search my research topics and gave me useful and practical advice. Most of our discussions, if not all of them, were very fruitful. Many practical applications of my idea have been pointed out and guided by her to make them finally into this thesis. Additionally, the flexibility on working schedule and her being easy to approach gave me the support I needed when some difficulties hit. It has been a really wonderful journey to work with Laura. And also I want to give a special thank to my hooder, the professor "Laura Waller" who put a wig on and perhaps did some extreme reading on MRI papers during the hooding ceremony. The vibe was always good when he was present.

I want to thank Lei Tian and Jingshan Zhong who worked with Waller lab as postdocs. They presented road models for being a good researcher and were always willing to give out their support and comments on research. I was so glad, and lucky, that they had established a good co-working space when I joined the lab. Also thanks to my fellow labmates, Li-Hao, Michael Chen, Michael Kellman, Regina, Grace, Nick, Nico, Zach, David, Gautam, Fanglin, Stuart, Kristina, Kyrollos, Nicole, Henry Pinkard, Henry Wang, Emrah, Shwetadwip, Dan and Aamod, without them this work would have been more difficult and even impossible to accomplish. I really enjoyed the discussion we had and hanging out after work.

Outside Ph.D. work, I want to thank my friends, especially those in the Bay Area. We had laughable, tearful, exciting and heartbreaking moments together. You enrich my life and have made me a better, happier and stronger person. Without you, I would not have understood my inner self better. That, in my opinion, is far more important than getting the Ph.D. degree. My sincere thank to all of you, the people I have met.

With all, I want to thank my parents, my sister and my family. You let me know that there is always a harbor to return, to prepare myself for challenging voyages of life. You teach me what an unreserved support is. I will always be grateful for your being in my life.

一路上有你們，真好！

Chapter 1

Introduction

Light is arguably the most useful way for human to perceive this world and to transmit information. Each moment of the world can be captured by light thanks to its tremendous speed and its straight-line propagation. For example, lighthouse keepers use light to inform far ships that the shore is within their reach. The straight-line propagation is of great importance as it is the primitive way for humans to obtain information from light. With the image captured by our eyes and the straight-line propagation nature, we can trace the image back to where the light comes from. Our eyes play an important role as the intensity and direction detectors of light. However, when the situation becomes complicated, the information obtained from our bare eyes may not be sufficient. As in the lighthouse example, if the shore is filled with fog, a light ray originating from the lighthouse does not travel in a straight line but rather is scattered in multiple directions. Some prior works aimed at filtering out the unscattered components in the measurement [185, 58], discarding a significant portion of the light. On the contrary, we want to utilize the scattered light to enhance the light signal for the inverse process. When scattering happens, we observe that a light ray is smeared into multiple directions around the original direction. Hence the knowledge of directional distributions of light sheds light on undoing the scattering.

To address imaging through scattering has led us to study the *phase space* of light. Phase space is a broad term in physics which describes the states of a physics system. The dimension of phase space can be higher than 3D, making it abstract. In optics, a similar entity is called a 4D light field, which describes the distribution of light rays over position and direction coordinates. Here we focus on the *Wigner function* [184] of stationary and quasi-monochromatic light as our phase space function. It incorporates the wave optical effects and completely describes stationary light outside quantum mechanics. This completeness is useful, for example, in analyzing the information captured by an imaging system [173]. Since coherence functions of light also describe the statistical property of light completely, the link between the coherence function and Wigner function can be established [173]. We will show that this connection helps the phase space analysis. One remark about the wave-optical phase space is that it unifies the treatment of coherent and partially coherent beams. For example, the propagation is a shearing operation in the phase space regardless of the

coherence property of the light. In this thesis, we study various aspects of the phase space of light, from its measurements to applications to 3D imaging theory and imaging in scattering environments. We aim at bridging the gap between the abstract high-dimensional phase space and actual experiments which the reconstruction of objects relies upon. We achieve our goal by 1) providing several means of measuring Wigner functions to make it easy to access and 2) developing phase space theories for experiments to demonstrate its usefulness.

One of our main contributions is to measure the phase space with coded apertures, along with an efficient code design that is able to fully reconstruct the phase space. The acquisition of the abstract phase space requires more than a direct imaging system. Previously, inserting a physical component, a lenslet array, can enable an imaging system to measure the phase space [176]. This method is, however, limited by the pixel counts of a 2D sensor. Spreading a 4D quantity onto a 2D sensor with ~ 1 Megapixels results in unacceptably low resolution. In order to release this limitation for sampling a high-dimensional phase space, we turn to computational imaging methods. A general introduction to them is given in the following section. In addition, due to wave optical interference, directly extending the lenlet array measurement to capture N images of N angular directions would not capture the full wave-optical phase space. Utilizing the equivalence between phase space and coherence functions, one naive way to capture the wave-optical phase space is to measure pairwise coherence, i.e. mutual intensity, of all pairs of points in either real space or angular space, resulting in $O(N^2)$ measurements. We refine this method by using coded aperture measurements. Our coded aperture approach, which has $O(N)$ codes, is able to both capture the wave effects and retain low measurement complexity. The phase space captured by our efficient method will help researchers study light by performing physics operations in a digital computer.

Our other contribution is to combine both the theory and the measurements of the phase space into imaging point light sources through scattering. Other than the measurements, theories built upon phase space provide insights to the imaging system and to the light from the objects. For example, The information captured by an imaging system can be visualized in the phase space theory. Using phase space to describe light enables us to formulate our intuition of volumetric scattering as multiple Gaussian blur operators on angular coordinates. With the model developed from this operator, the measured scattered light becomes a signal instead of unwanted noise. Our measurements contain the angular information of the light, which, together with the theory, enhances imaging through scattering. We perform two inverse scattering experiments to demonstrate our goal. Moreover, materializing an idea of physics as a phase-space operator, together with phase-space analysis, is a powerful tool to help researchers design and examine an optical system. The method mentioned above will serve as one such example. This approach for inverse scattering has been applied in [146] and we expect it to be broadly used to solve other problems in the future.

To understand the scattering to a greater degree, we further study how light scatters/interacts with objects. The complicated interaction is present only if there is a nonuniform distribution of refractive index. Therefore the wave scattering theories for nonuniform media, such as Born and Rytov approximations, are reviewed, and we propose a method that is able to compute the scattered electric field given the distribution and to reconstruct
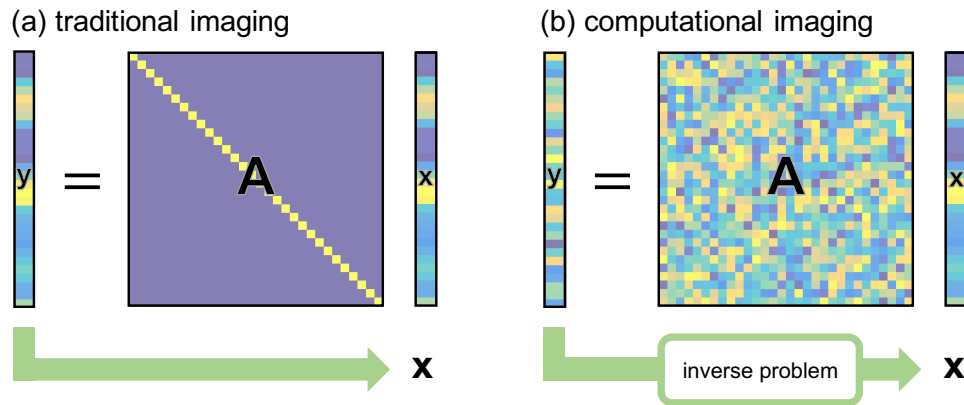
Figure 1.1: An example linear measurement for comparison between direct and computational imaging. The sensing matrix **A** converts the object we want to observe to an camera-captured image or images. In direct imaging, the images are close to the object in appearance while there is an inverse problem to solve in computational imaging. The matrix **A** for computational imaging mixes the object information from different pixels but needs not be fully populated.

the distribution from measured scattered fields.

1.1 Computational Imaging

Computational imaging (CI) is a methodology to reconstruct *labeled*, or high dimensional, images through a jointly designed measurement device and its associated computational reconstruction. The images here are defined as a 2D spatial distribution of physical quantities of interest so some measurements are not considered CI, for example measuring the temperature at a point with a digital thermometer, regardless of how much computation is involved in the process. The labeling is redundant if only a single image is reconstructed, but through labels such as a depth index or angular indexes, 3D imaging and light field imaging are included in this definition. To explain how CI works and how it distinguishes itself from traditional imaging, we show an example linear imaging scheme in Fig. 1.1.

There are many reasons to apply CI [129]. For example, it can be applied in order to satisfy constraints of applications, to reveal more information in 3D, to make it easier to perceive information than direct imaging, or simply because there is no direct method to capture an abstract physics quantity. DiffuserCam [5], which uses a diffuser instead of lenses to "focus" the light onto the camera, is an example for satisfying the constraints of an application. It release the length constraint for image formation under conventional lenses and collects unrecognizable and seemingly informationless images. The collected data then require an extra algorithm that has the knowledge of the diffuser to reconstruct the human-recognizable image. Another example is Fourier ptychography [195, 175], which is

a fast, high information-throughput CI method. Contrary to high-resolution field-scanning microscopy, Fourier ptychography scans illumination angles and uses a low resolution objective with a large field of view to collect the light from an (absorptive) phase object. It speeds up the measurement because the scanning happens in illumination electronics instead of mechanically scanning the object. On the other hand, it still retains the resolution in the reconstruction as the high-resolution information is contained in multiple low-resolution images from the multiple illuminations. The reconstruction algorithm, illumination pattern and the optical system are jointly designed to carry out Fourier ptychography. There are many other examples of applying CI such as compressed sensing magnetic resonance imaging [125] where shorter measurement time is desirable, computed tomography scan where multiple X-ray projections are used to form a 3D image [81], *etc.*

CI requires its designers to formulate forward problems for their cases, *i.e.* the relations between the measurements and the quantities of interest, in order for the follow-up computation to reconstruct. As shown in the examples, the computation is as important as the measurement because of the need to process the raw data. The forward problem formulation makes CI flexible such that it can be adapted for physical quantities measurable only indirectly, *e.g.* phase retrieval [62, 61]. Some metrics [6] may be used to judge the forward model for whether the measured data contain enough information for computation to reconstruct the unknown. The designers then solve a mathematical inverse problem to obtain the quantities of interest once the measurements are done. However, the inverse problem is not always easy to solve. Depending on the math of the forward model, the inverse problem, cast as an optimization problem, can have multiple local optima that are not close to the unknown ground truth [175, 191]. Fortunately, the inverse problem solving is able to incorporate prior knowledge of the measurements or unknowns, *e.g.* noise statistics or sparsity of the unknown, to help the solving. The prior knowledge can enhance CI methods to tackle the ill-posed inverse problem or some tough imaging conditions, such as using the Total Variation regularization for image formation [156] or Compressed Sensing with sparsity prior for limited measurement time [36, 125]. The above properties make CI an appealing approach to many problems and it will be used through out this thesis work.

1.2 Dissertation Overview

We combine our experimental and theoretical studies of phase space to image through scattering. By doing so, we show how the framework of phase space helps in a challenging imaging condition, hoping to bridge the gap between abstract phase space and actual experiments. In addition, scattering phenomenon is further studied outside of the scope of phase space. The rest of the thesis is organized as follows.

- Chapter 2 provides the necessary mathematical background for phase space and the definitions of notation and functions used throughout the thesis.

- Chapter 3 discusses a scanning method of measuring phase space. It is a direct extension of lenslet array light-field measurements and it samples the phase space much more densely than the former. The sample counts here are not limited by the pixels on a sensor but rather by the number of those pixels multiplied with the number of scanning positions.

- Chapter 4 discusses a multiplexed phase-space measurement and its application to 3D incoherent sources. We show that the random multiplex codes can achieve diffraction limited resolution of the 3D incoherent sources while avoiding reconstructing the tremendous 4D phase space. The multiplexing scheme incorporates the compressed sensing smoothly and can be further utilized to work with priors other than sparsity.

- Chapter 5 describes a coded aperture method to measure the phase space of samples of arbitrary coherence. We address here the code design problem opened by the random design in previous chapter. The mutual intensity is introduced in detail and heavily used for analyzing the code design as it graphically reveals the information of a measurement. Also, the mutual intensity is equivalent to the Wigner function. Hence, by capturing the mutual intensity, we show that the system and the designed codes are able to reconstruct the desired phase space.

- Chapter 6 applies the phase space theory to analyze scattering. The phase space itself is also measured and shown to be resistant to the scattering-induced blurring. The phase space theory is useful for formulating our intuition about scattering into mathematical formulae. The reconstruction of point sources and neural activity, both under volumetric scattering, have been experimentally performed successfully.

- Chapter 7 moves the study of scattering from the phase space of light to how the light interacts with scatterers. We propose an algorithm that expands the scattered field as a series of optimization solver update steps. This algorithm is also capable of providing the gradient update for estimating the scatterers if the scatterer is the unknown and the scattered field is measured. Experimental verification as well as comparisons to the analytical forms of some cases are provided. A total variation regularizer is used in the inverse problem solving for the scatterers, which enables substantial measurement data reduction.

Chapter 2

Background of Phase Space

In this chapter, we will review the concept of phase space briefly and introduce the mathematics and convention that we use to describe the phase space of light. As the discussion unrolls, we need to distinguish 1) coordinate vectors in physics and 2) matrices and vectors used in inverse problems. We use an italic bold font for coordinate vectors, such as $\boldsymbol{x}$ for the spatial dimensions, *e.g.* coordinates (x, y, z) for 3D cases and (x, y) for 2D cases, $\boldsymbol{r}$ for 2D transverse coordinates (x, y), and $\boldsymbol{u}$ for 2D spatial frequency coordinates (u_x, u_y). The matrices and vectors for computation are denoted in bold fonts. Some examples are: $\mathbf{A}$, $\mathbf{G}$, $\mathbf{I}$ for matrices and $\mathbf{x}$, $\mathbf{f}$, $\mathbf{u}$ for vectors.

2.1   Electric Field Convention

Throughout this thesis, we deal with a scalar electric field in phasor notation.[1] We assume the light to be statistically stationary and quasi-monochromatic, and when multiple frequencies exist, we consider the center frequency for an average behavior or repeat the outcomes of discussions in this work for each wavelength. The electric field is a function of 3D space $\boldsymbol{x}$ and time t

$$\Re\{E(\boldsymbol{x})e^{-\mathrm{j}2\pi ft}\} \tag{2.1}$$

where $\Re$ is taking the real part of the expression, f is the temporal frequency of the field, j is the imaginary unit, and $E(\boldsymbol{x})$ is the complex-valued phasor electric field (in Chapter 7, $u(\boldsymbol{x})$ is used in place of $E(\boldsymbol{x})$). Since the phasor is a complex quantity, the complex conjugate of a phasor E is denoted by symbol E^*. The symbol λ is used to denote the wavelength of the field in vacuum and equals to speed of light in vacuum divided by f. For simplicity, the time dependence is dropped in further discussion.

---

[1]Static electric fields are not discussed here.
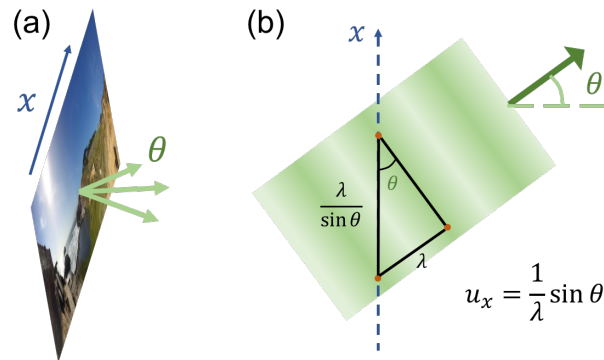
Figure 2.1: Propagation angle is related to the spatial frequency of monochromatic light that is locally approximated as a plane wave. (a) Ray optics models light as nondiverging beams, each of which is parametrized by a propagation direction and a point it passes through. In wave optics, light is modeled as an electric field which does not possess those parameters. (b) If we locally approximate an electric field as a plane wave and ignore its diffraction, we can define the propagation wavenumber vector as the propagation angles. Three points on the oscillation peaks of the approximated plane wave are marked, where the distance is λ along the propagation direction and $(\lambda/\sin\theta)$ along the transverse direction. The spatial frequency is the inverse of the period, which is $(\sin\theta/\lambda)$ along the transverse direction, and is approximately proportional to θ when θ is small.

2.2 From Phase Space in Physics to That of Light

In physics, phase space describes all possible states of a dynamic system. Consider the dynamics of a ball. It can be described by the ball's position and momentum in 3D space (ignoring rotation). A ball with a particular momentum and position is represented in the 6D phase space as a delta function. In addition, multiple balls are represented as a superposition of the phase space functions of individual balls. Integrating over this 6D space will give the number of delta functions, that is, the number of balls. Furthermore, if we consider each ball as a density distribution, each delta function will become a distribution with integrated value equal 1, indicating the "conservation" of the number of balls.

The phase space of light we refer to is a 4D function that describes the spatial and angular distributions of light [8, 173]. Like the ball example above, a light ray under ray optics can be described by 3 positional coordinates and 3 directional coordinates. However, when only forward propagating light is present, *e.g.* propagating in free space, the phase-space description can be reduced to 4D with two transverse spatial coordinates and two angular coordinates [173] because the 3D electric field (and intensity) can be determined uniquely from the 2D electric field at any plane. Examples of this are radiance in radiometry [184] and light fields which describe both position and angle of each ray. The measurement of phase-space enables digital refocusing, aberration removal and 3D reconstruction [140, 116]. To include phase and diffraction effects, a more general wave-optics descriptor of phase space

(a) coherent light $\qquad$ (b) partially coherent light
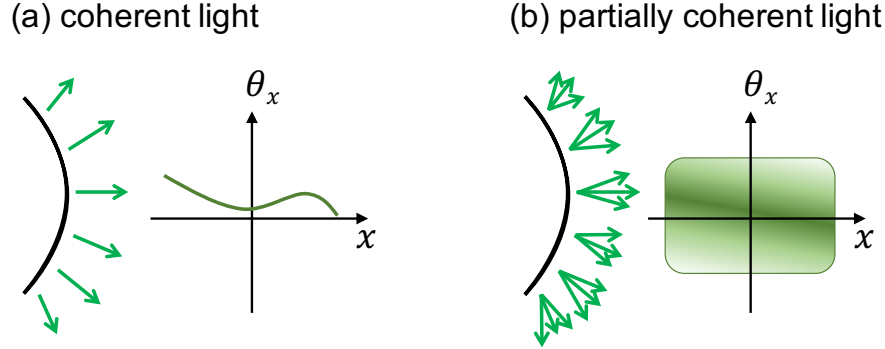
Figure 2.2: Phase space illustrations for coherent and partially coherent light. (a) A coherent light beam and its phase-space representation. (b) A partially coherent light beam, which fills up more of phase space.

(*e.g.* Wigner functions [184]) must be used and propagation angles are generalized to spatial frequencies [194, 4] (see Fig. 2.1). This function can further describe coherent and partially coherent light [130, 188, 173] in an intuitive way, since the local correlation length is inversely proportional to the width of the local spatial spectrum. One can think of partial coherence as the wavefront containing multiple directions at each point in space; the more directions, the less coherence. Hence a coherent beam usually has a narrow band in phase space while a partially coherent beam possesses a larger area (see Fig. 2.2).

2.3 The Wigner Function and Connections of Phase-Space Functions

We choose the *Wigner function* [184] as the phase-space description to use in this work because of its wave-optical nature, ability to describe coherent and partially coherent beams in a unified fashion, and its intuitive framework. We will also use its mathematical equivalent functions for some application. The 4D Wigner function $W(\boldsymbol{r}, \boldsymbol{u})$ has two transverse spatial coordinates and two spatial frequency coordinates which relate to propagation angles (θ_x, θ_y) by $(\sin\theta_x, \sin\theta_y) = \lambda\boldsymbol{u}/n_r$ (n_r is the refractive index of background medium). It is defined as [8, 173]

$$W(\boldsymbol{r}, \boldsymbol{u}) \triangleq \iint \left\langle \tilde{E}^* \left(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2} \right) \tilde{E} \left(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2} \right) \right\rangle e^{\mathrm{j}2\pi\boldsymbol{r}\cdot\Delta\boldsymbol{u}} \, \mathrm{d}^2(\Delta\boldsymbol{u}) \qquad (2.2)$$

$$= \iint \left\langle E^* \left(\boldsymbol{r} - \frac{\Delta\boldsymbol{r}}{2} \right) E \left(\boldsymbol{r} + \frac{\Delta\boldsymbol{r}}{2} \right) \right\rangle e^{-\mathrm{j}2\pi\boldsymbol{u}\cdot\Delta\boldsymbol{r}} \, \mathrm{d}^2(\Delta\boldsymbol{r}), \qquad (2.3)$$

where E is a partially coherent, stationary quasi-monochromatic electric field, the tilde $\tilde{\ }$ denotes Fourier transform, and $\langle\cdot\rangle$ denotes the ensemble average for multiple coherence modes if they exist. The ensemble average allows representation of both coherent and partially (spa-
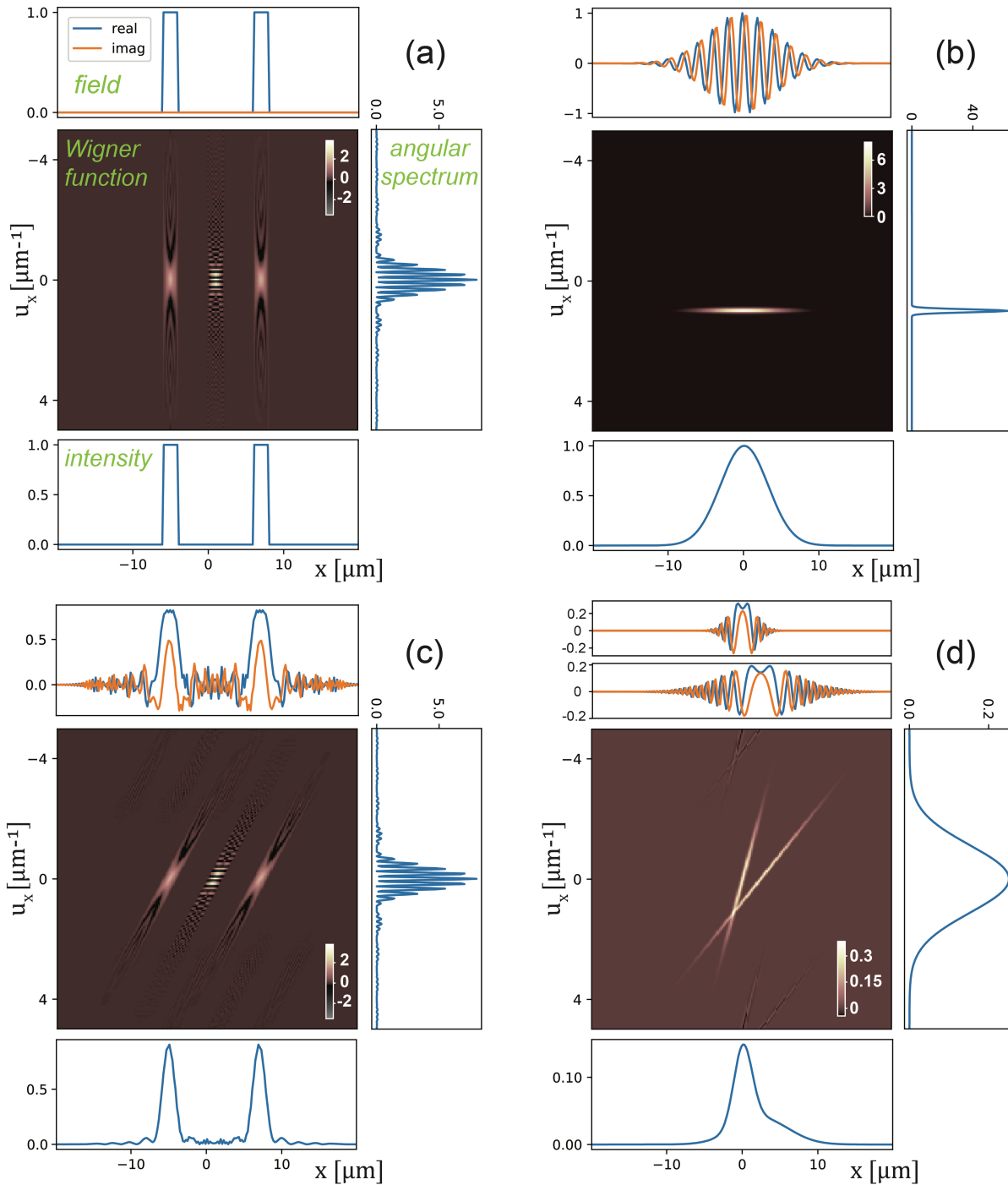
Figure 2.3: Examples of Wigner functions and their projections (Intensity and angular spectrum) for some sample 1D electric field distributions. (a) Plane-wave illuminated double slits. (b) An oblique Gaussian beam. (c) The double slits with a defocus. (d) Two incoherent point sources. The top plot of each subfigure is the electric field, the central square part is the Wigner function, the bottom plot is the intensity, and the right plot is the angular spectrum.

tially) coherent light. It is straightforward to prove that Eqs. (2.2) and (2.3) are equivalent by substituting

$$\tilde{E}(\boldsymbol{u}') = \iint E(\boldsymbol{r}')e^{-\mathrm{j}2\pi\boldsymbol{r}'\cdot\boldsymbol{u}'}\ \mathrm{d}^2\boldsymbol{r}',$$

and a change of variables. Derivation details are provided in Appendix A. A *real-space* image, *i.e.* a conventional 2D intensity image, is a projection of the Wigner function over all spatial frequencies,

$$\text{Intensity}(\boldsymbol{r}) = \langle E^*(\boldsymbol{r})E(\boldsymbol{r})\rangle = \iint W(\boldsymbol{r},\boldsymbol{u})\ \mathrm{d}^2\boldsymbol{u}, \tag{2.4}$$

and an angular spectrum is given by projecting along spatial coordinates

$$\text{Spectrum}(\boldsymbol{u}) = \langle \tilde{E}^*(\boldsymbol{u})\tilde{E}(\boldsymbol{u})\rangle = \iint W(\boldsymbol{r},\boldsymbol{u})\ \mathrm{d}^2\boldsymbol{r}. \tag{2.5}$$

See Fig. 2.3 for examples of Wigner functions and their corresponding intensity images and angular spectra.

The Wigner function handles the interference but still retains the geometric optics picture. The interference is manifested through negative values of the Wigner function [54]. From the definition Eq. (2.2), the complex conjugate of a Wigner function equals to itself so it is real-valued. However, it is not guaranteed to be positive definite and is considered a quasi-probability (quasi-intensity) distribution [85]. Dragoman [54] domonstrated that the negative values are a result of interference and diffraction. On the other hand, in macroscopic scale as the wavelength of the light becomes negligible relative to the size scales of interest, a Wigner function is locally averaged and as a result, those negative values are washed out and the Wigner function becomes a light field [194]. We further show in Sec. 2.4 that the shearing operation of digital refocusing in the light field [140] is also valid for Wigner functions.

In addition to describing the distribution of light rays, the Wigner function has a strong connection to coherence. As mentioned earlier, the Wigner function describes the local coherence in the spatial frequency (angular) content of the location. A more rigorous connection is seen in its definition, Eqs. (2.2) and (2.3). The quantities with the angled brackets are coordinate-transformed *Mutual Intensities* (a.k.a. Mutual-Coherence functions):

$$\tilde{\Gamma}(\Delta\boldsymbol{u},\boldsymbol{u}) \triangleq \left\langle \tilde{E}^*\left(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2}\right)\tilde{E}\left(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2}\right)\right\rangle \tag{2.6}$$

$$\Gamma(\boldsymbol{r},\Delta\boldsymbol{r}) \triangleq \left\langle E^*\left(\boldsymbol{r} - \frac{\Delta\boldsymbol{r}}{2}\right)E\left(\boldsymbol{r} + \frac{\Delta\boldsymbol{r}}{2}\right)\right\rangle \tag{2.7}$$

They record the coherence of each pair of points in the coordinates and are equivalent to the Wigner function since all of them are connected by Fourier transforms [173]. In this work, we do not study the coherence in depth but rather use these functions to analyze the phase space. Figure 2.4 summarizes the connections of Wigner functions to other phase space functions. The ambiguity function shown here is for a complete diagram and is out of the scope of this thesis. Interested readers can refer to [30] and [173] for more information about ambiguity functions.
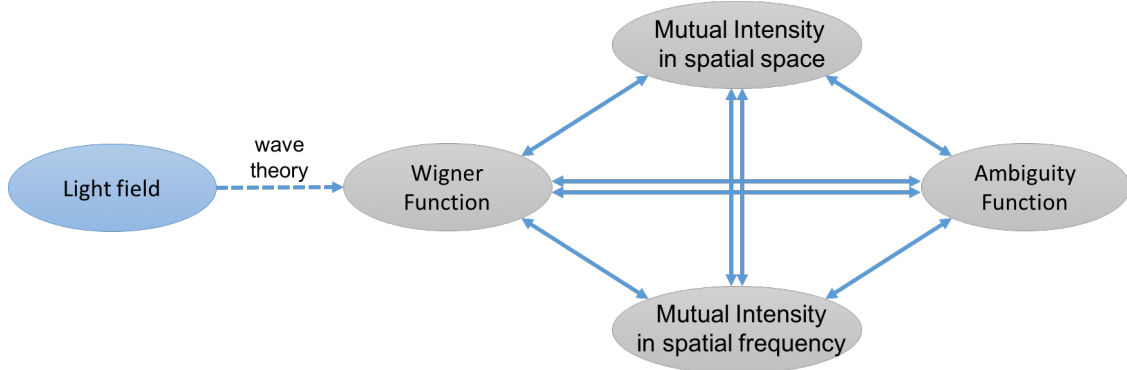
Figure 2.4: Connections of phase space functions. The light field is connected to the Wigner function by extending from geometrical optics to wave optics, indicated by the dashed arrow. A solid segment indicates a Fourier transform on one of the coordinates of the phase space functions while a double segment means two Fourier transforms on both of the coordinates. The two mutual-intensity functions are connected to the Wigner function through Fourier transforming one of its coordinates while the ambiguity function is connected through Fourier transforming both.

2.4 Propagation in Phase Space

One of the most appealing reasons for using the phase-space framework is that propagation becomes a geometric shearing operation, independent of the coherence properties of the light [173, 8]. Writing the propagation operator as $\mathcal{P}_{\Delta z}$, we describe light propagation of distance Δz using the Wigner function:

$$\mathcal{P}_{\Delta z}W(\boldsymbol{r},\boldsymbol{u}) = \iint W(\boldsymbol{r}',\boldsymbol{u})\delta(\boldsymbol{r}-\boldsymbol{r}'-\lambda\Delta z\boldsymbol{u}/n_r)\,\mathrm{d}^2\boldsymbol{r}', \tag{2.8}$$

where n_r is the refractive index of the material and δ is the Dirac delta function. Its proof is provided in Appendix A. This is analogous to the shift-and-add approach for digital refocusing of light-field data [140] but the Wigner function further accounts for wave-optical effects.

Moreover, describing a physical event (here it is propagation) as an operator inspires us to devise another phase-space operator for light interacting with scattering medium, which will be discussed in Chapter 6.

2.5 Equations for Phase-Space Measurements

Our work focuses on applying codes to the Fourier plane of the unknown electric field, it makes the mathematical derivations in following chapters short to introduce some common equations in this background chapter. The measured quantity is the intensity in $\boldsymbol{r}$ and the

phase space function we want to connect to the measurement is the Fourier-space mutual intensity $\tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u})$, which can be converted to the Wigner function when it is suitable.

We mainly use a $4f$ system to access the Fourier plane on which a code is applied. The code is represented as $M(\boldsymbol{r}_{\mathrm{m}})$ where $\boldsymbol{r}_{\mathrm{m}}$ is the spatial coordinate of the code in the experimental setup. The function M may have extra indexes for indexing the experimental code sequence. In addition, the coordinates $\boldsymbol{r}_{\mathrm{m}}$ is related to the frequency coordinate $\boldsymbol{u}$ of the electric field by $\boldsymbol{r}_{\mathrm{m}} = \lambda f \boldsymbol{u}$ where f is the front (between the object and the Fourier plane) focal length of the $4f$ system [72].

A real-space image I measured with a Fourier-space code M is represented as the following,

$$I(\boldsymbol{r}) = \left\langle \left| \iint \tilde{E}(\boldsymbol{u}')M(\lambda f\boldsymbol{u}')e^{\mathrm{j}2\pi\boldsymbol{u}'\cdot\boldsymbol{r}} \; \mathrm{d}^2\boldsymbol{u}' \right|^2 \right\rangle \tag{2.9}$$

$$= \left\langle \iint \tilde{E}(\boldsymbol{u}')M(\lambda f\boldsymbol{u}')e^{\mathrm{j}2\pi\boldsymbol{u}'\cdot\boldsymbol{r}} \; \mathrm{d}^2\boldsymbol{u}' \iint \tilde{E}^*(\boldsymbol{u}'')M^*(\lambda f\boldsymbol{u}'')e^{-\mathrm{j}2\pi\boldsymbol{u}''\cdot\boldsymbol{r}} \; \mathrm{d}^2\boldsymbol{u}'' \right\rangle \tag{2.10}$$

$$= \iiiint \left\langle \tilde{E}(\boldsymbol{u}')\tilde{E}^*(\boldsymbol{u}'') \right\rangle M(\lambda f\boldsymbol{u}')M^*(\lambda f\boldsymbol{u}'')e^{\mathrm{j}2\pi(\boldsymbol{u}'-\boldsymbol{u}'')\cdot\boldsymbol{r}} \; \mathrm{d}^2\boldsymbol{u}'\mathrm{d}^2\boldsymbol{u}'' \tag{2.11}$$

$$= \iiiint \tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u})M\left(\lambda f(\boldsymbol{u}+\frac{\Delta\boldsymbol{u}}{2})\right) M^*\left(\lambda f(\boldsymbol{u}-\frac{\Delta\boldsymbol{u}}{2})\right) e^{\mathrm{j}2\pi\Delta\boldsymbol{u}\cdot\boldsymbol{r}} \; \mathrm{d}^2\boldsymbol{u}\mathrm{d}^2(\Delta\boldsymbol{u}). \tag{2.12}$$

Here $\boldsymbol{u} = (\boldsymbol{u}' + \boldsymbol{u}'')/2$ and $\Delta\boldsymbol{u} = \boldsymbol{u}' - \boldsymbol{u}''$. Eqs. (2.9) to (2.12) will serve as the basis for the derivation in the following chapters.

Chapter 3

Fourier Spectrogram: Densely-Sampled Phase Space

A light field describes phase space in the geometrical optics regime where light is considered as rays. A popular way of measuring light fields is to use lenslet arrays [139], which capture both spatial and angular information about light rays in a single shot. Each of the lenlets converges local light to different point according to the light's direction. Those converged points, together with the locations of the lenslets, defines the 4D light-field coordinates. The intensity measured at those converged points are the light-field data. Though it is sufficient to use the light field information to compose a photograph, it suffers from low pixel counts and diffraction at the micrometer scale, which are important for microscopy applications. Hence we turn to a multishot method to improve the number of data counts and to include the wave-optical effect into our formulation. The measured quantity in this work is the Fourier Spectrogram, the Wigner function of light convolved with the 4D kernel of an aperture in spatial frequency space.

In previous work [183], an optical spectrogram system was developed for high-resolution phase-space imaging, by applying a windowing aperture in the x space and capturing u_x-space intensity images that correspond to the local spatial spectra information. A spatial light modulator (SLM) was programmed to act as a reconfigurable aperture in x space, blocking all light except a small local region. This aperture was then scanned across x space while capturing u_x-space intensity sequentially. The result is not the WDF of the beam, but rather a spectrogram, which is a 4D convolution of the beam's WDF with the aperture's WDF [7]. The choice of window function shape and size is important [29], as one must trade off spatial resolution for spatial frequency resolution. This measurement scheme can be compared to previous work in measuring coherence functions through interferometric or tomographic means [131, 65, 29, 153]. When the incoming light has a strong DC or other spatial frequency component, the spectrogram suffers from a dynamic range issue. Hence, we propose to switch x and u_x spaces, applying apertures in u_x space while capturing intensity images in x space. The resulting sequence is a Fourier spectrogram.

Both the spectrogram and the Fourier spectrogram systems are able to achieve 1000^4 pixel

data sets, without the requirement of inverting such a large matrix. However, the scanning of the aperture and relative low light efficiency required long acquisition times. Indeed, the capture time for a full 4D data with high resolution (pixel counts) in all four dimensions will be formidably time consuming, since it must be done with a 2D sensor and the final result will involve ~ 1 Terabyte of information (n^4, where $n = 1000$ is the number of pixels in a single dimension). Therefore in the proposed setup, we speed up the measurement through synchronization of hardware and the use of a fast digital micromirror device.

3.1 Fourier Spectrograms

The Fourier spectrogram, measured by putting a scanning aperture in the Fourier space of the light, is not directly the Wigner function of the light but its smoothed version [194]. The reason to measure the Fourier spectrogram is that it is relatively simple to measure and also more practical than measuring the Wigner function since the Wigner function can contain negative values while the Fourier spectrogram cannot. Nevertheless, the Fourier spectrogram, like the Wigner function, is a 4D function where the position of the scanning aperture gives the frequency coordinates. We will show that, similar to the spectrogram [7], the smoothing is done by a 4D convolution of the Wigner function of light with the 4D kernel of the aperture.

We let the aperture be square. Starting with Eqs. (2.9) to (2.12), the shifting aperture introduces a parameter $\boldsymbol{u}_{\mathrm{sh}}$ for shifting to the function M, resulting in

$$M(\lambda f \boldsymbol{u}; \boldsymbol{u}_{\mathrm{sh}}) = \mathrm{rect}\left(\frac{\lambda f(\boldsymbol{u} - \boldsymbol{u}_{\mathrm{sh}})}{a}\right), \tag{3.1}$$

where a is the side length of the square aperture and $\boldsymbol{u}_{\mathrm{sh}}$ is the shift vector in the spatial frequency. Hence the measured intensity corresponding to $\boldsymbol{u}_{\mathrm{sh}}$ is represented as

$$I(\boldsymbol{r}, \boldsymbol{u}_{\mathrm{sh}}) = \left\langle \left| \iint \tilde{E}(\boldsymbol{u}') \mathrm{rect}\left(\frac{\lambda f(\boldsymbol{u}' - \boldsymbol{u}_{\mathrm{sh}})}{a}\right) e^{\mathrm{j}2\pi\boldsymbol{u}'\cdot\boldsymbol{r}} \,\mathrm{d}^2\boldsymbol{u}' \right|^2 \right\rangle. \tag{3.2}$$

Following Eq. (2.12), we have

$$I(\boldsymbol{r}, \boldsymbol{u}_{\mathrm{sh}}) = \iiiint \tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u}) \mathrm{rect}\left(\frac{\lambda f(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2} - \boldsymbol{u}_{\mathrm{sh}})}{a}\right) \mathrm{rect}^*\left(\frac{\lambda f(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2} - \boldsymbol{u}_{\mathrm{sh}})}{a}\right) \cdot$$
$$e^{\mathrm{j}2\pi\Delta\boldsymbol{u}\cdot\boldsymbol{r}} \,\mathrm{d}^2\boldsymbol{u}\mathrm{d}^2(\Delta\boldsymbol{u}). \tag{3.3}$$

Substituting with the $\tilde{\Gamma}$-W connection, $\tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u}) = \iint W(\boldsymbol{r}', \boldsymbol{u})e^{-\mathrm{j}2\pi\boldsymbol{r}'\cdot\Delta\boldsymbol{u}} \,\mathrm{d}^2\boldsymbol{r}'$ and massag-

ing the formula a little, we have

$$I(\boldsymbol{r}, \boldsymbol{u}_{\text{sh}}) = \iiiint \mathrm{d}^2\boldsymbol{r}'\mathrm{d}^2\boldsymbol{u} \ W(\boldsymbol{r}', \boldsymbol{u}) \iint \text{rect}\left(\frac{\lambda f(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2} - \boldsymbol{u}_{\text{sh}})}{a}\right) \cdot$$

$$\text{rect}^*\left(\frac{\lambda f(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2} - \boldsymbol{u}_{\text{sh}})}{a}\right) e^{\mathrm{j}2\pi\Delta\boldsymbol{u}\cdot(\boldsymbol{r}-\boldsymbol{r}')} \ \mathrm{d}^2(\Delta\boldsymbol{u}). \tag{3.4}$$

If we define the Wigner function representation of the optical system with the aperture as

$$W_{\text{a}}(\boldsymbol{r}, \boldsymbol{u}; w) = \iint \text{rect}\left(\frac{\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2}}{w}\right) \text{rect}^*\left(\frac{\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2}}{w}\right) e^{\mathrm{j}2\pi\Delta\boldsymbol{u}\cdot\boldsymbol{r}} \ \mathrm{d}^2(\Delta\boldsymbol{u}), \tag{3.5}$$

Eq. (3.4) becomes

$$I(\boldsymbol{r}, \boldsymbol{u}_{\text{sh}}) = \iiiint \mathrm{d}^2\boldsymbol{r}'\mathrm{d}^2\boldsymbol{u} \ W(\boldsymbol{r}', \boldsymbol{u})W_{\text{a}}(\boldsymbol{r} - \boldsymbol{r}', \boldsymbol{u} - \boldsymbol{u}_{\text{sh}}; a/\lambda f)$$

$$= \iiiint \mathrm{d}^2\boldsymbol{r}'\mathrm{d}^2\boldsymbol{u} \ W(\boldsymbol{r}', \boldsymbol{u})W_{\text{a}}(\boldsymbol{r} - \boldsymbol{r}', \boldsymbol{u}_{\text{sh}} - \boldsymbol{u}; a/\lambda f). \tag{3.6}$$

Here in the last step we apply the fact that the rect function is real-valued and even. As shown in Eq. (3.6), a Fourier spectrogram is the result of a Wigner function convolving with the optical system's Wigner representation. For this reason, the spectrogram is often called a smoothed Wigner function [194].

3.2 Digital Micromirror Devices in Phase-Space Imaging

In this chapter we use a digital micromirror device (DMD) for our spatial light modulator. In previous work [183], the SLM was a liquid crystal on silicon (LCOS), which operates much slower ($\sim$ 60 Hz max frame rate) than the DMD ($\sim$kHz max frame rate). In addition, the LCOS reduced light efficiency due to its polarization selectivity, which required pre-filtering of the input light into the correct polarization. Since exposure time is a key limiting factor in our acquisition speed, any improvement in light efficiency also helps with acquisition speed. The DMD array is polarization insensitive, which not only improves light efficiency, but also makes the system more general for various applications. A larger problem with the LCOS SLM is that the polarization selectivity is imperfect. One must place crossed polarizers at the input and output of the SLM in order to block the unmodulated light. However, even with expensive high-quality polarization optics, a small percentage of light leaks through the crossed polarizers in the black parts of the coded image. This would normally create only minimal artifacts, but due to the nature of the imaging system collecting the light in the Fourier space, a large part of this unwanted leaked light shows up on the low-frequency part
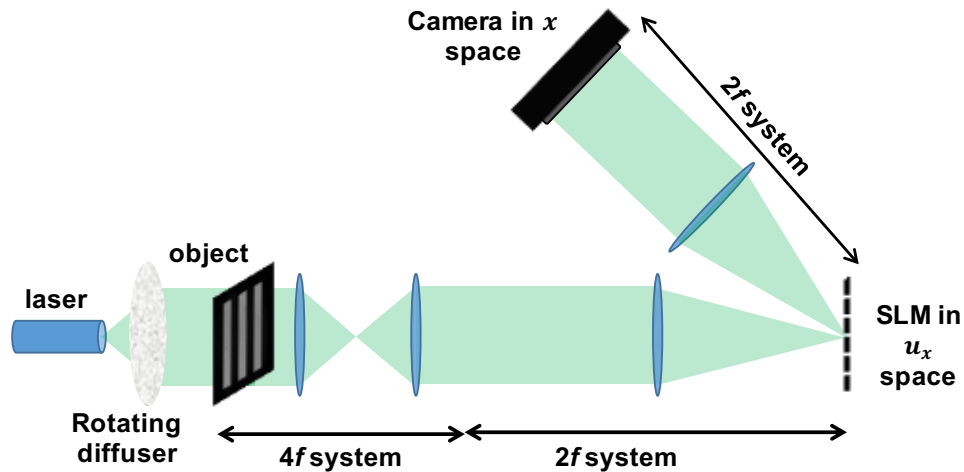
Figure 3.1: The Fourier spectrogram measurement setup for microscopic objects. The magnifying $4f$ system is formed by a 20x/0.4NA objective lens and a tube lens of focal length 200 mm. The lens in front of the DMD has a focal length of 225 mm and the one between DMD and camera has a focal length of 175 mm.

of the u_x space image, creating an unwanted streaking artifact near the DC term which can be seen on the right side of Fig. 3.2. In a DMD device, there is no leaked light, since 'off' pixels are fully reflected out of the imaging pathway, and so the result is much cleaner.

A further advantage of switching to a DMD SLM is in improving the speed of capture. To take advantage of the DMD's extremely fast switching rates, it is synchronized with a camera via on-board circuitry, which avoids a computer in the loop. We use the board controlling the DMD to generate a voltage signal each time it starts a new frame. This signal is fed to the trigger input of the camera directly. By this setting, the synchronization is accomplished through hardware entirely and is limited only by the required exposure time, giving speed improvements of several orders of magnitude over the setup in [183].

Our experimental setup is shown in Fig. 3.1. Laser light passes through a rotating diffuser, which makes it partially coherent (with Gaussian statistics, FWHM 5°). The object is illuminated entirely, and the scattered light is collected by a microscope objective. A $4f$ system acting as a microscope is used to magnify the image. Then a $2f$ system transforms the object light to Fourier space, where our DMD SLM is mounted. The light is then patterned by the DMD according to our desired coding scheme (in this case a scanned window function) and the filtered light further passes through another $2f$ system, turning into intensity images collected by the camera in x-space.

In Fig. 3.2 we show the result of an in-focus image illuminated by a partially coherent beam in our experimental setup, as compared to the results of [183]. The object here is a three-bar pattern, which is one dimensional. The object is illuminated by a partially coherent light source, created by inserting a rotating diffuser into the illumination pathway. One can see that the u_x axis information displays a notable spread, due to the finite coherence length

Phase-space measurement with new setup   Phase-space measurement with old setup
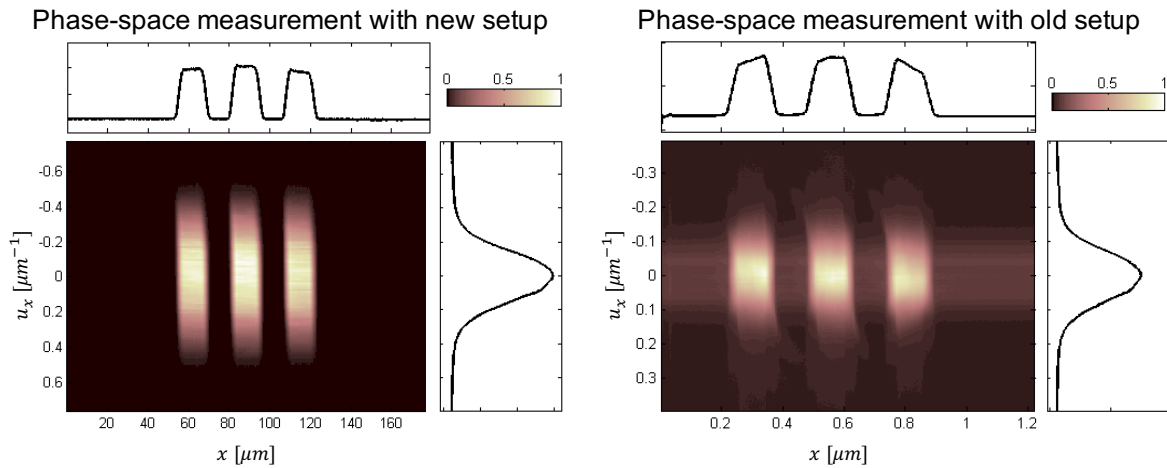
Figure 3.2: Experimental results for phase-space measurements of three-bar test pattern with partially coherent light, as compared to the results of [183]. Results are much cleaner due to elimination of polarization leakage, with no low-frequency artifacts present. Scan speeds are also significantly increased in the new scheme. Plots at top and side of each phase-space figure represent the projections onto x and u_x space, respectively.

of the illuminating light. We can see three strips in the plot of WDF that are in-focus, since they are vertical, without any tilt.

Further improvement in the quality of the phase-space image over our previous design comes from the modification of placing the scanning aperture in Fourier (u_x) space, rather than x space. While it is not guaranteed, we expect this scheme to have better performance than windowing in x-space because most signals are largely composed of low frequency information. We take the phase-space component images in x space; since our camera has a higher resolution than our DMD, we will always have high quality x space images than u_x space, which is likely preferred visually. Moreover, the u_x-space scanning can be efficient by paying more attention on lower frequency parts, while the x-space scanning requires going through every pixel. This leads to improved ability to test new coded aperture strategies which pattern the u_x space to have more than one pixel on at a time, further increasing the light throughput (See Chapters 4 and 5).

3.3   Fourier Spectrogram Experiments for Point Sources and Scattering

In the previous case, we use a rotating diffuser in front of a laser to reduce the optical coherence of the light source. Let us consider a different case that the sources are spatially incoherent, that is, point-like sources. They raise much interest as the fluorophores are
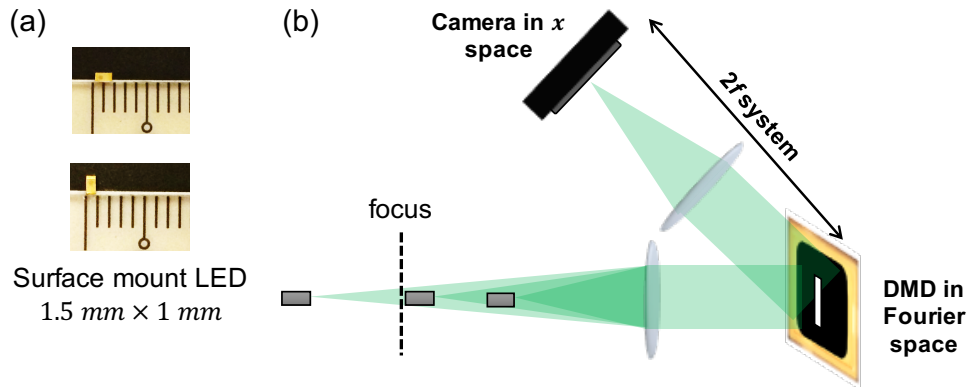
Figure 3.3: Fourier spectrogram setup for phase-space measurement experiments to 3D localize three point emitters. (a) The pictures of one LED on a ruler in order to show the dimensions of the LED used. The LED has a spectrum centered around 633 μm. (b) The Fourier spectrogram imaging setup with the LEDs around the front focal plane of the system. The lens in front of the DMD has a focal length of 225 mm and the one between DMD and camera has a focal length of 175 mm. The axial position of the three LEDs relative to the focal plane is: (from left to right) -58 mm, 0.91 mm and 120 mm. The scanning window width on the DMD is 1740.8 μm.

point sources and vastly used in biological studies. We use light emitting diodes (LEDs) as our point sources in this case and modify the setup to image millimeter scale objects (see Fig. 3.3). With this new scene, we can illustrate the property of the phase space of point sources, which reveals the position information of the points. Furthermore, a light scattering environment is emulated. The Fourier spectrogram retains structures about the sources and can be further utilized to counter the scattering in imaging.

As shown in the upper part of Fig. 3.4, the LEDs at different depths show as lines with different slopes in the Fourier spectrogram. This gives us a cue for defocused amounts which is explained in Sec. 2.4. The line structure in the Fourier spectrogram is also resistant to scattering. We inserted rotating diffusers in-between the point sources to simulate volumetric scattering and the measured Fourier spectrogram still retains three linear structures although they are widened (lower part of Fig. 3.4). As a comparison, the information of the point sources is not inferrable from the traditional image, shown in the center column of the figure. This comparison shows that Fourier spectrogram imaging has a better chance to reconstruct scenes behind scattering than traditional imaging. The reconstruction is further improved by considering how the widening relates to the scattering situation of the sample. A detailed model discussing this widening is provided in Chapter 6.
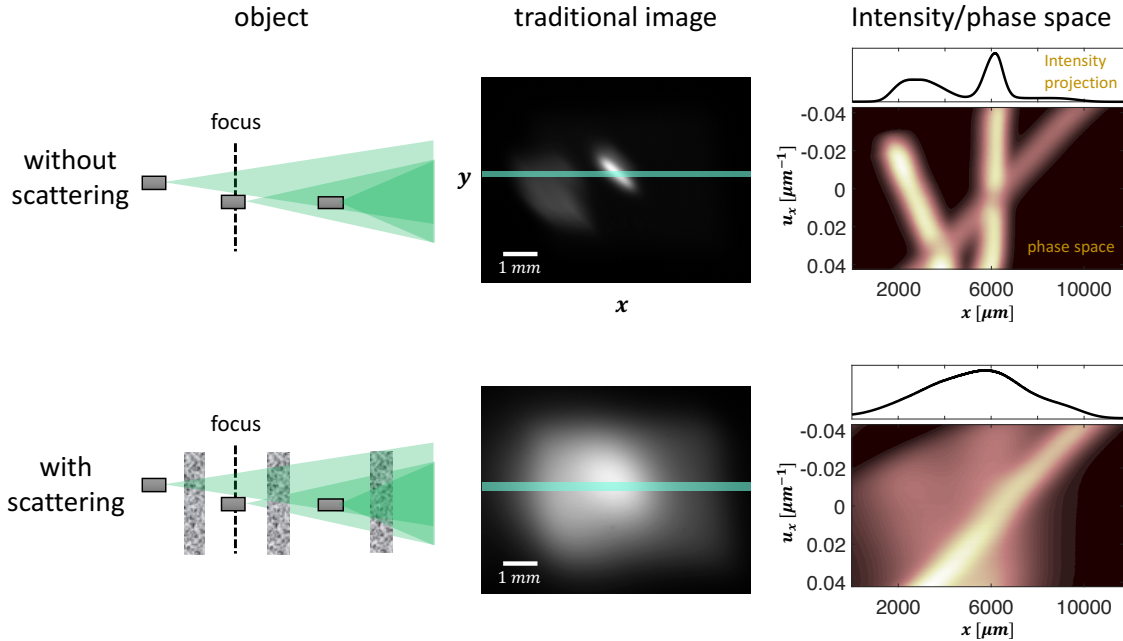
Figure 3.4: One slice of measured Fourier spectrogram for experiments with three point emitters at different 3D positions. The upper part shows the result in clear medium and the lower part in scatterers. The left column shows the object, the middle column shows the traditional intensity images taken without Fourier space modification, and the right column shows the Fourier spectrogram and its projection, the intensity. The cutlines in the middle column indicates the y slice from which the Fourier spectrogram is built (from the corresponding slice in aperture scanned images).

3.4   Limitation of the DMD Device

Though the use of the DMD in this chapter speeds up the measurements, the blazed grating of the DMD may cause dispersion problems in applications to nonmonochromatic light. If the application images the DMD, such as a DMD-based projector where the image of DMD is displayed on the screen, the dispersion caused by the DMD is insignificant because the dispersed light will be refocused back to the same pixel after the imaging system. However, in our setup, we put the DMD in the Fourier space and image the light in the real space. The light dispersed by the blazed grating of the DMD will then arrive at different points on the sensor.

The grating of a LCOS SLM does not cause severe dispersion because the LCOS SLM is designed to reflect the light off perpendicularly to its plane. The beam we use in such setup is the zeroth order beam, which does not change its direction among all colors. On the other hand, in a DMD SLM setup, we use high-order beams where the blazed grating of DMD enhances their light efficiency. These beams changes directions more rapidly than lower order beams when their color changes. For example, the DMD we used causes a red

light source with a bandwidth of 30 nm to be imaged as a diagonal strip across a one-inch square camera in an imaging system of unity magnification ratio.

Mathematically, we can also reveal this behavior. For simplicity, we consider a 1D blaze grating. The grating is put in the front focal plane of a $2f$ system and is illuminated by a normal-incident wave. We let θ denote the blaze angle ($\theta = 0$ means a flat plane), p the pitch of the grating, and f the focal length of the $2f$ system. The electric field on the back focal plane of the $2f$ system is

$$E(x) = pe^{-\frac{\mathrm{j}\pi p}{\lambda f}(x - 2f\tan\theta)}\mathrm{sinc}\left(\frac{p}{\lambda f}(x - 2f\tan\theta)\right)\mathrm{comb}\left(\frac{px}{\lambda f}\right). \tag{3.7}$$

Here $\mathrm{comb}(\xi) = \sum_{k=-\infty}^{\infty} \delta(\xi - k)$ and $\mathrm{sinc}(\xi) = \sin(\pi\xi)/\pi\xi$. By taking absolute value squares, plugging the values for a typical system ($\theta = 13°$, $f = 100$ mm, $p = 13\sqrt{2}$ μm) and sweeping the wavelength within a 30 nm bandwidth around the 600 nm, we get a 3.7 mm central lobe of dispersion. This corresponds to ~ 570 pixels of a typical sCMOS camera (pixel size 6.5 μm) and thus the dispersion is not negligible. A quick fix to this problem is to apply a narrow-band color filter which loses the light drastically. Therefore in later chapters, we turn back to LCOS SLM if the source light is broadband and dim to mitigate this dispersion problem.

3.5 Conclusion

We have demonstrated phase-space imaging by windowed scanning of Fourier space, using a digital micromirror device, while taking the x-space pictures corresponding to each window. The synchronization between DMD and a camera gives fast acquisition and flexible control. The new system configuration becomes similar in concept to pupil segmentation microscopy [88], which has been proven useful in imaging through scattering, but offers significantly better resolution in the pupil plane. The spectrogram contains more information than the traditional imaging and we will show how it can be applied to imaging through scattering in Chapter 6.

Chapter 4

Coded Aperture for 3D Fluorescence Microscopy

In the previous chapter, we discussed how the Fourier spectrogram measurement enhances the number of sampling points of phase space functions. The amount of data is so large ($\approx 1000^4$ points) that it becomes hard to handle and redundant in order to reconstruct a 2D scene or to recover positions of several point sources. Though this redundancy can help in a harsh imaging situation, *e.g.* in a scattering environment, a large amount of data raises some concerns. The speed of spectrogram measurement is impacted as it is proportional to the number of data points. Therefore, we seek to reduce the number of data by aiming at certain less general applications than a full spectrogram. The application we will focus on in this chapter is 3D microscopy of fluorescence samples.

3D fluorescence microscopy is a critical tool for bioimaging, since most samples are thick and can be functionally labeled. High-resolution 3D imaging typically uses confocal [46], two-photon [80] or light sheet microscopy [148]. Because these methods all involve scanning, they are inherently limited in terms of speed or volume. Light field microscopy [117, 33], on the other hand, achieves single-shot 3D capture, but sacrifices resolution because it distributes the 4D spatial-angular information across a 2D spatial sensor. High resolution and single-shot capture are possible with coded aperture microscopy [145, 157, 160]; however, this requires an extremely sparse sample. Here, we describe a multi-shot coded aperture microscopy method for high-resolution imaging of large and dense volumes with an efficient data capture rate.

Coded aperture is a multiplexed way to probe the phase space of fluorescent samples and we seek to reconstruct the 3D flourescence distribution at the end. As in the Fourier spectrogram setup, we use a spatial light modulator (SLM) in the pupil plane of a microscope in order to sequentially pattern multiplexed coded apertures while capturing images in real space (see the experimental setup in Fig. 4.1). Having multiple apertures in a pattern increases the light throughput of the system, giving us flexibility to trade-off between photon noise suppression and measurement speed. A compressed sensing scheme, which fits naturally with coded aperture methods, can further reduce the number of measurements. Then, we reconstruct the 3D fluorescence distribution of our sample by solving an inverse prob-

lem via regularized least squares with a proximal accelerated gradient descent solver. We experimentally reconstruct a 101 Megavoxel 3D volume ($1010 \times 510 \times 500\mu$m with NA 0.4), demonstrating improved measurement qualities compared to scanning aperture methods.

One factor that makes multiplexing preferable is the enhancement of image resolution compared with the spectrogram methods. Multiple apertures opened at once allow the light in those areas to interfere. This interference provides finer features and higher-order correlation information than a spectrogram in the images. Even though the aperture in a Fourier spectrogram measurement scans through all positions in Fourier space, the information of the phase space function is lost in the process and cannot be recovered in general. We discuss the detail about this information loss in next chapter, leaving the current one to focus on the experiment.

4.1 Our Method

Our experimental setup consists of a widefield fluorescence microscope with a spatial light modulator in Fourier space (the pupil plane). The SLM implements a series of quasi-random coded aperture patterns, while collecting real space images for each (Fig. 4.1) [121]. The 4D phase space can have very large pixel counts ($\sim 10^{12}$, the product of the pixel counts of the SLM and the sensor). Compared to scanning aperture methods [122, 183], the new scheme has three major benefits. *First*, it achieves better resolution by capturing high-frequency interference effects (high-order correlations). This enables diffraction-limited resolution at the microscope's full numerical aperture (NA). *Second*, we achieve higher light throughput by opening up more of the pupil in each capture; this can be traded for shorter exposure time and faster acquisition. *Third*, the multiplexed nature of the measurements means that we can employ compressed sensing approaches (when samples are sparse) in order to capture fewer images without sacrificing resolution. This means that the number of required images scales not with the reconstructed number of resolved voxels, but rather with the sparsity of the volume.

Our method can be thought of as a multi-shot coded aperture scheme for diffraction-limited 3D fluorescence microscopy. It is analogous to coded aperture photography [115, 118, 77, 37]; however, we use a wave-optical model to account for diffraction effects, so intensity measurements are nonlinear with complex-field. Fluorescent imaging allows a simplification of the forward model, since each fluorophore is spatially coherent with itself but incoherent with all other fluorophores. Our reconstruction algorithm then becomes a large-scale inverse problem akin to multi-image 3D deconvolution, formulated as a convex ℓ_1 regularized least-squares problem and solved by a fast iterative shrinkage-thresholding algorithm (FISTA) [12].

4.1.1 Forward model

Here, we assume that the object is a 3D volume of incoherent emitters with no occlusions. The emitters are also assumed to have identical emission spectra. Our forward model relates
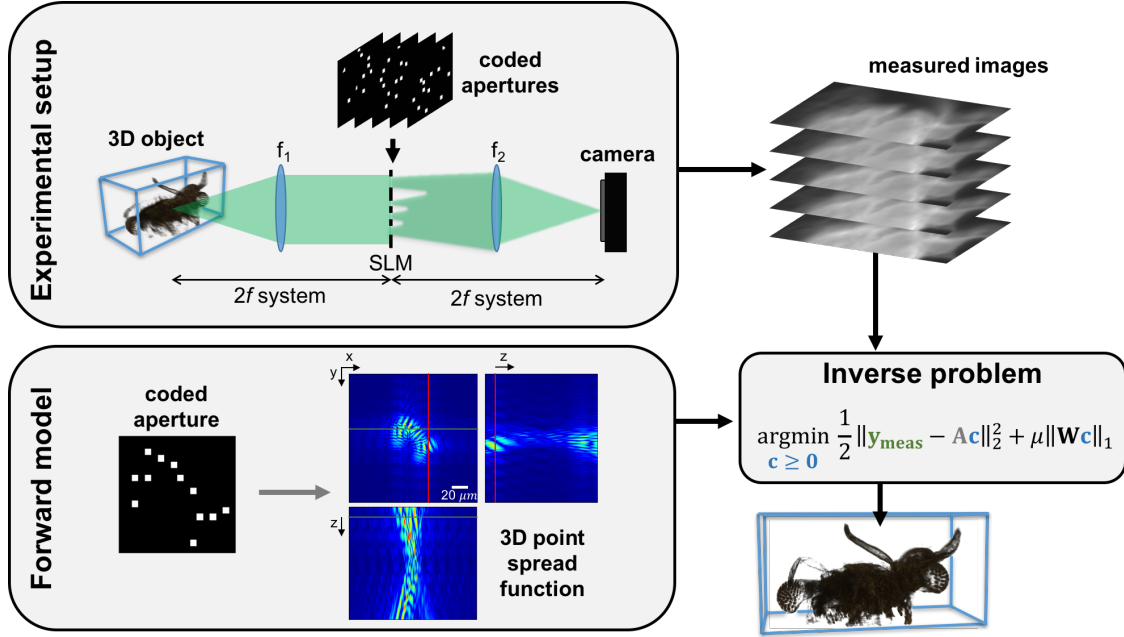
Figure 4.1: Phase-space multiplexing for 3D fluorescence microscopy. The microscope (a $4f$ system) uses a Spatial Light Modulator (SLM) in Fourier space to implement multiple coded apertures, while capturing 2D intensity images in real space for each. Our wave-optical forward model $\mathbf{A}$ relates the object $\mathbf{c}$ to the measured images $\mathbf{y}_{\text{meas}}$ for each pattern. The inverse problem recovers the object, subject to sparsity priors where applicable.

each coded aperture's captured image to the 3D object's intensity. Each image contains information from multiple spatial frequencies and their interference terms (which are the key to resolution enhancement and will be discussed in Chapter 5). Our setup falls in the same category as that introduced in Sec. 2.5. Therefore we start from Eq. (2.12) with a subscript n introduced to denote the codes,

$$I_n(\mathbf{r}) = \iiiint \tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u}) M_n\left(\lambda f(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2})\right) M_n^*\left(\lambda f(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2})\right) e^{j2\pi\Delta\boldsymbol{u}\cdot\boldsymbol{r}} \, \mathrm{d}^2\boldsymbol{u}\mathrm{d}^2(\Delta\boldsymbol{u}).$$

(4.1)

Consider a complex electric field at the front focal plane $E_s(\boldsymbol{r}_1; \alpha)$, generated by a point source, where α denotes properties of the point source (*e.g.* location and wavelengths). The field acts like a unique coherent mode — it interferes coherently with itself but not with other modes. Hence the Fourier-space mutual intensity $\tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u})$ in Eq. (4.1) is a weighted sum of that of each $E_s(\boldsymbol{r}_1; \alpha)$ where the weights, denoted by $C(\alpha)$, are proportional to the emission intensity of each point source. We have

$$\tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u}) = \sum_\alpha C(\alpha)\tilde{E}_s\left(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2}; \alpha\right) \tilde{E}_s^*\left(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2}; \alpha\right).$$

(4.2)

We specify α to be $(\boldsymbol{r}_s, z_s, \lambda)$ and further decompose $C(\alpha)$ into the product of the emission strength $C(\boldsymbol{r}_s, z_s)$ and the spectrum of the source $S(\lambda)$, where $\boldsymbol{r}_s$ is the transverse coordinate of the source, z_s is its defocus distance, and λ indicates a single color in the emission spectrum of the source. The field $\tilde{E}_s$ in Eq. (4.2) can be represented by

$$\tilde{E}_s(\boldsymbol{u}_1; \boldsymbol{r}_s, z_s, \lambda) = \begin{cases} e^{j\frac{2\pi}{\lambda}(-z_s)\sqrt{1-\lambda^2|\boldsymbol{u}_1|^2}-j2\pi\boldsymbol{r}_s\cdot\boldsymbol{u}_1}, & |\boldsymbol{u}_1| < \frac{NA}{\lambda} \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

where angular spectrum propagation [72] is used to account for defocusing. Plugging $C(\alpha)$ and Eq. (4.3) into Eq. (4.2), we have

$$\tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u}) = \iiint \sum_{\lambda} S(\lambda) e^{j\frac{2\pi z_s}{\lambda}\left(\sqrt{1-\lambda^2|\boldsymbol{u}-\frac{\Delta\boldsymbol{u}}{2}|^2}-\sqrt{1-\lambda^2|\boldsymbol{u}+\frac{\Delta\boldsymbol{u}}{2}|^2}\right)} C(\boldsymbol{r}_s, z_s) \, \mathrm{d}^2\boldsymbol{r}_s \, \mathrm{d}z_s. \tag{4.4}$$

Hence the formula for the intensity images, Eq. (4.1), can be written as

$$I_n(\boldsymbol{r}) = \iiint \left(\iint \sum_{\lambda} S(\lambda) K_{M_n, z_s}(\Delta\boldsymbol{u}; \lambda) e^{j2\pi(\boldsymbol{r}-\boldsymbol{r}_s)\cdot\Delta\boldsymbol{u}} \, \mathrm{d}^2\Delta\boldsymbol{u} \right) C(\boldsymbol{r}_s, z_s) \, \mathrm{d}^2\boldsymbol{r}_s \, \mathrm{d}z_s, \tag{4.5}$$

where

$$K_{M_n, z_s}(\Delta\boldsymbol{u}; \lambda) = \iint M_n^*\left(\lambda f(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2})\right) M_n\left(\lambda f(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2})\right)$$
$$e^{j\frac{2\pi z_s}{\lambda}\left(\sqrt{1-\lambda^2|\boldsymbol{u}-\frac{\Delta\boldsymbol{u}}{2}|^2}-\sqrt{1-\lambda^2|\boldsymbol{u}+\frac{\Delta\boldsymbol{u}}{2}|^2}\right)} \, \mathrm{d}^2\boldsymbol{u} \tag{4.6}$$

is the kernel for code M_n at depth z_s.

Equation (4.5) describes the forward model for a 3D fluorescent object $C(\boldsymbol{r}_s, z_s)$ with no occlusions. The term in parentheses is a convolution kernel describing the 3D point spread function for code M_n (shown in Fig. 4.1). For simplicity, we assume here no scattering, though incorporating the scattering forward model in [122, 146] is straightforward.

4.1.2 Inverse problem

Based on the raw data and forward model, the inverse problem is formulated as a nonlinear optimization. Our goal is to reconstruct the 3D intensity distribution $C(\boldsymbol{r}_s, z_s)$ from the measured images. To do so, we aim to minimize data mismatch, with an ℓ_1 regularizer to mitigate the effects of noise (and promote sparsity where applicable). The mismatch is defined as the least-squares error between the measured intensity images and the intensity predicted by our forward model (Eq. (4.5)). This formulation has a smooth part and a non-smooth part in the objective function and is efficiently solved by a proximal gradient descent solver (FISTA [12]).

To formulate the inverse problem, we first discretize the forward model in Eq. (4.5) to be

$$\mathbf{y} = \mathbf{Ac}. \tag{4.7}$$

Here $\mathbf{y} \in \mathbb{R}^{MP \times 1}$ corresponds to predicted images on the sensor; each small chunk ($\in \mathbb{R}^{P \times 1}$) of $\mathbf{y}$ is a vectorized image $I_n(\boldsymbol{r})$. We discretize $\boldsymbol{r}$ into P pixels, and the number of codes is M, so $n = 1 \ldots M$. Similarly, we discretize $\boldsymbol{r}_s$ and z_s into P' pixels and L samples, respectively, to obtain a vectorized version of $C(\boldsymbol{r}_s, z_s)$, denoted by $\mathbf{c} \in \mathbb{R}^{LP' \times 1}$. The matrix $\mathbf{A} \in \mathbb{R}^{MP \times LP'}$, which is not materialized in implementation, represents the summation and convolution in Eq. (4.5) where the latter uses 2D Fast Fourier Transforms (FFTs) for each subvector ($\in \mathbb{R}^{P' \times 1}$) of $\mathbf{c}$, with zero-padding to avoid periodic boundary condition errors. The convolution kernel is precomputed and stored for speed.

The inverse problem becomes minimizing a data fidelity term plus an ℓ_1 regularization with parameter μ:

$$\arg\min_{\mathbf{c} \geq 0} \frac{1}{2} \|\mathbf{y} - \mathbf{y}_{\text{meas}}\|_2^2 + \mu \|\mathbf{W}\mathbf{c}\|_1, \tag{4.8}$$

where $\mathbf{y}_{\text{meas}} \in \mathbb{R}^{MP \times 1}$ is the measured intensity. We also use a diagonal matrix $\mathbf{W} \in \mathbb{R}^{LP' \times LP'}$ to lower the weight of point sources near the borders of images whose light falls off the sensor. Each diagonal entry of $\mathbf{W}$ is obtained by summing the corresponding column in $\mathbf{A}$. The value of μ is $5 \times 10^{-4} \|\mathbf{W}^{-1}\mathbf{A}^{\text{T}}\mathbf{y}_{\text{meas}}\|_\infty$. Note that $\mathbf{c} = \mathbf{0}$ wherever $\mu \geq \max(\mathbf{W}^{-1}\mathbf{A}^{\text{T}}\mathbf{y}_{\text{meas}})$. Outside point sources may also contribute to the measured intensity due to defocus; hence, we use an extended field-of-view method [16] to solve for more 2D points in $\mathbf{c}$ than $\mathbf{y}$ (*i.e.* $P' > P$).

4.2   Design of Coded Apertures

In the scanning-aperture scheme [183, 122], smaller apertures give better frequency sampling of the 4D phase space, at a cost of: 1) lower resolution, 2) lower signal-to-noise ratio (SNR) and 3) large data sets. Our multiplexing scheme mitigates all of these problems. Multiplexing achieves diffraction-limited resolution by additionally capturing interference terms, which cover the full NA-limited bandwidth. This is evident in the Fourier transform of the captured images (Fig. 4.2). The SNR improvement is also visible; the multiplexed image is less noisy.

Our codes are chosen by quasi-random non-replacement selection. This is not an ideal method and we will study the design in detail in Chapter 5. We section the SLM plane into 18×18 square blocks and keep only the 240 blocks that are inside the system NA. For each code, we open 12 blocks, selected randomly from the blocks remaining after excluding ones that were open in previous sequences. Here the numbers 18 and 12 are heuristic such that the point spread function is not bigger than the field of view if only one square block is open and the degree of multiplexing, $12/240 = 5\%$, is on the same order of that in multiplexed Fourier Ptychography [175]. In this scheme, the full NA can be covered by 20 codes. To allow for both diversity and redundancy, we choose to cover the entire pupil 5 times, resulting in 100 multiplexed aperture patterns. Examples of the codes and their corresponding measured intensity images are shown in Fig. 4.3.

Importantly, the number of multiplexed patterns can be flexibly chosen to trade off accuracy for speed of capture. For instance, by increasing the number of openings in each
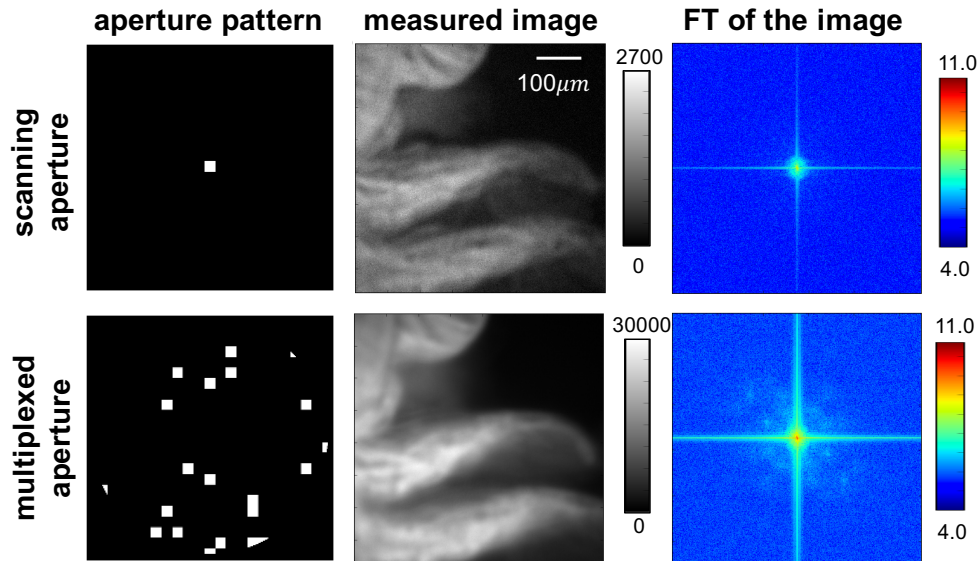
Figure 4.2: Multiplexed phase-space measurements contain more information than scanning-aperture measurements. (Left) A sample aperture pattern, (Middle) the corresponding measured intensity image (same exposure time), and (Right) its log-scale Fourier transform. The multiplexed measurements have better light throughput and more high-frequency content.

pattern, we can cover the entire pupil with fewer patterns. This means that we may be able to reconstruct the object from fewer measurements, if the inverse problem is solvable. By using *a priori* information about the object (such as sparsity in 3D) as a constraint, we can solve under-determined problems with fewer total measured pixels than voxels in the reconstruction [36]. Hence the number of openings in each pattern (the amount of multiplexing) should be chosen according to both the sparsity of the object and the importance of capture speed.

4.3 Simulation Validation

In order to test the proposed 3D reconstruction method, we simulate a synthetic data set measured on a $4f$ imaging system. Note that the design here is different from the experiments and the simulation is just to check if the multiplexed phase-space method in general is viable. The object in the simulation is a 96 μm $\times$ 96 μm $\times$ 10 μm 3D volume, which is stored in a $200 \times 200 \times 10$ 3D array. Within the 3D volume, 1600 mutually incoherent point sources are randomly distributed (Fig. 4.4(a)). Each of the point sources radiates a spherical wave with wavelength of 650 nm. The middle layer of the 3D volume is set to the front focal plane of the $4f$ imaging system. In the frequency space of the imaging system, we place a binary amplitude SLM. The size of a single SLM block (the square block in Fig. 4.4(b)) is 0.063 μm^{-1}, which characterizes the spatial frequency bandwidth (but not the physical size).
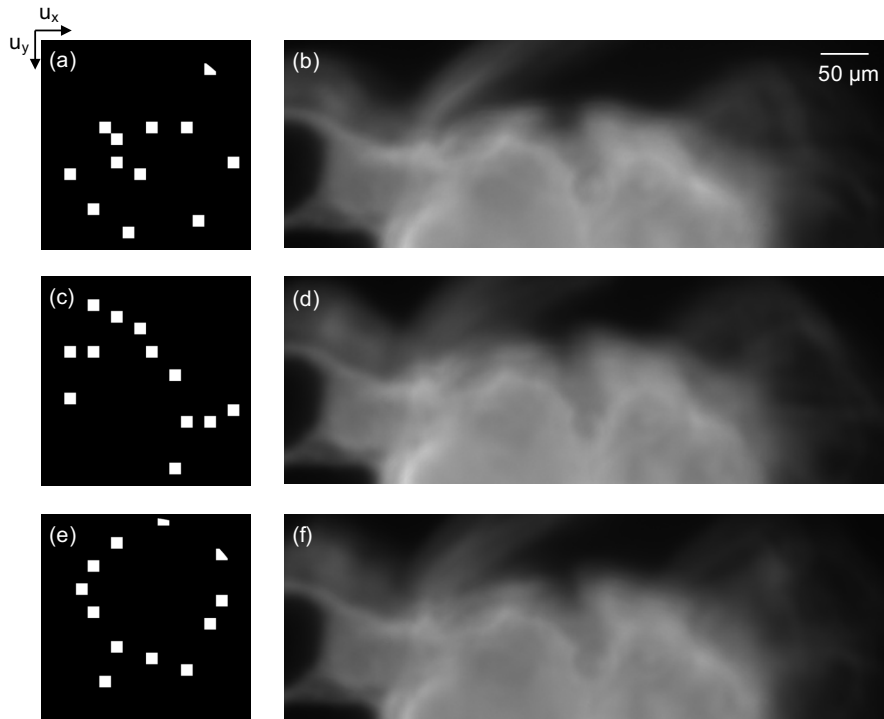
Figure 4.3: The measured intensity images corresponding to different codes shows the diversity in measured data. (a)(c)(e) Three multiplex codes applied on the SLM and (b)(d)(f) their corresponding measured intensity images of the fluorescent sample. The variation due to the change of codes can be observed on the right most parts of the intensity images.

Due to the limitation of the numerical aperture of the imaging system, only the signals of spatial frequency within ± 0.63 μm^{-1} are allowed to pass through the system. We randomly generate 20 SLM binary codes, one instance of which is shown in Fig. 4.4(b). For each code, we record a 2D intensity image on the back focal plane of the $4f$ imaging system.

We test our sparse 3D reconstruction algorithm with this synthetic data set. The ℓ_1 regularized least square error optimization problem in Eq. (4.8) is supposed to return a sparse estimation of 3D point sources distribution that resemble the distribution of the 1600 simulated point sources. The solver in question is FISTA [12]. We first use all of the 20 intensity images as measured input in the optimization problem to verify the correct implementation of the solver and the ability of our algorithm. The recovered 3D volume has a Root Mean Square Error (RMSE) of 0.0006 to the truth (Fig. 4.4(c)), which is small compared to the source amplitude of 1. Therefore, the result asserts the correct implementation and the ability of our algorithm to reconstruct sparse sources.

Next, we test the performance of the sparse 3D reconstruction algorithm with fewer images captured, in order to speed up the acquisition. We use 5 intensity images as input and the recovered 3D reconstruction result is still accurate but with larger RMSE (Fig. 4.4(d))
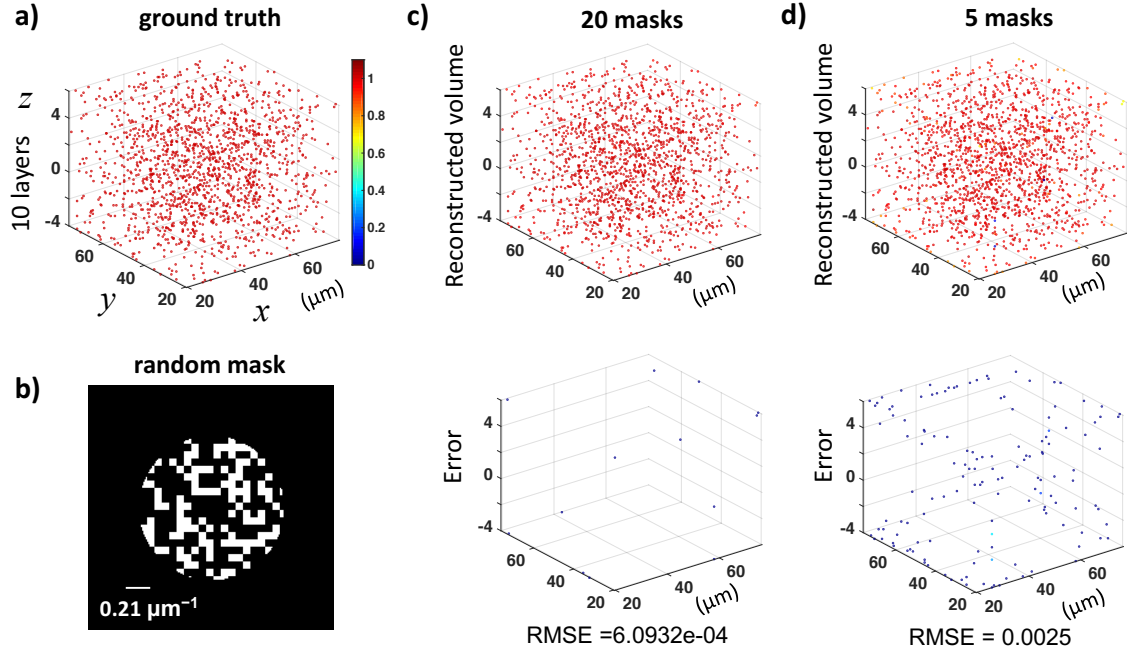
Figure 4.4: Simulation result of 3D sparse reconstruction of fluorescent points with coded aperture. (a) $\sim$ 1600 fluorescent point sources randomly distributed in a 96 μm $\times$ 96 μm $\times$ 10 μm 3D volume. (b) One of the random coded apertures displayed on the SLM. (c) A 3D sparse reconstruction result using intensity measurements of 20 codes as input and (d) that of 5 codes.

as expected. Without the ℓ_1 regularization, the least square error optimization problem is ill-posed, since the number of measured variables ($200 \times 200 \times 10$) is less than the dimension of the unknown 3D volume ($200 \times 200 \times 20$). The ℓ_1 regularization helps solve the ill-posed least squares problem by promoting sparsity of a 3D reconstruction.

4.4 Experiments

Our experimental setup consists of the $4f$ system ($f_1 = 250\ mm$, $f_2 = 225\ mm$) shown in Fig. 4.1, with an additional $4f$ system in front, made of an objective lens (20$\times$ NA 0.4) and a tube lens ($f_t = 200\ mm$) to relay the sample (see Fig. 4.5). The SLM (1400$\times$1050 pixels of size 10.3μm) is a liquid crystal chip from a 3-LCOS projector (Canon SX50) which is reflective and polarization-sensitive, so we fold the optical train with a polarization beam splitter and insert linear polarizers.[1] Our sensor (Hamamatsu ORCA-Flash4.0 V2) captures the multiplexed images and synchronizes with the SLM via computer control.

---

[1]The use of LCOS SLM is to mitigate the dispersion problem that can be caused by a DMD in a $2f$ setup. See Sec. 3.4 for more detail.
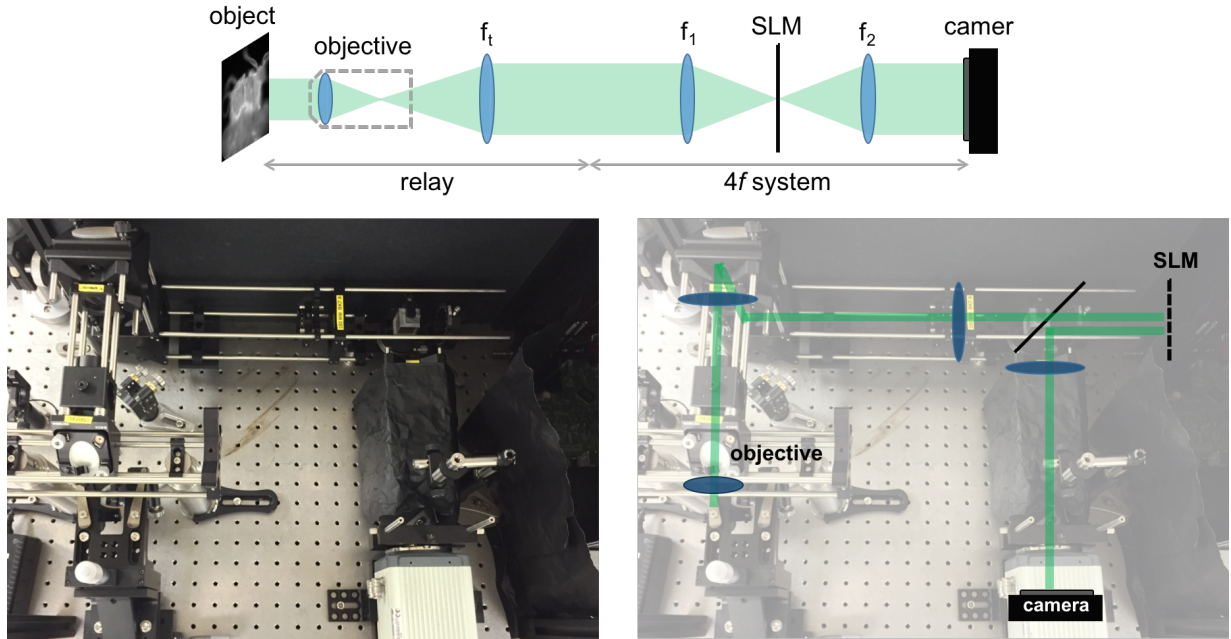
Figure 4.5: Detailed setup for multiplexed phase space imaging with fluorescent samples. (Top) a schematic of the optical train of our system with mirrors not shown, (Bottom left) a snapshot of the experimental setup and (Bottom right) the snapshot overlaid with the optical train.

Our sample is a fixed brine shrimp stained with fluorophore Eosin Y (Carolina Biological). It is relatively dense, yet does not fill the entire 3D volume so still has significant sparsity. The reconstructed 3D intensity (Fig. 4.6(a), 4.6(b) and 4.6(f)-4.6(h)) is stitched from five volume reconstructions, each with $640 \times 640 \times 120$ voxels to represent the sample volume of $455 \times 455 \times 600$ μm. The reconstruction is cropped to the central part of our extended field of view, so the final volume contains $1422 \times 715 \times 100$ voxels corresponding to $1010 \times 510 \times 500 \mu$m. The dataset size is large (9 GB), and since the size of the 3D array is $\sim 5 \times 10^7$ without the extended field-of-view and the measured data is $\sim 4 \times 10^7$, the number of operations for evaluating Eq. (4.5) is on the order of 3×10^{10}. This takes 4 seconds to compute on a computer with 48-core 3.0 GHz CPUs and requires 94 GB memory to store the kernel (Eq. (4.6)).

The reconstructed 3D intensity is shown in Fig. 4.6, alongside images from a confocal microscope and a widefield focus stack, for comparison. Both our method and the focus stack use a 0.4 NA objective and widefield illumination, while the confocal uses a 0.25 NA one for illumination and capture; hence, the confocal results should have slightly better resolution. Our reconstructed slices appear to have slightly lower resolution than the defocus stack and confocal, possibly due to the missing information in the Fourier space mutual intensity that will be discussed in the next chapter (Fig. 5.1(g)), where the mutual intensity is undersampled. As expected, each depth slice of our reconstruction has better rejection of information from other depths, similar to the confocal ones. The undersampling mention
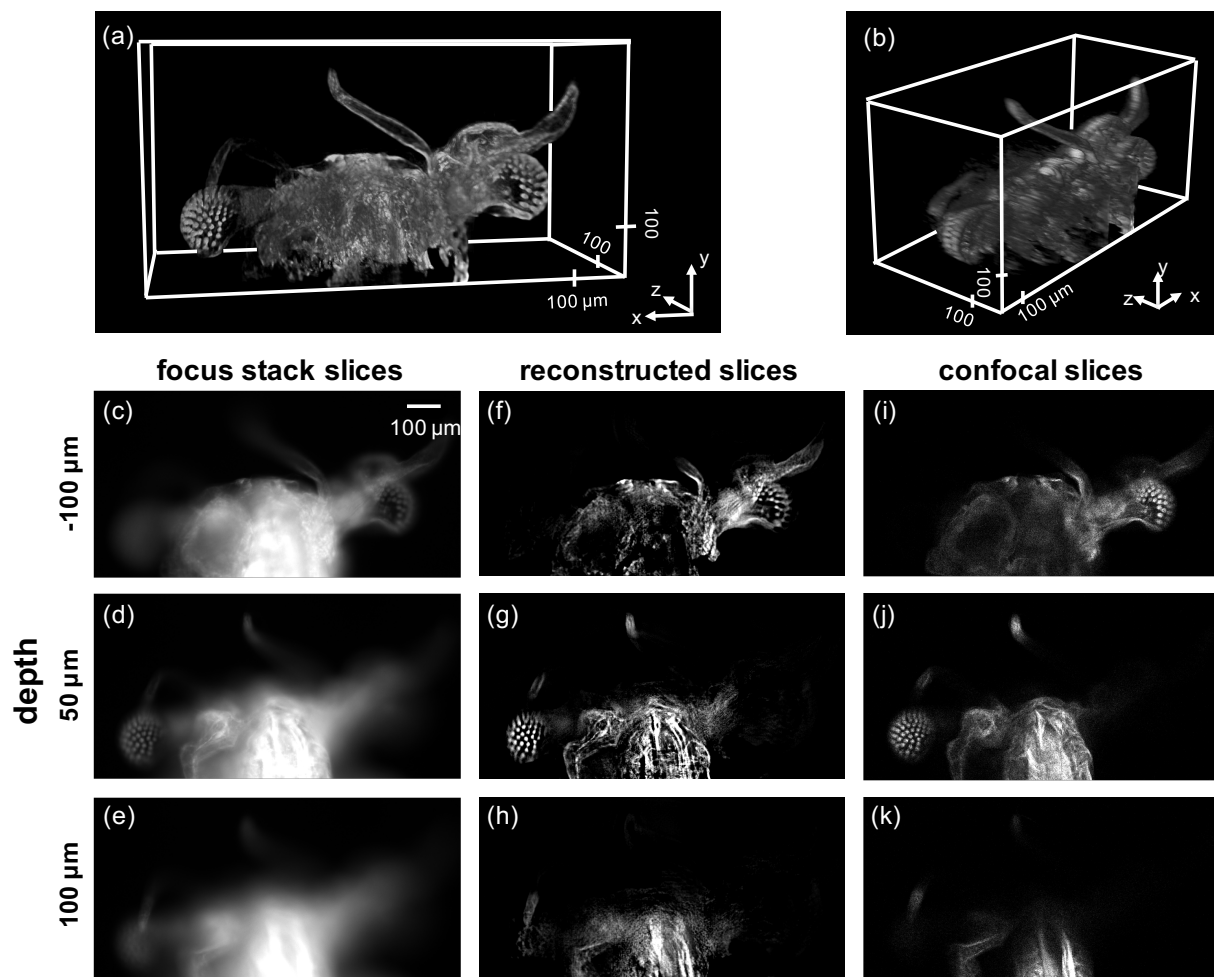
Figure 4.6: 3D reconstruction of a fluorescent brine shrimp sample as compared to focus stack and confocal microscopy. (a) and (b) 3D renderings of the reconstructed fluorescence intensity distribution ($1010 \times 510 \times 500 \mu$m) from different perspectives. (c)-(e) 2D widefield images at three different focus planes. (f)-(h) Slices of our reconstructed volume at the same depth planes. (i)-(k) Confocal microscopy slices at the same depth planes for comparison.
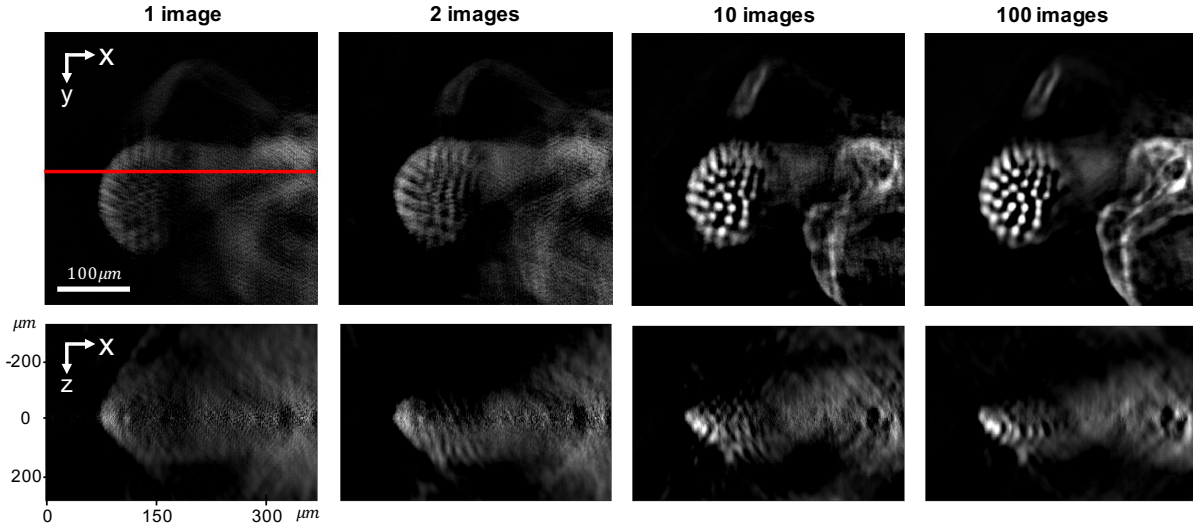
Figure 4.7: Image quality can be traded for capture speed (number of coded aperture images). The x-y and x-z slices from 3D reconstructions from increasing numbers of images with different coded apertures show that this object is too dense to be accurately reconstructed by a single coded-aperture image, but gives a reasonable reconstruction with 10 or more images, due to sparsity of the sample. The ratios of the measured data to the reconstructed voxels are 0.8% (1 image), 1.6% (2 images), 8% (10 images) and 80% (100 images).

above can also be seen from the measured data count for each subvolume $\sim 4 \times 10^7$ and that of the corresponding 3D array to reconstruct $\sim 4 \times 10^7$. A regularizer helps the inverse problem with the under-determined $\mathbf{A}$ and leads us to consider reconstruction with data reduction.

4.5 Compressed Sensing for Faster Capture

To illustrate the flexible tradeoff between capture time (number of coded apertures used) and quality, we show in Fig. 4.7 reconstructions with fewer coded aperture images. The case of only 1 image corresponds to a single coded aperture and gives a poor result, since the sample here is relatively dense to the sample applicable to the single-shot methods [157, 145]. However, with as few as 10 images we obtain a reasonable result, despite the fact that we are solving a severely under-determined problem. This is possible because the measurements are multiplexed and so the ℓ_1 regularizer acts as a sparsity promoter.

4.6 Conclusion

We demonstrated 3D reconstruction of a large-volume high-resolution and relatively dense fluorescent object from multiplexed phase-space measurements, which is similar to coded aperture measurements in photography [118]. An SLM in Fourier space dynamically implements quasi-random coded apertures while intensity images are collected in real space for each coded aperture. In this chapter we mainly focus on the experiment. The remaining phase-space theory of multiplexing and the design of codes will be discussed in the next chapter. Our 3D reconstruction is formulated as an ℓ_1-regularized least-squares problem. This method enables diffraction-limited 3D imaging with high resolution across large volumes, efficient data capture and a flexible acquisition scheme for samples of different type and size, without the need of a mechanical scan.

Chapter 5

Phase Space Measurement for Samples of Arbitrary Coherence

Phase space provides tremendous data that can be used in applications such as 3D scene reconstruction and aberration removal. We have discussed two phase space measurement methods in detail so far, the Fourier spectrogram and the coded aperture. The Fourier spectrogram is a smoothed version of the Wigner function. While it increases the pixel counts compared to a light field, it might still lose information of the object due to the low-pass filtering. The coded aperture method has been shown able to reconstruct 3D incoherent samples. The random codes open two or more scanning positions of the Fourier spectrogram measurement at once, allowing the light in those areas to interfere and thus potentially capturing more information than Fourier spectrogram. While the random code design is a good choice, it opens a question of what is the best set of codes to use, especially for samples with arbitrary coherence instead of incoherent samples. In this chapter, we will use a close relative to the phase space function, the Wigner function, to address these concerns and propose a code design that efficiently captures the information of samples of arbitrary coherence.

5.1 Introduction

The Mutual Intensity (MI) contains identical information as the Wigner function because the Fourier transform connecting them is unitary. The MI further sheds light on understanding graphically the information captured by the experimental system, which can also be used to help design the code sequence. For the information (and its loss) in the Fourier spectrogram measurement, the convolution theorem is applied since the spectrogram is a 4D convolution of the Wigner function with the 4D measurement kernel. The MI, which is a 2D Fourier transform instead of the 4D one on the two spatial coordinates of the Fourier spectrogram, is sufficient to reveal the information loss.[1] The scanning aperture in the Fourier spectrogram

---

[1] We refer readers interested in applying a 4D Fourier transform to [173, 194] for the Ambiguity Function.

measurement selects and allows a small region in the MI to be measured, which results in most of the MI being missing and thus the information loss.

Measuring the MI without loss of information is important in measuring the phase space of light of arbitrary coherence. Arbitrary coherence makes the MI function have more freedom than that of a coherent light or incoherence sources. In contrast to the Fourier spectrogram, coded aperture measurements are able to measure the entire MI. A code sequence that has the following two properties satisfies this statement: 1) each of its codes opens a pair of apertures and 2) all possible pairs are traversed in the sequence. This sequence is equivalent to applying Young's two-pinhole interference experiments to all pairs. For a field with N sample points in 2D, the number of required pair measurements is $N \times (N-1)/2$. For instance, for a 2D field with dimension 100×100 which has a MI with 10^8 sample points, this method would require 5×10^7 measurements. Even if each measurement takes 1 ms, the entire measurement would still take 14 hours. We need to modify this naive sequence in order to make measuring the MI practical.

To design a more efficient sequence than the naive one, we review the coded aperture imaging. Coded aperture imaging was proposed to overcome the impracticality of lenses in applications like X-ray imaging and astronomy in the early days [166]. Various types of codes have been introduced [59, 166, 64, 71] but they mostly focus on single-shot applications. If light from different points in an in-focus scene is incoherent, the optical transfer function (OTF) of an imaging system can describe the measurement of the source light [79]. The code is related to the OTF through the autocorrelation. Those single-shot codes aim at improving the OTF of the imaging system by, for example, reducing the photon noise. Uniformly redundant arrays [59] and modified uniformly redundant arrays [76] provide flattened OTFs, that is, the frequency space of an intensity image has sampling weights as equal as possible [166]. It prevents the photon noise of larger-weight contents to beat the signals of lesser ones.

Among many useful arrays [64], we find that nonredundant arrays [70, 64, 71] are suitable for our MI acquisition. A nonredundant array code opens more than two apertures, and any pair of the apertures in the code has a unique difference vector. These properties make the acquisition able to recover the MI (thanks to the unique difference vector) and faster than the naive sequence above (because of > 2 apertures). Those difference vectors will be explained in the following section, which plays a central role in MI measurements. We then apply a design based on nonredundant arrays to simulated experiments for reconstructing MI. The result shows that we are able to capture most of the features of the ground truth MI. A constraint on the MI for fluorescence can be further applied to both simulations and experiments. The constraint, which contains prior information of the light, helps us reduce the required number of coded aperture measurements substantially, without losing the phase space information.
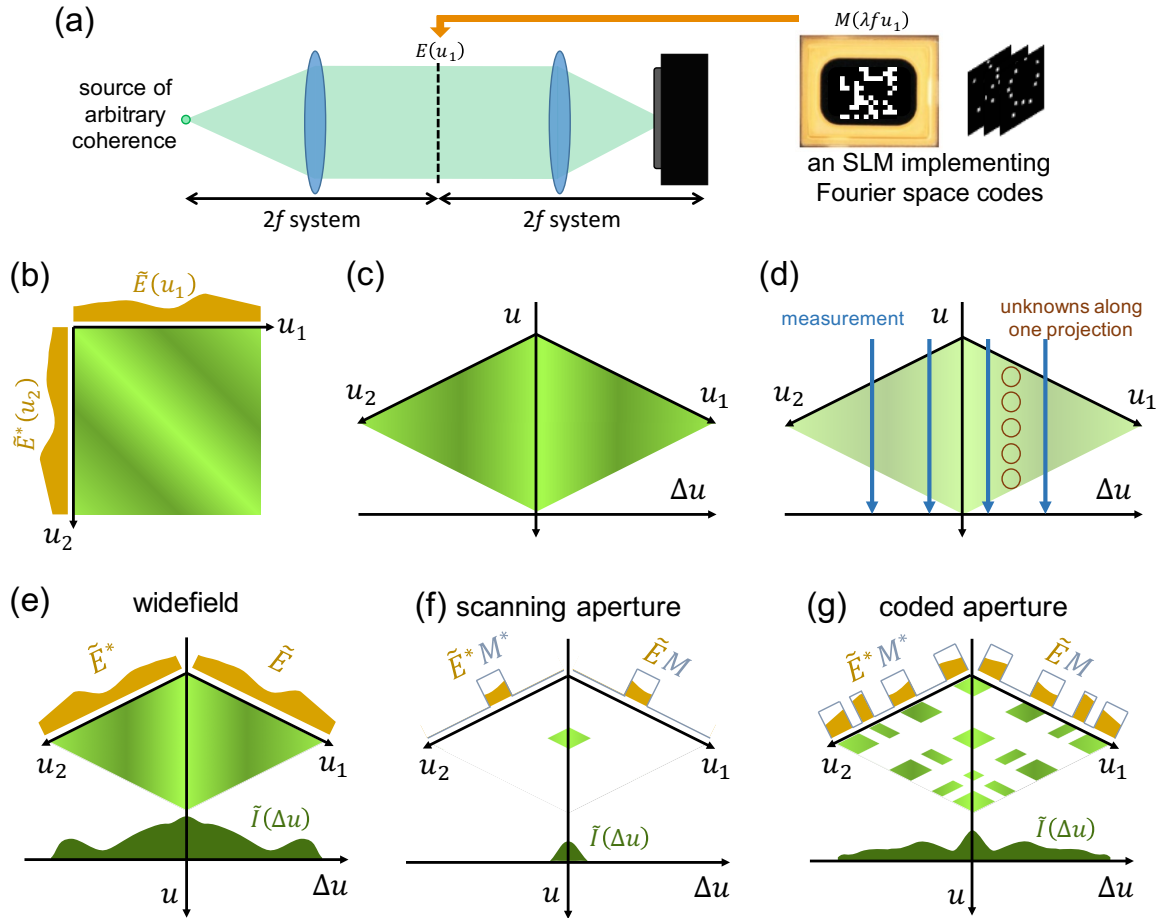
Figure 5.1: Schematic for mutual intensity measurement. (a) An example experimental
setup that measures the MI of sources of arbitrary coherence. The source has wavelength of
λ and the lenses have focal lengths f. In the middle of the $4f$ system, we have access to the
Fourier domain of the field at the front focal plane. A spatial light modulator (SLM) applies
codes in this domain. A camera captures intensity images at the back focal plane. (b) The
MI is constructed by the outer product of the field and its complex conjugate, followed by an
ensemble average (not shown) of the field. Coordinate u_2 is a duplication of u_1 of the electric
field. For illustration purpose, the field shown here is 1D which results in a 2D MI. (c) The
coordinate transformed version of the MI. (d) The projection of MI along $\boldsymbol{u}$ axis describes
the measurement. The intensity image captured by a camera is the inverse Fourier transform
of the projection. We define unknown MI samples in this rotated coordinate. A subset of
the unknowns along a projection line is indicated by brown circles. Multiple projections
with different codes is able to distinguish these unknowns. (e) The Fourier transform of
the widefield measurement, $\tilde{I}$, is depicted as a direct projection in the MI domain while
(f) the scanning aperture measurement and (g) the coded aperture measurement apply an
additional mask (outer product of the Fourier space code M) before the projection.

5.2 The Fourier Domain Mutual Intensity

As introduced in Sec. 2.3, the MI is an ensemble sum of the outer product of an electric field and its complex conjugate:

$$\tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u}) \triangleq \left\langle \tilde{E}^* \left(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2} \right) \tilde{E} \left(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2} \right) \right\rangle. \tag{5.1}$$

The angle bracket denotes the ensemble average and can be dropped if the field is coherent such as laser light. The MI is equivalent to the Wigner function by a Fourier transform as shown in the definition of the Wigner function (Eq. (2.2)). The MI can be defined either in spatial space or in spatial frequency space, which are related by Fourier transforms (see Fig. 2.4). We will focus on the MI in spatial frequency space since we apply aperture coding there. Unlike $E(\boldsymbol{u}_2)^* E(\boldsymbol{u}_1)$, Eq. (5.1) introduces a coordinate transform to the two arguments of the outer product. The transforming to the center-of-mass and differential coordinates helps reveal the local properties such as local coherence, revealed through the Wigner function, and the differential coordinate will be shown connecting the MI to intensity measurements. Figs. 5.1(b) and (c) show how a typical MI is constructed from an electric field and the coordinate transform of the MI.

It is important to represent the intensity measurements using the MI in order to measure the MI and to see how the MI helps analyze an imaging system. As shown in Eq. (2.4), the intensity image $I(\boldsymbol{r})$ is a projection of the Wigner function over the spatial frequency coordinate. Since the MI has both of its coordinates in spatial frequency, we Fourier transform the intensity image, arriving at

$$\tilde{I}(\Delta\boldsymbol{u}) = \iint I(\boldsymbol{r}) e^{-\mathrm{j}2\pi\Delta\boldsymbol{u}\cdot\boldsymbol{r}} \, \mathrm{d}^2\boldsymbol{r} = \iint \left(\iint W(\boldsymbol{r}, \boldsymbol{u}) \, \mathrm{d}^2\boldsymbol{u} \right) e^{-\mathrm{j}2\pi\Delta\boldsymbol{u}\cdot\boldsymbol{r}} \, \mathrm{d}^2\boldsymbol{r}. \tag{5.2}$$

Using Eq. (2.2) to undo the Fourier transform over $\boldsymbol{r}$ in the above formula, we have

$$\tilde{I}(\Delta\boldsymbol{u}) = \iint \left\langle \tilde{E}^* \left(\boldsymbol{u} - \frac{\Delta\boldsymbol{u}}{2} \right) \tilde{E} \left(\boldsymbol{u} + \frac{\Delta\boldsymbol{u}}{2} \right) \right\rangle \, \mathrm{d}^2\boldsymbol{u}$$

$$= \iint \tilde{\Gamma}(\Delta\boldsymbol{u}, \boldsymbol{u}) \, \mathrm{d}^2\boldsymbol{u}. \tag{5.3}$$

Eq. (5.3) states that the Fourier transform of the intensity image is a projection of the MI along $\boldsymbol{u}$ axis. Fig. 5.1(d) depicts the measurement as a projection of the MI.

5.2.1 Intensity measurement with coded apertures

When a code M is applied in the Fourier space (see Fig. 5.1(a)), the electric field is modified as

$$\tilde{E}(\boldsymbol{u}_1) M(\lambda f \boldsymbol{u}_1). \tag{5.4}$$

Here the coordinate of the code is $\lambda f \boldsymbol{u}_1$ according to the theory of Fourier optics [72]. By following Eqs. (2.9) to (2.12), we can have the relation between the intensity image and the MI. We further Fourier transform the intensity image and assign an index n for the codes $(I \rightarrow I_n, M \rightarrow M_n)$, arriving at, from Eq. (2.12),

$$\tilde{I}_n(\Delta \boldsymbol{u}) = \iint \tilde{\Gamma}(\Delta \boldsymbol{u}, \boldsymbol{u}) M_n \left(\lambda f(\boldsymbol{u} + \frac{\Delta \boldsymbol{u}}{2}) \right) M_n^* \left(\lambda f(\boldsymbol{u} - \frac{\Delta \boldsymbol{u}}{2}) \right) \mathrm{d}^2 \boldsymbol{u}. \qquad (5.5)$$

Without applying a code, the intensity image is a widefield measurement (Fig. 5.1(e)). If we apply an aperture in Fourier space as shown in Fig. 5.1(f), the projection captures a patch of the MI. If two (or more) apertures are applied, the projection will capture four (or more) patches of the MI because of the outer-product nature of the MI (see Fig. 5.1(g) and Figs. 5.2(b)-(d)). Note that the Fourier spectrogram measurement is depicted in Fig. 5.1(f), whose projection comes from only the area near the $\boldsymbol{u}$ axis. Hence scanning an aperture in the entire Fourier space does not cover areas far from the $\boldsymbol{u}$ axis and the Fourier spectrogram loses the MI information of those areas. In contrast, the project of a coded aperture can capture those areas which contain higher $\Delta \boldsymbol{u}$ components of the MI (See Fig. 5.1(g)), resulting in enhanced resolution. Nevertheless, it does not guarantee invertibility to merely contain the information. The codes need to be designed carefully in order to reconstruct the MI from its lower dimensional projections.

5.2.2   Unknown samples in the mutual intensity

The projection property of the real space intensity measurement leads us to define the complex-valued unknown samples on the projection line (or in the projection area when discussing 4D MI). See Fig. 5.1(d) for a subset of the unknowns that are on the a projection line and Fig. 5.2(a) for those of a 2D MI. However the actual unknowns to solve depends on the incoming light. For example, the MI is constant along the $\boldsymbol{u}$-axis if we image in-focus incoherent sources only [87], which makes each projection line have only one known to solve for. Uniformly redundant arrays [59, 76] are useful in this case because it has a stronger signal than scanning aperture and provides almost-equal weights on all projection lines, which avoids the noise from one projection line to beat the signal of another. The redundancy means how many samples on a projection line are sampled and uniformity means each line has about the same number of sampling. Fig. 5.2(b) shows an example of applying an aperture code designed from a modified uniformly redundant array. The bottom part of Fig. 5.2(b) shows the weight of each projection line. These weights are the OTF of the system for imaging in-focus sources and can be generalized to 3D OTF when defocus is considered in the codes [173].

When the MI varies along $\boldsymbol{u}$-axis and the incoming light has arbitrary coherence, the actual unknowns may be all the samples. Note that the reverse in $\Delta \boldsymbol{u}$ and the complex conjugation make the MI equal to itself. Hence the actual number of unknown samples is $N(N-1)/2$ where N is the number of samples in the electric field. We can use two-point
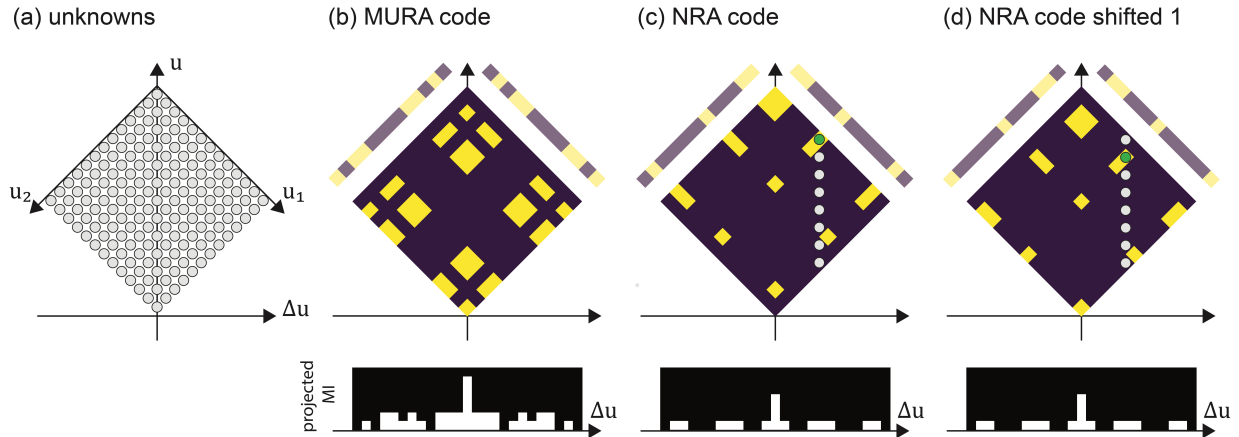
Figure 5.2: Measuring the MI of 1D fields with uniformly redundant array (URA) and nonredundant array (NRA). (a) The MI schematic for the other 3 subfigures. The unknowns of the MI function are indicated by circles. Here we divide the 1D field into 13 "squares," resulting in 169 unknowns in the MI. (b) A diagram indicates the measurement with a code generated from a modulus-13 modified uniformly redundant array (MURA). The two copies of the 1D MURA are shown as the rotated strips. The transparent part of the code is yellow. The MI is shown as the rotated square image with the MI patches selected by the code being yellow. The measurement as a projection along $\boldsymbol{u}$ axis is shown in the bottom part. (c) For a code generated from a nonredundant array (Singer(13,4,1)). The unknowns along a projection line are overlaid with the green dot showing the one that is measured. (d) When shifting the pattern by 1 square, we can capture a different unknown along the projection line.

scanning, *i.e.* Young's 2-pinhole experiments, to probe all the unknown samples, which results in $O(N^2)$ pairwise measurements. We find that the nonredundant array is useful for this task since it measures one sample per projection line from most of the lines. Because each projection line is independent, the measurement with nonredundant arrays corresponds to simultaneously measuring multiple samples at once (see Fig. 5.2(c)). Furthermore, we can capture different unknown samples of the MI by shifting such code by one sample, as shown Fig. 5.2(d). We will discuss the code design with nonredundant arrays in the next section.

5.3   Nonredundant Array Designs

In order to introduce the design without ambiguity, we define the following two terms:

- A *square* is a segment on a 1D electric field or a square area on a 2D one that is transparent to light, thus allowing us to capture light in its area.

- A *patch* is a small 2D square area (or a 4D cube) in the MI, which is selected by two

1D (2D) squares. The outer product in the definition of the MI requires two squares to select a patch.

The unknowns referred to in the following always belong to patches.

Let us first assume each patch contains one complex-valued unknown. Nonredundant arrays can capture one patch along each of the projection lines because any pair of squares in the array has a unique positional difference and each projection line corresponds to a unique difference coordinate $\Delta \boldsymbol{u}$. A type of sets called *cyclic difference sets* can have this uniqueness. A difference set has three parameters, (v, k, λ), where v is the modulus of the set, k is the number of elements in the set and λ is the number of occurrence of each nonzero difference. We assume the difference set parameter $\lambda = 1$ in this chapter unless otherwise noted. The difference operation among the k elements is subject to modulo v, thus the allowed differences range from 0 to $v - 1$. Singer proposed a classical example of this type of sets [163] (termed Singer sets) and those sets can be used in coded aperture measurements [64]. We can find these sets computed in [74, 75]. An example of the Singer set with parameters $(v, k, \lambda) = (91, 10, 1)$ is given below:

$$\text{Singer}(91, 10, 1) = \{0, 1, 3, 9, 27, 49, 56, 61, 77, 81\} \text{ with modulus 91.}$$

Readers can check for themselves that the difference of each pair is unique, including negative differences modulo 91 (*e.g.* $(1 - 9) \bmod 91 = (-8) \bmod 91 = 83$). Moreover, every number strictly less than the modulus appears exactly once in the difference [92], except 0 which obviously appears k times. We will use this type of set to design code sequences in both 1D and 2D cases.

The code sequence for measuring the MIs of 1D fields is designed by following procedure:

1. Pick a Singer set with modulus v and set it to be $\mathcal{S}$.

2. Section the 1D pupil into n_x equally sized squares where $n_x \leq v$.

3. Label them $0, \ldots, n_x - 1$ serially.

4. Generate a code with squares set open if their serial numbers are in $\mathcal{S}$.

5. Add 1 to each element in $\mathcal{S}$ and replace each by itself modulo v. This results in circular shift of the code and can probe different unknowns of the MI (see Figs. 5.2(c) and (d))

6. Repeat steps 4 and 5 until there are v codes.

Figs. 5.2(c) and (d) illustrate what in the MI are captured by these circular-shift patterns. Two observations are made: 1) the design traverses all MI projection lines because every difference less than the modulus appears in the set, and 2) all blocks on a line are visited by the shifting. They together guarantee the coverage of the entire MI. The v^2 unknown samples of MI that arise from pupil sectioning and MI outer product, are recoverable by these v codes.
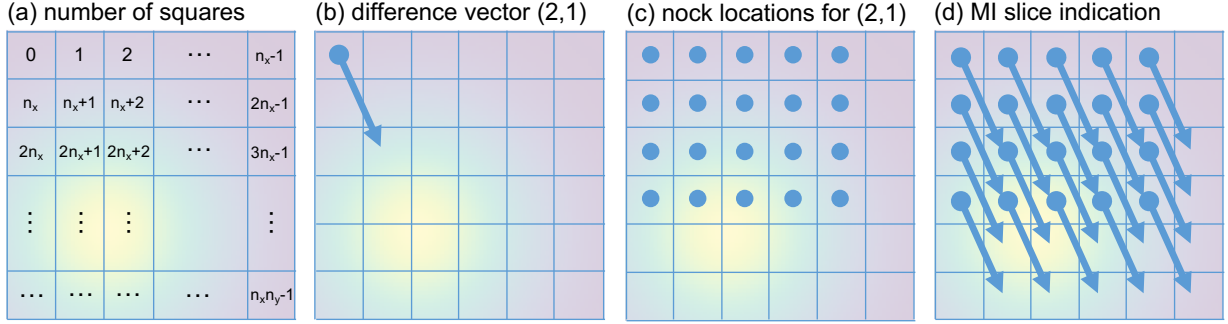
Figure 5.3: Illustration of a slice of the 4D MI with the 2D field coordinates. Four subfigures are in $\boldsymbol{u}$-coordinates of electric fields and the displayed are 2D electric fields, not to be confused with the 2D MI of 1D electric fields. (a) Serial numbers for squares used in code design. (b) Shows a difference vector (2,1). (c) Circles show the locations of the roots of difference vector (2,1) such that MI is nonzero if coordinate $\boldsymbol{u}_2$ of MI is at one of the circles and $(\boldsymbol{u}_1 - \boldsymbol{u}_2)$ equals to the difference (2,1). (d) Visualization of the 4D MI slice selected by difference vector (2,1). The slice is the extension to a projection line in Fig. 5.1(d) which corresponds to a fixed difference between $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$.

5.3.1   2D code design with a nonredundant array

The 2D code design is largely identical to its 1D counterpart:

1. Pick a Singer set with modulus v and set it to be $\mathcal{S}$.

2. Section the 2D pupil into $n_x n_y$ squares where n_x (n_y) is the number of squares in x (y) direction, $n_x n_y \leq v$, and n_x is not necessarily equal to n_y.

3. Label them $0, \ldots, n_x n_y - 1$ serially. A square at row i and column j, both starting from 0, has its serial number $i \times n_x + j$. (see Fig. 5.3(a))

4. Generate a code with squares set open if their serial numbers are in $\mathcal{S}$.

5. Add 1 to each element in $\mathcal{S}$ and replace each by itself modulo v.

6. Repeat steps 4 and 5 until there are v codes.

The difference vector $\Delta\boldsymbol{u}$ of a pair in a code becomes 2D and each difference vector is still unique in that code. The uniqueness can be proven by contradicting the assumption that there are two identical difference vectors from two pairs. The two pairs would have the same serial-number difference, thanks to the rectangle lattice of the squares in step 3. Identical serial number differences contradict the property of difference sets so a difference vector must be unique in the code containing it.

In 2D field cases, a projection line is generalized to a MI slice corresponding to a certain 2D difference coordinate $\Delta\boldsymbol{u}$. An example is illustrated in Figs. 5.3(b) to (d); note that the
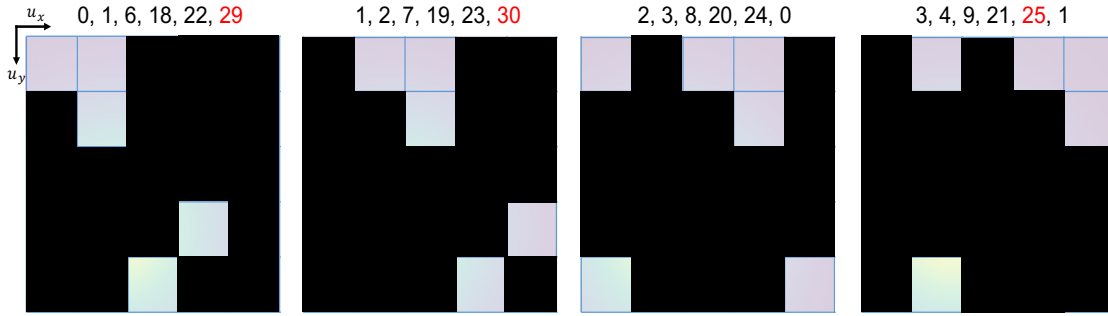
Figure 5.4: Four out of 31 codes designed with Singer(31,6,1)=$\{0, 1, 6, 18, 22, 29\}$. The axes
of each of the 4 subfigures are u_y vertically and u_x horizontally for 2D electric fields, not
to be confused with the 2D MI of 1D electric fields. The area of 2D field is sectioned into
25 squares, resulting in $25 \times 24/2 = 300$ complex-valued unknowns in the MI. The numbers
above each code is the set used in generating the code (they are related by shifting 1 and
modulo 31). The red numbers are out of range (0 to 24) and hence there are no corresponding
squares in the designed codes.

MI is 4D and abstract so we use the 2D electric field and arrows to indicate the MI slice in
the illustration. Two observations similar to those in the 1D case can still be made: 1) All
the serial-number differences are traversed and hence, after rasterizing, all the 2D differences
are visited. 2) Consider a certain 2D coordinate difference. If a pair of apertures has the
that difference, the circular shift of the design will make every arrow in Fig. 5.3(d) visited
by this aperture pair or its shifted version. A 2D example of codes designed with a Singer
set is shown in Fig. 5.4.

5.3.2   Design for multiple unknowns in a patch

However, it is possible that a square can be large enough to allow significant variation of
MI values within itself. One way to reduce this variation is to have finer squares, but this
results in increased number of codes, impacting the measurement time. We choose another
approach to address this variation, assuming there are multiple samples in a square. The
multiple subsquare samples result in multiple unknowns in a patch. The unknowns within
the same patch and along the same projection line are not distinguishable if we do not
redraw the squares. Therefore we lump such unknowns into one unknown where the lumped
unknown is the sum of those. See the five blue boxes in Fig. 5.5(b) for an example of five
lumped unknowns.

For cases with a single unknown in each patch, the code design above is suitable since
the linear equations directly assign the constants of a linear system to the unknowns except
those on the zero-difference projection line. However the design is only suitable for lumped
unknowns that have no crosstalk with their neighboring MI patches (see red and green boxes
in Fig. 5.5 for crosstalk and non-crosstalk, respectively). We define crosstalk as the situation
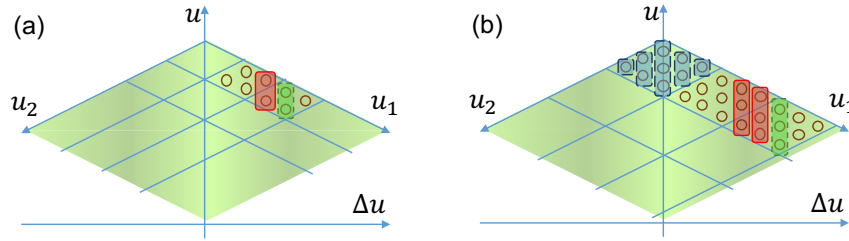
Figure 5.5: Diagrams showing cases of multiple unknown samples behind a patch in MI. A subset of unknown samples are shown as brown circles. (a) 2 unknown samples in a square, resulting in 4 samples in a patch. (b) 3 unknown samples in a square, resulting in 9 samples in a patch. The red boxes (solid boundary) indicate where the mixture between two patches that have different central differences ($\Delta\boldsymbol{u}$'s) happens while the green boxes (short dashed boundary) indicate unknowns without such mixture. Each blue box (long dashed boundary) represents a lumped unknown since its ingredients are not distinguishable.

that *the unknowns in patches with different central differences $\Delta\boldsymbol{u}$ are present in the same system of linear equations.* Here the central difference of a patch refers to the average of $\Delta\boldsymbol{u}$'s of all samples of the patch. The existence of crosstalk requires the code design above to be modified.

Despite the fact that lumping the unknown samples makes a linear system have fewer unknowns to solve, crosstalk emerges from having multiple samples in a square and resulting in poor conditioning of the inverse problem for those crosstalking lumped unknowns. We use a separating approach to resolve this crosstalk problem. On top of the codes introduced in Sec. 5.3.1, which will be termed *all-square-used* type, we add other codes that avoid such crosstalk by using alternate squares for 1D fields (*alternate-square* type) or squares in alternate columns/rows/both for 2D fields (*alternate-column, alternate-row, alternate-both* types) in the design. Only those selected squares are given serial numbers when generating the codes of the same type. The idea is to solve for the fully separated lumped unknowns first (those in alternate-square in the 2D MI or those in alternate-both in the 4D MI), and then solve linear systems that include others. This method by no means is the most efficient way but it is more efficient than the $O(N^2)$ two-point measurements and always provides full-rank systems of linear equations. For illustration purposes, let us define a square to be odd-number-labeled if it has an odd serial number in the design given in Sec. 5.3 (*i.e.* the all-square-used design), and similarly even-number-labeled; for 2D fields, we have those with suffixes '-in-x' and '-in-y': odd-number-labeled-in-y, even-number-labeled-in-y (see squares in Fig. 5.3 with coefficients of n_x being odd/even), odd-number-labeled-in-x and even-number-labeled-in-x (see squares in Fig. 5.3 with offset terms being odd/even). With alternate-square and alternate-both, the patches having one of its squares being odd-number-labeled and the other even-number-labeled are absent, making patches that share boundaries not concurrently present. Hence the patches in alternate-square in 1D field cases (alternate-both in 2D cases) do not have crosstalk and their lumped unknowns can be solved by assigning

the measured values to the lumped unknowns. After that, we can solve for the unknowns in the other codes, all-square-used in 1D field cases (alternate-column, alternate-row and all-square-used in 2D field cases). Note that the assigning above is conceptual, just to show the the solving ability, and we use matrix inversion to solve small linear equations, each of which has unknowns on a projection line (a projection plane in 2D field cases) only. We summarize the code design extension in the following list. For measuring the MIs of 1D fields:

- (all-square-used) Identical to the procedure introduced in Sec. 5.3.

- (alternate-square) Select only odd-number-labeled squares in step 3 of the all-square-used procedure and generate codes. Select only even-number-labeled squares and repeat the generation.

For measuring the MIs of 2D fields:

- (all-square-used) Identical to the procedure introduced in Sec. 5.3.1.

- (alternate-column) Select only odd-number-labeled-in-x squares (regardless of their labels in y) in step 3 of the all-square-used procedure and generate codes. Select only even-number-labeled-in-x squares (regardless of their labels in y) and repeat the generation.

- (alternate-row) Select only odd-number-labeled-in-y squares (regardless of their labels in x) in step 3 of the all-square-used procedure and generate codes. Select only even-number-labeled-in-y squares (regardless of their labels in x) and repeat the generation.

- (alternate-both) Select only 1) odd-number-labeled-in-x and odd-number-labeled-in-y squares in step 3 of the all-square-used procedure and generate codes. Repeat the generation for 2) even-number-labeled-in-x and odd-number-labeled-in-y, 3) odd-number-labeled-in-x and even-number-labeled-in-y, and 4) even-number-labeled-in-x and even-number-labeled-in-y.

Note that because the numbers of squares in different cases above are different, we can pick different Singer sets for reducing the number of measurements. Furthermore, since the linear systems other than alternate-both and alternate-square are no longer assigning unknowns to to their constant terms, we can use sets with $\lambda > 1$ to improve light throughput as long as the systems remain full-rank.

5.4 MI Reconstruction with Singer Set Designs

The reconstruction is to solve systems of linear equations for all projection lines where the unknowns of a linear system are the lumped unknowns on a line. Each row of the matrix of a linear system corresponds to a code while each column corresponds to a lumped
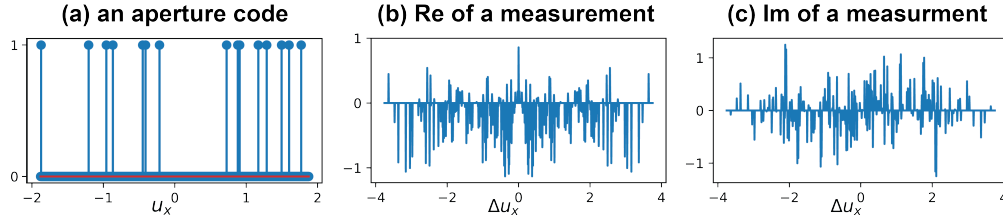
Figure 5.6: One of the simulated measurements for the 1D field MI. The MI is simulated with a pseudo random generator. (a) An aperture code used for probing the simulated MI. (b) The real part of the MI projection where the latter stands for intensity measurement in real space. (c) The imaginary part of the MI projection. Note that the projection in (b) is even and that in (c) is odd in order to have a real-valued inverse-Fourier-transformed image. The vertical axes of plots have arbitrary units while u_x and Δu_x have units of $(\mu\mathrm{m})^{-1}$.

unknown. The matrix coefficients are determined by the codes. If the difference vector associated with unknown indexed by j is present in code i, then the (i,j)-th component of the matrix is 1, otherwise 0. The i'th constant term in the linear system is obtained by Fourier transforming the intensity image corresponding to code i and taking the frequency component corresponding to the difference vector of this linear system. The MI is reconstructable if linear systems of all projection lines are full-rank. After solving for the lumped unknowns, the unknowns constituting a lumped unknown share the averaged value of their lumped one. We apply reconstruction with the code design introduced in Sec. 5.3.2 to simulated measurements in 1D and 2D and perform digital refocusing to demonstrate that the reconstructed MI is equivalent to the Wigner function.

5.4.1   1D field simulation

We simulate the 2D MI of a 1D field that has multiple statistical components in its ensemble sum. We directly simulate the MI randomly without specifying the field components.[2] The axis of the electric field is divided into 576 squares that can be coded, resulting in $331776 = 576 \times 576$ patches in the MI. Each square contains 2 sample points for the field, therefore $4 = 2 \times 2$ sample points under a MI patch. The spacing for the samples is 3.263×10^{-3} $\mu\mathrm{m}^{-1}$. The 4 samples in a patch can reveal how the crosstalk affects the reconstruction. The code design follows the discussion in Sec. 5.3.2 with Singer(651,26,1) for the all-square-used type and Singer(307,18,1) for the alternate-square types, resulting in 1265 measurements. An aperture code and its associated projection are shown in Fig. 5.6.

---

[2]The MI is simulated by using a pseudo random generator to draw two 1152×1152 grids from Uniform$(0,1)$, one for real and the other for imaginary parts, followed by a circular convolution of a 2D Gaussian function which has unity amplitude and a standard deviation of 30 pixels. The real and imaginary parts of the MI are scaled and offset to have a range of -0.5 to 0.5. The pseudo random MI is generated from NumPy [144], with Python programming language codes `numpy.random.seed((13,14,7,41))` and `numpy.random.rand(1152,1152)+1j*numpy.random.rand(1152,1152)`.
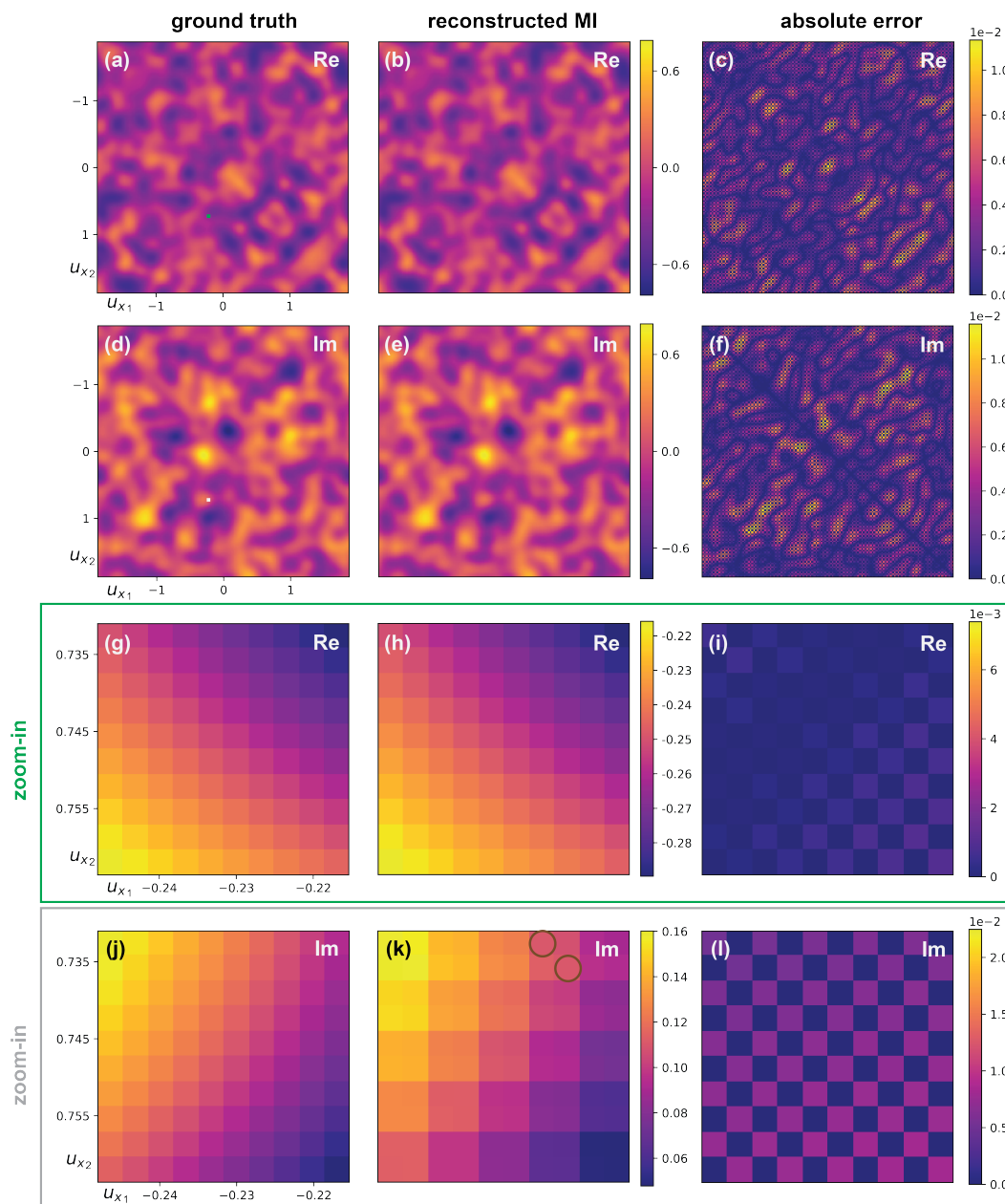
Figure 5.7: Reconstruction for the 1D field simulation. The left column shows the ground truth of the simulation, the middle one shows the reconstructed result and the right one shows the absolute values of the reconstruction errors for real and imaginary parts. (a)(b)(c) Show the real parts in the whole simulation space. (d)(e)(f) Show the imaginary parts in the whole simulation space. (g)(h)(i) Show the real parts of the zoomed-in region of the green square in (a). (j)(k)(l) Show the imaginary parts of the zoomed-in region of the white square in (d). We further use the two brown circles in (k) to indicate an example of unknown samples whose values are not resolved due to the projection nature of measurement. All the axes have units of $(\mu m)^{-1}$.

The reconstruction result is shown in Fig. 5.7. We can see that the reconstruction is visually the same as the ground truth. We indicate two samples that are along the measurement projection by brown circles in Fig. 5.7(k). The reconstruction errors for these two samples are not zero because they cannot be distinguished under the code scheme. However for their common neighboring samples, there is no ambiguity and thus the reconstruction has zero error. The above explains the checkerboard feature in Figs. 5.7(i) and (l) The root-mean-square of pixel-wise percentage errors is 5.4%.

5.4.2   2D field simulation

We simulate a 2D field that has two statistical components in its ensemble sum. The two axes, u_x and u_y, of the electric field are divided into 24 parts in each, resulting in 576 squares for the entire field that can be coded. Each square contains $9 = 3 \times 3$ samples, which results in a 4D MI of shape $(72 = 24 \times 3, 72, 72, 72)$ and a MI patch of shape $(3,3,3,3)$. The spacing for the samples is 5.221×10^{-2} μm^{-1}. The field and a slice of its MI are shown in the upper part of Fig. 5.8. We also show in Figs. 5.8(d) and (e) the measurement from a code. The sets used to generate the code sequence are not all nonredundant because the openings in a code are too few and might impact the experimental measurement signal-to-noise ratio. As the lumped unknowns are introduced, the reconstruction becomes solving a linear equation for each projection line instead of directly measuring a MI patch. We can use arrays with mild redundancy as long as each linear system is nonsingular. The set we use for the all-square-used codes is Singer(651,26,1), which is a nonredundant array, those for the alternate-column and alternate-row codes are Singer(341,85,21) which has a redundancy of 21, and those for alternate-both codes are Singer(156,31,6) which has redundancy of 6. This results in 2639 measurements[3]. The MI becomes 4D in 2D field cases so we show in Fig. 5.9 the reconstruction of the slice in 5.8(c). Another reconstruction with different 2D slicing way of the 4D MI (a slice of $\boldsymbol{u}_2$) is shown in Fig. 5.10.

The reconstruction takes 25 minutes on a 2-GHz 24-core Intel® Xeon® CPU (E5-2620) but the computation is not fully parallelized here. The reconstruction error is higher because there are more lumped unknowns per patch than the 1D field case. The root-mean-square of pixel-wise percentage error is 5.47%. From Fig. 5.9(c) we can see that the reconstructed MI holds constant within a u_x square. The constant value is the average value of the MI patch to which the square belongs. However this behavior is not shown along other cut lines that is not parallel to the measurement projection (see Figs. 5.10(d) and (e)).

5.4.3   2D field simulation for point sources in 3D

We apply the coded aperture measurements to the field generated by a collection of point sources. We want to demonstrate that our method can reconstruct the Wigner function and

---

[3]as a result of 651 for all squares, 341×2 for alternate columns, 341×2 for alternate rows and 156×4 for alternate both columns and rows
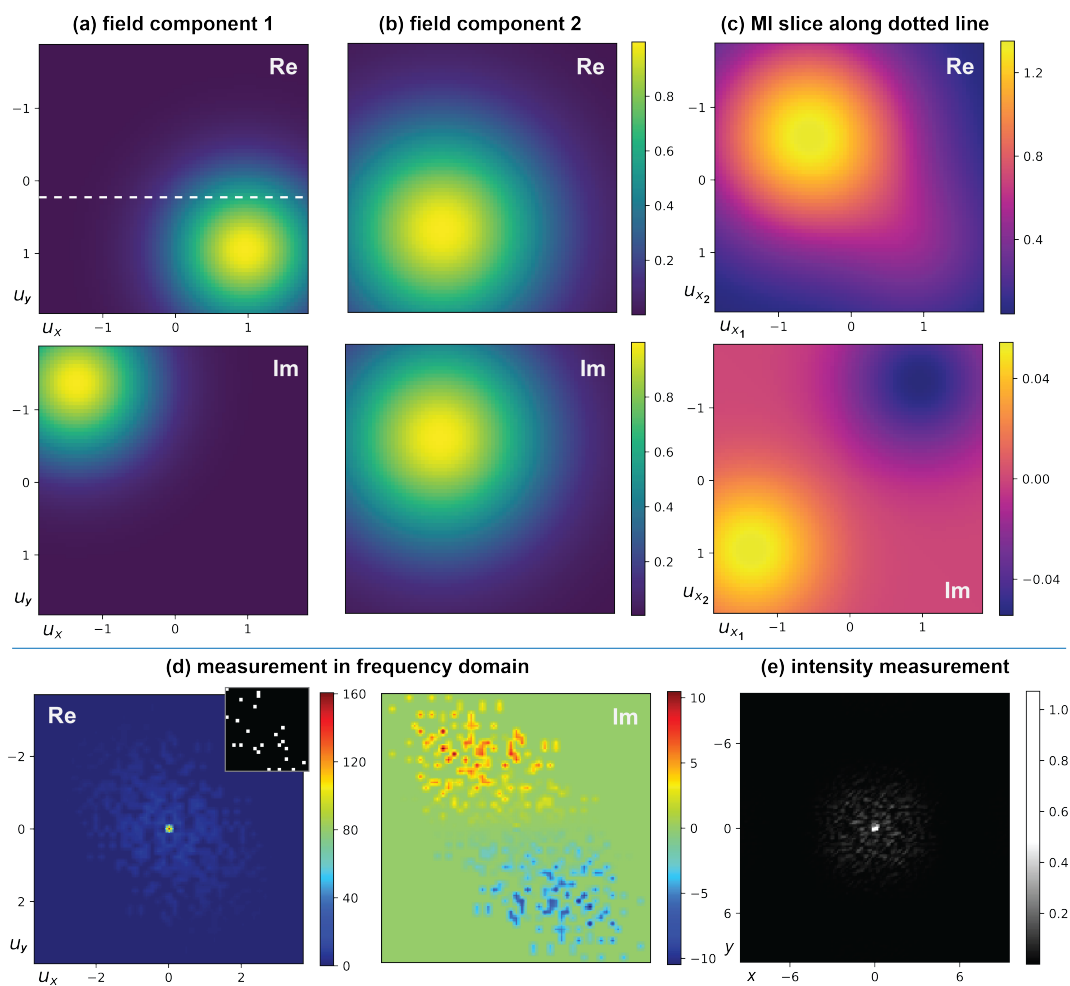
Figure 5.8: Simulation for a 2D field case. (a)(b) The two statistical components of the 2D field. (c) The MI of the field along the dotted line. (d) The measurement in frequency domain for a code used (inset). (e) The intensity image of the measurement, which is the inverse Fourier transform of (d). All color bars are with arbitrary units. The units of x, y are μm and those of u_x and u_y are $(\mu$m$)^{-1}$.
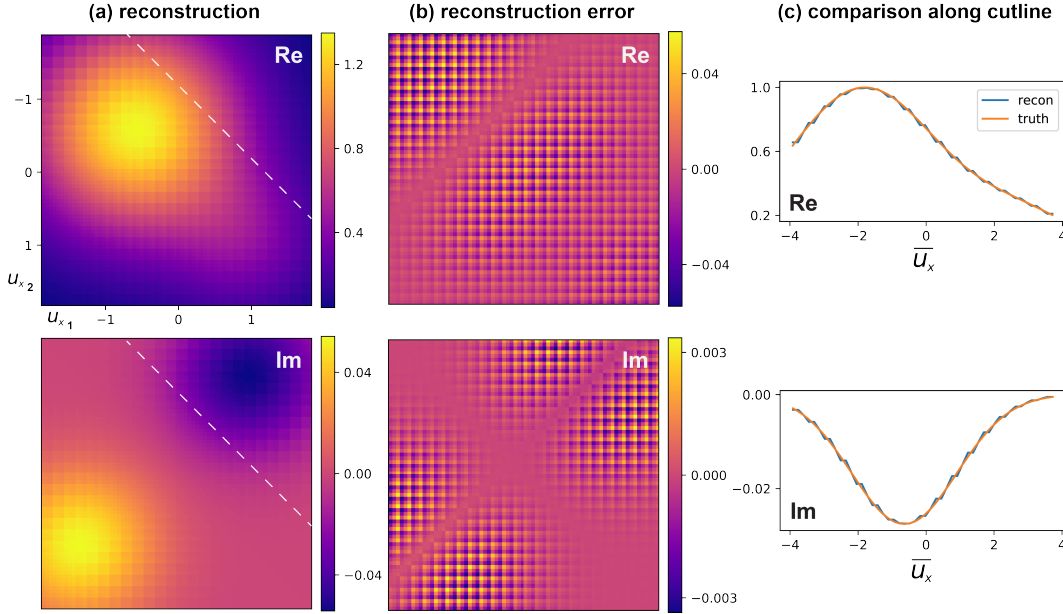
Figure 5.9: Reconstruction for the 2D field simulation. The reconstructed MI slice showed here is the one in Fig. 5.8. Column (a) the real and imaginary parts of the reconstructed MI slice. Column (b) the the reconstruction error. All color bars are with arbitrary units. (c) Plots for comparing the reconstructed MI to the ground truth along the cutline in column (a). The cutline is along the measurement projection of the MI. The units of u_x's and $\overline{u_x}$ are $(\mu\mathrm{m})^{-1}$.

the reconstructed phase space can be used in refocusing, as a light field can. We simulate 29 point sources (wavelength=0.5 μm) within a 10 μm $\times$ 10 μm $\times$ 10 μm space as shown in Fig. 5.11(a). In the same figure we also show some simulated measurements. The three simulation spaces, r for intensity images, u for Fourier transform of the field and Δu for the Fourier transforms of intensity images are set as follows: 229 samples in each direction of r space with sampling spacing 0.435 μm, 115 samples in each direction of u space with sampling spacing 0.01 μm^{-1}, and 229 samples in each direction of Δu with sampling spacing 0.01 μm^{-1}. The Fourier transforms of the intensity images are used in the reconstruction algorithm. The codes are designed under the setting that each of u_x and u_y is divided into 23 parts, resulting in 529 squares and $25 = 5 \times 5$ samples under each square. We increase the light throughput of the codes by allowing mild redundancy in the difference sets and also modify one set in order to have full rank in each linear system. The sets used to generate the codes are Singer(553,24,1) for all-square-used codes, Singer(307,18,1) for alternate-column and alternate-row codes, and a modified Singer(156,31,6), crossing out an element for full-rank, for alternate-both codes,

$$\{0, 1, 2, \cancel{4}, 14, 18, 21, 22, 30, 31, 37, 42, 45, 49, 51, 55, 56,$$
$$60, 76, 82, 85, 87, 88, 93, 95, 98, 108, 110, 117, 134, 142\}.$$
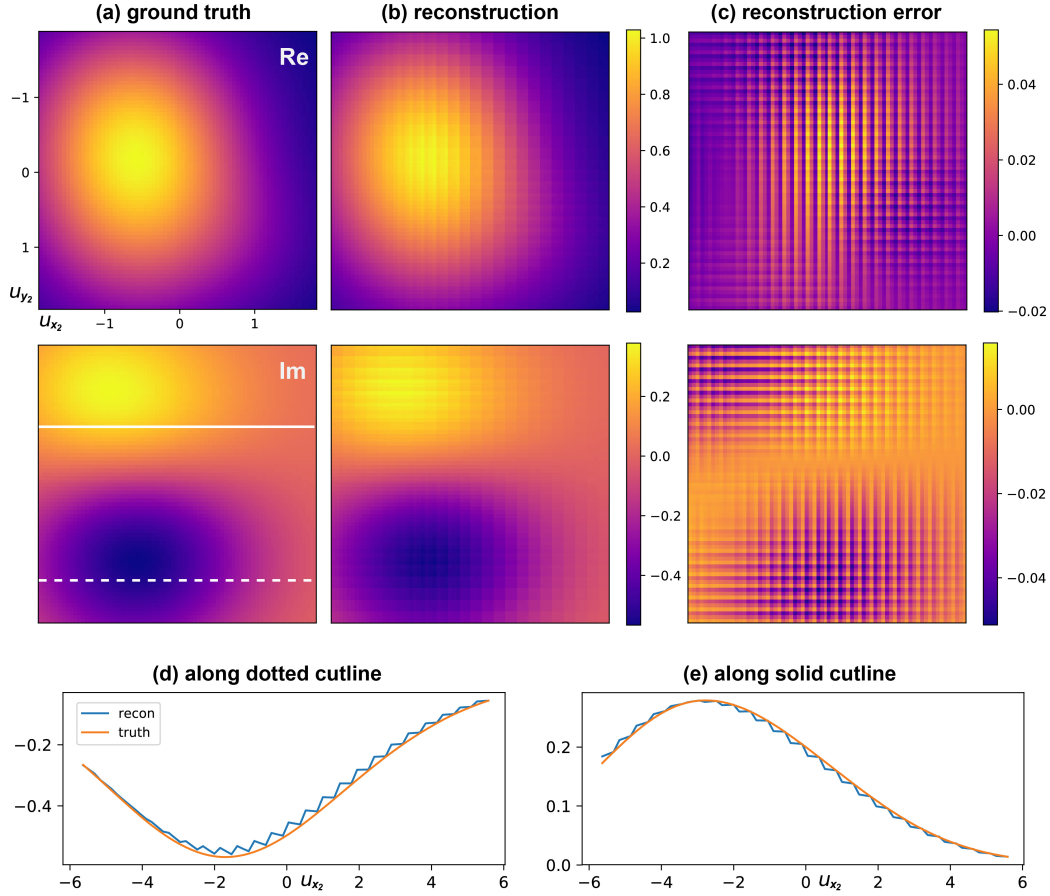
Figure 5.10: Another reconstruction for the 2D field simulation. Column (a) the real and
imaginary parts of the ground truth. (b) Those of the reconstructed MI slice. Column (c)
the the reconstruction error. All color bars are with arbitrary units. (d)(e) Comparison of
the imaginary part of the reconstructed MI and the ground truth along two cut lines shown
in (a). The unit of u_x and u_y are $(\mu\mathrm{m})^{-1}$

This results in 2405 measurements. One remark to be noted is that our method above
without priors, is applicable to cases with much more point sources and with multiple co-
herent/partially coherent fields, all present at the same time.

In the reconstruction, we additionally apply the prior knowledge that the fields are gener-
ated by 3D point sources, which restricts the bandwidth of the MI along each measurement
projection line [87]. Within each projection line, the prior is applied through a 400-iteration
Gerchberg-Saxton algorithm between the bandwidth prior and the solved MI patches. We
use PyWren [91] for parallel computation on the Amazon cloud server.

The reconstructed MI is further transformed into Wigner functions, as shown in Fig. 5.12.
The reconstruction error is relatively small compared to the value of the Wigner function.
As shown in Fig. 5.13, we apply the Wigner function, converted from the measured MI, to
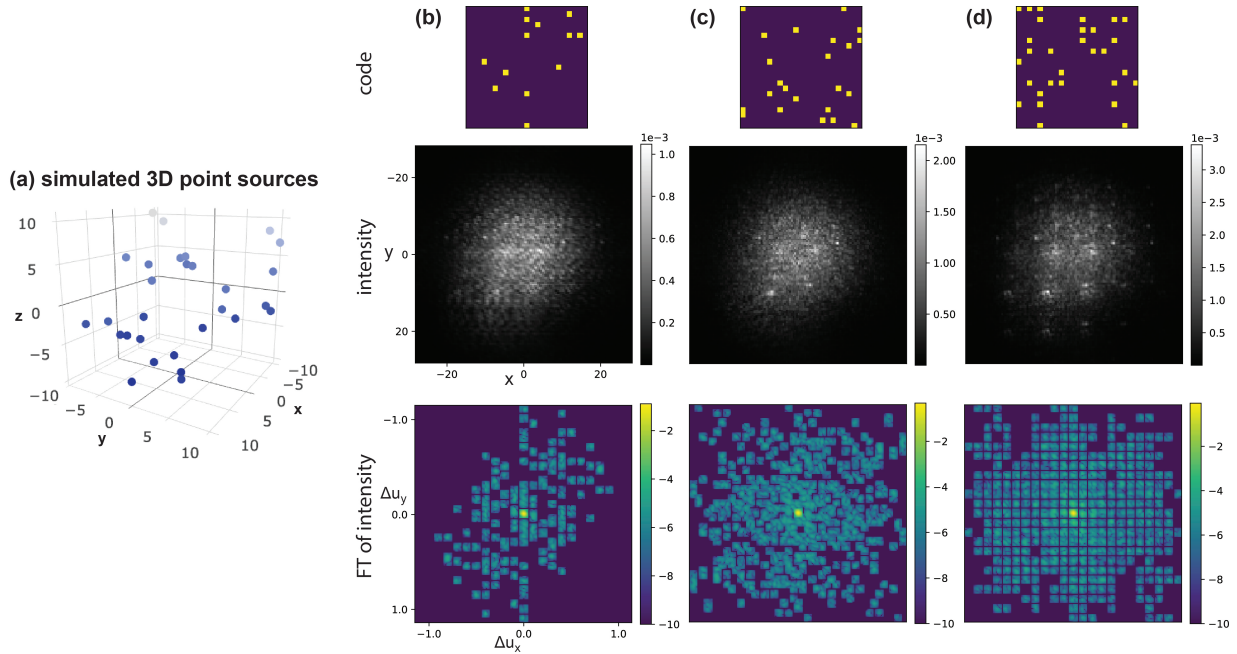
Figure 5.11: Simulation for a 2D field generated by a collection of 3D point sources. (a) The simulated distribution of the point sources. Columns (b)(c)(d) show some codes, the measurements in $x - y$ space and the Fourier transform (FT) of the measurements: (b) An example of alternate-column case, (c) that of all-square-used case, and (d) that of alternate-both. All color bars are with arbitrary units. The units of x, y and z are μm and those of Δu_x and Δu_y are $(\mu \text{m})^{-1}$.
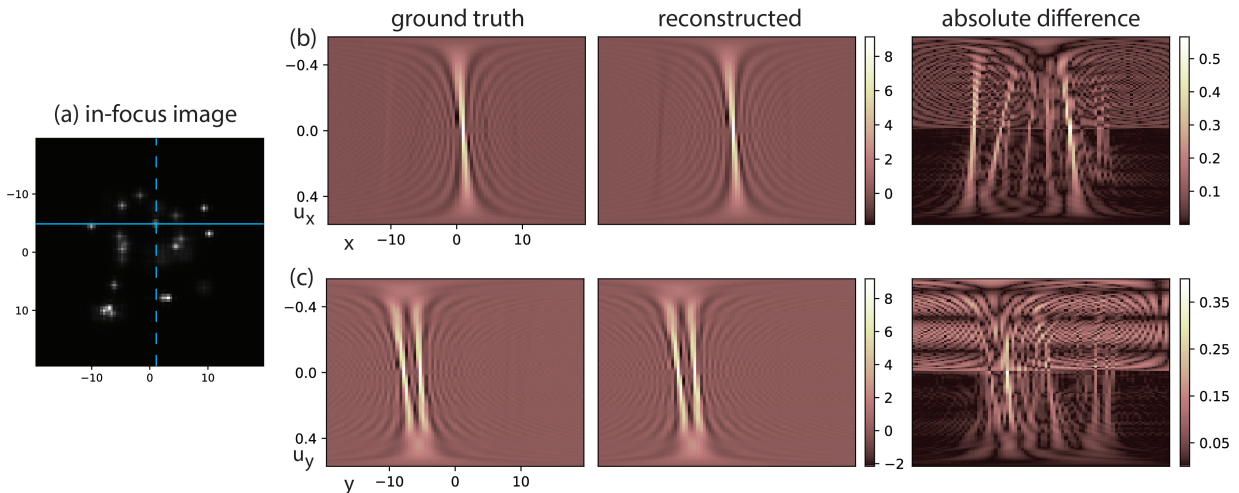
Figure 5.12: Slices of the Wigner function of the field generated by 3D point sources. (a) In-focus intensity image of the 3D point sources. Row (b) the ground truth, reconstructed Wigner function and their difference along the solid cutline and row (c) those along the dashed cutline. The units of x, y are μm and those of u_x, u_y are $(\mu \text{m})^{-1}$.
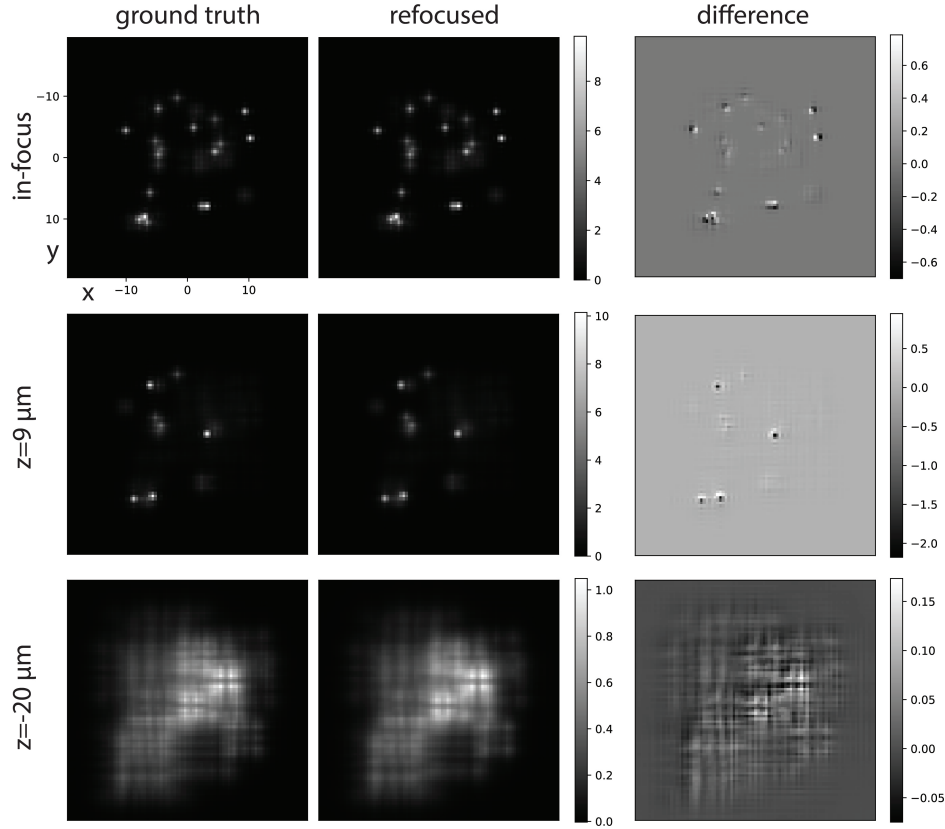
Figure 5.13: Refocused images from the reconstructed phase space function for the simulated 3D point sources. The left column lists the ground truth images at different focuses while the middle column lists the refocused images at the corresponding focuses. The each row of the two columns shares the color bar. The right column lists the error between the ground truth image and the refocused one. Note that the total aperture space is rectangular instead of circular as in a microscope so the ground-truth far-defocused image, $z = -20\mu$m, appears to have rectangular features.

digitally refocusing an image. This further confirms that capturing 4D phase space can be done by using the codes designed above to capture the MI. With the codes designed above, we are able to capture most of the features of the MI. Further utilizing the constraints of MI, such as the bandwidths for 3D incoherent sources or other smoothing constraints, can refine the results shown here and will be studied in the future.

5.5 Conclusion

In this chapter, we demonstrate the full 4D phase space measurement with coded apertures and a code design method with difference sets. We are able to reconstruct the MI patchwise with full-rank linear equations. The math structure of MIs and intensity measurements is

discussed, which has a strong relation with coordinate difference in the aperture space. This relation makes the difference sets from group theory useful in the design of codes. Though the difference set is 1D, we generalize it to 2D and successfully apply the generated 2D codes to reconstruct two simulated 2D-field cases. In the last simulated example, we also demonstrate the use of prior knowledge to help reconstruct a more accurate phase space than reconstruction without prior knowledge. We can potentially use this prior in the code design process to reduce the number of measurements.

Chapter 6

Scattered Light in Phase Space

We have demonstrated various high resolution measurements of phase space in previous chapters. This densely sampled phase space is redundant for us to image a 3D scene. We will take this redundant information to other use: imaging through scattering, as the scattering blurs out a traditional image and scrambles light rays. We want to show that the massive phase-space information provides robustness against scattering. In this chapter, we will demonstrate the use of phase-space imaging for 3D localization of multiple point sources inside volumetric scattering material. The effect of scattering is to spread angular (spatial frequency) information, which can be measured by phase space imaging. We derive a multi-slice forward model for homogenous volumetric scattering, then develop a reconstruction algorithm that exploits sparsity in order to further regulate the problem. By using 4D measurements for 3D reconstruction, the dimensionality mismatch provides significant robustness to multiple scattering, with either static or dynamic diffusers. Experimentally, our high-resolution 4D phase-space data is collected by a spectrogram setup, with results successfully recovering the 3D positions of multiple LEDs embedded in turbid scattering media. Later we apply the phase-space forward model to help record neural activity of a zebrafish brain with a lenslet array microscope setup.[1]

6.1 Introduction

Imaging through scattering remains one of the most important problems in optics. Examples are ubiquitous: imaging in foggy weather, detecting objects behind diffused glass and *in vivo* biological studies. In many situations, scattering media does not absorb light, but scatters it many times, causing images to look diffused. The scattered images still contain significant information about the object (see Fig. 3.4), however, so one has a chance to undo scattering either optically or computationally. Some approaches aim to filter out the scattered light [185, 51, 84, 58, 102], leaving a very weak signal. Instead, we wish to *use* the

---

[1]Neural activity tracking and the study of resolvability are the joint work [146] with Nicolas Pégard when he was with Hillel Adesnik's lab at UC Berkeley.

scattered light in reconstructing the signal [56]. Phase conjugation does this [48, 83, 132], but only works at a single point in space. Related methods use the coherent transmission matrix [69, 103, 150, 181] or adaptive optics [88] to undo scattering, either computationally or physically. Here, we demonstrate a new framework for 3D reconstruction of incoherent point objects (*e.g.* fluorophores, LEDs) well beyond the single-scattering regime, by measuring phase space.

Although the scattering process is completely deterministic, exact inversion through turbid media would require full characterization of all scattering events across 3D space and time. Since this is infeasible with current technology, we turn to statistical methods. In contrast to previous studies, we work in the 4D phase-space domain, which relates to 2nd order correlations of the wave-field [8]. Phase space describes light not only by its 2D positional information (x, y), but also by its 2D distributions of spatial frequency (u_x, u_y) for each position. Spatial frequency and angle of propagation are directly related, so phase-space can be thought of as the wave-optical generalization of light fields [140, 116]. The effect of scattering is to locally spread light into a statistical distribution of angles (spatial frequencies); hence, it is not surprising that angle-resolved measurements are useful for imaging through scattering.

The key idea here is that each point source in 3D traces out a unique 2D plane within the 4D phase space. Just a few parameters (intercepts and slope of the plane) fully define a point source's 3D position, so 4D phase-space measurements are extremely redundant. This redundancy can be exploited to mitigate the effects of volumetric scattering, which blurs phase space in all dimensions. By modeling the blur as a convolution in 4D phase space and using it to constrain the image reconstruction procedure, we show that information about the object remains long after the image appears completely diffused.

Previous studies have used lenslet arrays to recover 2D objects behind thin scattering screens by utilizing multiple perspectives [154, 134, 1], requiring coherent (laser) illumination and moving diffusers. Other works have used phase retrieval to image through thin diffusers by making a single-scattering approximation [73, 107, 66, 167, 165, 99, 100], including recent work using phase-space measurements for calibrating depth [170]. However, most scattering material is volumetric and dynamic, so single or static scattering approximations are not applicable. Here, we demonstrate the validity of our method both for thick and thin diffusers and both static and dynamic scattering.

The assumptions made here are that our object consists of a sparse set of point sources (*i.e.* fewer point sources than the number of 4D pixels) and that the scattering material is homogenous. This holds for situations including fluorophores in biological tissue or vehicle detection in fog. The measurement system captures phase space with very high resolution [183] by using spectrogram measurements [7] taken with a digital micromirror device (DMD) array. The choice of spectrogram instead of coded aperture is to keep the reconstruction direct and simple. The concepts presented here, however, are general and can be used with any phase-space dataset (*e.g.* light-field cameras), as we will show in Section 6.7.

Our inverse algorithm consists of sparse signal recovery in 4D phase space via constrained optimization procedures. A multi-slice forward model [45, 127, 174] includes multiple scat-
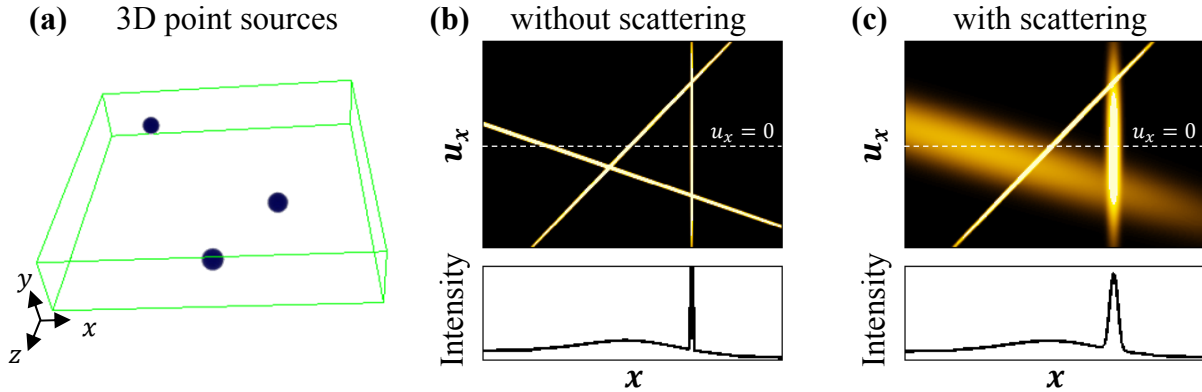
Figure 6.1: Illustrating phase space measurements for three point sources at different depths, with and without scattering media. (a) The 3D positions of the point sources. (b) A phase-space slice and intensity cut-line for the case of no scattering. Each point source creates a line whose intercept with $u_x = 0$ defines its lateral position and slope defines its depth. (c) A phase-space slice and intensity cut-line for the case of point sources in volumetric scattering material. The lines blur in proportion to point source's depth inside the scattering material.

tering and propagation effects, leading to a well-defined inverse problem. In addition to solving for the 3D particle locations, our method is also able to estimate the (scalar) scattering coefficient. Success degrades with increased scattering or decreased sparsity. The limits are difficult to quantify, since they will have a complicated dependence on the object, geometry and the amount of scattering. The important point is that the volume and resolution of data that we can reconstruct will scale not with the number of measurements, but rather with the sparsity, such that this method could in future be scaled up to very large volumes with high-resolution recovery.

6.2 3D Localization of Point Sources in Scattering

Consider the phase-space description of an in-focus point source (dropping the y dimensions for simplicity). The point source emits light at a single point along x and isotropically in all directions, so its phase space is a vertical line covering all frequencies (angles). Since propagation shears phase space (Eq. (2.8)), moving the point source out of focus will cause the line to tilt proportionately. Thus, the intersection of the line with $u_x = 0$ defines the transverse position of the point source and the slope of the line defines its depth above the focal plane (z position):

$$\frac{\mathrm{d}u_x}{\mathrm{d}x} = \frac{n_r}{\lambda z},\tag{6.1}$$

where z is the distance between the point source and the focus plane of the imaging system. This simple model for 3D positions of point sources is easily extended to multiple point

sources that are incoherent with each other (*e.g.* fluorophores). Each point source creates a distinct line in phase space that linearly superimposes with the others (see Fig. 6.1(b)). Since any line can be fully represented by just two parameters (intercept and slope), measuring the full phase space provides a significant amount of redundant data for 3D localization; this is important for undoing the scattering. Note that when we consider both x and y dimensions, phase space extends from 2D to 4D and the line becomes a hyperplane in the higher-dimensional space, creating even more redundancy. Thus, determining the 3D positions of point sources becomes a plane-fitting problem which is highly overdetermined when using 4D phase-space measurements.

When volumetric scattering is introduced, light will be spread along the angle dimension as it propagates, causing the phase space function to blur along both r and u. Thus, point sources at deeper depths will be both more tilted and more spread out (see Fig. 6.1(c)). In the following section, we derive an analytical phase-space model for this volumetric scattering of a point source and use it as a forward model for our inverse algorithm, which attempts to identify the 3D locations of many point sources simultaneously.

6.3   Algorithm

To reconstruct the 3D positions of the point sources from a 4D phase-space dataset, we first develop an analytic forward model for scattering, which is used in the inverse problem. By assuming a point source in homogeneous scattering media with Gaussian spreading statistics, our forward model becomes a tilted 4D Gaussian in phase space. It is derived with a multi-slice approach, in order to account for multiple scattering. Next, we use this forward model in inverse problem solving, where we implement an atomic norm optimization procedure subject to constraints of sparsity and the physics of our forward model.

Our experiments measure the Fourier spectrogram introduced in Sec. 3.1, which is considered a smoothed Wigner function and is always nonnegative. The forward model for scattering is also nonnegative (will be shown shortly); therefore, we ignore the difference between our measured Fourier spectrogram and the forward-model Wigner function. The forward model can be further improved for future experiments but it is sufficient to show the proof of concept experiments here.

6.3.1   Forward model derivation

The volumetric scattering medium is modeled as a stack of infinitely many transverse planes of thin dynamic diffusers separated by homogeneous refractive index material. The Wigner function of the emitted light from a point source is thus propagated and scattered repeatedly as it passes through the scattering material. First, it propagates through a small distance, Δz, in the homogeneous material, as described in Eq. (2.8). Next, it is scattered by a

diffusing plane, denoted by operator $\mathcal{D}_{\sigma,N}$,

$$\mathcal{D}_{\sigma,N}W(\boldsymbol{r},\boldsymbol{u}) = \iint W(\boldsymbol{r},\boldsymbol{u}')\frac{N}{6\pi\sigma^2}\exp\left(-\frac{N}{6\sigma^2}(\boldsymbol{u}-\boldsymbol{u}')^2\right)\mathrm{d}^2\boldsymbol{u}'. \qquad (6.2)$$

Here σ is the scattering coefficient (defined as the standard deviation of the angular spreading) and $W(\boldsymbol{r},\boldsymbol{u})$ is the Wigner function of the wave-field immediately before the diffuser. Thus, scattering is modeled as an in-place Gaussian spreading of the angle (frequency) information, which accurately describes dynamic diffusers and is a good approximation for static diffusers. Note that $\iint \mathcal{D}_{\sigma,N}W(\boldsymbol{r},\boldsymbol{u})d^2\boldsymbol{u} = \iint W(\boldsymbol{r},\boldsymbol{u}')d^2\boldsymbol{u}'$, which means that the light spreads without changing its position or overall intensity. In our multi-slice approach, this process of propagate-and-scatter is repeated for each z step through the scattering material until it reaches the exit plane. Multi-slice approaches [45, 127, 174] have been used previously for coherent wave-fields and nonlinear propagation [3]. Here, to apply these ideas in phase space, we seek an analytical expression for the limit of infinitesimal z steps.

Consider a point source located at $(\boldsymbol{r}_s, z_s)$, described by the Wigner function, $W_s = \delta(\boldsymbol{r}-\boldsymbol{r}_s)$. The light then propagates a total distance ℓ through the volumetric scattering medium, divided into N layers. Letting $W_N(\boldsymbol{r},\boldsymbol{u})$ denote the Wigner function of the 2D field after the last layer of scattering material (exit plane), we have

$$W_N(\boldsymbol{r},\boldsymbol{u}) = \left(\mathcal{D}_{\sigma,N}\mathcal{P}_{\Delta z=\ell/N}\right)^N W_s(\boldsymbol{r},\boldsymbol{u}). \qquad (6.3)$$

Here $\mathcal{P}_{\Delta z}$ is the propagation operator given by Eq. (2.8). In order to derive the analytical phase-space model for the case of $N \to \infty$, $W_\infty(\boldsymbol{r},\boldsymbol{u})$, we use Feynman path integrals [126] (see Appendix B for details),

$$W_\infty(\boldsymbol{r},\boldsymbol{u}) = \lim_{N\to\infty} W_N(\boldsymbol{r},\boldsymbol{u}) = \frac{n_r^2}{2\pi\lambda^2\ell^2\sigma^2}\exp\left(-\frac{n_r^2}{2\lambda^2\ell^2\sigma^2}(\boldsymbol{r}-\boldsymbol{r}_s-\lambda\ell\boldsymbol{u}/n_r)^2\right). \qquad (6.4)$$

Equation (6.4) is our forward model describing the phase space at the exit of the scattering medium from a point source at depth z_s. In our actual imaging system, however, the exit plane is at axial distance z_d, which is not necessarily the native focus plane of the microscope. Therefore, the phase space that we measure may have an extra propagation operation due to the imaging system. The light still passes through a distance $\ell = z_d - z_s$ of scattering material, and then is subject to an extra free-space back propagation operation on $W_\infty(\boldsymbol{r},\boldsymbol{u})$ from the exit plane, $z = z_d$, to the actual focus, $z = 0$. Note that the interface of different refractive indices does not affect the phase space function, unlike the light field, since the spatial frequencies do not change across the interface while propagation angles do. As a result, the Wigner function at our measurement plane is described by $W(\boldsymbol{r},\boldsymbol{u}) = \mathcal{P}'_{-z_d}W_\infty(\boldsymbol{r},\boldsymbol{u})$, where $\mathcal{P}'$ represents propagation in air. Accounting for the difference in the refractive index from air to the scattering medium, the expression becomes,

$$W(\boldsymbol{r},\boldsymbol{u}) = \frac{n_r^2}{2\pi\lambda^2(z_d-z_s)^2\sigma^2}\exp\left(-\frac{n_r^2}{2\lambda^2(z_d-z_s)^2\sigma^2}\left(\boldsymbol{r}-\boldsymbol{r}_s+\lambda\left(z_d-\frac{z_d-z_s}{n_r}\right)\boldsymbol{u}\right)^2\right). \tag{6.5}$$

The above formula is our main theoretical result.

6.3.2   Reconstruction

The goal for reconstruction is to find both the transverse position $\boldsymbol{r}_s$ and the depth z_s of many point sources simultaneously. Our method is based on minimizing the atomic norm [18], which promotes sparsity in the solution, in a similar fashion to compressed sensing. Using Eq. (6.5), we can calculate the phase space for each possible point source in 3D. The collection of all of these phase-space functions form an atom set. The phase-space measurement from multiple point sources will then be a nonnegative linear combination of the atoms in this set. Our algorithm aims to decompose the measurements into a sparse combination of these atoms by minimizing the reconstruction error while jointly optimizing for a sparse number of atoms.

First, we define the atom set based on Eq. (6.5) with a modification for numerical stable

$$\mathcal{A} = \{a(\boldsymbol{r}, \boldsymbol{u}; \boldsymbol{r}_s, z_s)\} \tag{6.6}$$

where the atom $a(\boldsymbol{r}, \boldsymbol{u}; \boldsymbol{r}_s, z_s)$ is the 4D function of $(\boldsymbol{r}, \boldsymbol{u})$ in Eq. (6.5) with $(z_d - z_s)^2\sigma^2 \to ((z_d - z_s)\sigma + 1)^2$. Each of them is parametrized by the 3D position $(\boldsymbol{r}_s, z_s)$. The decomposed weighted sum of atoms is

$$\hat{I}(\boldsymbol{r}, \boldsymbol{u}) = \sum_{\boldsymbol{r}_s, z_s} c(\boldsymbol{r}_s, z_s) a(\boldsymbol{r}, \boldsymbol{u}; \boldsymbol{r}_s, z_s), \tag{6.7}$$

where $c(\boldsymbol{r}_s, z_s)$ is a nonnegative coefficient describing the radiating power of a point source at position $(\boldsymbol{r}_s, z_s)$. Equation (6.7) can be regarded as a phase-space prediction of the expected measurement data, given our current estimate of the object. To find the sparse $c(\boldsymbol{r}_s, z_s)$ that minimizes the difference between the prediction and the measured data, we solve the following optimization problem

$$\min_{c \geq 0} \sum_{\boldsymbol{r}, \boldsymbol{u}} \left| I(\boldsymbol{r}, \boldsymbol{u}) - \hat{I}(\boldsymbol{r}, \boldsymbol{u}) \right|^2 + \mu \sum_{\boldsymbol{r}_s, z_s} |c(\boldsymbol{r}_s, z_s)|, \tag{6.8}$$

where μ is a tuned regularization constant and $I(\boldsymbol{r}, \boldsymbol{u})$ is the measured phase space. Here the sum is over the continuous parameter space of all possible locations of sources. In this work, we discretize the set of candidate positions. After discretization, the non-smooth convex optimization reduces to the popular ℓ_1-minimization problem, also known as the LASSO [178]. This formulation encourages sparsity in the observed point sources, and performs well in our experiments. We use an accelerated proximal gradient method [189] which iteratively minimizes the reconstruction error (the squared term) and "shrinks" the c vector towards a sparse solution. We then threshold the recovered $c(\boldsymbol{r}_s, z_s)$ by τ and spatially cluster it [171], before estimating the position of each point source as the centroid of each cluster.

While the parameters n_r and z_d in Eq. (6.5) can be reasonably assumed to be known (or measured), the amount of scattering is often unknown and cannot be assumed a priori. Thus, our algorithm implements an outer loop optimization over different scattering coefficients in

order to solve for σ as well as the point source positions. We generate multiple $\mathcal{A}$ sets with different σ and solve Eq. (6.8) with the $\hat{I}$ given by each $\mathcal{A}$, one at a time. The σ that minimizes Eq. (6.8) and the detected points in that optimization are chosen to be the final outcome of the algorithm.

6.4   Experiments

To measure the Fourier spectrogram (Eq. (3.2)), we require a programmable aperture in the Fourier domain of the object [183]. Building a folded $4f$ system after the object (see Fig. 6.2(a) with FL_1=225 mm, FL_2=175 mm), we place a reflective DMD (DLP$^{®}$ Discovery 4100, .7" XGA) in the Fourier plane in order to rapidly create and scan the apertures, while capturing real-space intensity images. Our sCMOS camera (PCO.edge 5.5) is placed at the output of the $4f$ system and fully synchronized with the DMD controller via hardware link. Acquisition time can achieve camera-limited maximum frame rates (~ 1 kHz), but are typically limited by light efficiency to tens of seconds per dataset. Conveniently, phase-space sampling density can be flexibly traded off for acquisition time and noise performance. We capture data at the highest resolution possible, though it may be excessive for the simple objects used here. Alternatively, the optical system could be replaced by a single-shot light-field camera [140], at a cost of lower resolution.

We perform two sets of experiments, using LEDs placed at various locations within a 3D volume. First, we scan a 1D window (strip) across the DMD and collect 2D phase-space data for the case of multiple discrete rotating (dynamic) diffusers. Second, we scan a 2D window across the DMD and collect 4D phase-space data for the case of LEDs embedded in scattering gelatin, to demonstrate the applicability of the algorithm to volumetric scattering effects.

6.4.1   1D object with multiple thin diffusers

The object for the 1D experiment consists of 3 LEDs (dimension 0.9 mm×1.5 mm) in a line, which serve as point sources with different transverse (x) and depth (z) positions. The LEDs have center wavelength 657 nm and bandwidth 38.8 nm; we use a bandpass color filter (650 nm $\pm$ 5 nm) to suppress dispersion effects from the DMD. The window has a width of 1.74 mm and is stepped across 225 positions in angle space (step size 54 μm) to use the full area of the DMD. The 2D phase space slices are computed along the green line (averaged over a small area) and plotted at the corresponding u_x in Fig. 6.2. The phase space image is then input to our algorithm to recover the positions and depths of the point sources. In the algorithm, we setup up 1000 sample points in the x coordinate and 200 sample planes in z. The sampling size is 11.7 μm in x and 1.5 mm in z. The scattering parameter, σ, was found to be 3×10^{-4} μm^{-1} for the case of no scattering and 1.1×10^{-2} μm^{-1} for both the 1-diffuser and 3-diffuser cases.
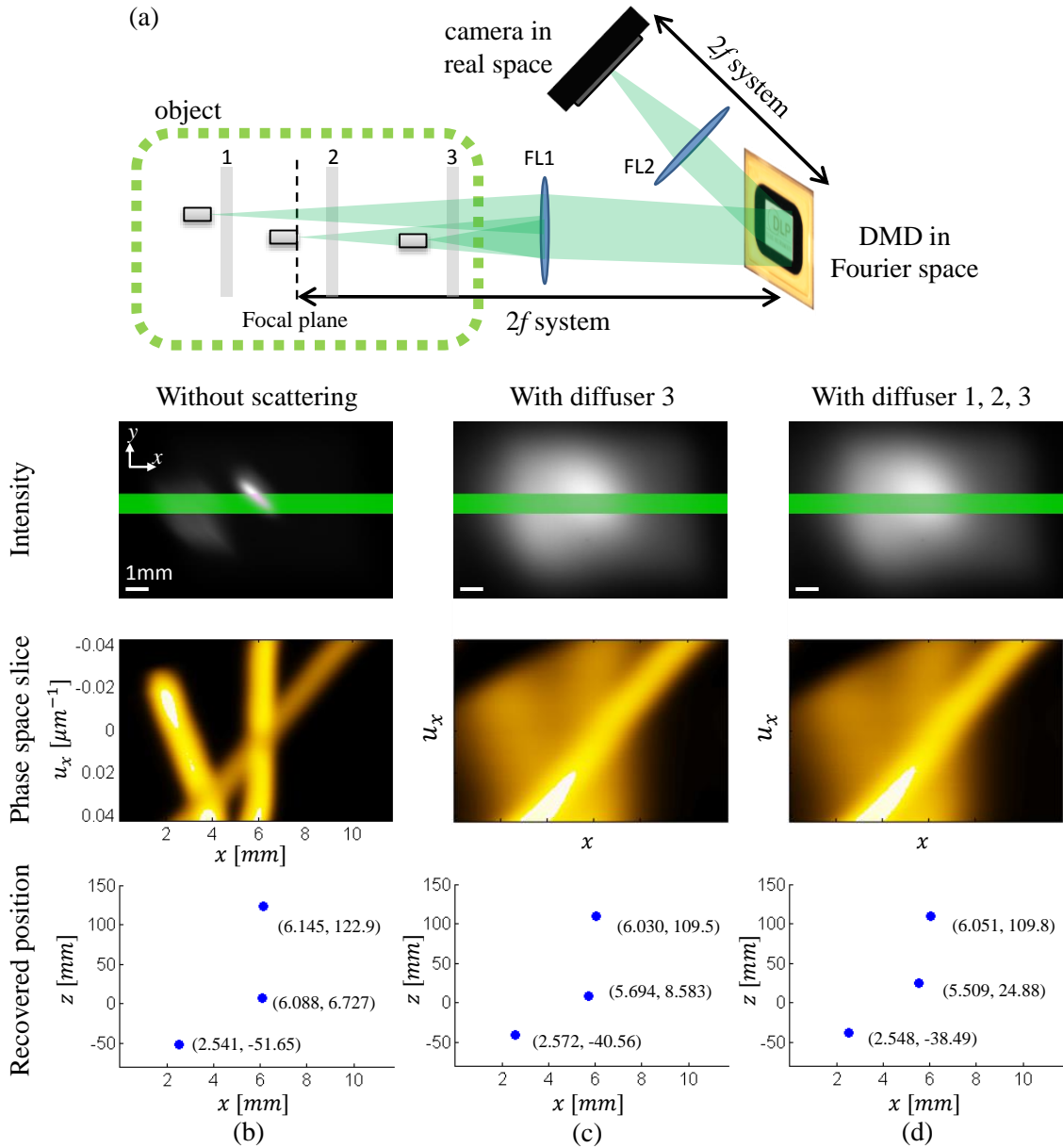
Figure 6.2: Experimental localization of LEDs at different depths, with and without rotating diffusers between them. (a) The setup uses an object consisting of 3 LEDs and a spectrogram measurement system that employs a DMD in Fourier space to scan the aperture. (b-d) Intensity, phase space (along green line) and recovered LED positions for the case (b) without scattering, (c) with diffuser 3 only, and (d) with all three diffusers. The algorithm successfully recovers the position and depth of each LED in all cases.

With no diffusers (non-scattering case), the LED positions are easily determined by the phase space measurements (Fig. 6.2(b)). Each LED prescribes a distinct line through

the phase-space measurement, allowing us to find its lateral and depth position using our algorithm. Notice that the conventional 2D intensity image (taken with an open window in Fourier space) does not contain sufficient information for depth detection, even in the non-scattering case.

When a thin rotating diffuser is placed after the LEDs (diffuser 3 in Fig. 6.2(a)), the phase space lines become blurred, with LEDs further away from the camera blurring more, as expected (Fig. 6.2(c)). In this case, the intensity image is completely diffused. However, the three lines in phase space are still distinguishable, and so the three point sources can still be located with good accuracy. Interestingly, as additional diffusers are added between LEDs to create a multiple scattering situation, the phase space blurring only becomes marginally worse (Fig. 6.2(d)). This demonstrates that the scattering material closest to the exit plane will play the largest role, since more light passes through it. Even with three rotating diffusers placed between the LEDs, we are able to successfully recover the position and depth of the LEDs.

6.4.2   2D object with volumetric scattering material

Next, we extend our experiments to 4D acquisition and a truly volumetric scattering. The target consists of 8 LEDs (center wavelength 524 nm, bandwidth 32.0 nm) imaged both with and without a surrounding scattering medium of chopped gelatin in air. The air in the gaps between the diffusing gelatin creates a refractive index mismatch which provides a strong forward-scattering situation intended to mimic that of biological tissue. The scattering is assumed to be homogeneous across the volume since the gelatin is distributed approximately uniformly. The data collection procedure scans a 2D square window with side length 1.09 mm across 26×26 positions (step size 365 μm) on the DMD, collecting 676 images (16 bit) in total, each with 5.53M pixels. This large dataset ($\sim$ 7 GB) demonstrates the ability of our system to capture very high resolution phase-space data and by implementing our algorithm in a distributed computing environment, we are able to solve the computational inverse problem. However, since the density of point sources is small for this experiment, we find that it is sufficient to down-sample the 4D data to $128 \times 128 \times 8 \times 8$ before running our algorithm. We reconstruct a 3D result of $128 \times 128 \times 128$ voxels and recover a scattering parameter σ of 5×10^{-3} μm^{-1} for the case without scattering and 1×10^{-2} μm^{-1} for the case with scattering.

The results of this experiment are shown in Fig. 6.3 for measurements taken both before adding the scattering gelatin and after. We illustrate the 4D data by showing only a few 2D slices along u_x-x and u_y-y of the 4D phase space acquired. Each line in the 2D slice corresponds to a portion of the plane generated by a point source. After adding the scattering material, the intensity images blur beyond distinction, and the phase-space slices also blur. However, the phase-space images still retain sufficient information for localizing each LED. Note that even when the 2D phase-space slices seem completely blurred, there may still be sufficient information for localization when considering the entire 4D dataset. Thus, the 3D recovery succeeds, even with highly multiply scattering situations. Of course, the algorithm
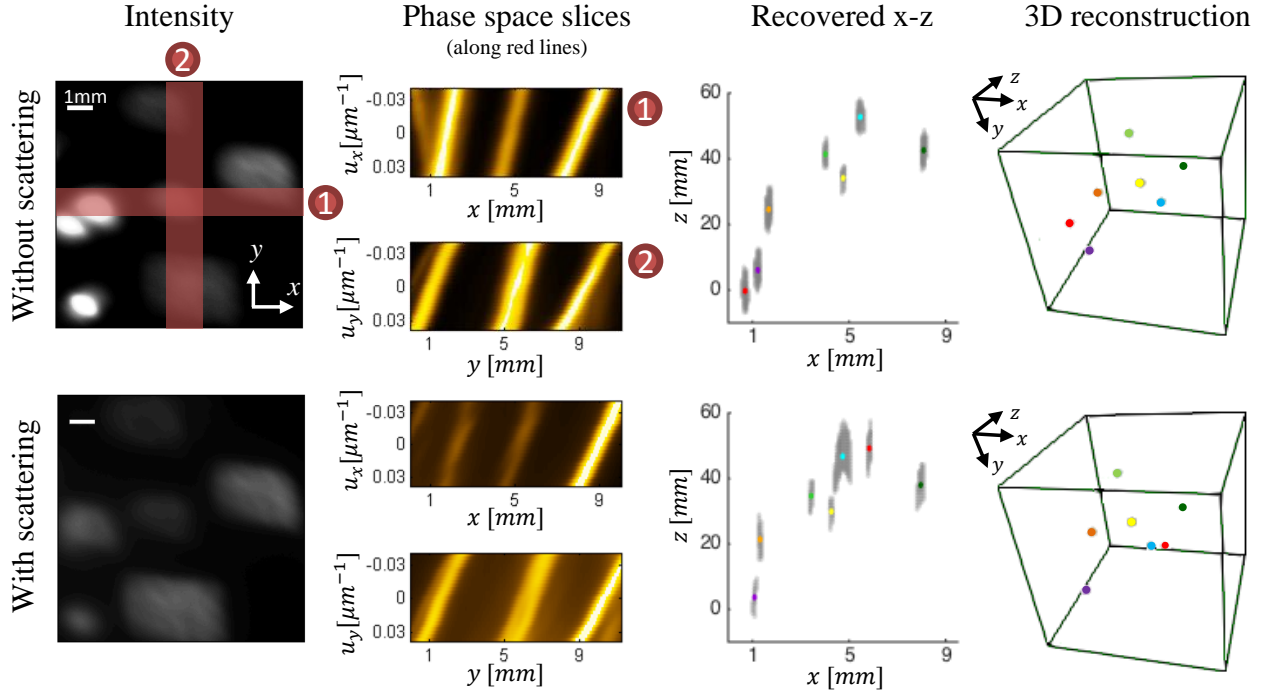
Figure 6.3: 4D phase-space experiments with and without volumetric scattering. The first column contains the 2D intensity images that are captured by a conventional camera and the second column shows some example phase space 2D slices (along the red lines in the intensity images). The third column shows the recovered $x - z$ positions of point sources (y not shown) and the last column shows reconstructed 3D positions of the LEDs. Each color dot corresponds to a successful LED position recovery, while red dots are failures due to occlusions by the finite size of LEDs.

will fail with too much scattering or with reduced phase-space sampling. In our resulting 3D localization, one can see that the variance of the estimate in the z direction is somewhat worse than the lateral direction, due to the small range of angles captured. One of the LEDs is mistakenly localized in an errant position, which we believe is due to artifacts caused by the blocking of some light by the wire leads on the LEDs. For non-occluded LEDs, however, the algorithm correctly localizes the LEDs with good accuracy and recovers a scattering parameter of σ of 5×10^{-3} μm^{-1} for the case without scattering and 1×10^{-2} μm^{-1} for the case with scattering, consistent with previous measurements.

6.5 Discussion

Our setup is analogous to that presented in [183], with several key differences. Instead of placing the camera in Fourier space and scanning a real-space window, here we choose to place the camera in real space and scan a Fourier-space window. This leads to better real-

space resolution and a larger field of view, even when down-sampling the aperture positions to reduce acquisition time. Additionally, since most objects typically have a large DC term, capturing images in real space (as opposed to Fourier space) tends to mitigate the dynamic range problems encountered in [183]. Finally, instead of an LCOS modulator, a DMD is used here because it is orders of magnitude faster in modulating the light. Since the DMD is not polarization sensitive, we further obtain $2\times$ better photon efficiency for unpolarized objects and avoid crossed polarizer leakage. The disadvantages of the DMD array are that it suffers color dispersion artifacts (see Sec.3.4) and that it must be implemented in a folded geometry which can cause aberrations at high angles.

The experiments presented above offer a proof of concept for our proposed method using 4D phase-space measurements to robustly localize a sparse set of point sources through both dynamic and volumetric scattering media. The method does not require coherent or active illumination, so works with existing imaging systems. The depth dependence of widening of phase space hyperlines is general and can be applied to other phase space measurements. In the rest of this chapter, we discuss the application of the phase space scattering theory developed in Section 6.3 to a light-field measurement of zebrafish brain activity.

6.6   Brain Imaing and Resolvability of Phase-Space Imaging Through Scattering

Brains play a core role in animal's perception and consciousness. It is a scattering tissue filled with a dense network of neurons that exchange information by means of electrical signals called action potentials. Understanding the mechanisms by which a brain processes information requires the ability to detect action potentials from many individual neurons simultaneously across large volumes of tissue. Engineered calcium-sensitive proteins [41] and voltage-sensitive dyes [147] enable optical detection of action potentials without disturbing the neuron's physiology. The major obstacle for the optical approaches falls onto undoing the scattering. Many methods have been proposed to achieve this goal [14, 51, 158, 98, 101, 9, 26], all of which involve scanning so frame rates are limited for large-volume imaging. To address the demand of speed and the issue of scattering, we turn to light-field imaging of phase space. Light-field imaging [117], together with wave-optical theory [33, 174], provides a single shot way to capture 3D information of a microscopic scene in question. It has provided promising results for functional brain imaging with 3D volume image reconstruction to quantify the fluorescence levels of individual neurons [151]. As it is a phase space imaging, we want to analyze its ability to extract information from a scattering scene. In this section, we examines such ability for a particular application: imaging brain neurons, which has a typical cell body of $\approx 10~\mu$m. Our analysis is done by inspecting measured raw light-field images (without any processing such as refocusing or applying the model introduced above) for the distinguishability (defined below) of two emulated point sources under a mouse brain tissue with various thickness (varied from 100 μm to 400 μm). We also compare them
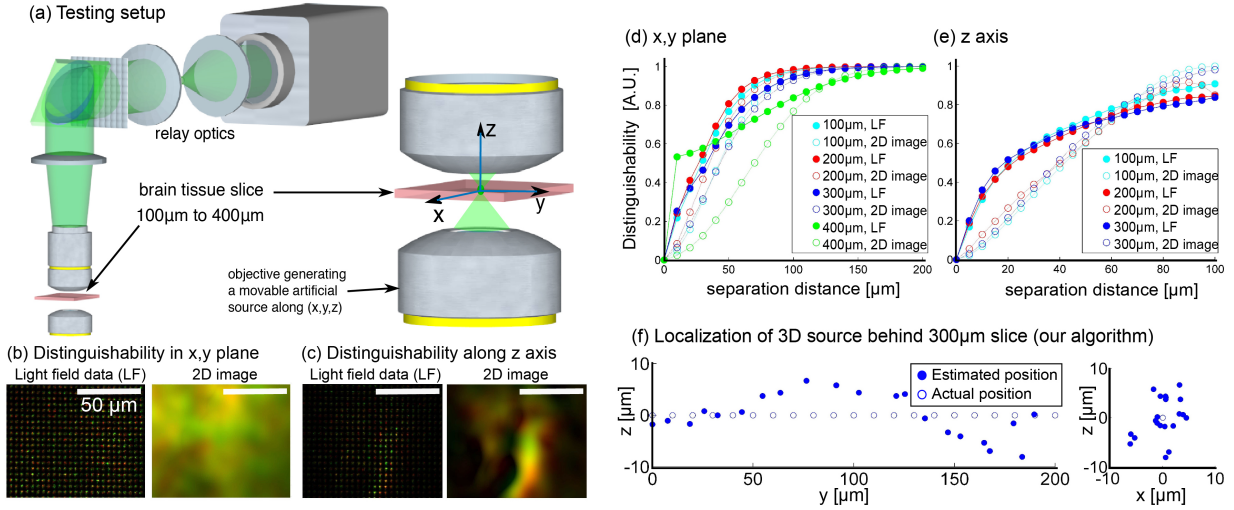
Figure 6.4: Experimentally testing distinguishability of two emulated source under light-field microscopy. (a) Experimental setup for spatial resolution measurements. Slices of mouse brain tissue with varying thickness are placed above an artificial source (created by a second microscope objective) intended to mimic the fluorescence in an active neuron. The artificial source can be precisely positioned at any 3D location. (b,c) Comparison of distinguishability for light-field data versus 2D fluorescence data. The red and green channels of the colormap are used to show the measurement images from two point sources simultaneously in one image. The sources are separated (b) in the (x, y) plane and (c) along the z axis. (d,e) Distinguishability of the two captured images as a function of separation distance between two sources, (d) in the lateral plane and (e) along the optical axis. (f) Other than pure distinguishability analysis, we further use our algorithm (Eq. (6.8)) to estimate the position of the source through a 300 μm slice for controlled source displacements along the y axis in strong scattering.

to conventional 2D fluorescence microscopy by removing the light-field lenslet array from our system. At last, the resulting distinguishability of our light-field system shows sufficient resolution (defined via distinguishability) and the resolvable density of point sources to image through scattering tissue.

Fig. 6.4(a) shows the experimental setup. The point source with controllable location is emulated by focusing light with a second objective to a spot with a size of a typical neuron cell body ($\approx 10~\mu$m); here the light $\lambda = 532$ nm is in the emission spectral range of GCamp, a fluorescent protein used in neural activity tracking. A two-point-source measurement is mimicked by digitally adding two measured images for a source at two different positions. A slab of scattering mouse brain tissue is inserted between the point-emulating objective and the measurement objective; the tissue itself does not express any particular fluorescence. Some important features of the setup are: (a detail description is given in Sec. 6.7.1)

- The pitch of demagnified lenslet array in the sample space is 4 μm. (hence the spatial

coordinate sampling of light-field is 4 μm)

- The measurement objective has NA_{obj} 0.5 and 40 camera pixels are used to sample this NA in each of the x, y dimensions.

In traditional light field microscopy where volume image reconstruction is the goal, *resolution* can be obtained by measuring the size of the point-spread function [117] or spatial bandwidth [33]. In brain tissue, resolution is further complicated by a dependence on scattering and density of neurons. Hence we adopt an experimental approach with a distinguishability metric. We define the spatial resolution along a given axis (x, y or z) as the minimum allowed separation distance δ_x, δ_y or δ_z between two neurons for them to be identified as two separate sources. Consider two sources, 1 and 2, and the corresponding measurements, I_1 and I_2. Distinguishability D, a useful metric in defining the resolution, is given by:

$$D(I_1, I_2) = 1 - \frac{\int I_1(\boldsymbol{\xi}) I_2(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}}{\sqrt{\int I_1^2(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \int I_2^2(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}}}, \tag{6.9}$$

where $\boldsymbol{\xi}$ denotes the 2D coordinates on the sensor. By definition, $D = 1$ (*fully distinguishable*) when the recorded images give two disjoint sets of pixels and $D = 0$ (*not distinguishable*) for the case of identical light field signatures, as a direct consequence of Cauchy-Schwarz's inequality. Before defining the resolution with distinguishability, let us look at the distinguishability from light-field measurements and 2D conventional fluorescence imaging. On left hand sides of Figs. 6.4(b) and (c), we display the measured light-field images from two source positions (40 μm separation and thus yielding two different measured images) simultaneously in two color channels, with one in red and the other in green. The same color-mixing display for 2D intensity images are shown on the right hand sides of Figs. 6.4(b) and (c). The distinguishability is higher if a color-mixing image has more green and red pixels, while it is lower if more yellow pixels. Light field measurements provide better distinguishability than 2D fluorescence images, particularly in the axial dimension. Figs. 6.4(d) and (e) plot the experimentally measured distinguishability as the separation distance between two sources is increased, for both the lateral and axial dimensions and with both light field and 2D fluorescence data. To get a sense for how distinguishability relates to localization error in our algorithm (Eq. (6.8)), we additionally recover 3D positions as the source is moved along y behind 300 μm of scattering brain tissue, using the light field data (see Fig. 6.4(f)).

Resolution and maximum resolvable density

In the absence of noise, it is theoretically sufficient to resolve sources 1 and 2 with a single pixel difference ($D > 0$) between I_1 and I_2, provided that a forward model can accurately decompose the measured data and produce different outputs for any two different neurons. Practically, at full frame rate and in low light conditions, a conservative condition for identification [124] is to compare the distinguishability to the signal-noise-ratio (SNR) in the light
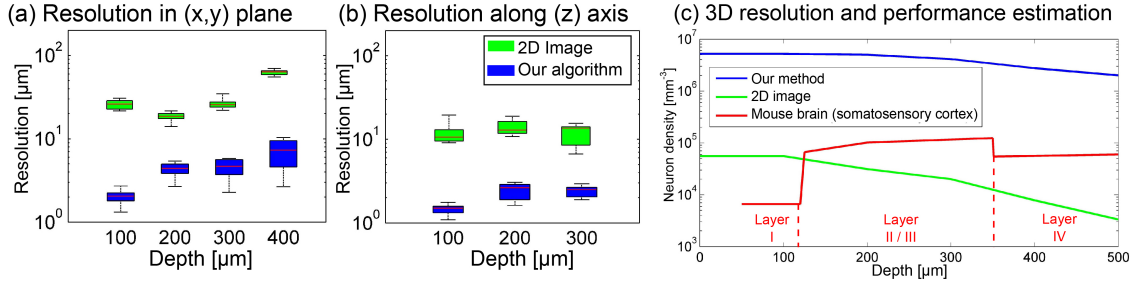
Figure 6.5: Spatial resolution analysis for our method, according to the minimal distance between two sources required for correct identification as separate neurons. (a) In the lateral (x, y) plane and (b) along the optical axis through a given depth of mouse brain tissue. Fluorescence microscopy (green) and light field microscopy (blue) are compared on the same scale and show a ten-fold difference in performance along all axes. (c) Estimated maximum resolvable neuron density as a function of depth in mouse brain tissue, as compared to typical neuron density observed in the mouse barrel cortex.

field measurements:

$$D(I_1, I_2) > \frac{1}{\text{SNR}}. \tag{6.10}$$

For an application targeting at 100 Hz sampling rate and without significant photo-bleaching, the fluorophore are excited such that the fluorescent light has SNR ≈ 3. The minimal separation distances in the focal plane of $\delta_x = \delta_y$, as well as along the optical axis δ_z that make the measured distinguishability satisfy the inequality are defined as the spatial resolution. The experiment is repeated in several locations and for various thicknesses of brain tissue, with results summarized in Fig. 6.5(a) and (b). Overall, the light field data provides $\sim$ 10× better localization resolution in all dimensions, as compared to 2D fluorescence images parsed by the same algorithm. Light field deconvolution methods [33] are expected to have performance somewhere in between these two.

To understand how our resolution metric relates to functional imaging capabilities, we estimate the maximum resolvable density of neurons as one neuron per resolved volume $(\delta_x \delta_y \delta_z)$:

$$\text{maximum resolvable density} = \frac{1}{\delta_x \delta_y \delta_z}.$$

We compare the density for our method and that for 2D intensity images to the density of neurons typically observed in layers I to IV of mouse brain (primary somatosensory cortex) [133] (see Fig. 6.5(c)). This plot confirms experimental observations: 2D fluorescence microscopy is unable to identify neurons located below Layer I in the barrel cortex. However, light field data enables a 1000-fold improvement (10-fold improvement along each axis) in the neuron density that can be resolved, as compared to 2D fluorescence, so is a promising avenue towards neural activity tracking in all layers.
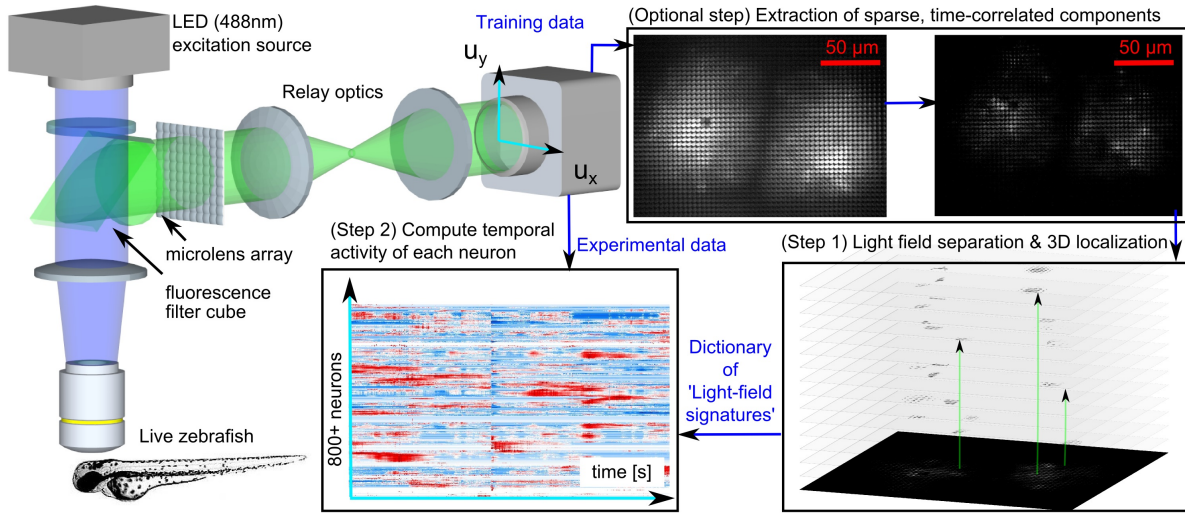
Figure 6.6: Experimental setup and computational imaging for brain activity tracking of samples tagged with engineered fluorescent proteins. A micro-lens array is inserted into a fluorescence microscope for light-field data acquisition. We adopt a dictionary based approach to track the neural activity where the dictionary contains time-independent features of present neurons. The setup is then used to capture two data sets: a training video of sparse frame and a measurement video of neural activity. Note that our algorithm has the flexibility to use the measurement video as the training video. Detail of the flow is described in the main text.

6.7 Application of Scattering Theory to Zebrafish

Light-field microscopy shows the ability to extract information from a single shot of a scattering scene as discussed above. Here we further apply the light-field microscope setup to brain activity tracking, with a sparsity-based algorithm developed from the phase space theory in Sec. 6.3. The setup and algorithm together enable monitoring neurons with high temporal and spatial resolution inside a large scattering volume, provided that only a sparse set of neurons are active at once. In contrast to the traditional light-field microscopy [151] whose 3D reconstruction fidelity is limited by the 2D sensor pixel counts, we skip the step of explicitly reconstructing a 3D image and instead attempt to simply *distinguish* and *localize* each neuron in 3D. Hence the data collection requirements scale not with the number of voxels to be reconstructed, but rather with the number of active neurons at a particular time. We demonstrate our method experimentally for zebrafish neural activity tracking with 800+ neural structures at 100 fps, which includes a demonstration of the ability of 3D localization of our theory.

6.7.1 Experimental setup and flow

The experimental setup (shown in Fig. 6.6) is a fluorescence microscope that has been modified by introducing a micro-lens array at the imaging plane, with the sensor placed at the relayed back focal plane of the array. On the vertical arm, the objective (water immersion, NA_{obj} 0.5) and the tube lens form a 40x magnification $4f$ system. On the horizontal arm after the green color filter, a square-lattice micro-lens array (lens pitch 150 μm, $f = 5.2$ mm, NA_{ML} 0.014) is placed at the imaging plane of the $4f$ system above. The back focal plane of the micro-lens array is relayed onto the camera by a 1.7x maginification $4f$ system.

The micro-lens is chosen such that NA_{ML} matches and is slightly greater than NA_{obj} after magnification; its pitch corresponds to sampling size of $p = 4$ μm at the sample plane, which is comparible to the size of a neuron. This design enables us to map the camera pixels to the 4D light-field as the pupil of each micro-lens records a local angular distribution. We call the 2D intensity measurement at the sensor plane, $I(\boldsymbol{\xi}, t)$, *light-field measurements* at each time, t:

$$I(\boldsymbol{\xi}, t) = \mathcal{I}\left(x, y, \theta_x, \theta_y, t\right).$$

The sampling of the 4D light field on a 2D plane of pixels is given by

$$\xi_x = N_p\left(\left\lfloor \frac{x}{p} \right\rfloor + \frac{\theta_x}{2NA_{obj}}\right) \quad \text{and} \quad \xi_y = N_p\left(\left\lfloor \frac{y}{p} \right\rfloor + \frac{\theta_y}{2NA_{obj}}\right),$$

where $\boldsymbol{\xi} = (\xi_x, \xi_y)$ are the lateral coordinates at the sensor in the unit of number of camera pixels, and $\lfloor\ \rfloor$ is the floor function. $N_p = 40$ corresponds to the number of camera pixels under each micro-lens in each dimension (total N_p^2 pixels under a micro-lens). This achieves good angular (and hence, axial) resolution by sacrificing lateral resolution, which will be improved in post-processing. The resulting field of view (at the sample) is a 200 μm square, with $N_l = 50$ micro-lenses in each direction, for a total of $N_p^2 N_l^2 = 4 \times 10^6$ pixels on the sCMOS sensor (Andor Zyla 4.2). Each acquired frame contains full-volume fluorescence data, with temporal resolution equal to the camera's frame rate, $1/\delta_t = 100$ fps.

The ultimate goal is to reconstruct $a_j(t)$, the time-dependent magnitude of fluorescence of neural j in

$$I(\boldsymbol{\xi}, t) = \sum_{j=1}^{N} I_j(\boldsymbol{\xi}) a_j(t), \tag{6.11}$$

where $I_j(\boldsymbol{\xi})$ is the *light-field signature* of neuron j, that is, the measurement we would capture if only the jth neuron were active constantly. Conveniently, the signature naturally encodes any shape variations and effects of aberrations and scattering. We assume here that the light-field signatures do not change over time. Hence the task is split into two parts: build a dictionary of light-field signatures for each neuron and reconstructing the neural activity. The other part of Fig. 6.6 summarizes the flow to achieve our goal of neural activity tracking, which we detailed below.

- Step 1: Build a dictionary of light-field signatures, $\{I_j(\boldsymbol{\xi})\}$

1. Record a training video $I_{\text{tr}}(\boldsymbol{\xi}, t)$, ideally with as few neurons firing at once as possible.

2. (optional) Decompose $I_{\text{tr}}(\boldsymbol{\xi}, t)$ into independent components, resulting in $I_{\text{tr}}^k(\boldsymbol{\xi})$ where k indexes the components. If not performing this step, set each k corresponding to different time t.

3. Solve the following ℓ_1-regularized optimization problem for c^k

$$\min_{c^k \geq 0} \|I_{\text{tr}}^k(\boldsymbol{\xi}) - \hat{I}^k(\boldsymbol{\xi})\|^2 + \mu \sum_{\boldsymbol{x}_i} c^k(\boldsymbol{x}_i), \tag{6.12}$$

where $\boldsymbol{x}_i = (x_i, y_i, z_i)$ discretizes the volume of interest with finer gridding than $p = 4$ μm, μ is a hand-tuned regularization constant which enforces sparsity and

$$\hat{I}^k(\boldsymbol{\xi}) = \sum_{\boldsymbol{x}_i} c^k(\boldsymbol{x}_i)\mathcal{A}(\boldsymbol{\xi}, \boldsymbol{x}_i) \tag{6.13}$$

is the predicted measurement, given by applying our forward model $\mathcal{A}$ (Eq. (6.5)) to $c^k(\boldsymbol{x}_i)$, the estimate of the spatial distribution of sources in the sample.

4. Optimize over σ the scattering parameter in forward model $\mathcal{A}$. Use the result from the minimum-giving σ.

5. Obtain the normalized light-field signature $I_j(\boldsymbol{\xi})$ of neuron j at r_j from each k,

$$I_j(\boldsymbol{\xi}) = \frac{1}{b_j^k} \frac{c^k(\boldsymbol{x}_j)\mathcal{A}(\boldsymbol{\xi}, \boldsymbol{x}_j)}{\hat{I}^k(\boldsymbol{\xi})} I_{\text{tr}}^k(\boldsymbol{\xi}) \tag{6.14}$$

where $b_j^k = \int \frac{c^k(\boldsymbol{x}_j)\mathcal{A}(\boldsymbol{\xi}, \boldsymbol{x}_j)}{\hat{I}^k(\boldsymbol{\xi})} I_{\text{tr}}^k(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$ is for normalization such that $\int I_j(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} = 1$.

6. Merge the light-field signatures from the same neuron by evaluating the mutual distances between identified neurons and detecting overlaps within the typical size of one neuron (here we use a 4 μm mutual distance threshold).

- Step 2: Neural activity tracking

1. Record the measurement video $I(\boldsymbol{\xi}, t)$.

2. Solve the nonnegative least square problem for the activity $a_1(t), \ldots, a_N(t)$ of neurons

$$\min_{a_1, \ldots, a_N \geq 0} \|I(\boldsymbol{\xi}, t) - \sum_{j=1}^N a_j(t) I_j(\boldsymbol{\xi})\|^2. \tag{6.15}$$

Independent component analysis

The optional step above is an independent component analysis (ICA). The video $I_{\text{tr}}(\boldsymbol{\xi}, t)$ is represented in as a nonnegative matrix $\mathbf{I}$ of dimension $N_l^2 N_p^2 \times N_t$ where N_t is the number of time frames, and we want to factor it as a product of two matrices

$$\mathbf{I} = \mathbf{ST}$$
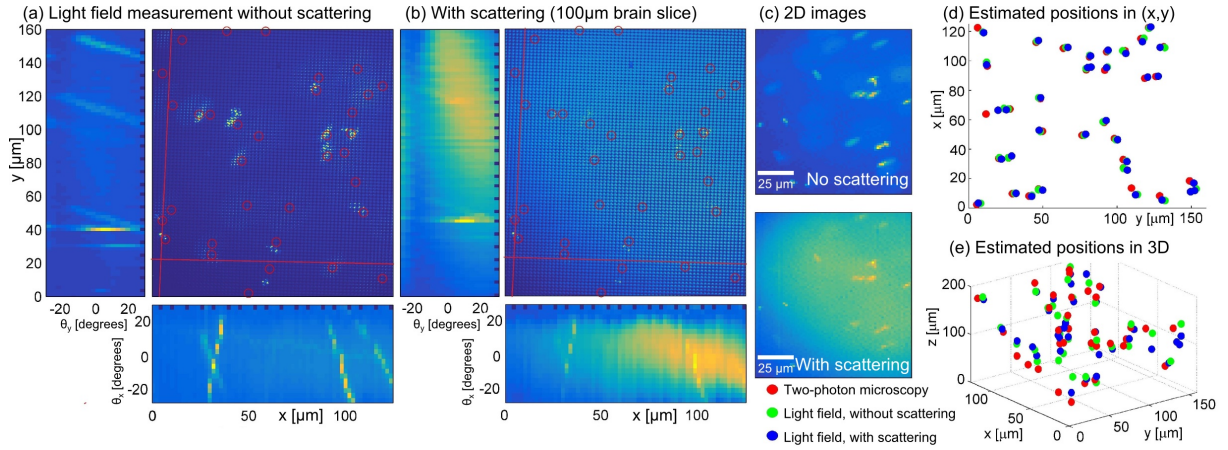
Figure 6.7: Single-shot experimental detection and 3D localization of sparsely-distributed fluorescent beads, with and without scattering, as compared to two-photon microscopy scanned images. (a) Single-shot light-field measurement and several space-angle slices (along the red lines) *without scattering*. (b) Dataset recorded after placing a 100 μm slice of wild-type mouse brain tissue directly above the beads so as to introduce realistic scattering conditions without displacing the volume of interest. (c) 2D intensity images become blurred by scattering. (d,e) Comparison of localization capabilities for two-photon and our light-field microscopy, with and without scattering. (d) Estimated source positions are projected onto the x, y plane for visualization and (e) shown in 3D.

where $\mathbf{S} \in R_+^{N_l^2 N_p^2 \times N_k}$ and $\mathbf{T} \in R_+^{N_k \times N_t}$. The number of individual components in the training data, N_k, is determined by decomposing matrix $\mathbf{I}$ into singular values. The analysis is done via a nonnegative matrix factorization optimization [110, 180], modified by adding a lasso regularization on the temporal component [178, 138]:

$$\min_{\mathbf{S}_{i,k}>0, \mathbf{T}_{k,j}>0} \|\mathbf{I} - \mathbf{S}\mathbf{T}\|^2 + \lambda_1 \sum_{k,j} |\mathbf{T}_{k,j}|, \tag{6.16}$$

where λ_1 depends on the level of spontaneous neural activity in the calibration dataset and is manually adjusted here to fit the data. See Supplement 1 of [146] for details on the ICA step. The regularization of the temporal component may be improved by taking into account the specific temporal dynamics of calcium fluorescence [149].

The components $I_{\mathrm{tr}}^k(\xi)$ are then formed from the columns of matrix $\mathbf{S}$.

6.7.2 3D localization and light-field signature extraction

We take extra standard single-shot light-field experiments to demonstrate the 3D detection and localization capabilities of our system. We show experimental results for a simple static fluorescent test object *with* and *without* scattering. Our non-scattering sample is a static suspension (5.0×10^3 μL^{-1}) of 1 μm sparsely-distributed fluorescent beads in a 200 μm

slice of agarose gel. As a proxy to ground-truth knowledge of the bead positions, we use two-photon microscopy to scan the imaging volume. We then record a single-shot light-field frame, which is shown in Fig. 6.7(a) along with several space-angle slices of the 4D light field. Each bead traces out a tilted line in the space-angle plot, as expected. After estimating the 3D bead positions using Eq. (6.12) (k here traverses only 1 index because of only single shot), we compare the results to our two-photon data (Figs. 6.7(d) and (e)). Both detect the same set of beads, with a median difference in position of 1.3 μm in the (x, y) plane and 12.8 μm along the z axis. Assuming that the two-photon result is accurate, the error in our scheme is small enough to distinguish individual neurons and localize them.

The next step is to test our method with scattering tissue. We do this by repeating the same experiment, after placing a 100 μm slice of mouse brain tissue on top of the sample. This emulates conditions that would normally prevent good depth reconstruction. To get a sense of the amount of scattering, we show 2D intensity images in Fig. 6.7(c). The scattered image is degraded, yet there is still structure in the 4D light-field measurement. Despite scattering, the median difference in detected positions between our algorithm and the two-photon data is 1.8 μm in the (x, y) plane and 15.5 μm along the z axis, slightly worse than the non-scattering case. A video that shows localization with light-field refocusing and thresholding fails can be found in Visualization 2 of [146]. These experiments show that the presence of optical scattering in this case does not significantly affect our ability to localize sparse sources.

Finally, we take the light-field signature extraction to the experimental data. The Step 1 in Sec. 6.7.1, including the ICA step, is performed here, and the resulting light-field signatures are shown as colored pixels in Fig. 6.8(c). The result shows that our method can successfully distiguish neurons from ICA sparsified frames and can further build a dictionary of light-field signatures for activity tracking.

6.7.3   Result of neural activity tracking

After the completion of the training step, the dictionary of light-field signatures can be used to efficiently decompose any single-shot measurement acquired by the light-field microscope (including the training data) into a linear positive combination of elements of the dictionary (Step 2 in Sec. 6.7.1). The number of active neurons, N, in one frame should be smaller than the number of sensor pixels ($N_p^2 N_l^2 = 4 \times 10^6$) but not necessarily sparse. Experimental results for neural activity tracking with both the ICA step and the compressive detection and localization are shown in Fig. 6.9. A five-day-old Tg(NeuroD:GCaMP6f) zebrafish expressing GCaMP6s in the telencephalon is placed in the microscope. The fish is live, awake, and immobilized in 2% low-temperature melting agarose. 40 independent components are extracted from 500 diverse frames from the calibration step (the parameter σ is found to be 0.02 μm^{-1}). Each independent component is then separated into single neuron signatures. The final dictionary contains a set of 802 light-field signatures, as well as an estimated position in 3D for each corresponding calcium source (see Fig. 6.9(c)). We then record 10 seconds of spontaneous brain activity at 100 fps. The solution to Eq. (6.15) provides a quantitative
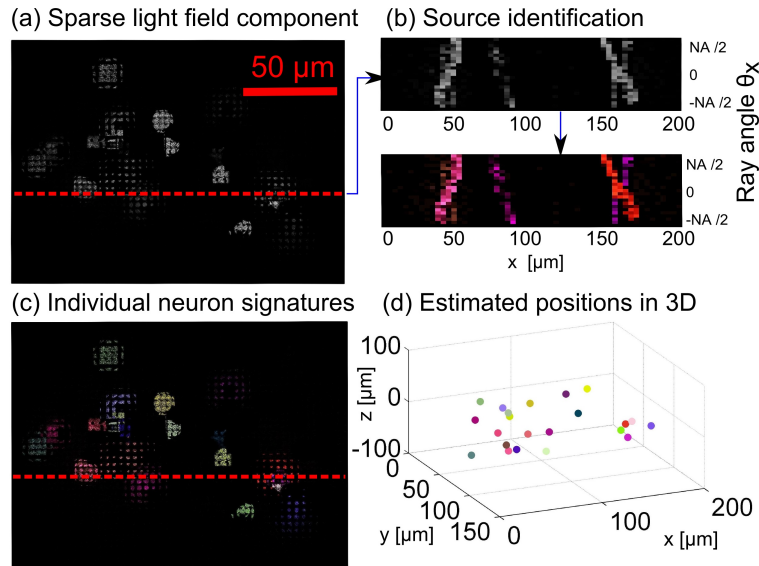
Figure 6.8: Extracting individual neuron light-field signatures and 3D positions. (a) One example from the 40 sparse light-field components. (b) Light-field slice along the red dashed line. Each distinct neuron prescribes a line in the space-angle plot, whose position and tilt indicates lateral position and depth, respectively. Individual neuron structures are distinguished and localized in light-field space – shown here as different colors. (c) Overlay of extracted light-field signatures for multiple neural structures, each with a different color. (d) Estimated 3D positions for each of the neurons in this component.
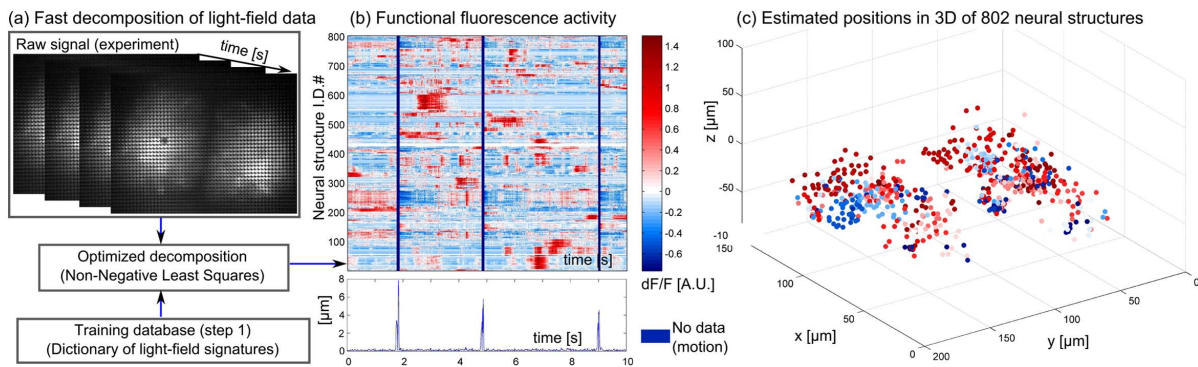
Figure 6.9: Neural activity tracking in the telencephalon of a five-day-old live zebrafish restrained in agarose. (a) Light-field signatures were extracted for 802 neural structures and 10 seconds of spontaneous activity was recorded at 100 Hz. (b) The normalized change of fluorescence, dF/F, is displayed for each neuron as a function of time. Bottom part shows the fish's motion, which is quantified by digitally tracking the first moment of the 2D image, with visible motion artifacts at $t = 1.9$ s, $t = 4.9$ s, and $t = 9$ s. (c) For each identified neuron, the position in 3D space is estimated, with color showing time-averaged fluorescence activity across both telencephalic lobes of the fore-brain.

measurement of fluorescence for all neurons in the field of view for which a light-field signature has been identified. Fig. 6.9(b) shows color-coded lines representing the normalized change of fluorescence, F, as a function of time, given by:

$$\left(\frac{\mathrm{d}F}{F_0}\right)_j (t) = \frac{a_i(t)}{(1/T)\int_{t'=0}^{T} a_j(t')\,\mathrm{d}t'} - 1, \tag{6.17}$$

where T is the time duration of the video and j indexes different neural structures. We display activity in each neuron as a function of time in Fig. 6.9(b) and show a video reconstruction of the 3D activity in Visualization 4 of [146].

We have demonstrated compressive light-field microscopy as a path toward directly addressing the needs of neuroscience for accurate, quantitative measurement of fluorescence activity in the living brain. The phase space scattering theory and the optimization problem formulation here enable single-shot capture of volumetric brain activity with neuron-scale resolution. With the phase space model, the data requirements scale with the number of active neurons in a single frame, not the number of sample voxels, and the tracking does not ever involve 3D image reconstruction. We believe that this method can scale to extremely large networks of neurons and be amenable to use with patterned stimulation, enabling functional activity mapping of the entire mouse brain cortex.

6.8 Verification of the Scattering Model

We have shown that the scattering model Eq. (6.5) is successfully applied to reconstruct point sources within a scattering environment. We conduct another theoretical analysis on the path integral to verify the correctness of the model. However, the finding in the analysis suggests the following modification to Eq. (6.5) (changes revealed in red color):

$$W(\boldsymbol{r}, \boldsymbol{u}) = \frac{n_r{}^2}{2\pi\lambda^2(z_d - z_s)^3\bar{\sigma}^2} \exp\left(-\frac{n_r{}^2}{2\lambda^2(z_d - z_s)^3\bar{\sigma}^2}\left(\boldsymbol{r} - \boldsymbol{r}_s + \lambda\left(z_d - \frac{z_d - z_s}{n_r}\right)\boldsymbol{u}\right)^2\right). \tag{6.18}$$

That is, the functional relation between the variance (the square of the width) of the Gaussian function in Eq. (6.5) and the depth $z_d - z_s$ inside scattering medium should be cubic instead of quadratic. This change stems from the freedom to normalize the path integral. As a path integral conducts infinitely many integrals [126], there is at least one parameter that approaches zero when the number of integrals goes to infinity, in order to make the path integral converge. It has some freedom to determine this parameter if convergence is the only requirement. Note that the normalization constant is purely in the operators $\mathcal{D}$ and $\mathcal{P}$ in Eq. (6.3), independent of which source Wigner function W_s is being operated on. Hence we can consider a simpler case in order to reveal the constant: the diffusion of an impinging plane wave. A cascade of the volumetric diffusing operations, that is $(\mathcal{D}\mathcal{P})^N(\mathcal{D}\mathcal{P})^N$, reveals the diffusing parameter's dependence on the diffusing depth, against to the previous result Eq. (6.5) where σ does not depend on it. A detailed derivation is in Appendix B.2.

name	λ	n_r	$\Delta x, \Delta y$	N_x, N_y	Δz	bump width	bump height
value	0.68	1.0	0.33	6144	5.0	10.0	2.6587
unit	μm	none	μm	none	μm	μm	radian (as phase delay)

Table 6.1: Simulation parameters for a point source propagating through finitely many multiple diffusing screens.

6.8.1   Simulation verification

In order to test whether the proposed update is valid, we simulate a point source propagating through multiple diffusing slices. A 2D space of shape (N_y, N_x) is used with pixel sizes $\Delta y, \Delta x$. A point source with wavelength λ generates a delta function at the center of the simulation space. The light then propagates Δz to a diffusing layer, interacts with the layer, and propagates to another and so on. The propagation is done by circularly convolving the field with Rayleigh-Sommerfeld diffraction. We have ensured that N_y, N_x are large enough to avoid the wrapping around problem. The light propagates equal distances between adjacent diffusing layers. The diffusing layer serves as a random phase object whose transmittance is multiplied with the field directly in x-y space. After the field propagates to one of the 25 scattering depths $z_{\text{meas}} = 20, 40, \ldots, 500$ μm and interacts with the diffusing screen there, we record a refocused intensity image for each (refocusing to the plane of the point source). The parameters used in the simulation are listed in Table 6.1.

The random phase object is generated by convolving a Gaussian-bump phase function with delta functions at the center locations for the bumps. The formula for the Gaussian bump is

$$\text{(bump height)} \cdot \exp\left[-0.5(x^2 + y^2)/(\text{bump width})^2\right].$$

The center locations are generated from random Poisson disk sampling [31] such that the distance between any two of the center locations is at least a bump width. After the convolution, the random phase object becomes a transmittance by multiplying the imaginary unit and being exponentiated.

A subset of the refocused intensity images are shown in Fig. 6.10. As expected, even with refocusing to undo the diffraction, the field does not become a delta function due to the random scattering. The size of the refocused spot is larger when the light propagates deeper in the scattering media. We fit two Gaussian peaks to each of the intensity images, where they fits well and thus it suggest a model mismatch to both path integral formulae introduced above. A possible cause of this mismatch is the ballistic photons that are not diffused and are not considered by our phase space model. The fitting process first converts each of the intensity images to a function of radial coordinates from the center, averaging among all angles, and then uses a weighted squared difference as the fitting-cost function. The weight for each radial sample is proportional to the square root of the number of the intensity image pixels at that radial distance. Some of the fitted results are shown in Fig. 6.11. Except the
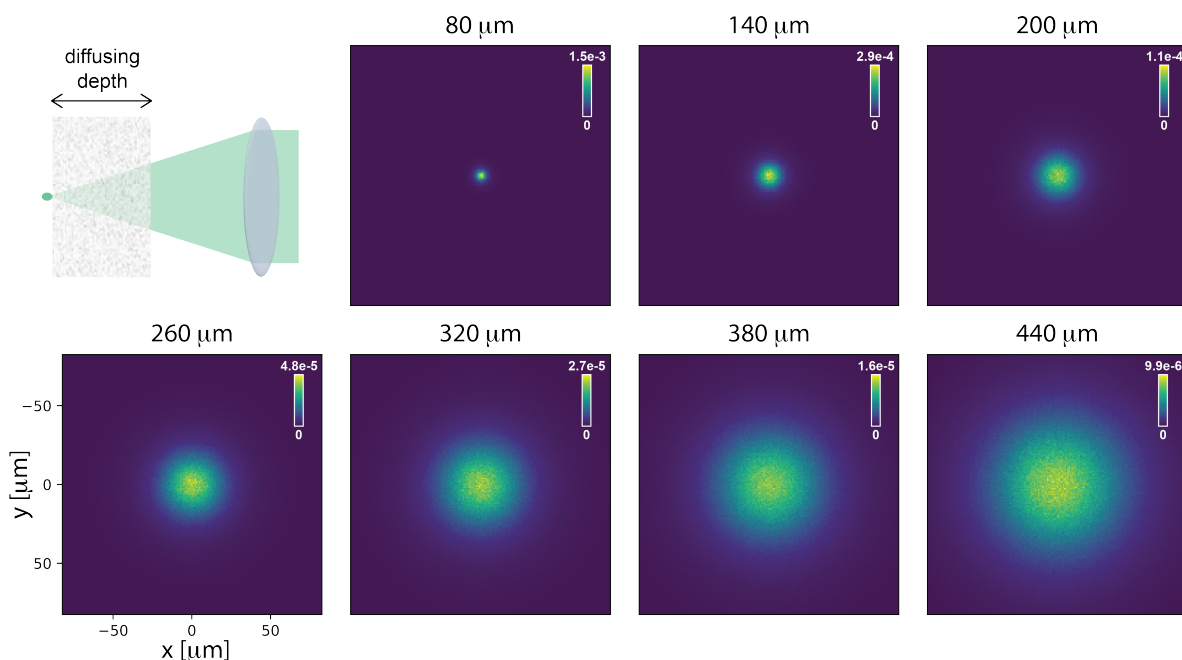
Figure 6.10: Simulated volumetric scattering and refocusing for model verification. In the top-left corner we put the simulation setup before the first lens of the system (usually a microscope objective). The light from the point source immediately goes into the volumetric scattering media, which is simulated by multiple random phase screens. After propagating through the scattering media, the light is focused back to where the point source is, which would form a sharp point image if the media is absent. The rest of the figure shows the focused images at different diffusing depths (labeled on top of each subfigures).
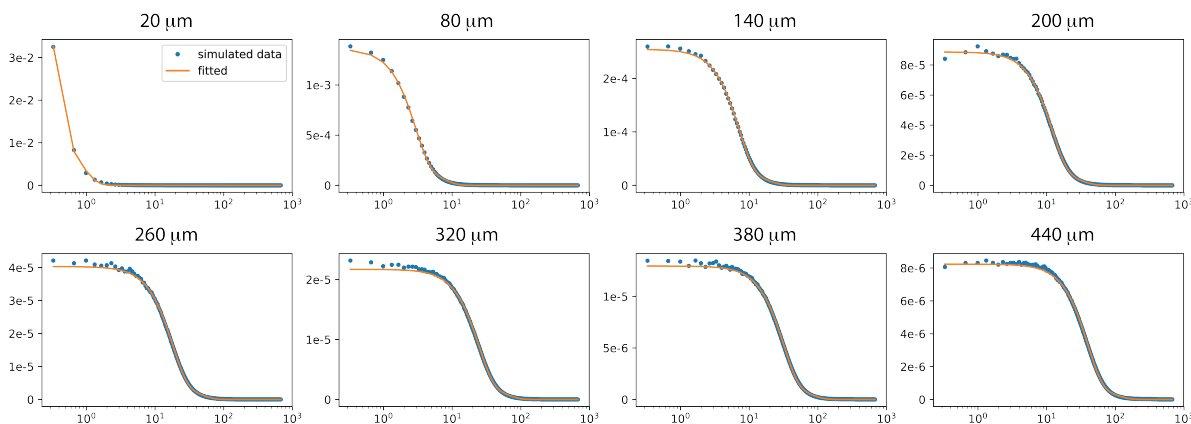
Figure 6.11: The result of two-Gaussian-peak fitting for refocused intensity images of the scattered point source. The diffusing depth is labeled on top of each subfigure.
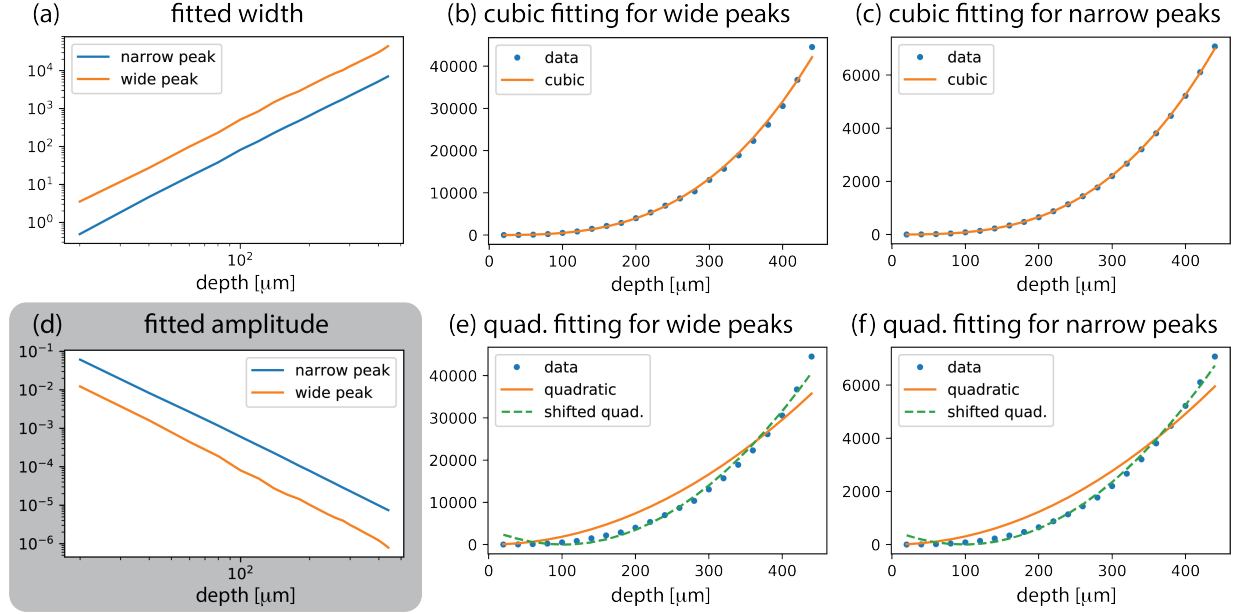
Figure 6.12: Dependence of the fitted Gaussian widths and the amplitude on the scattering depth. The vertical axes have units of μm except that of (d) which has an arbitrary unit for amplitude. (a) The fitted widths for the wide and narrow peaks v.s. the diffusing depth. (b) Showing the fitting result for the widths of wide peaks using a cube function. (c) Showing the fitting result using a quadratic function and a shifted quadratic function. The cubic function has a form of ax^3, the quadratic one has a form of ax^2, and the shifted quadratic one has a form of $a(x - c)^2$ where a is the coefficient of the polynomial and c is horizontal shift used only in the shifted quadratic function. (e)(f) Showing the result for the widths of narrow peaks, similar to (b) and (c). (d) The amplitude v.s. scattering depth. The slopes in the two log-log plots suggest negative correlation between the width and the amplitude.

region of small radii where pixel counts are too few to have reliable averaged radial function values, the two-Gaussian-peak fitting is acceptable and accurate enough for analysis.

Next we take the fitted widths from all z_{meas} to figure out the dependence of the width on the depth. We apply another polynomial fitting for the fitted widths (Gaussian variance instead of standard deviation) as a function of z_{meas}. The widths of the wide and narrow peaks are fitted separately. We found that a cubic function fits better than a quadratic one (see Fig. 6.12). From the fitted polynomial, we can derive the values for our models (Eq. (6.5) and Eq. (6.18)). The fitted values for wide peaks are $\bar{\sigma} = 0.03270$ (cubic), $\sigma = 0.6320$ (quadratic) and those for narrow peaks are $\bar{\sigma} = 0.01332$, $\sigma = 0.2578$. This confirms Eq. (6.18)'s amendment to Eq. (6.5). However, we have seen from previous sections that the model (Eq. (6.5)) works for the experimental data. To address that, we consider a horizontal shift in the quadratic fitting function, which corresponds to a poorly estimated thickness of the scattering media z_d. We find out that a shifted quadratic function can fit the data well (see Figs. 6.12(e)(f)), except the cases where the point source propagates through
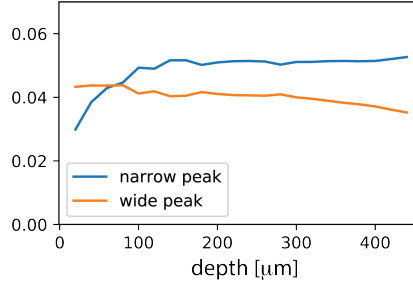
Figure 6.13: Product of the fitted width and the fitted amplitude as a function of depth. The vertical axis has an arbitrary unit for the product of the fitted width and amplitude.

a small depth of scattering media, that is, near the surface of the the scattering media. The fitted parameters of the shifted quadratic function are $\sigma = 0.8755$, shift= 101.2 μm for wide peaks and $\sigma = 0.3525$, shift= 97.78 μm for narrow peaks.

Furthermore, the negative correlation between the width and the amplitude, as shown in Figs. 6.12(a) and (d), motivates us to examine if the product of them is constant across different depths. This behavior is suggested by our models. Multiplying the width with the amplitude for each depth shows that the products are close for the wide peak and for the narrow peak (see Fig. 6.13).

The above multi-screen simulation suggests that our model captures the depth dependence of the width and amplitude of the peaks. The unexplained feature, the two-peak behavior, can potentially be addressed by considering ballistic photons at deriving the phase space model. This will mount a modification to the path integral derived in Appendix B for future work.

Chapter 7

SEAGLE: A Light Scattering Model

Phase space reveals information that is hidden in an intensity image under scattering situations. It is helpful to understand how the scatterers interact with the illuminating light, in order to extract more information out of the measurements. Hence we turn to study this scattering process through the wave equation. The scatterers usually form a nonuniform distribution of index of refraction, which bends the light in a complicated way. The study of the scattering process aims at creating a model that describes this complex operation, and using the model to estimate the distribution of scatterers given measured data. This is then equivalent to reconstructing the spatial permittivity distribution of the scatterers.

7.1   Introduction

Reconstruction of the spatial permittivity distribution of an unknown object from the measurements of the scattered waves at different illumination angles is common in numerous applications. Traditional formulations of the problem are based on linearizing the relationship between the permittivity and the measured wave. For example, if one assumes straight-ray propagation of waves, the phase of the transmitted wave can be interpreted as a line integral of the permittivity along the propagation direction. This approximation leads to an efficient reconstruction with the filtered back-projection algorithm [93]. Diffraction tomography uses a more refined linear scattering model based on the first Born or Rytov approximations [187, 52, 21]. It establishes a Fourier transform-based relationship between the measured wave and the permittivity, and thus enables the reconstruction of the latter with a direct numerical application of the inverse Fourier transform.

Recent research in compressive sensing and sparse signal processing has established that sparse regularization can dramatically improve the quality of reconstructed images, even when the amount of measured data is severely limited [35, 53]. This has popularized optimization-based inverse scattering approaches that combine linear forward models with regularizers that mitigate ill-posedness by promoting solutions that are sparse in a suitable transform domain. One class of such regularizers is total variation (TV) [155], which
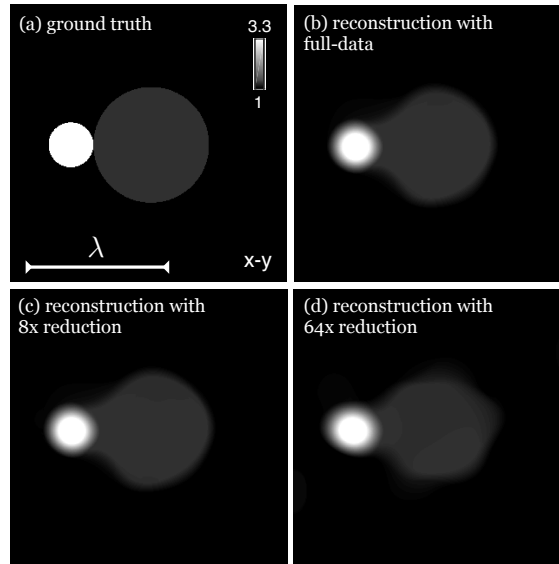
Figure 7.1: SEAGLE can be used to reconstruct the spatial distribution of dielectric permittivity from measurements [68] of complex scattered waves at different illumination angles. Illustration using experimental data at 3 GHz: (a) ground truth; (b) using full data; (c) 8× data reduction; (d) 64× data reduction.

substantially reduces undesired artifacts due to missing data [32, 119, 168].

The main advantage of imaging with linear forward models is that the reconstruction can be reduced to a convex optimization problem that is relatively simple and efficient to solve [27, 142, 19, 11]. However, multiple scattering of waves limit the validity of the Born and Rytov approximations when the objects are relatively large or have high permittivity contrast compared to the background [40]. Multiple scattering is a fundamental problem in diffraction tomography and its complete resolution would enable imaging through strongly scattering objects such as human tissue [143]. As multiple scattering leads to nonlinear forward models, the challenge is in finding computationally tractable methods that can account for the nonlinearity while also making the image reconstruction tractable. To that end, we propose a new method that efficiently combines a nonlinear forward model with the TV regularizer, thus enabling high-quality imaging from a limited number of measurements. Figure 7.1 provides an example, illustrating the quality of reconstruction for a strongly scattering object from a public experimental dataset [68]. In particular, the 320×320 image in Figure 7.1(d) was obtained from only 16 experimentally collected measurements at different illumination angles.

7.1.1 Contributions

Our work builds upon prior work on inverse scattering that has been applied to a variety of practical problems in optical, microwave, and radar imaging. The proposed method—called Series Expansion with Accelerated Gradient Descent on Lippmann-Schwinger Equation (SEAGLE)—further extends this work by considering an iterative forward model that still enables efficient sparsity-driven inversion. The performance of SEAGLE is robust to large permittivity contrast, data reduction, and measurement noise.

The key contribution of this paper is a novel image reconstruction strategy based on the explicit evaluation of the gradient of an iterative forward model with respect to the unknown parameters that correspond to the permittivity of the object. Specifically, we rely on the Nesterov's accelerated-gradient method (AGM) [137] to iteratively approximate the scattered waves. The key benefit is the guaranteed convergence of the latter even for objects with large permittivity contrast, even for general nonlinear functions of the scattered wave. However, the solution to AGM may not be unique if the functions are not strongly convex. We therefore additionally present extensive validation of our approach on analytical, simulated, and experimental data. The experimental data used in our evaluations comes from a public dataset [68], which enables easy comparison with several other related approaches.

7.1.2 Related Work

Imaging systems, such as optical projection tomography (OPT), diffraction tomography, optical coherence tomography (OCT), digital holography, and subsurface radar, rely on the linearization of object-wave interaction [159, 44, 32, 108, 169, 168, 104, 119, 152, 50, 28, 177, 42, 90, 114, 120]. Early work in microwave imaging has shown the promise of accounting for the nonlinear nature of scattering [179, 186, 43, 106]. These have been extended by a large number of techniques incorporating the nonlinear nature of scattering. Several recent publications have reviewed these methods [78, 135, 89], which include the conjugate gradient method (CGM) [39, 15], the contrast source inversion method (CSIM) [55], the hybrid method (HM) [135], and the recursive Born method (RBM) [95]. Some recent work has explored the idea of statistically modeling multiple scattering for imaging through diffusive or turbid media [38, 164, 122]. Other work has explored the combination of nonlinear scattering with a sparsity regularization [122, 193, 34, 97]. This paper extends our preliminary work [123] by including key mathematical derivations, as well as more extensive validation on experimentally collected data.

Recently, nonlinear inverse scattering in transmission attracts public attention and the beam propagation method (BPM) was proposed for performing it [105, 127, 174, 97, 96]. BPM-based methods numerically propagate the field slice-by-slice through the object. The Jacobian matrix of BPM can be efficiently computed with error backpropagation algorithms, which enables fast image reconstruction. The nonlinear model presented here is based on the Lippmann-Schwinger equation [22]. The main advantage of the proposed formulation is

that it accounts for both transmitted and reflected waves. This makes the proposed method more suitable than BPM when reflections are important.

The Lippmann-Schwinger equation, also known as the Foldy-Lax multiple scattering model, has been extensively used in the inverse scattering literature for imaging under wave scattering [190, 57], diffuse optical tomography [113, 112], impedance tomography [111], and for elastic wave scattering problems [192]. The common theme involves first estimating the contrast source reflectivity of the target by exploiting the joint sparsity across multiple illuminations. Then, the Lippmann-Schwinger equation is used to estimate the total field, which in turn is used to separate the target permittivity from the estimated contrast source reflectivity. Our proposed method differs from these works in that we jointly estimate the total field as well as the target permittivity in a closed loop framework that allows us to exploit the spatial structure of the target through regularization.

The problem setting in this paper is also related to that in full-waveform inversion (FWI) [172, 182, 82] used in geophysical applications. We propose an alternative approach for solving such problems, namely modeling the forward scattering process using AGM and estimating the material parameters. FWI-based methods often use a differential form of Helmholtz equation, while we rely on integral-domain formulation given by the Lippmann-Schwinger equation. Moreover, established frameworks in FWI-based methods utilize Krylov based solvers along with the adjoint state method [141] to estimate the gradient of the forward wave propagation model with respect to the permittivity. However, these methods rely on the linearity of the wave equation as a function the scattered field. Otherwise, for nonlinear functions of the scattered field, these methods require iterative linearization which can become slow and lacks convergence guarantees. Our AGM-based forward model on the other hand is guaranteed to converge for general nonlinear convex functions of the scattered field, and has the same asymptotic convergence rate as Krylov-based methods for strictly convex functions. Moreover, a key difference of our method is in the combination of the AGM-based forward model and sparsity-driven image reconstruction using the fast iterative shrinkage/thresholding algorithm (FISTA) [12]. Our experiments show that our formulation is promising, as it enables fast, stable, and reliable convergence when working with a limited amount of data.

The experimental data used in this paper comes from a public dataset of complex wave-field measurements of several objects at various illumination angles and frequencies [68]. Several other methods have been tested on this dataset [2, 10, 47, 55, 60]. This enables qualitative evaluation of the performance of the proposed technique against other algorithms. While many of the methods tested on the data use multiple frequencies, the results here rely on a single frequency. However, the method uses the latest techniques in large-scale optimization with sparse regularization, which enables subsampling and leads to improvements in imaging performance. We expect the performance of the proposed method to improve further if multi-frequency measurements are incorporated.
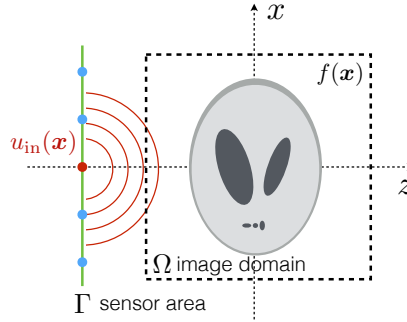
Figure 7.2: Schematic representation of the scattering experiment. An object with a real scattering potential $f(\boldsymbol{x})$, $\boldsymbol{x} \in \Omega$, is illuminated with an input wave u_{in}, which interacts with the object and leads to the scattered wave u_{sc} measured at the sensing region Γ represented with a green line.

7.2 Forward Model

The forward problem computes the scattered field given a distribution of inhomogeneous permittivity, while the inverse problem reconstructs this distribution. The model we propose here can be interpreted as a series expansion based on the iterates of the gradient method. This expansion can be made arbitrarily accurate and is stable for high permittivity objects. Additionally, it enables efficient computation of the gradient of the cost function, which is essential for fast image reconstruction. In this section we focus on introducing the forward model, leaving the gradient evaluation and inversion to the next section. Our derivations are for the scenario of a single illumination, but the generalization to an arbitrary number of illuminations is straightforward.

7.2.1 Problem formulation

Consider the scattering problem in Figure 7.2, where an object of the permittivity distribution $\epsilon(\boldsymbol{x})$ in the bounded domain $\Omega \subseteq \mathbb{R}^D$, with $D \in \{2, 3\}$, is immersed into a background medium of permittivity ϵ_b, and illuminated with the incident electric field $u_{\text{in}}(\boldsymbol{x})$. We assume that the incident field is monochromatic and coherent, and it is known inside Ω and at the locations of the sensors Γ. The result of object-wave interaction is measured at the location of the sensors as a scattered field $u_{\text{sc}}(\boldsymbol{x})$. The scattering of light can be accurately described by the Lippmann-Schwinger equation inside the image domain [22]

$$u(\boldsymbol{x}) = u_{\text{in}}(\boldsymbol{x}) + \int_{\Omega} g(\boldsymbol{x} - \boldsymbol{x}')\, f(\boldsymbol{x}')\, u(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}', \quad (\boldsymbol{x} \in \Omega) \tag{7.1}$$

where $u(\boldsymbol{x}) = u_{\text{in}}(\boldsymbol{x}) + u_{\text{sc}}(\boldsymbol{x})$ is the total electric field, $f(\boldsymbol{x}) \triangleq k^2(\epsilon(\boldsymbol{x}) - \epsilon_b)$ is the scattering potential, which is assumed to be real, and $k = 2\pi/\lambda$ is the wavenumber in vacuum. The

function $g(\boldsymbol{x})$ is the Green's function defined as

$$
g(\boldsymbol{x}) \triangleq \begin{cases} \dfrac{\mathrm{j}}{4} H_0^{(1)}(k_b \|\boldsymbol{x}\|_{\ell_2}) & \text{in 2D} \\[2ex] \dfrac{\mathrm{e}^{\mathrm{j}k_b\|\boldsymbol{x}\|_{\ell_2}}}{4\pi\|\boldsymbol{x}\|_{\ell_2}} & \text{in 3D,} \end{cases} \tag{7.2}
$$

where $k_b \triangleq k\sqrt{\epsilon_b}$ is the wavenumber of the background medium and $H_0^{(1)}$ is the zero-order Hankel function of the first kind. Note that the Green's function satisfies the Helmholtz equation

$$
\left(\nabla^2 + k_b^2\right) g(\boldsymbol{x}) = -\delta(\boldsymbol{x}),
$$

as well as the outgoing-wave boundary condition. The knowledge of the total-field u inside the image domain Ω enables the prediction of the scattered field at the sensor area

$$
u_{\mathrm{sc}}(\boldsymbol{x}) = \int_{\Omega} g(\boldsymbol{x} - \boldsymbol{x}')\, f(\boldsymbol{x}')\, u(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}'. \quad (\boldsymbol{x} \in \Gamma) \tag{7.3}
$$

The computation of the scattered wave is equivalent to solving Eq. (7.1) for $u(\boldsymbol{x})$ inside the image and evaluating Eq. (7.3) for $u_{\mathrm{sc}}(\boldsymbol{x})$ at the sensor locations. Note that u is present on both sides of Eq. (7.1) and that the relation between u and the scattering potential f is nonlinear. The first Born and the Rytov approximations [187, 52, 21] are linear models that replace u in Eq. (7.3) with a suitable approximation that decouples the nonlinear dependence of u on f. However, such linearization imposes a strong assumption that the object is weakly scattering, which makes the corresponding reconstruction methods inapplicable to a large variety of imaging problems [40]. Our forward model described next is a fast nonlinear method for solving Eq. (7.1) that overcomes these weakly scattering object assumptions.

The results presented in this paper rely on the scalar theory of diffraction, which yields accurate results when two conditions are met: (*i*) objects are sufficiently large compared with the wavelength; (*ii*) the measurements are taken sufficiently far from the object. Experimental demonstration of this was provided by Silver [162] and the applicability of scalar theory to instrumentation was extensively discussion in the classical literature in electromagnetics and optics [23, 72]. Note also that both of these conditions are generally met in optical imaging, which is an important 3D application area for the method in this paper.

7.2.2 Algorithmic Expansion of the Scattered Waves

We separate the computation of the electric field into two parts: the total field $u(\boldsymbol{x})$ in the image domain and the scattered field $u_{\mathrm{sc}}(\boldsymbol{x})$ at the sensors. The discretization and combination of Eq. (7.1) and Eq. (7.3) leads to the following matrix-vector description of the forward problem

$$
\mathbf{y} = \mathbf{H}(\mathbf{u} \bullet \mathbf{f}) + \mathbf{e} \tag{7.4a}
$$

$$
\mathbf{u} = \mathbf{u}_{\mathrm{in}} + \mathbf{G}(\mathbf{u} \bullet \mathbf{f}), \tag{7.4b}
$$

Algorithm 1 Forward model computation.

> **intput:** Image $\mathbf{f} \in \mathbb{R}^N$, maximum number of iterations K, tolerance δ_{tol}, and initial-
> ization $\mathbf{u}_{\mathrm{init}} = \mathbf{u}_{\mathrm{in}}$.
> **set:** $\mathbf{u}^{-1} \leftarrow \mathbf{u}_{\mathrm{init}}$, $\mathbf{u}^0 \leftarrow \mathbf{u}_{\mathrm{init}}$, $t_0 \leftarrow 0$

1: **for** $k \leftarrow 1$ to K **do**
2:   $t_k \leftarrow (1 + \sqrt{1 + 4t_{k-1}^2})/2$,
3:   $\mu_k \leftarrow (1 - t_{k-1})/t_k$
4:   $\mathbf{s}^k \leftarrow (1 - \mu_k)\mathbf{u}^{k-1} + \mu_k \mathbf{u}^{k-2}$
5:   $\mathbf{g} \leftarrow \mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\mathrm{in}})$ ▷ gradient at $\mathbf{s}^k$
6:   **if** $\|\mathbf{g}\|_2 < \delta_{\mathrm{tol}}$ **then** $K \leftarrow k$, break the loop
7:   $\gamma_k \leftarrow \|\mathbf{g}\|_2^2 / \|\mathbf{A}\mathbf{g}\|_2^2$
8:   $\mathbf{u}^k \leftarrow \mathbf{s}^k - \gamma_k \mathbf{g}$
9: $\widehat{\mathbf{u}} \leftarrow \mathbf{u}^K$
10: $\mathbf{z} \leftarrow \mathbf{H}(\widehat{\mathbf{u}} \bullet \mathbf{f})$
> **return:** predicted scattered wave $\mathbf{z}$, as well as $\widehat{\mathbf{u}}$, $\{\mathbf{s}^k\}$, $\{\gamma_k\}$, and $\{\mu_k\}$.

---

where $\mathbf{f} \in \mathbb{R}^N$ is the discretized scattering potential f, $\mathbf{y} \in \mathbb{C}^M$ is the measured scattered field u_{sc} at Γ, $\mathbf{u}_{\mathrm{in}} \in \mathbb{C}^N$ is the input field u_{in} inside Ω, $\mathbf{H} \in \mathbb{C}^{M \times N}$ is the discretization of the Green's function at Γ, $\mathbf{G} \in \mathbb{C}^{N \times N}$ is the discretization of the Green's function inside Ω, the symbol $\bullet$ denotes a component-wise multiplication between two vectors, and $\mathbf{e} \in \mathbb{C}^M$ models the random noise at the measurements. Using the shorthand notation $\mathbf{A} \triangleq \mathbf{I} - \mathbf{G}\mathrm{diag}(\mathbf{f})$, where $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity matrix and $\mathrm{diag}(\cdot)$ is an operator that forms a diagonal matrix from its argument, we can represent the forward scattering in Eq. (7.4b) as a minimization problem

$$\widehat{\mathbf{u}}(\mathbf{f}) \triangleq \underset{\mathbf{u} \in \mathbb{C}^N}{\arg\min} \{\mathcal{S}(\mathbf{u})\} \tag{7.5}$$

$$\text{with} \quad \mathcal{S}(\mathbf{u}) \triangleq \frac{1}{2}\|\mathbf{A}\mathbf{u} - \mathbf{u}_{\mathrm{in}}\|_{\ell_2}^2,$$

where the matrix $\mathbf{A}$ is a function of $\mathbf{f}$. The gradient of $\mathcal{S}$ can be computed as

$$\nabla\mathcal{S}(\mathbf{u}) = \mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{u} - \mathbf{u}_{\mathrm{in}}). \tag{7.6}$$

Since Eq. (7.5) corresponds to the optimization of a differentiable function, it is possible to compute the total field $\widehat{\mathbf{u}}$ iteratively using Nesterov's AGM [137]

$$t_k \leftarrow \frac{1}{2}\left(1 + \sqrt{1 + t_{k-1}^2}\right) \tag{7.7a}$$

$$\mathbf{s}^k \leftarrow \mathbf{u}^{k-1} + ((t_{k-1} - 1)/t_k)(\mathbf{u}^{k-1} - \mathbf{u}^{k-2}) \tag{7.7b}$$

$$\mathbf{u}^k \leftarrow \mathbf{s}^k - \nu\mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\mathrm{in}}), \tag{7.7c}$$
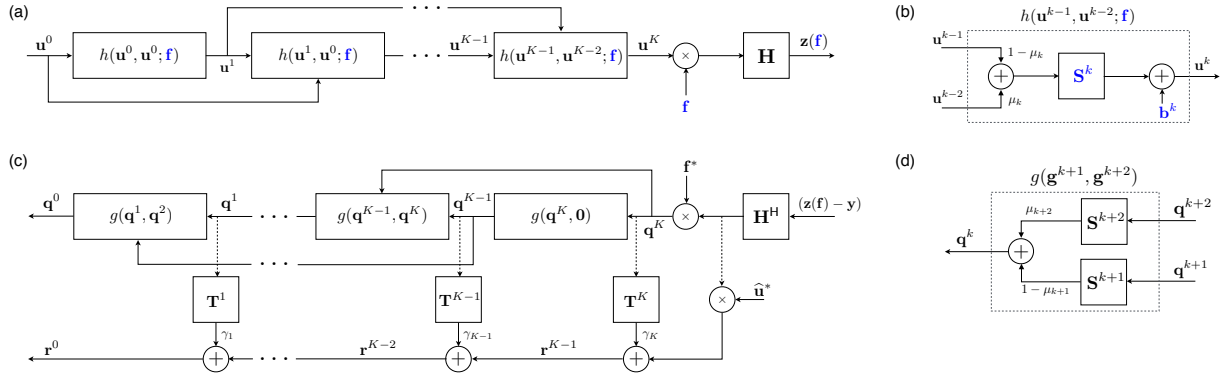
Figure 7.3: A schematic representation of the method with adaptable parameters $\mathbf{S}^k \triangleq \mathbf{I} - \gamma_k \mathbf{A}^\mathsf{H} \mathbf{A}$ and $\mathbf{b}^k \triangleq \gamma_k \mathbf{A}^\mathsf{H} \mathbf{u}^0$ that depend on the scattering potential $\mathbf{f}$. (a) Forward model computation corresponding to Algorithm 1 for K forward iterations. Adaptable parameters are marked in blue. (b) The schematic view of a single forward iteration. (c) Error backpropagation corresponding to Algorithm 3 for K iterations. (d) The schematic view of a single backward iteration. The notation $\mathbf{v}^*$ indicates the elementwise complex conjugation of the vector $\mathbf{v}$. Note that the algorithm does not require physical storage of matrices, as they can be efficiently implemented as convolutions using FFT.

for $k = 1, 2, \ldots, K$, where $\mathbf{u}^0 = \mathbf{u}^{-1} = \mathbf{u}_{\text{in}}$, $q_0 = 1$, and $\nu > 0$ is the step-size. At any moment, the predicted scattered field can be set to $\mathbf{z}^k = \mathbf{H}(\mathbf{u}^k \bullet \mathbf{f})$ with $\mathbf{u}^k$ given by Eq. (7.7c). Note that the resulting set of fields $\{\mathbf{u}^k\}_{k \in [1\ldots K]}$ and $\{\mathbf{z}^k\}_{k \in [1\ldots K]}$ can be interpreted as a K-term series expansion of the wave-fields inside the object and at the sensor locations, respectively. The full procedure for forward computation with a convenient adaptive step-size is summarized in Algorithm 1.

There are strong parallels between the AGM-based formulation of scattering and the popular Born series expansion [22, 95]. Both approaches produce a sequence of wave-field vectors $\{\mathbf{u}_k\}_{k \in [1,\ldots,K]}$, starting from the initial $\mathbf{u} = \mathbf{u}_{\text{in}}$. The final field $\widehat{\mathbf{u}}$ is the linear combination of all the intermediate field vectors, which indicates that this is an expansion of the field $\widehat{\mathbf{u}}$ with respect to K-terms, where each term brings additional information about scattering. The traditional Born series and SEAGLE are thus identical for $K = 0$, but yield different intermediate fields for any other K. Crucially, while Born series is known to diverge for strong scatterers, AGM is guaranteed to converge for sufficiently large K, as Eq. (7.5) is a smooth and convex optimization problem [137, 136, 12]

7.3   Inverse Problem

We now present the overall image reconstruction algorithm, based on the state-of-the-art fast iterative shrinkage/thresholding algorithm (FISTA) [12]. The application of FISTA to nonlinear inverse scattering is, however, nontrivial due to the requirement of the gradient of

---

Algorithm 2 Image formation with FISTA.

> **input:** scattered field $\mathbf{y}$, initial guess $\mathbf{f}^0$, initial step $\gamma_0 > 0$, step reduction rate $0 < \eta < 1$, and regularization parameter $\tau > 0$.
> **set:** $t \leftarrow 1$, $\tilde{\mathbf{f}}^0 \leftarrow \mathbf{f}^0$, $q_0 \leftarrow 1$

1: **repeat**
2: $\quad \gamma_t \leftarrow \gamma_{t-1}/\eta$
3: $\quad$ **repeat** $\hfill \triangleright$ line search
4: $\quad\quad \gamma_t \leftarrow \eta \gamma_t$
5: $\quad\quad \mathbf{f}^t \leftarrow \operatorname{prox}_{\gamma\tau\mathcal{R}}(\tilde{\mathbf{f}}^{t-1} - \gamma_t \nabla\mathcal{D}(\tilde{\mathbf{f}}^{t-1}))$
6: $\quad$ **until** $\mathcal{D}(\mathbf{f}^t) + \tau\mathcal{R}(\mathbf{f}^t) \leq Q_{\gamma_t}(\mathbf{f}^t, \tilde{\mathbf{f}}^{t-1})$
7: $\quad q_t \leftarrow \frac{1}{2}\left(1 + \sqrt{1 + 4q_{t-1}^2}\right)$
8: $\quad \tilde{\mathbf{f}}^t \leftarrow \mathbf{f}^t + ((q_{t-1} - 1)/q_t)(\mathbf{f}^t - \mathbf{f}^{t-1})$
9: $\quad t \leftarrow t + 1$
10: **until** stopping criterion
> **return:** estimate of the scattering potential $\mathbf{f}^t$.

Algorithm 3 Error backpropagation for $\nabla\mathcal{D}(\mathbf{f})$.

> **intput:** Image $\mathbf{f} \in \mathbb{R}^N$, measurements $\mathbf{y} \in \mathbb{C}^M$, input wave field $\mathbf{u}_{\text{in}} \in \mathbb{C}^N$.

1: $(\mathbf{z}, \widehat{\mathbf{u}}, \{\mathbf{s}^k\}, \{\gamma_k\}, \{\mu_k\}) \leftarrow$ run Algorithm 1
2: $\mathbf{q}^{K+1} \leftarrow \mathbf{0}$
3: $\mathbf{q}^K \leftarrow \operatorname{diag}(\mathbf{f})^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}(\mathbf{z} - \mathbf{y})$
4: $\mathbf{r}^K \leftarrow \operatorname{diag}(\widehat{\mathbf{u}})^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}(\mathbf{z} - \mathbf{y})$
5: **for** $k \leftarrow K$ to 1 **do**
6: $\quad \mathbf{S}^k \triangleq \mathbf{I} - \gamma_k\mathbf{A}^{\mathsf{H}}\mathbf{A}$
7: $\quad \mathbf{T}^k \triangleq \operatorname{diag}(\mathbf{G}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}))^{\mathsf{H}} + \operatorname{diag}(\mathbf{s}^k)^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{A}$
8: $\quad \mathbf{q}^{k-1} \leftarrow (1 - \mu_k)\mathbf{S}^k\mathbf{q}^k + \mu_{k+1}\mathbf{S}^{k+1}\mathbf{q}^{k+1}$
9: $\quad \mathbf{r}^{k-1} \leftarrow \mathbf{r}^k + \gamma_k\mathbf{T}^k\mathbf{q}^k$
> **return:** $\nabla\mathcal{D}(\mathbf{f}) = \operatorname{Re}\{\mathbf{r}^0\}$ the gradient in Eq. (7.10).

---

the scattered field with respect to the object. We solve this by providing an explicit formula, based on error backpropagation [20].

7.3.1 Image Reconstruction

We formulate image reconstruction as the following optimization problem

$$\widehat{\mathbf{f}} = \underset{\mathbf{f}\in\mathbb{R}^N}{\arg\min}\left\{\mathcal{D}(\mathbf{f}) + \tau\mathcal{R}(\mathbf{f})\right\}, \tag{7.8a}$$

where

$$\mathcal{D}(\mathbf{f}) \triangleq \frac{1}{2}\|\mathbf{y} - \mathbf{z}(\mathbf{f})\|_{\ell_2}^2 \quad \text{and} \tag{7.8b}$$

$$\mathcal{R}(\mathbf{f}) \triangleq \sum_{n=1}^{N} \|[\mathbf{Df}]_n\|_{\ell_2} = \sum_{n=1}^{N} \sqrt{\sum_{d=1}^{D} |[\mathbf{D}_d\mathbf{f}]_n|^2}. \tag{7.8c}$$

The data-fidelity term $\mathcal{D}$ measures the discrepancy between the actual measurements $\mathbf{y}$ and the ones predicted by our scattering model $\mathbf{z}$. The function $\mathcal{R}$ is the isotropic TV regularizer and the parameter $\tau > 0$ controls the strength of the regularization, where $\mathbf{D} : \mathbb{R}^N \to \mathbb{R}^{N \times D}$ is the discrete gradient operator with matrix $\mathbf{D}_d$ denoting the finite difference operation along dimension d.

The image can then be formed iteratively using a first order method such as ISTA [63, 49, 13]

$$\mathbf{f}^t \leftarrow \text{prox}_{\gamma\tau\mathcal{R}} \left(\mathbf{f}^{t-1} - \gamma\nabla\mathcal{D}(\mathbf{f}^{t-1}) \right), \tag{7.9}$$

for $t = 1, 2, 3, \ldots$ or its accelerated variant FISTA [12] summarized in Algorithm 2. Note that the algorithm relies on the definition of the quadratic upper bound

$$Q_\gamma(x, y) \triangleq \mathcal{D}(y) + \nabla\mathcal{D}(y)^{\mathrm{T}}(x - y) + \frac{1}{2\gamma}\|x - y\|_2^2 + \tau\mathcal{R}(x)$$

for setting the step-size parameter $\gamma > 0$ using the line search. The operator $\text{prox}_{\gamma\tau\mathcal{R}}$ denotes the proximity operator, and for isotropic TV it can be efficiently evaluated [11, 94]. Finally, an efficient implementation of the imaging algorithm requires the gradient of the data-fidelity term

$$\nabla\mathcal{D}(\mathbf{f}) = \text{Re}\left\{ \left[\frac{\partial\,\mathbf{z}(\mathbf{f})}{\partial\mathbf{f}}\right]^{\mathsf{H}} (\mathbf{z}(\mathbf{f}) - \mathbf{y}) \right\}, \tag{7.10}$$

which can be evaluated explicitly using Algorithm 3.

The mathematical derivation of Algorithm 3 is given in Appendix C. It is similar to the derivation of the standard backpropagation used in deep learning [20, 109]. Figure 7.3 visually illustrates the steps required for the forward model and backpropagation computations in Algorithm 1 and Algorithm 3, respectively. In particular, Figure 7.3(a) and Figure 7.3(b) illustrate the way intermediate iterates are combined during the K forward iterations and the schematic of a single iteration, respectively. Similarly, Figure 7.3(c) and Figure 7.3(d) illustrate the computation of intermediate iterates in K backward iterations and the schematic of a single such iteration, respectively. Note that the algorithm does not need to explicitly store the matrices, as they can be implemented as convolutions using the fast Fourier transform (FFT) algorithm. Thus the memory required for the algorithm only includes the storage of iterate vectors in Algorithm 3. The overall per iteration complexity of SEAGLE for each illumination is $\mathcal{O}(KN \log N)$ where K is the number of AGM terms and N is the dimension of the imaging domain.

One of the key benefits of SEAGLE is that it offers high levels of parallelism making it well suited for GPU implementations. In particular, computations can be treated independently in parallel for each illumination, which greatly reduces the computational cost of image formation. It is possible to further reduce the cost of each imaging iteration, by considering the incremental variant of Algorithm 2 that processes only a subset of illuminations at each iteration [17].

While the theoretical convergence of FISTA is difficult to analyze for nonconvex functions, it is often used as a faster alternative to the standard gradient-based methods in the context deep learning and broader machine learning [24, 128, 25]. In fact, we observed that our method reliably converges and achieves excellent results on a wide array of problems, as reported in Section 7.4.

7.4   Experimental Evaluation

We now present the results of validating our method on analytically obtained scattering data for simple scenarios, scattering data obtained with a high-fidelity simulator, and experimentally collected data from the public dataset [67]. Note that all the image reconstruction results reported in this section rely on TV regularization.

7.4.1   Validation on analytic data

In the first set of experiments, we validated our forward model for two simple objects where analytic expressions of the scattered wave exist: a two-dimensional point source scattered by a cylinder, and a three-dimensional point source scattered by a sphere. The expressions are derived following the mathematical formalism in [86], which we review in Appendix D for completeness. As illustrated in Figures 7.4(a) and (b), in both cases, the objects have diameters equal to 6 wavelengths. The wavelength is set to 74.9 mm, the source is placed 1 m away from the center of the objects, the grid size is set to 4.8 mm (6 mm), and there are 250 points (128 points) along each axis in 2D (3D). The contrast of an object is defined as $\max(|\mathbf{f}|)/k_b^2$. In Figures 7.4(b) and (d), we quantitatively evaluate the performance of our forward model with the normalized error defined as

$$\text{normalized error} \triangleq \frac{\|\widehat{\mathbf{u}} - \mathbf{u}_{\text{true}}\|_{\ell_2}^2}{\|\mathbf{u}_{\text{true}}\|_{\ell_2}^2}, \tag{7.11}$$

where $\widehat{\mathbf{u}}$ is the solution of Eq. (7.5) and $\mathbf{u}_{\text{true}}$ is the analytic expression. For comparison, we additionally provide the performance of the first-Born (FB) and Rytov approximations at 20% contrast. In Fig. 7.4(e) and 7.4(f), we demonstrate a visual comparison between the analytic expression and the result of our model. In 2D, the forward computations took 0.03, 0.18, 0.62, 1.3, 4.4, and 14.4 seconds for 5%, 10%, 20%, 30%, 50%, and 100% contrast levels, respectively. Similarly, in 3D, the computations took 2.1, 7.7, 32, 200, and 981.7 seconds for the same contrast levels. Overall, we observed that, by allowing for a large enough value
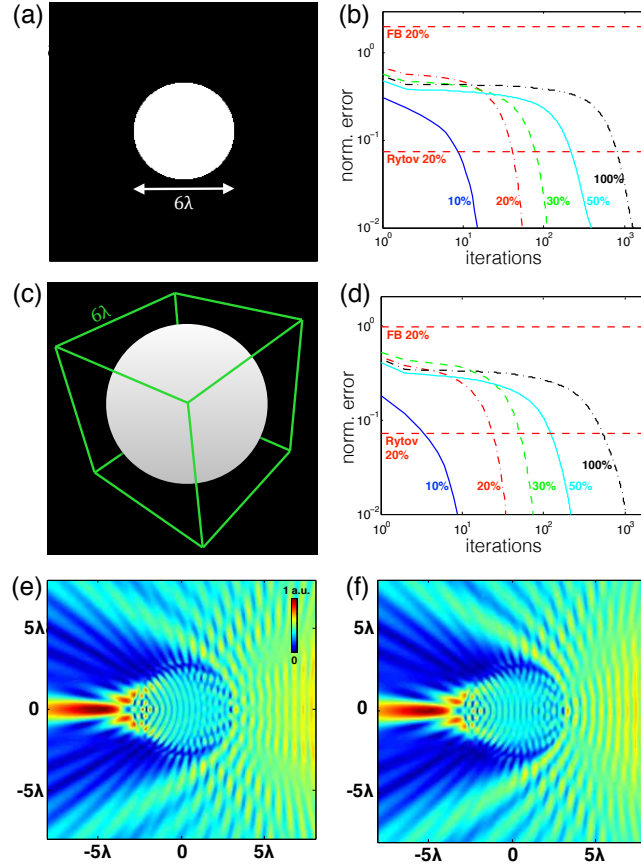
Figure 7.4:   Analytical validation of the forward model: (a) cylinder of diameter 6λ; (b) normalized errors in scattering for different contrast levels; (c) sphere of diameter 6λ; (d) normalized errors in scattering for different contrast levels; (e) analytic field for a cylinder at a contrast level of 100%; (f) corresponding field computed by our forward model.

of K, our forward model can match the analytically obtained field with arbitrarily high precision. The actual value of K depends on the severity of multiple scattering and must be adapted on the basis of the application of interest. For example, we observed that for objects closely resembling biological samples, one generally requires $10 \leq K \leq 30$.

7.4.2   Validation on simulated data

We next validated the proposed technique for reconstructing the Shepp-Logan phantom in an ill-posed, strongly scattering, and compressive regime ($M = 25 \times 338$ and $N = 250 \times 250$). Specifically, we consider the setup in Fig. 7.2 where the scattered wave measurements are generated with an FDTD simulator. The object is of size 84.9 cm × 113 cm. We place two linear detectors on either side of the phantom at a distance of 95.9 cm from the center of the object. Each detector has 169 sensors placed with a spacing of 3.84 cm. The transmitters are
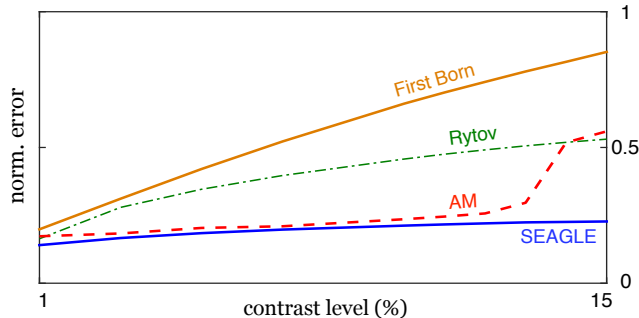
Figure 7.5: Quantitative evaluation of normalized reconstruction error against the contrast level for four methods: first-Born, Rytov, AM, and SEAGLE. All the results were obtained by using TV regularization.

positioned on a line 48.0 cm left to the left detector. They are spaced uniformly in azimuth with respect to the center of the phantom (every $5°$ within $\pm 60°$). We set up a 120 cm $\times$ 120 cm square area for reconstructing the object, with pixel size 0.479 cm. The wavelength of the illuminating light is 7.49 cm.

We compare results of our approach against three alternative methods. We regularize the solution of all the methods with TV. As the first reference method, we consider the solution of the linearized model based on FB, which is known to be valid only for weakly scattering objects. Additionally, we consider an inverse scattering approach that is based on the Rytov approximation, which is known to be more robust to moderate levels of scattering. Finally, we consider a popular optimization scheme extensively used in optical imaging and FWI, denoted AM for *alternating minimization*, for strongly scattering objects that iteratively alternates between updating the contrast function for a fixed field and updating the field for a fixed contrast function [186, 39, 15]. All three methods minimize the same error functional; however, each method relies on a distinct forward model. Image reconstruction in all the approaches was done using FISTA with TV regularizer that was empirically set for the best performance. The order of SEAGLE's forward model is set to $K = 120$, but Algorithm 1 may terminate earlier when the objective function Eq. (7.5) is below $\delta_{\text{tol}} = 5 \times 10^{-7} \|\mathbf{u}_{\text{in}}\|_{\ell_2}^2$.

Figure 7.5 summarizes the normalized error performance $\|\widehat{\mathbf{f}} - \mathbf{f}\|_{\ell_2}^2 / \|\mathbf{f}\|_{\ell_2}^2$, where $\mathbf{f}$ and $\widehat{\mathbf{f}}$ denote the true and estimated object, of all the methods for various contrast levels between 1% and 15%. The results confirm that while all the methods yield good performance at low contrasts, the performance of linearized methods, FB and Rytov, degenerate as contrast levels increase. One can also observe that the performance of AM is similar to SEAGLE for low to moderate contrast levels, but SEAGLE outperforms AM for higher contrasts. In fact, we generally observed that the performance of AM rapidly degenerates for very high contrast levels, while the performance of SEAGLE is relatively stable.

Figure 7.6 summarizes the performance of the same methods for the contrast level of 20%. We have omitted the result of AM as the method was not able to reconstruct the image at this high contrast level. The figure additionally provides quantitative performance
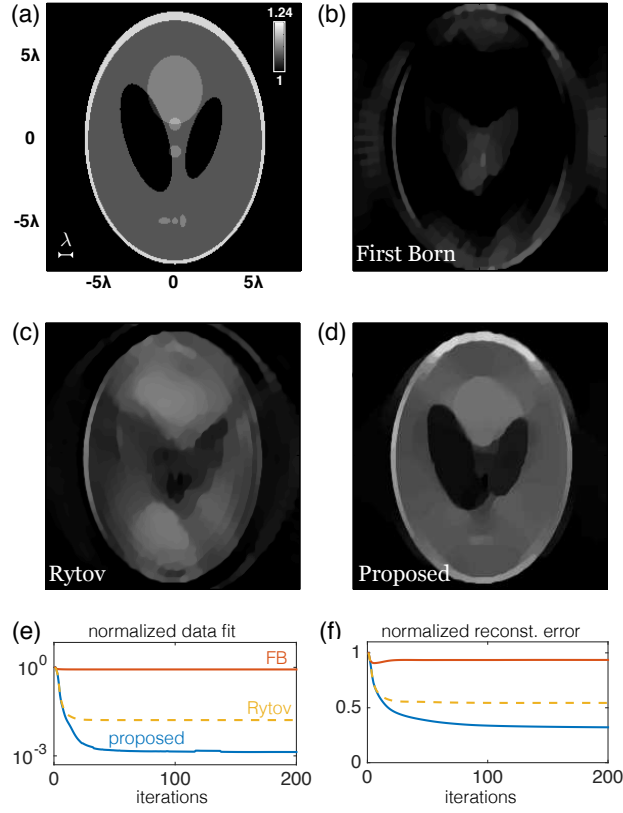
Figure 7.6: Comparison of the proposed approach with baseline methods on simulated data. (a) Shepp-Logan at 20% contrast; the reconstructed results with (b) the first Born approximation; (c) the Rytov approximation, and (d) our method; (e) per-iteration data fit; (f) per-iteration error.

evaluation in terms of data fit

$$\text{normalized data fit} \triangleq \frac{\mathcal{D}(\widehat{\mathbf{f}})}{\mathcal{D}(\mathbf{0})} = \frac{\|\mathbf{z}(\widehat{\mathbf{f}}) - \mathbf{y}\|_{\ell_2}^2}{\|\mathbf{y}\|_{\ell_2}^2}. \tag{7.12}$$

Simulation results corroborate the benefit of using the proposed method for strongly scattering objects. It can be seen that, due to the ill-posed nature of the measurements, the reconstructed images suffer from missing frequency artifacts [161]. However, the proposed method is still able to accurately capture most features of the object while the linear methods fail to do so. Note also, that our method was initialized with the background value of the dielectric permittivity, $\epsilon_{\mathrm{b}} = 1$, and that it takes fewer than 50 FISTA iterations for converging to a stationary point (see convergence plot in Figure 7.6(f)). It took SEAGLE about 1.5 seconds on average to process an illumination at each FISTA iteration.
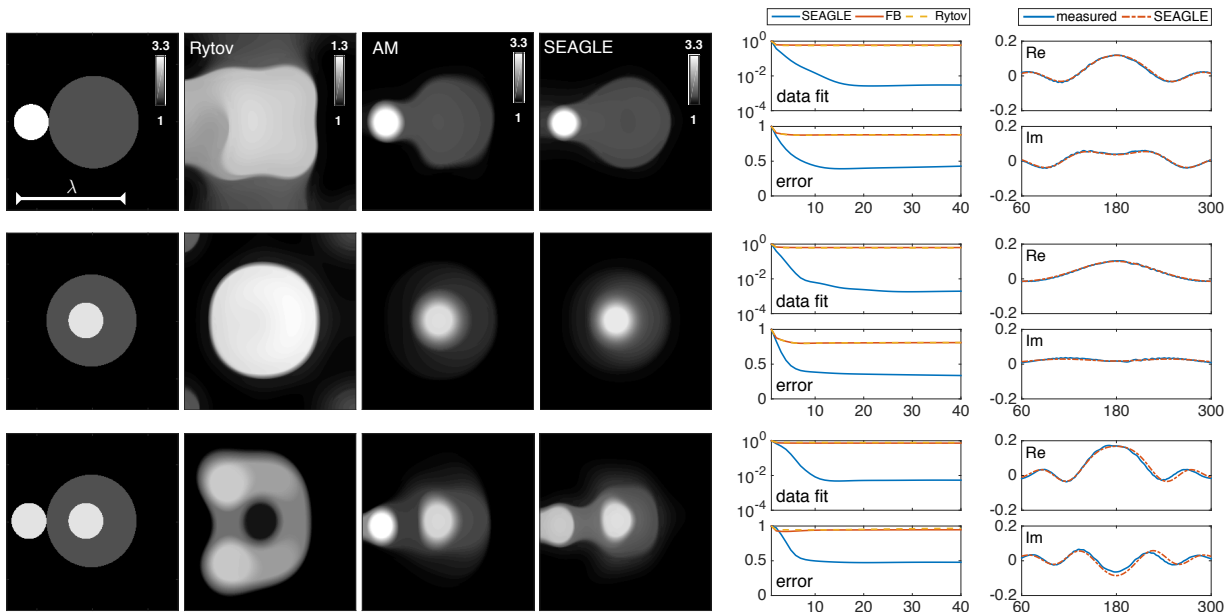
Figure 7.7: Reconstruction from an experimentally measured objects at 3 GHz, from top to bottom: *FoamDielExtTM*, *FoamDielIntTM*, and *FoamTwinDielTM*. From left to right: ground truth; reconstruction using the Rytov approximation; reconstruction based on alternating minimization; reconstruction with SEAGLE; evolution of normalized data-fit (top) and the normalized reconstruction error (bottom); and the true and predicted measurements for the transmission angle zero.

7.4.3   Validation on experimental data

We apply our method to three objects from the public dataset provided by the Fresnel institute [68]: *FoamDielExtTM*, *FoamDielIntTM*, and *FoamTwinDielTM*. These objects are placed in a region of size 15 cm × 15 cm at the center of a circular rim of radius 1.67 m and measured using 360 detectors and 8 transmitters evenly distributed on the rim. The number of transmitters is increased to 18 for *FoamTwinDielTM* and are also uniformly spaced. In all cases, only one transmitter is turned on at a time, while 241 detectors are used for each transmitter by excluding 119 detectors that are closest to the transmitter. While the full data contains multiple frequency measurements, we only use the data corresponding to the 3 GHz. As before, we compare the result of our method with the first-Born and Rytov approximations, as well as AM, all regularized with TV. We set the highest order of SEAGLE terms to $K = 200$ and the TV regularization to $\tau = 0.25 \times 10^{-8} \|\mathbf{y}\|_{\ell_2}^2$, and run the image formation algorithm for 40 FISTA iterations. The reconstruction was initialized with the background value of the dielectric permittivity, which in this case corresponds to $\epsilon_{\mathrm{b}} = 1$.

Figure 7.7 summarizes the imaging results on the experimental data. The quantitative evaluation is performed using the same metrics as before. The results show that our method successfully captures the shape of the objects, as well as the value of the permittivity. Both
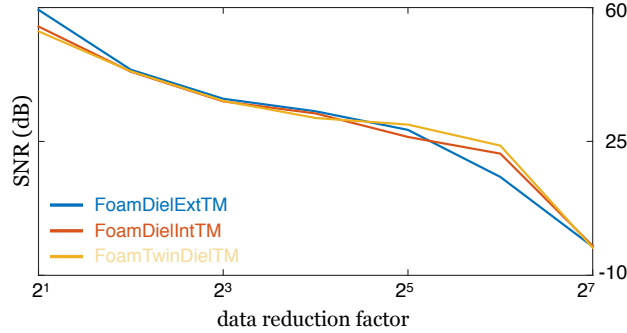
Figure 7.8: Reconstruction quality (see text) of the proposed method at various values of the data-reduction factor.

first Born and Rytov approximations underestimate the permittivity. One can also see that the data-fit error for both of the linear forward models remain high as iterations progress. On the other hand, the object reconstructed by the proposed method closely agrees with the measured data (see rightmost column in Figure 7.7). It took the proposed method about 6.59 seconds on average to process an illumination at each FISTA iteration.

Figure 7.8 illustrates the performance of our method when using a limited number of measured data. In particular, we consider the reconstruction of the same three objects, but reduce the number of measurements for each transmission using regular downsampling by factors of $2, 4, 8, 16, 32, 64$, and 128. The full dataset consists of 8 transmissions with 241 measurements each; a factor of 128 downsampling reduces to 8 transmissions with 2 measurements each. The size of the reconstructed image is set to 320×320 pixels. The reconstruction performance is quantified as

$$\text{SNR (dB)} \triangleq 10 \log_{10} \left(\frac{\|\mathbf{f}_{\text{ref}}\|_{\ell_2}^2}{\|\widehat{\mathbf{f}} - \mathbf{f}_{\text{ref}}\|_{\ell_2}^2} \right), \tag{7.13}$$

where $\mathbf{f}_{\text{ref}}$ is the reconstructed image with all the measured data (see Figure 7.7). The visual illustration is provided for *FoamDielExtTM* in Figure 7.1. This result highlights the stability of the proposed method to subsampling and experimental noise, even at highly nonlinear scattering scenarios.

Note that several other methods have been tested on this dataset [2, 10, 47, 55, 60]. Qualitative comparison of our results in Figure 7.7 with the results of those methods indicates that our approach achieves comparable performance using only a fraction of data (*i.e.*, a single frequency with possible subsampling). Additionally, we observe a reliably stable and fast convergence starting from the initialization to the background permittivity, which is desirable in strongly scattering regimes.

7.5 Conclusion

In conclusion, we have demonstrated a nonconvex optimization technique for solving nonlinear inverse scattering problems. We have applied the technique to simulated and experimentally measured data in microwave frequencies. The scattering was modeled iteratively as a series expansion with Nesterov's accelerated-gradient method. By structuring the expansion as a recursive feedforward network, we derived a backpropagation formula for evaluating the gradient that can be used for fast iterative image reconstruction. The algorithm yields images of better quality than methods using linear forward models and is competitive with state-of-the-art inverse scattering approaches, tested on the same dataset. While the optimization problem is not convex, we have observed that the algorithm converges reliably within 100 iterations from a constant initialization of the permittivity. Our approach provides a promising framework for active correction of scattering in various applications and has the potential of significantly increasing the resolution and robustness when imaging strongly scattering objects.

Appendix A

Derivation of Wigner Function Formulae

A.1 Definitions of the Wigner Function

To transform Eq. (2.2) into Eq. (2.3), we first plugging the Fourier transform of the field into Eq. (2.2), resulting in

$$W(\boldsymbol{r}, \boldsymbol{u}) = \iint \left\langle \iint E^*(\boldsymbol{r}'') e^{\mathrm{j}2\pi \boldsymbol{r}'' \cdot (\boldsymbol{u} - \frac{\Delta \boldsymbol{u}}{2})} \, \mathrm{d}^2 \boldsymbol{r}'' \iint E(\boldsymbol{r}') e^{-\mathrm{j}2\pi \boldsymbol{r}' \cdot (\boldsymbol{u} + \frac{\Delta \boldsymbol{u}}{2})} \, \mathrm{d}^2 \boldsymbol{r}' \right\rangle e^{\mathrm{j}2\pi \boldsymbol{r} \cdot \Delta \boldsymbol{u}} \, \mathrm{d}^2(\Delta \boldsymbol{u})$$

$$= \iiiint \left(\iint e^{\mathrm{j}2\pi(\boldsymbol{r} - \frac{\boldsymbol{r}' + \boldsymbol{r}''}{2}) \cdot \Delta \boldsymbol{u}} \, \mathrm{d}^2(\Delta \boldsymbol{u}) \right) \langle E^*(\boldsymbol{r}'') E(\boldsymbol{r}') \rangle \, e^{-\mathrm{j}2\pi(\boldsymbol{r}' - \boldsymbol{r}'') \cdot \boldsymbol{u}} \, \mathrm{d}^2 \boldsymbol{r}'' \mathrm{d}^2 \boldsymbol{r}'$$

$$= \iiiint \delta \left(\boldsymbol{r} - \frac{\boldsymbol{r}' + \boldsymbol{r}''}{2} \right) \langle E^*(\boldsymbol{r}'') E(\boldsymbol{r}') \rangle \, e^{-\mathrm{j}2\pi(\boldsymbol{r}' - \boldsymbol{r}'') \cdot \boldsymbol{u}} \, \mathrm{d}^2 \boldsymbol{r}'' \mathrm{d}^2 \boldsymbol{r}' \tag{A.1}$$

where

$$\delta(\boldsymbol{r}) = \iint \exp(\mathrm{j}2\pi \boldsymbol{u} \cdot \boldsymbol{r}) \, \mathrm{d}^2 \boldsymbol{u}$$

is the 2D Dirac delta function. After a change of variables $\boldsymbol{r}''' = (\boldsymbol{r}' + \boldsymbol{r}'')/2$ and $\Delta \boldsymbol{r} = \boldsymbol{r}' - \boldsymbol{r}''$ (the Jacobian of this change is 1), we have

$$W(\boldsymbol{r}, \boldsymbol{u}) = \iiiint \delta(\boldsymbol{r} - \boldsymbol{r}''') \left\langle E^* \left(\boldsymbol{r}''' - \frac{\Delta \boldsymbol{r}}{2} \right) E \left(\boldsymbol{r}''' + \frac{\Delta \boldsymbol{r}}{2} \right) \right\rangle e^{-\mathrm{j}2\pi \boldsymbol{u} \cdot \Delta \boldsymbol{r}} \, \mathrm{d}^2 \boldsymbol{r}''' \mathrm{d}^2(\Delta \boldsymbol{r})$$

$$= \iint \left\langle E^* \left(\boldsymbol{r} - \frac{\Delta \boldsymbol{r}}{2} \right) E \left(\boldsymbol{r} + \frac{\Delta \boldsymbol{r}}{2} \right) \right\rangle e^{-\mathrm{j}2\pi \boldsymbol{u} \cdot \Delta \boldsymbol{r}} \, \mathrm{d}^2(\Delta \boldsymbol{r}), \tag{A.2}$$

which is the desired Eq. (2.3).

The connection to Intensity($\boldsymbol{r}$) is easily seen from Eq. (2.3) whose exponential term becomes $\delta(\Delta \boldsymbol{r})$ after taking the integral over $\boldsymbol{u}$, while the connection to Spectrum($\boldsymbol{u}$) is easily seen from Eq. (2.2) whose exponential term becomes $\delta(\Delta \boldsymbol{u})$ after taking the integral over $\boldsymbol{r}$.

A.2   Propagation of the Wigner function

In Sec. 2.4 we introduce the propagation operator of the Wigner distribution $W(\boldsymbol{r}, \boldsymbol{u})$. Let $W_{\Delta z}(\boldsymbol{r}, \boldsymbol{u})$ denote the Wigner distribution of a partially coherent field propagated by Δz, and let $W_0(\boldsymbol{r}, \boldsymbol{u})$ denote the original one. The propagation operator states

$$W_{\Delta z}(\boldsymbol{r}, \boldsymbol{u}) = \mathcal{P}_{\Delta z} W_0(\boldsymbol{r}, \boldsymbol{u}) = \iint W_0(\boldsymbol{r}', \boldsymbol{u}) \delta(\boldsymbol{r} - \boldsymbol{r}' - \lambda \Delta z \boldsymbol{u}/n_r) \, \mathrm{d}^2 \boldsymbol{r}'. \qquad \text{(A.3)}$$

Proof:

From the definition of the Wigner function given by Eq. (2.2)

$$W(\boldsymbol{r}, \boldsymbol{u}) = \iint \left\langle \tilde{E}^* \left(\boldsymbol{u} - \frac{\Delta \boldsymbol{u}}{2} \right) \tilde{E} \left(\boldsymbol{u} + \frac{\Delta \boldsymbol{u}}{2} \right) \right\rangle e^{\mathrm{j} 2\pi \boldsymbol{r} \cdot \Delta \boldsymbol{u}} \, \mathrm{d}^2(\Delta \boldsymbol{u}),$$

we propagate each of the coherent mode of the electric field by a distance Δz. The Fresnel kernel [72] is used to propagate each of the coherence modes

$$\tilde{E}(\boldsymbol{u}; z = \Delta z) = \tilde{E}(\boldsymbol{u}; z = 0) \exp \left(\mathrm{j} 2\pi \frac{\Delta z}{\lambda/n_r} - \mathrm{j}\pi (\lambda/n_r) \Delta z \|\boldsymbol{u}\|^2 \right). \qquad \text{(A.4)}$$

Here n_r is the refractive index of background medium and λ is the wavelength in vacuum. The propagated field is connected to the original field by

$$
\begin{aligned}
&\tilde{E}^* \left(\boldsymbol{u} - \frac{\Delta \boldsymbol{u}}{2}; z = \Delta z \right) \tilde{E} \left(\boldsymbol{u} + \frac{\Delta \boldsymbol{u}}{2}; z = \Delta z \right) \\
=& \tilde{E}^* \left(\boldsymbol{u} - \frac{\Delta \boldsymbol{u}}{2}; z = 0 \right) \tilde{E} \left(\boldsymbol{u} + \frac{\Delta \boldsymbol{u}}{2}; z = 0 \right) \exp \left(-\mathrm{j}\pi(\lambda/n_r)\Delta z(-\|\boldsymbol{u} - \frac{\Delta \boldsymbol{u}}{2}\|^2 + \|\boldsymbol{u} + \frac{\Delta \boldsymbol{u}}{2}\|^2) \right) \\
=& \tilde{E}^* \left(\boldsymbol{u} - \frac{\Delta \boldsymbol{u}}{2}; z = 0 \right) \tilde{E} \left(\boldsymbol{u} + \frac{\Delta \boldsymbol{u}}{2}; z = 0 \right) \exp \left(-\mathrm{j}\pi(\lambda/n_r)\Delta z(2\boldsymbol{u} \cdot \Delta \boldsymbol{u}) \right)
\end{aligned}
\qquad \text{(A.5)}
$$

Plug Eq. (A.5) to Eq. (2.2) and notice that we can pull the transfer function of the Fresnel kernel out of the ensemble average because it is the same for all coherence component. We have the Wigner function at $z = \Delta z$ expressed as

$$
\begin{aligned}
W_{\Delta z}(\boldsymbol{r}, \boldsymbol{u}) &= \iint \left\langle \tilde{E}^* \left(\boldsymbol{u} - \frac{\Delta \boldsymbol{u}}{2}; z = 0 \right) \tilde{E} \left(\boldsymbol{u} + \frac{\Delta \boldsymbol{u}}{2}; z = 0 \right) \right\rangle e^{\mathrm{j} 2\pi(\boldsymbol{r} - (\lambda \Delta z/n_r)\boldsymbol{u}) \cdot \Delta \boldsymbol{u}} \, \mathrm{d}^2(\Delta \boldsymbol{u}) \\
&= W_0 \left(\boldsymbol{r} - \frac{\lambda \Delta z}{n_r} \boldsymbol{u}, \boldsymbol{u} \right) \\
&= \iint W_0 \left(\boldsymbol{r}', \boldsymbol{u} \right) \delta \left(\boldsymbol{r} - \frac{\lambda \Delta z}{n_r} \boldsymbol{u} - \boldsymbol{r}' \right) \, \mathrm{d}^2 \boldsymbol{r}'.
\end{aligned}
\qquad \text{(A.6)}
$$

Q.E.D.

Appendix B

Volumetric Scattering Model

B.1 Deriving the model

In this appendix, we will derive an important formula that accounts for the volumetric scattering, Eq. (6.4),

$$W_\infty(\boldsymbol{r}, \boldsymbol{u}) = \lim_{N \to \infty} W_N(\boldsymbol{r}, \boldsymbol{u}) = \frac{n_r{}^2}{2\pi\lambda^2\ell^2\sigma^2} \exp\left(-\frac{n_r{}^2}{2\lambda^2\ell^2\sigma^2}(\boldsymbol{r} - \boldsymbol{r}_s - \lambda\ell\boldsymbol{u}/n_r)^2\right).$$

Here we use the propagation operator (Eq. (2.8))

$$\mathcal{P}_{\Delta z} W(\boldsymbol{r}, \boldsymbol{u}) = \iint W(\boldsymbol{r}', \boldsymbol{u})\delta(\boldsymbol{r} - \boldsymbol{r}' - \lambda\Delta z\boldsymbol{u}/n_r) \, \mathrm{d}^2\boldsymbol{r}',$$

and the in-place diffusing operator (Eq. (6.2))

$$\mathcal{D}_{\sigma,N} W(\boldsymbol{r}, \boldsymbol{u}) = \iint W(\boldsymbol{r}, \boldsymbol{u}') \frac{N}{6\pi\sigma^2} \exp\left(-\frac{N}{6\sigma^2}(\boldsymbol{u} - \boldsymbol{u}')^2\right) \mathrm{d}^2\boldsymbol{u}'.$$

However, in order to avoid the seemingly arbitrary number 6 in Eq. (6.2) and to show the necessity of model correction, we use a parameter α in the diffusing operator instead. It will later be identified to a proper number such that we obtain the current form of Eq. (6.4). The diffusing operator used in the derivation is then

$$\mathcal{D}_\alpha W(\boldsymbol{r}, \boldsymbol{u}) = \iint W(\boldsymbol{r}, \boldsymbol{u}') \frac{1}{2\pi\alpha^2} \exp\left(-\frac{1}{2\alpha^2}(\boldsymbol{u} - \boldsymbol{u}')^2\right) \mathrm{d}^2\boldsymbol{u}'. \tag{B.1}$$

Derivation:

Let us start with Eq. (6.3) and use λ' to denote λ/n_r.

$$W = \underbrace{\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}} \mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}} \dots \mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}}_{N \text{ pairs of } \mathcal{DP}} W_s \tag{B.2}$$

Here ℓ is the depth of diffusing volume. We evaluate this step-by-step to discover the pattern.

$$\left[\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}} W_s\right](\boldsymbol{r}'', \boldsymbol{u}'')$$

$$= \int \frac{1}{2\pi\alpha^2} \exp\left(-\frac{\|\boldsymbol{u}'' - \boldsymbol{u}'\|^2}{2\alpha^2}\right) \int \delta(\boldsymbol{r}'' - \boldsymbol{r}' - \lambda'(\ell/N)\boldsymbol{u}')W_s(\boldsymbol{r}', \boldsymbol{u}') \ \mathrm{d}^2\boldsymbol{r}' \mathrm{d}^2\boldsymbol{u}'$$

$$= \int \frac{1}{2\pi\alpha^2} \exp\left(-\frac{\|\boldsymbol{u}'' - \boldsymbol{u}'\|^2}{2\alpha^2}\right) W_s(\boldsymbol{r}'' - \lambda'(\ell/N)\boldsymbol{u}', \boldsymbol{u}') \ \mathrm{d}^2\boldsymbol{u}' \quad \text{(B.3)}$$

$$\left[\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}} \mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}} W_s\right](\boldsymbol{r}''', \boldsymbol{u}''')$$

$$= \int \frac{1}{2\pi\alpha^2} \exp\left(-\frac{\|\boldsymbol{u}''' - \boldsymbol{u}''\|^2}{2\alpha^2}\right) \int \delta(\boldsymbol{r}''' - \boldsymbol{r}'' - \lambda'(\ell/N)\boldsymbol{u}'') \int \frac{1}{2\pi\alpha^2} \exp\left(-\frac{\|\boldsymbol{u}'' - \boldsymbol{u}'\|^2}{2\alpha^2}\right) \cdot$$
$$W_s(\boldsymbol{r}'' - \lambda'(\ell/N)\boldsymbol{u}', \boldsymbol{u}') \ \mathrm{d}^2\boldsymbol{u}' \mathrm{d}^2\boldsymbol{r}'' \mathrm{d}^2\boldsymbol{u}''$$

$$= \iint \left(\frac{1}{2\pi\alpha^2}\right)^2 \exp\left(-\frac{\|\boldsymbol{u}'' - \boldsymbol{u}'\|^2 + \|\boldsymbol{u}''' - \boldsymbol{u}''\|^2}{2\alpha^2}\right) W_s\left(\boldsymbol{r}''' - \lambda'(\ell/N)(\boldsymbol{u}' + \boldsymbol{u}''), \boldsymbol{u}'\right) \ \mathrm{d}^2\boldsymbol{u}' \mathrm{d}^2\boldsymbol{u}''$$
$$\text{(B.4)}$$

After applying N pairs of operations, we arrive

$$\left(\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}\right)^N W_s = \int \cdots \int \left(\frac{1}{2\pi\alpha^2}\right)^N \exp\left(-\frac{\sum_{i=1}^N \|\boldsymbol{u}^{(i+1)} - \boldsymbol{u}^{(i)}\|^2}{2\alpha^2}\right) \cdot$$
$$W_s\left(\boldsymbol{r}^{(N+1)} - \frac{\lambda'\ell}{N} \sum_{i=1}^N \boldsymbol{u}^{(i)}, \boldsymbol{u}'\right) \ \mathrm{d}^2\boldsymbol{u}' \mathrm{d}^2\boldsymbol{u}'' \cdots \mathrm{d}^2\boldsymbol{u}^{(N)} \quad \text{(B.5)}$$

The last expression $\left(\mathcal{D}_\alpha \mathcal{P}_{\ell/N}\right)^N W_s$ is a function of $\boldsymbol{r}^{(N+1)}$ and $\boldsymbol{u}^{(N+1)}$, which we redefined as

$$\boldsymbol{r} = \boldsymbol{r}^{(N+1)}$$
$$\boldsymbol{u} = \boldsymbol{u}^{(N+1)}.$$

Next we plug in the Wigner function for a point source at $\boldsymbol{r}_s$

$$W_s(\boldsymbol{r}', \boldsymbol{u}') = \delta(\boldsymbol{r}' - \boldsymbol{r}_s) \quad \text{(B.6)}$$

We have

$$\left(\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}\right)^N W_s = \left(\frac{1}{2\pi\alpha^2}\right)^N \int \cdots \int \exp\left(-\frac{\sum_{i=1}^N \|\boldsymbol{u}^{(i+1)} - \boldsymbol{u}^{(i)}\|^2}{2\alpha^2}\right) \cdot$$
$$\delta\left(\boldsymbol{r} - \boldsymbol{r}_s - \frac{\lambda'\ell}{N} \sum_{i=1}^N \boldsymbol{u}^{(i)}\right) \mathrm{d}^2\boldsymbol{u}' \mathrm{d}^2\boldsymbol{u}'' \cdots \mathrm{d}^2\boldsymbol{u}^{(N)} \quad \text{(B.7)}$$

The expression above is separable for x and y direction. So we focus on x part first and multiply the result for y at the end. The x part of Eq. (B.7), denoted by I_x, is

$$I_x = \sqrt{\frac{1}{2\pi\alpha^2}}^N \int \cdots \int \exp\left(-\frac{\sum_{i=1}^N (u_x^{(i)} - u_x^{(i+1)})^2}{2\alpha^2}\right) \delta\left(x - x_s - \frac{\lambda'\ell}{N}\sum_{i=1}^N u_x^{(i)}\right) \mathrm{d}u_x' \mathrm{d}u_x'' \cdots \mathrm{d}u_x^{(N)}$$
(B.8)

$$= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^N \int \cdots \int \exp\left(-\frac{\sum_{i=1}^N (u_x^{(i)} - u_x^{(i+1)})^2}{2\alpha^2}\right) \delta\left(\frac{N}{\lambda'\ell}(x - x_s) - \sum_{i=1}^N u_x^{(i)}\right) \cdot$$
$$\mathrm{d}u_x' \mathrm{d}u_x'' \cdots \mathrm{d}u_x^{(N)}$$
(B.9)

For simplicity, we let ξ denote

$$\xi = \frac{N}{\lambda'\ell}(x - x_s).$$
(B.10)

In order to solve this integral, we also make a change of variables

$$w^{(i)} = u_x^{(i)} - u_x^{(i+1)}$$

for $i = 2 \ldots N$. Note that $w^{(N)} = u_x^{(N)} - u_x$ contains only an independent variable $u_x^{(N)}$ before this transformation and that the Jacobian of this transformation is 1.

$$I_x = \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^N \int \cdots \int \exp\left(-\frac{\sum_{i=1}^N (w^{(i)})^2}{2\alpha^2}\right) \delta\left(\xi - \sum_{i=1}^N u_x^{(i)}\right) \mathrm{d}u_x' \mathrm{d}u_x'' \cdots \mathrm{d}u_x^{(N)}$$

$$= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^N \int \cdots \int \exp\left(-\frac{\sum_{i=2}^N (w^{(i)})^2}{2\alpha^2}\right) \exp\left(-\frac{(u_x' - u_x'')^2}{2\alpha^2}\right) \delta\left(\xi - \sum_{i=1}^N u_x^{(i)}\right) \cdot$$
$$\mathrm{d}u_x' \mathrm{d}u_x'' \cdots \mathrm{d}u_x^{(N)}$$

$$= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^N \int \cdots \int \exp\left(-\frac{\sum_{i=2}^N (w^{(i)})^2}{2\alpha^2}\right) \exp\left(-\frac{(\xi - (\sum_{i=2}^N u_x^{(i)}) - u_x'')^2}{2\alpha^2}\right) \cdot$$
$$\mathrm{d}u_x'' \cdots \mathrm{d}u_x^{(N)}$$

$$= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^N \int \cdots \int \exp\left(-\frac{\sum_{i=2}^N (w^{(i)})^2}{2\alpha^2}\right) \exp\left(-\frac{(\xi - 2u_x'' - (\sum_{i=3}^N u_x^{(i)}))^2}{2\alpha^2}\right) \cdot$$
$$\mathrm{d}w'' \cdots \mathrm{d}w^{(N)}$$

$$= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^N \int \cdots \int \exp\left(-\frac{\sum_{i=2}^N (w^{(i)})^2}{2\alpha^2}\right) \exp\left(-\frac{(\xi - 2w'' - 3u_x''' - (\sum_{i=4}^N u_x^{(i)}))^2}{2\alpha^2}\right) \cdot$$
$$\mathrm{d}w'' \cdots \mathrm{d}w^{(N)}$$

We then replace all the $u_x^{(i)}$ by the change of variable and arrive at the following formula

$$
\begin{aligned}
I_x &= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{N}\int\cdots\int\exp\left(-\frac{\sum_{i=2}^{N}(w^{(i)})^2}{2\alpha^2}\right)\cdot\\
&\quad \exp\left(-\frac{(\xi-2w''-3w'''-\ldots-Nw^{(N)}-Nu_x^{(N+1)})^2}{2\alpha^2}\right)\mathrm{d}w''\cdots\mathrm{d}w^{(N)}\\
&= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{N}\int\cdots\int\exp\left(-\frac{\sum_{i=2}^{N}(w^{(i)})^2}{2\alpha^2}\right)\exp\left(-\frac{(\xi-Nu_x-\sum_{\ell=2}^{N}\ell w^{(\ell)})^2}{2\alpha^2}\right)\cdot\\
&\quad \mathrm{d}w''\cdots\mathrm{d}w^{(N)} \tag{B.11}
\end{aligned}
$$

Integral over w''

Now let us consider the integration over w'' to discover the pattern that show up in the integration. Here we use color to keep track the factor that is due to the coefficient of w'' inside the summation.

$$
\begin{aligned}
&= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{N}\int\cdots\int\ \mathrm{d}w''\cdots\mathrm{d}w^{(N)}\\
&\quad \exp\left(-\frac{\sum_{i=3}^{N}(w^{(i)})^2}{2\alpha^2}\right)\exp\left(-\frac{(w'')^2+(\xi-Nu_x-\sum_{\ell=3}^{N}\ell w^{(\ell)}-2w'')^2}{2\alpha^2}\right)\\
&= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{N}\int\cdots\int\ \mathrm{d}w''\cdots\mathrm{d}w^{(N)}\exp\left(-\frac{\sum_{i=3}^{N}(w^{(i)})^2}{2\alpha^2}\right)\\
&\quad \exp\left(-\frac{(2^2+1)(w'')^2-2\cdot 2w''(\xi-Nu_x-\sum_{\ell=3}^{N}\ell w^{(\ell)})+(\xi-Nu_x-\sum_{\ell=3}^{N}\ell w^{(\ell)})^2}{2\alpha^2}\right)\\
&= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{N}\int\cdots\int\ \mathrm{d}w''\cdots\mathrm{d}w^{(N)}\exp\left(-\frac{\sum_{i=3}^{N}(w^{(i)})^2}{2\alpha^2}\right)\\
&\quad \exp\left(-(2^2+1)\frac{(w'')^2-2\cdot\frac{2w''}{2^2+1}(\xi-Nu_x-\sum_{\ell=3}^{N}\ell w^{(\ell)})+\frac{1}{2^2+1}(\xi-Nu_x-\sum_{\ell=3}^{N}\ell w^{(\ell)})^2}{2\alpha^2}\right)
\end{aligned}
\tag{B.12}
$$

Complete the square for w'' in the numerator of the exponent, the numerator becomes

$$\left(w'' - \frac{2}{2^2+1}(\xi - Nu_x - \sum_{\ell=3}^{N} \ell w^{(\ell)})\right)^2 + \left(-(\frac{2}{2^2+1})^2 + \frac{1}{2^2+1}\right)\left(\xi - Nu_x - \sum_{\ell=3}^{N} \ell w^{(\ell)}\right)^2$$

$$= \left(w'' - \frac{2}{2^2+1}(\xi - Nu_x - \sum_{\ell=3}^{N} \ell w^{(\ell)})\right)^2 + \left(\frac{-2^2+2^2+1}{(2^2+1)^2}\right)\left(\xi - Nu_x - \sum_{\ell=3}^{N} \ell w^{(\ell)}\right)^2$$

$$= \left(w'' - \frac{2}{2^2+1}(\xi - Nu_x - \sum_{\ell=3}^{N} \ell w^{(\ell)})\right)^2 + \frac{1}{(2^2+1)^2}\left(\xi - Nu_x - \sum_{\ell=3}^{N} \ell w^{(\ell)}\right)^2 \tag{B.13}$$

Hence the integration over w'' is separated and can be carried out as

$$\int \exp\left[-\frac{(2^2+1)}{2\alpha^2}\left(w'' - \frac{2}{2^2+1}(\xi - Nu_x - \sum_{\ell=3}^{N} \ell w^{(\ell)})\right)^2\right] \, dw'' = \sqrt{\frac{2\pi\alpha^2}{2^2+1}} \tag{B.14}$$

Plugging Eq. (B.14) into I_x (Eq. (B.12)) we have,

$$I_x = \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{-N-1}\sqrt{\frac{1}{2^2+1}}\int\cdots\int \, dw''' \cdots dw^{(N)}$$
$$\exp\left(-\frac{\sum_{i=3}^{N}(w^{(i)})^2}{2\alpha^2}\right)\exp\left(-\frac{(\xi - Nu_x - \sum_{\ell=3}^{N} \ell w^{(\ell)})^2}{2\alpha^2(2^2+1)}\right) \tag{B.15}$$

Mathematical induction

Equation Eq. (B.15) sheds light on the form of completing the square of one of $w^{(\ell)}$'s. Let us rewrite Eq. (B.11) for a comparison with Eq. (B.15)

$$I_x = \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{-N}\int\cdots\int \, dw'' \cdots dw^{(N)}$$
$$\exp\left(-\frac{\sum_{i=2}^{N}(w^{(i)})^2}{2\alpha^2}\right)\exp\left(-\frac{(\xi - Nu_x - \sum_{\ell=2}^{N} \ell w^{(\ell)})^2}{2\alpha^2}\right)$$

From the comparison, we hypothesize that after $dw^{(k)}$ is integrated, the formula of I_x becomes

$$I_x = \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{-N+1-k}\sqrt{\frac{1}{\sum_{i=1}^{k} i^2}}\int\cdots\int \, dw^{(k+1)} \cdots dw^{(N)}$$
$$\exp\left(-\frac{\sum_{i=k+1}^{N}(w^{(i)})^2}{2\alpha^2}\right)\exp\left(-\frac{(\xi - Nu_x - \sum_{\ell=k+1}^{N} \ell w^{(\ell)})^2}{2\alpha^2(\sum_{i=1}^{k} i^2)}\right) \tag{B.16}$$

Eq. (B.15) is an instance $k = 2$ of Eq. (B.16).

Let us derive the case for $k + 1$, that is, when $dw^{(k+1)}$ is integrated. We consider the following useful square completing

$$\left(\sum_{i=1}^{K-1} i^2\right) w^2 + (A - Kw)^2$$

$$= \left(\sum_{i=1}^{K} i^2\right) w^2 - 2KwA + A^2$$

$$= \left(\sum_{i=1}^{K} i^2\right) \left(w^2 - \frac{2KA}{\left(\sum_{i=1}^{K} i^2\right)} w + \frac{K^2 A^2}{\left(\sum_{i=1}^{K} i^2\right)^2} - \frac{K^2 A^2}{\left(\sum_{i=1}^{K} i^2\right)^2} + \frac{1}{\left(\sum_{i=1}^{K} i^2\right)} A^2\right)$$

$$= \left(\sum_{i=1}^{K} i^2\right) \left(w - \frac{KA}{\left(\sum_{i=1}^{K} i^2\right)}\right)^2 + \frac{-K^2 + \sum_{i=1}^{K} i^2}{\sum_{i=1}^{K} i^2} A^2$$

$$= \left(\sum_{i=1}^{K} i^2\right) \left(w - \frac{KA}{\left(\sum_{i=1}^{K} i^2\right)}\right)^2 + \frac{\sum_{i=1}^{K-1} i^2}{\sum_{i=1}^{K} i^2} A^2 \tag{B.17}$$

Taking $w^{(k+1)}$ from the first exponent to the second one in Eq. (B.16), we have

$$I_x = \left(\frac{N}{\lambda'\ell}\right) \sqrt{\frac{1}{2\pi\alpha^2}}^{N+1-k} \sqrt{\frac{1}{\sum_{i=1}^{k} i^2}} \int \cdots \int dw^{(k+1)} \cdots dw^{(N)} \exp\left(-\frac{\sum_{i=k+2}^{N} (w^{(i)})^2}{2\alpha^2}\right)$$

$$\exp\left(-\frac{\left(\sum_{i=1}^{k} i^2\right)(w^{(k+1)})^2 + (\xi - Nu_x - (\sum_{\ell=k+2}^{N} \ell w^{(\ell)}) - (k+1)w^{(k+1)})^2}{2\alpha^2(\sum_{i=1}^{k} i^2)}\right) \tag{B.18}$$

Applying Eq. (B.17) with

$$w = w^{(k+1)}, \quad K = k+1, \quad A = \xi - Nu_x - \sum_{\ell=k+2}^{N} \ell w^{(\ell)},$$

to Eq. (B.18), we have

$$I_x = \left(\frac{N}{\lambda'\ell}\right) \sqrt{\frac{1}{2\pi\alpha^2}}^{N+1-k} \sqrt{\frac{1}{\sum_{i=1}^{k} i^2}} \int \cdots \int dw^{(k+1)} \cdots dw^{(N)} \exp\left(-\frac{\sum_{i=k+2}^{N} (w^{(i)})^2}{2\alpha^2}\right) \cdot$$

$$\exp\left(-\frac{1}{2\alpha^2(\sum_{i=1}^{k} i^2)} \cdot \left(\sum_{i=1}^{k+1} i^2\right) \left(w^{(k+1)} - \frac{(k+1)(\xi - Nu_x - \sum_{\ell=k+2}^{N} \ell w^{(\ell)})}{\left(\sum_{i=1}^{k+1} i^2\right)}\right)^2\right) \cdot$$

$$\exp\left(-\frac{1}{2\alpha^2(\sum_{i=1}^{k} i^2)} \cdot \frac{\sum_{i=1}^{k} i^2}{\sum_{i=1}^{k+1} i^2} (\xi - Nu_x - \sum_{\ell=k+2}^{N} \ell w^{(\ell)})^2\right). \tag{B.19}$$

Here $w^{(k+1)}$ is singled out and only appears in the middle row. Hence the integral over $w^{(k+1)}$ is to evaluate the integral of the middle row.

$$\int \exp\left(-\frac{(\sum_{i=1}^{k+1} i^2)}{2\alpha^2(\sum_{i=1}^{k} i^2)}\left(w^{(k+1)}-\ldots\right)^2\right) \mathrm{d}w^{(k+1)} = \sqrt{\frac{2\pi\alpha^2(\sum_{i=1}^{k} i^2)}{(\sum_{i=1}^{k+1} i^2)}}.$$

We have I_x

$$I_x = \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{-N+1-k}\sqrt{\frac{1}{\sum_{i=1}^{k} i^2}}\int\cdots\int \mathrm{d}w^{(k+2)}\cdots\mathrm{d}w^{(N)}\exp\left(-\frac{\sum_{i=k+2}^{N}(w^{(i)})^2}{2\alpha^2}\right).$$

$$\sqrt{\frac{2\pi\alpha^2(\sum_{i=1}^{k} i^2)}{(\sum_{i=1}^{k+1} i^2)}}\exp\left(-\frac{1}{2\alpha^2(\sum_{i=1}^{k} i^2)}\cdot\frac{\sum_{i=1}^{k} i^2}{\sum_{i=1}^{k+1} i^2}(\xi - Nu_x - \sum_{\ell=k+2}^{N}\ell w^{(\ell)})^2\right)$$

$$= \left(\frac{N}{\lambda'\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}^{-N+1-(k+1)}\sqrt{\frac{1}{\sum_{i=1}^{k+1} i^2}}\int\cdots\int \mathrm{d}w^{(k+2)}\cdots\mathrm{d}w^{(N)}$$

$$\exp\left(-\frac{\sum_{i=k+2}^{N}(w^{(i)})^2}{2\alpha^2}\right)\exp\left(-\frac{(\xi - Nu_x - \sum_{\ell=k+2}^{N}\ell w^{(\ell)})^2}{2\alpha^2(\sum_{i=1}^{k+1} i^2)}\right). \tag{B.20}$$

Eq. (B.20) is the hypothesis Eq. (B.16) with k replaced by $k+1$. Therefore by mathematical induction, the hypothesis holds for $k = 2 \ldots N$. We have the final expression for I_x (ξ defined in Eq. (B.10) and $\lambda' = \lambda/n_r$) to be

$$I_x = \left(\frac{Nn_r}{\lambda\ell}\right)\sqrt{\frac{1}{2\pi\alpha^2}}\sqrt{\frac{1}{\sum_{i=1}^{N} i^2}}\exp\left(-\frac{(\frac{Nn_r}{\lambda\ell}(x-x_s) - Nu_x)^2}{2\alpha^2(\sum_{i=1}^{N} i^2)}\right). \tag{B.21}$$

Letting $N \to \infty$

The identity of $\sum_{i=1}^{N} i^2$ is simplified

$$\sum_{i=1}^{N} i^2 = \frac{N}{6}(N+1)(2N+1) \to \frac{N^3}{3}. \tag{B.22}$$

Thus,

$$I_x = \frac{n_r}{\lambda\ell}\sqrt{\frac{N^2}{2\pi\alpha^2}\cdot\frac{3}{N^3}}\exp\left(-\frac{N^2(\frac{n_r}{\lambda\ell}(x-x_s) - u_x)^2\cdot 3}{2\alpha^2 N^3}\right) \tag{B.23}$$

$$= \sqrt{\frac{3n_r^2}{2\pi\lambda^2\ell^2 N\alpha^2}}\exp\left(-\frac{3n_r^2(x-x_s - (\lambda/n_r)\ell u_x)^2}{2\lambda^2\ell^2 N\alpha^2}\right). \tag{B.24}$$

Combine the y part, we have the resulting Wigner function for a point source diffused by a volume

$$W(\boldsymbol{r}, \boldsymbol{u}) = \frac{3n_r^2}{2\pi\lambda^2\ell^2 N\alpha^2} \exp\left(-\frac{3n_r^2(\boldsymbol{r} - \boldsymbol{r}_s - (\lambda/n_r)\ell\boldsymbol{u})^2}{2\lambda^2\ell^2 N\alpha^2}\right). \tag{B.25}$$

Note that we have not determine α yet. If the value of α is set to

$$\alpha = \sqrt{\frac{3}{N}}\sigma, \tag{B.26}$$

Eq. (B.25) becomes

$$W(\boldsymbol{r}, \boldsymbol{u}) = \frac{n_r^2}{2\pi\lambda^2\ell^2\sigma^2} \exp\left(-\frac{n_r^2(\boldsymbol{r} - \boldsymbol{r}_s - (\lambda/n_r)\ell\boldsymbol{u})^2}{2\lambda^2\ell^2\sigma^2}\right), \tag{B.27}$$

which is the goal formula Eq. (6.4) and the diffusing operator Eq. (B.1) becomes Eq. (6.2). *Q.E.D.*

B.2 Correcting the model

The model we derive here has a desirable property that for a longer scattering depth, ℓ, the Gaussian appears wider. However, setting $\alpha = \sqrt{3/N}\sigma$ is totally arbitrary. This way can achieve convergence of the formula but does not determine the dependence of σ on some variables such as ℓ. If we consider another way to take the number of scattering layers to infinity by controlling the propagation distance $\Delta z \to 0$, the number of pairs in Eq. (B.2) becomes

$$N = \left\lfloor \frac{\ell}{\Delta z} \right\rfloor.$$

We will have Eq. (B.25) becoming

$$W(\boldsymbol{r}, \boldsymbol{u}) = \frac{3n_r^2\Delta z}{2\pi\lambda^2\ell^3\alpha^2} \exp\left(-\frac{3n_r^2(\boldsymbol{r} - \boldsymbol{r}_s - (\lambda/n_r)\ell\boldsymbol{u})^2\Delta z}{2\lambda^2\ell^3\alpha^2}\right). \tag{B.28}$$

Here the floor operation is dropped as it has negligible effect after taking the limit $\Delta z \to 0$. In Eq. (B.28), we can set α to

$$\alpha = \sqrt{3\Delta z}\bar{\sigma} = \sqrt{\frac{3\ell}{N}}\bar{\sigma} \tag{B.29}$$

(using $\bar{\sigma}$ to avoid confusion with σ), which results in an ℓ-dependent α and leads to

$$W(\boldsymbol{r}, \boldsymbol{u}) = \frac{n_r^2}{2\pi\lambda^2\ell^3\bar{\sigma}^2} \exp\left(-\frac{n_r^2(\boldsymbol{r} - \boldsymbol{r}_s - (\lambda/n_r)\ell\boldsymbol{u})^2}{2\lambda^2\ell^3\bar{\sigma}^2}\right). \tag{B.30}$$

This model also has the desirable property of more scattering if source is deeper. It only differs from the previous one by an extra ℓ multiplied. Hence the derivation itself is not enough to finalize the formula. We turn to another physical case in order to find out which case, Eq. (B.26) or Eq. (B.29), is correct.

B.2.1 Plane wave case

A plane wave is represented by a delta function in the Wigner function representation [173]. We substitute the delta function with W_s in Eq. (B.2), resulting in

$$W(\boldsymbol{r}, \boldsymbol{u}) = \underbrace{\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}} \mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}} \ldots \mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}}_{N \text{ pairs of } \mathcal{DP}} \delta(\boldsymbol{u} - \boldsymbol{u}_s) \tag{B.31}$$

where $\boldsymbol{u}_s$ is the transverse spatial frequency of the incident wave. We substitute $W_s(\boldsymbol{r}, \boldsymbol{u}) = \delta(\boldsymbol{u} - \boldsymbol{u}_s)$ with Eq. (B.5) and obtain

$$\left(\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}\right)^N W_s = \int \cdots \int \left(\frac{1}{2\pi\alpha^2}\right)^N \exp\left(-\frac{\sum_{i=1}^N \|\boldsymbol{u}^{(i+1)} - \boldsymbol{u}^{(i)}\|^2}{2\alpha^2}\right) \delta(\boldsymbol{u}' - \boldsymbol{u}_s) \, \mathrm{d}^2\boldsymbol{u}' \cdots \mathrm{d}^2\boldsymbol{u}^{(N)} \tag{B.32}$$

Here $\boldsymbol{u}^{(N+1)}$ is redefined as $\boldsymbol{u}$ as in the previous section. Instead of carrying out the computation ourselves, we find that two unnumbered equations following Eq. (14) of [126]

$$K = \lim_{N\to\infty} \left(\frac{m}{2\pi i\delta}\right)^{N/2} \int \cdots \int \mathrm{d}q_1 \cdots \mathrm{d}q_{N-1} \exp\left(i\frac{m}{2\delta}\sum_{j=1}^N (q_j - q_{j-1})^2\right)$$

and

$$K = \lim_{N\to\infty} \left(\frac{m}{2\pi i N\delta}\right)^{1/2} e^{im(q'-q)/2N\delta}$$

provide the result of integral. Here $q' = q_N$ and $q = q_0$. Note that the terms inside $\lim_{N\to\infty}$ equal to each other exactly and no simplification due to the limit is applied. After making the following identification,

$$\frac{m}{i\delta} = \frac{1}{\alpha^2}, \quad u_x^{(j)} = q_{j-1}, \quad u_y^{(j)} = q_{j-1},$$

and applying the result from [126] twice (for u_x and u_y), Eq. (B.32) becomes

$$\left(\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}\right)^N W_s = \int \left(\frac{1}{2\pi N\alpha^2}\right) \exp\left(-\frac{\|\boldsymbol{u} - \boldsymbol{u}'\|^2}{2N\alpha^2}\right) \delta(\boldsymbol{u}' - \boldsymbol{u}_s) \, \mathrm{d}^2\boldsymbol{u}' \tag{B.33}$$

Hence the volumetric diffusing operation on a space-independent input Wigner function is **convolving a Gaussian in spatial frequency domain**. However, this equation does not have explicit dependence on the scattering depth! This suggests that

$$\alpha = \alpha(\ell)$$

is indeed a function of scattering depth ℓ. Fortunately, the resulting formula is still space-independent so we can apply the operation again as if the input wave travels twice as long

distance.

$$\left(\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}\right)^N \left(\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}\right)^N W_s = \iint \left(\frac{1}{2\pi N\alpha^2}\right)^2 \exp\left(-\frac{\|\boldsymbol{u}-\boldsymbol{u}''\|^2 + \|\boldsymbol{u}''-\boldsymbol{u}'\|^2}{2N\alpha^2}\right) \delta(\boldsymbol{u}'-\boldsymbol{u}_s)$$
$$\mathrm{d}^2\boldsymbol{u}'\mathrm{d}^2\boldsymbol{u}''. \tag{B.34}$$

After completing the square for $\boldsymbol{u}''$, we have

$$\left(\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}\right)^N \left(\mathcal{D}_\alpha \mathcal{P}_{\frac{\ell}{N}}\right)^N W_s$$
$$= \iint \left(\frac{1}{2\pi N\alpha^2}\right)^2 \exp\left(-\frac{2\|\boldsymbol{u}''-(\boldsymbol{u}+\boldsymbol{u}')/2\|^2}{2N\alpha^2}\right) \exp\left(-\frac{\|\boldsymbol{u}-\boldsymbol{u}'\|^2}{2N\alpha^2 \cdot 2}\right) \delta(\boldsymbol{u}'-\boldsymbol{u}_s) \, \mathrm{d}^2\boldsymbol{u}'\mathrm{d}^2\boldsymbol{u}''$$
$$= \int \left(\frac{1}{2\pi N\alpha^2}\right)^2 \left(\pi N\alpha^2\right) \exp\left(-\frac{\|\boldsymbol{u}-\boldsymbol{u}'\|^2}{2N\alpha^2 \cdot 2}\right) \delta(\boldsymbol{u}'-\boldsymbol{u}_s) \, \mathrm{d}^2\boldsymbol{u}'$$
$$= \int \left(\frac{1}{2\pi N\alpha^2 \cdot 2}\right) \exp\left(-\frac{\|\boldsymbol{u}-\boldsymbol{u}'\|^2}{2N\textcolor{red}{\alpha^2 \cdot 2}}\right) \delta(\boldsymbol{u}'-\boldsymbol{u}_s) \, \mathrm{d}^2\boldsymbol{u}'. \tag{B.35}$$

By comparing Eq. (B.33) and Eq. (B.35), we have $\alpha(2\ell)^2$ be the part colored red in Eq. (B.35)

$$\alpha(2\ell)^2 = 2\alpha^2 = 2\alpha(\ell)^2.$$

Hence we see that Eq. (B.29) satisfies this relation and is the proper formula to use. This suggests that the model associated with the α in Eq. (B.29), *i.e.* Eq. (B.30), should be used in the place of Eq. (6.4).

Appendix C

Derivation of Error Backpropagation

In this appendix, we provide the derivation of error backpropagation applied to our method. The method essentially computes the gradient of the data-fidelity term. This gradient is a key step of updating the scattering potential in solving the inverse problem. We now present the mathematical derivation of the gradient computation and relate it to Algorithm 3.

The inputs of the error-back propagation are the data mismatch and the intermediate variables ($\{\mathbf{s}^k\}$, $\{\gamma_k\}$, $\{\mu_k\}$, $\widehat{\mathbf{u}} = \mathbf{u}^K$) of the forward model computation, and the output is the gradient. Here we follow the differentiation conventions for vectors: $(\frac{\partial \mathbf{u}}{\partial \mathbf{f}})_{ij} = \frac{\partial \mathbf{u}_i}{\partial \mathbf{f}_j}$ and $(\nabla_{\mathbf{f}}\mathbf{u})_{ij} = [(\frac{\partial \mathbf{u}}{\partial \mathbf{f}})^{\mathsf{H}}]_{ij} = \frac{\partial \mathbf{u}_j^*}{\partial \mathbf{f}_i}$. All boldface lower-case variables are column vectors.

Let us begin with the gradient of $\mathcal{D} = \frac{1}{2}\|\mathbf{z} - \mathbf{y}\|_2^2$.

$$
\begin{aligned}
\nabla_{\mathbf{f}}\mathcal{D} =& \frac{1}{2}\nabla_{\mathbf{f}}[(\mathbf{z} - \mathbf{y})^{\mathsf{H}}(\mathbf{z} - \mathbf{y})] \\
=& \frac{1}{2}[(\nabla_{\mathbf{f}}\mathbf{z})(\mathbf{z} - \mathbf{y}) + ((\mathbf{z} - \mathbf{y})^{\mathsf{H}}(\nabla_{\mathbf{f}}\mathbf{z})^{\mathsf{H}})^{\mathsf{T}}] \\
=& \mathrm{Re}\left\{(\nabla_{\mathbf{f}}\mathbf{z})(\mathbf{z} - \mathbf{y})\right\}.
\end{aligned}
\tag{C.1}
$$

This can be evaluated by applying the chain rule to $\nabla_{\mathbf{f}}\mathbf{z}$ and all the variables composing $\mathbf{z}$. The equations leading from the initialization all the way to $\mathbf{z}$ are listed below:

$$
\begin{aligned}
\mathbf{z} &= \mathbf{u}_{\mathrm{in}} + \mathbf{H}\mathrm{diag}(\mathbf{f})\mathbf{u}^K \\
\mathbf{s}^k &= (1 - \mu_k)\mathbf{u}^{k-1} + \mu_k\mathbf{u}^{k-2} \\
\mathbf{u}^k &= \mathbf{s}^k - \gamma_k\mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\mathrm{in}}),
\end{aligned}
$$

for $k = 1, \ldots, K$, where $\mathbf{A} \triangleq \mathbf{I} - \mathbf{G}\mathrm{diag}(\mathbf{f})$, and $\mathbf{u}^{-1} = \mathbf{u}^0$. It is worth noting that, while the step-size γ_k also depends on $\mathbf{f}$, we ignore this dependency to simplify the computation. The rationale for this simplification is that the step-size can be replaced by a fixed one. Furthermore, in practice, γ_k attains a stationary value for large enough k, which indicates that this simplification has a negligible effect on backpropagation.

C.1 Initialization of backpropagation

The initialization in Algorithm 3 is obtained by differentiating the first of the above equations with respect to $\mathbf{f}$. With $\text{diag}(\mathbf{f})\mathbf{u} = \text{diag}(\mathbf{u})\mathbf{f}$, we have

$$\nabla_{\mathbf{f}}\mathbf{z} = \left[\mathbf{H}\frac{\partial\mathbf{f}}{\partial\mathbf{f}}\text{diag}(\mathbf{u}^K) + \mathbf{H}\text{diag}(\mathbf{f})\frac{\partial\mathbf{u}^K}{\partial\mathbf{f}}\right]^{\mathsf{H}}$$
$$= \text{diag}(\mathbf{u}^K)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} + (\nabla_{\mathbf{f}}\mathbf{u}^K)\text{diag}(\mathbf{f})^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} \tag{C.2}$$

The first term gives the remainder that contributes to the final result while the second term gives the vector that multiplies with $\nabla_{\mathbf{f}}\mathbf{u}^K$. For convenience, we define two sets of vectors:

- $\mathbf{q}^k$: the vector that multiplies with $\nabla_{\mathbf{f}}\mathbf{u}^k$

- $\mathbf{r}^k$: the remainder before computing $(\nabla_{\mathbf{f}}\mathbf{u}^k)\mathbf{q}^k$.

In addition, due to the acceleration step in the forward computation, we expect subsequent $\mathbf{q}^{k-1}$ to have a contribution from $(\nabla_{\mathbf{f}}\mathbf{u}^{k+1})\mathbf{q}^{k+1}$ in addition to the contribution from its direct neighbor $(\nabla_{\mathbf{f}}\mathbf{u}^k)\mathbf{q}^k$. This leads to the third set of vectors:

- $\mathbf{p}^k$: the explicit contribution of $(\nabla_{\mathbf{f}}\mathbf{u}^{k+1})\mathbf{q}^{k+1}$ to $\mathbf{q}^{k-1}$.

Finally, multiplying (C.2) with $(\mathbf{z} - \mathbf{y})$ we identify

$$\mathbf{r}^K = \text{diag}(\mathbf{u}^K)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}(\mathbf{z} - \mathbf{y}) \tag{C.3}$$
$$\mathbf{q}^K = \text{diag}(\mathbf{f})^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}(\mathbf{z} - \mathbf{y}). \tag{C.4}$$

Since there is no term multiplying with $\nabla_{\mathbf{f}}\mathbf{u}^{K-1}$ explicitly (hence nothing to pass to $\mathbf{q}^{K-1}$), we have

$$\mathbf{p}^K = \mathbf{0}. \tag{C.5}$$

C.2 Recursive updates for $\mathbf{u}^k$

The computation of $(\nabla_{\mathbf{f}}\mathbf{u}^k)\mathbf{q}^k$ is the key step in error-back propagation. We evaluate the gradient $\nabla_{\mathbf{f}}\mathbf{u}^k$ by taking the Hermitian of the derivative, and the multiplication with $\mathbf{q}^k$ follows. The result should be passed onto another gradient with smaller k. Before we start, let us write out the gradient of $\mathbf{s}^k$ which is straightforward from its definition,

$$\nabla_{\mathbf{f}}\mathbf{s}^k = (1 - \mu_k)\nabla_{\mathbf{f}}\mathbf{u}^{k-1} + \mu_k\nabla_{\mathbf{f}}\mathbf{u}^{k-2}. \tag{C.6}$$

The derivative of $\mathbf{u}^k$ is

$$\frac{\partial\mathbf{u}^k}{\partial\mathbf{f}} = -\gamma_k\frac{\partial\mathbf{A}^{\mathsf{H}}}{\partial\mathbf{f}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}) - \gamma_k\mathbf{A}^{\mathsf{H}}\frac{\partial\mathbf{A}}{\partial\mathbf{f}}\mathbf{s}^k$$
$$+ (\mathbf{I} - \gamma_k\mathbf{A}^{\mathsf{H}}\mathbf{A})\frac{\partial\mathbf{s}^k}{\partial\mathbf{f}}. \tag{C.7}$$

The first term becomes

$$-\gamma_k \frac{\partial \mathbf{A}^{\mathsf{H}}}{\partial \mathbf{f}}(\mathbf{As}^k - \mathbf{u}_{\text{in}}) = \gamma_k \left(\frac{\partial}{\partial \mathbf{f}} \text{diag}(\mathbf{f}) \right) \mathbf{G}^{\mathsf{H}}(\mathbf{As}^k - \mathbf{u}_{\text{in}})$$
$$= \gamma_k \text{diag}(\mathbf{G}^{\mathsf{H}}(\mathbf{As}^k - \mathbf{u}_{\text{in}})),$$

and the second term becomes

$$-\gamma_k \mathbf{A}^{\mathsf{H}} \frac{\partial \mathbf{A}}{\partial \mathbf{f}} \mathbf{s}^k = \gamma_k \mathbf{A}^{\mathsf{H}} \mathbf{G} \left(\frac{\partial}{\partial \mathbf{f}} \text{diag}(\mathbf{f}) \right) \mathbf{s}^k$$
$$= \gamma_k \mathbf{A}^{\mathsf{H}} \mathbf{G} \text{diag}(\mathbf{s}^k).$$

By taking Hermitian transpose of Eq. (C.7), we have

$$\nabla_{\mathbf{f}} \mathbf{u}^k = \gamma_k \mathbf{T}^k + (\nabla_{\mathbf{f}} \mathbf{s}^k) \mathbf{S}^k \tag{C.8}$$

where

$$\mathbf{T}^k = \text{diag}(\mathbf{G}^{\mathsf{H}}(\mathbf{As}^k - \mathbf{u}_{\text{in}}))^{\mathsf{H}} + \text{diag}(\mathbf{s}^k)^{\mathsf{H}} \mathbf{G}^{\mathsf{H}} \mathbf{A} \tag{C.9}$$
$$\mathbf{S}^k = (\mathbf{I} - \gamma_k \mathbf{A}^{\mathsf{H}} \mathbf{A})^{\mathsf{H}} = \mathbf{I} - \gamma_k \mathbf{A}^{\mathsf{H}} \mathbf{A}. \tag{C.10}$$

By multiplying Eq. (C.8) with $\mathbf{q}^k$ and substituting with Eq. (C.6), we obtain the expression for $(\nabla_{\mathbf{f}} \mathbf{u}^k) \mathbf{q}^k$,

$$\begin{aligned}
(\nabla_{\mathbf{f}} \mathbf{u}^k) \mathbf{q}^k = &\gamma_k \mathbf{T}^k \mathbf{q}^k \\
&+ (\nabla_{\mathbf{f}} \mathbf{u}^{k-1}) \left[(1 - \mu_k) \mathbf{S}^k \mathbf{q}^k \right] \\
&+ (\nabla_{\mathbf{f}} \mathbf{u}^{k-2}) \left[\mu_k \mathbf{S}^k \mathbf{q}^k \right].
\end{aligned} \tag{C.11}$$

Note that because we set $\mathbf{u}^{-1} = \mathbf{u}^0$,

$$(\nabla_{\mathbf{f}} \mathbf{u}^1) \mathbf{q}^1 = \gamma_1 \mathbf{T}^1 \mathbf{q}^1 + (\nabla_{\mathbf{f}} \mathbf{u}^0) \mathbf{S}^1 \mathbf{q}^1. \tag{C.12}$$

C.3   Error backpropagation equations

From equations (C.2) to (C.5), we have

$$(\nabla_{\mathbf{f}} \mathbf{z})(\mathbf{z} - \mathbf{y}) = \mathbf{r}^K + (\nabla_{\mathbf{f}} \mathbf{u}^K) \mathbf{q}^K + (\nabla_{\mathbf{f}} \mathbf{u}^{K-1}) \mathbf{p}^K \tag{C.13}$$

Substituting Eq. (C.11) into Eq. (C.13), we have the following expressions

$$\begin{aligned}
(\nabla_{\mathbf{f}} \mathbf{z})(\mathbf{z} - \mathbf{y}) = &\mathbf{r}^K + (\nabla_{\mathbf{f}} \mathbf{u}^K) \mathbf{q}^K + (\nabla_{\mathbf{f}} \mathbf{u}^{K-1}) \mathbf{p}^K \\
= &\mathbf{r}^{K-1} + (\nabla_{\mathbf{f}} \mathbf{u}^{K-1}) \mathbf{q}^{K-1} + (\nabla_{\mathbf{f}} \mathbf{u}^{K-2}) \mathbf{p}^{K-1} \\
= &\dots \\
= &\mathbf{r}^1 + (\nabla_{\mathbf{f}} \mathbf{u}^1) \mathbf{q}^1 + (\nabla_{\mathbf{f}} \mathbf{u}^0) \mathbf{p}^1
\end{aligned} \tag{C.14}$$

and the recursion relations for $k = 2 \ldots K$

$$\mathbf{r}^{k-1} = \mathbf{r}^k + \gamma_k \mathbf{T}^k \mathbf{q}^k \tag{C.15}$$

$$\mathbf{q}^{k-1} = \mathbf{p}^k + (1 - \mu_k) \mathbf{S}^k \mathbf{q}^k \tag{C.16}$$

$$\mathbf{p}_{k-1} = \mu_k \mathbf{S}^k \mathbf{q}^k. \tag{C.17}$$

For the case $k = 1$, or namely $\mathbf{r}_0$ and $\mathbf{q}_0$ (note that $\mathbf{p}_0$ does not exist due to $\mathbf{u}^{-1} = \mathbf{u}^0$), we plug Eq. (C.12) into Eq. (C.14),

$$(\nabla_{\mathbf{f}} \mathbf{z})(\mathbf{z} - \mathbf{y}) = \mathbf{r}^1 + \gamma_1 \mathbf{T}^1 \mathbf{q}^1 + (\nabla_{\mathbf{f}} \mathbf{u}^0) \left[\mathbf{S}^1 \mathbf{q}^1 + \mathbf{p}^1 \right].$$

Hence we have

$$\mathbf{r}^0 = \mathbf{r}^1 + \gamma_1 \mathbf{T}^1 \mathbf{q}^1 \tag{C.18}$$

$$\mathbf{q}^0 = \mathbf{p}^1 + \mathbf{S}^1 \mathbf{q}^1. \tag{C.19}$$

In the initialization of our forward model, $\mathbf{u}^0$ is the incident field and does not depend on $\mathbf{f}$. Therefore $\nabla_{\mathbf{f}} \mathbf{u}^0 = \mathbf{0}$ and the gradient of data-fidelity is

$$\nabla_{\mathbf{f}} \mathcal{D} = \mathrm{Re} \left\{ (\nabla_{\mathbf{f}} \mathbf{z})(\mathbf{z} - \mathbf{y}) \right\} = \mathrm{Re} \left\{ \mathbf{r}^0 \right\}. \tag{C.20}$$

We summarize these recursion relations of error backpropagation in Algorithm 3 of the main text.

Appendix D

Analytic Solutions to Special Scattering Cases

In this section, our aim is to present the analytic expressions for scalar electric fields resulting from a point source outside a dielectric sphere in 2D and 3D (strictly speaking, the 2D case should be understood as an infinitely long line source illuminating a cylinder parallel to it and looking at the cross-section). A sketch of the derivation is provided after the actual expressions. A more complete description can be found in a number of standard textbooks such as [86].

D.1 Expressions

Consider a sphere of a radius r_{sph} and a refractive index $n = \sqrt{\epsilon}$. The source is located r_{s} distance away from the center of the sphere and the wavenumber of the source outside the sphere is k_b.

2D case:
We consider the polar coordinates

$$\boldsymbol{x} = (r\cos\theta, r\sin\theta),$$

and, without loss of generality, assume that the source is at $\theta_{\text{s}} = 0$. The field can be expressed as

$$E(\boldsymbol{x}; r_{\text{s}}) = \sum_{m=-\infty}^{\infty} R_m(r, r_{\text{s}}) \frac{e^{jm\theta}}{2\pi} \tag{D.1}$$

where $\rho = k_b r$, $\rho_{\text{sph}} = k_b r_{\text{sph}}$ and $\rho_{\text{s}} = k_b r_{\text{s}}$,

$$R_m(r, r_{\text{s}})$$
$$= \begin{cases} a_m J_m(n\rho) H_m^{(1)}(\rho_{\text{s}}), & r < r_{\text{sph}} \\ (b_m J_m(\rho) + c_m Y_m(\rho)) H_m^{(1)}(\rho_{\text{s}}), & r_{\text{sph}} \le r < r_{\text{s}} \\ (b_m J_m(\rho_{\text{s}}) + c_m Y_m(\rho_{\text{s}})) H_m^{(1)}(\rho), & r_{\text{s}} \le r \end{cases} \tag{D.2}$$

$$a_m = \frac{-1}{\rho_{\text{sph}} \Delta_m} \tag{D.3}$$

$$b_m = \frac{-\pi}{2\Delta_m} \begin{vmatrix} J_m(n\rho_{\text{sph}}) & nJ_{m-1}(n\rho_{\text{sph}}) \\ Y_m(\rho_{\text{sph}}) & Y_{m-1}(\rho_{\text{sph}}) \end{vmatrix} \tag{D.4}$$

$$c_m = \frac{\pi}{2\Delta_m} \begin{vmatrix} J_m(n\rho_{\text{sph}}) & nJ_{m-1}(n\rho_{\text{sph}}) \\ J_m(\rho_{\text{sph}}) & J_{m-1}(\rho_{\text{sph}}) \end{vmatrix} \tag{D.5}$$

$$\Delta_m = \begin{vmatrix} J_m(n\rho_{\text{sph}}) & nJ_{m-1}(n\rho_{\text{sph}}) \\ H_m^{(1)}(\rho_{\text{sph}}) & H_{m-1}^{(1)}(\rho_{\text{sph}}) \end{vmatrix} \tag{D.6}$$

and J_m and Y_m are the m'th order Bessel functions of the first kind and the second kind, and $H_m^{(1)} = J_m + jY_m$ is the m'th order Hankel's function of the first kind.

3D case:
We consider the spherical coordinates

$$\boldsymbol{x} = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta),$$

and, without loss of generality, assume that the source has zenith angle $\theta_{\text{s}} = 0$ and azimuthal angle $\phi_{\text{s}} = 0$. The field then reads

$$E(\boldsymbol{x}; r_{\text{s}}) = \sum_{l=0}^{\infty} R_l(r, r_{\text{s}}) \left(\frac{2l+1}{4\pi}\right) P_l(\cos\theta) \tag{D.7}$$

where, with $\rho = k_b r$, $\rho_{\text{sph}} = k_b r_{\text{sph}}$ and $\rho_{\text{s}} = k_b r_{\text{s}}$,

$$R_l(r, r_{\text{s}}) = \begin{cases} A_l j_l(n\rho) h_l^{(1)}(\rho_{\text{s}}), & r < r_{\text{sph}} \\ (B_l j_l(\rho) + C_l n_l(\rho)) h_l^{(1)}(\rho_{\text{s}}), & r_{\text{sph}} \le r < r_{\text{s}} \\ (B_l j_l(\rho_{\text{s}}) + C_l n_l(\rho_{\text{s}})) h_l^{(1)}(\rho), & r_{\text{s}} \le r \end{cases} \tag{D.8}$$

$$A_m = \frac{k_b}{\rho_{\text{sph}}^2 D_m} \tag{D.9}$$

$$B_m = \frac{-k_b}{D_m} \begin{vmatrix} j_l(n\rho_{\text{sph}}) & nj_{l+1}(n\rho_{\text{sph}}) \\ n_l(\rho_{\text{sph}}) & n_{l+1}(\rho_{\text{sph}}) \end{vmatrix} \tag{D.10}$$

$$C_m = \frac{k_b}{D_m} \begin{vmatrix} j_l(n\rho_{\text{sph}}) & nj_{l+1}(n\rho_{\text{sph}}) \\ j_l(\rho_{\text{sph}}) & j_{l+1}(\rho_{\text{sph}}) \end{vmatrix} \tag{D.11}$$

$$D_m = \begin{vmatrix} j_l(n\rho_{\text{sph}}) & nj_{l+1}(n\rho_{\text{sph}}) \\ h_l^{(1)}(\rho_{\text{sph}}) & h_{l+1}^{(1)}(\rho_{\text{sph}}) \end{vmatrix} \tag{D.12}$$

and j_l and n_l are the l'th order spherical Bessel function of the first kind and the second kind, $h_l^{(1)} = j_l + \mathrm{j}n_l$ is the l'th order of spherical Hankel function of the first kind, and $P_l(x)$ is the Legendre polynomial defined as

$$P_l(x) = \frac{1}{2^l l!} \frac{\mathrm{d}^l}{\mathrm{d}x^l}(x^2 + 1)^l. \tag{D.13}$$

D.2   Helmholtz equation

The Helmholtz equation for a point source is

$$\nabla_{\boldsymbol{x}}^2 E(\boldsymbol{x}, \boldsymbol{x}_\mathrm{s}) + k^2(\boldsymbol{x})E(\boldsymbol{x}, \boldsymbol{x}_\mathrm{s}) = -\delta(\boldsymbol{x} - \boldsymbol{x}_\mathrm{s}) \tag{D.14}$$

where E is the complex electric field at position $\boldsymbol{x}$ when the point source is at $\boldsymbol{x}_\mathrm{s}$ and $k^2(\boldsymbol{x})$ is defined as

$$k^2(\boldsymbol{x}) = k^2(\|\boldsymbol{x}\|_2) = \begin{cases} n^2 k_b^2, & \text{for } \|\boldsymbol{x}\|_2 < r_\mathrm{sph} \\ k_b^2, & \text{for } \|\boldsymbol{x}\|_2 > r_\mathrm{sph} \end{cases}. \tag{D.15}$$

Note that both 2D and 3D cases follow the same form. Their difference is that $\boldsymbol{x}$ and $\boldsymbol{x}_\mathrm{s}$ in them are 2D or 3D.

D.3   Derivation for 2D case

We consider the polar coordinates, assume $\theta_\mathrm{s} = 0$ without loss of generality, and use an ansatz for the electric field in Eq. (D.1). With the Laplacian in the polar coordinate and the following expansion of a 2D delta function [86]

$$\delta(\boldsymbol{x} - \boldsymbol{x}_\mathrm{s}) = \frac{1}{r}\delta(r - r_\mathrm{s})\frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{\mathrm{j}m(\theta - \theta_\mathrm{s})}, \tag{D.16}$$

eq. Eq. (D.14) becomes a sequence of equations on $R_m(r, r_\mathrm{s})$

$$\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}R_m(r, r_\mathrm{s})\right) + \left(rk^2(r) - \frac{m^2}{r}\right)R_m(r, r_\mathrm{s}) = -\delta(r - r_\mathrm{s}) \tag{D.17}$$

for each m. Each equations is a Bessel differential equation so the solution can be composed of Bessel functions of order m. The boundary conditions for R_m are as follows

1. finite value at $r = 0$

2. only outgoing component at $r = \infty$

3. continuous and first-derivative-continuous at $r = r_\mathrm{sph}$

4. continuous at $r = r_\mathrm{s}$

5. $\left.\frac{\partial R_m}{\partial r}\right|_{r_\mathrm{s}^+} - \left.\frac{\partial R_m}{\partial r}\right|_{r_\mathrm{s}^-} = -\frac{1}{r_\mathrm{s}}$ (integrate (D.17) around r_s)

The above condition and equations lead to Eqs. (D.1)-(D.6).

D.4   Derivation for 3D case

We consider the spherical coordinates and assume that the source lies on the zenith axis. The ansatz for the electric field is Eq. (D.7), the expansion of a 3D delta function is

$$
\delta(\boldsymbol{x} - \boldsymbol{x}_{\text{s}}) = \frac{1}{r^2}\delta(r - r_{\text{s}})\sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_l^m(\theta, \phi)Y_l^m(0, 0)
$$

$$
= \frac{1}{r^2}\delta(r - r_{\text{s}})\sum_{l=0}^{\infty}\left(\frac{2l + 1}{4\pi}\right)P_l(\cos\theta) \tag{D.18}
$$

and eq. Eq. (D.14) becomes

$$
\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}R_l(r, r_{\text{s}})\right) + \left(k^2(r)r^2 - l(l + 1)\right)R_l(r, r_{\text{s}}) = -\delta(r - r_{\text{s}}) \tag{D.19}
$$

for each l. These equations are spherical Bessel equations and there are corresponding spherical Bessel functions to compose the solution. The boundary conditions for the solution are the same as listed above except the last one becoming

$$
\left.\frac{\partial R_l}{\partial r}\right|_{r_{\text{s}}^+} - \left.\frac{\partial R_l}{\partial r}\right|_{r_{\text{s}}^-} = -\frac{1}{r_{\text{s}}^2}. \text{ (integrate Eq. (D.19) around } r_{\text{s}})
$$

Bibliography

[1]   David Abookasis and Tomer Moshe. "Feasibility study of hidden flow imaging based on laser speckle technique using multiperspectives contrast images". In: *Optics and Lasers in Engineering* 62 (2014), pp. 38–45.

[2]   A. Abubakar, P. M. van den Berg, and T. M. Habashy. "Application of the multiplicative regularized contrast source inversion method TM- and TE-polarized experimental Fresnel data". In: *Inv. Probl.* 21.6 (2005), S5–S14.

[3]   G. Agrawal. *Nonlinear Fiber Optics*. 3rd. Academic Press, 2001.

[4]   Miguel A. Alonso. "Wigner functions in optics: describing beams as ray bundles and pulses as particle ensembles". In: *Adv. Opt. Photon.* 3.4 (Dec. 2011), pp. 272–365. DOI: `10.1364/AOP.3.000272`.

[5]   Nick Antipa et al. "DiffuserCam: lensless single-exposure 3D imaging". In: *Optica* 5.1 (Jan. 2018), pp. 1–9. DOI: `10.1364/OPTICA.5.000001`.

[6]   Anthony Atkinson, Alexander Donev, and Randall Tobias. *Optimum experimental designs, with SAS*. Vol. 34. Oxford University Press, 2007.

[7]   H. Bartelt, K. Brenner, and A. Lohmann. "The Wigner distribution function and its optical production". In: *Opt. Commun.* 32.1 (1980), pp. 32–38. DOI: `DOI:10.1016/0030-4018(80)90308-9`.

[8]   M. Bastiaans. "The Wigner distribution function applied to optical signals and systems". In: *Opt. Commun.* 25.1 (1978), pp. 26–30.

[9]   Eugen Baumgart and Ulrich Kubitscheck. "Scanned light sheet microscopy with confocal slit detection". In: *Optics Express* 20.19 (2012), pp. 21805–21814.

[10]   A. Baussard. "Inversion of multi-frequency experimental data using an adaptive multiscale approach". In: *Inv. Probl.* 21.6 (2005), S15–S32.

[11]   A. Beck and M. Teboulle. "Fast Gradient-Based Algorithm for Constrained Total Variation Image Denoising and Deblurring Problems". In: *IEEE Trans. Image Process.* 18.11 (Nov. 2009), pp. 2419–2434.

[12]   Amir Beck and Marc Teboulle. "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems". In: *SIAM Journal on Imaging Sciences* 2.1 (2009), pp. 183–202. DOI: `10.1137/080716542`.

[13] J. Bect et al. "A ℓ_1-unified variational framework for image restoration". In: *Proc. ECCV*. Ed. by Springer. Vol. 3024. New York, 2004, pp. 1–13.

[14] Aurélien Bègue et al. "Two-photon excitation in scattering media by spatiotemporally shaped beams and their application in optogenetic stimulation". In: *Biomedical Optics Express* 4.12 (2013), pp. 2869–2879.

[15] K. Belkebir, P. C. Chaumet, and A. Sentenac. "Superresolution in total internal reflection tomography". In: *J. Opt. Soc. Am. A* 22.9 (Sept. 2005), pp. 1889–1897.

[16] M. Bertero and P. Boccacci. "A simple method for the reduction of boundary effects in the Richardson-Lucy approach to image deconvolution". In: *Astronomy and Astrophysics* 437 (July 2005), pp. 369–374. DOI: 10.1051/0004-6361:20052717.

[17] D. P. Bertsekas. "Incremental proximal methods for large scale convex optimization". In: *Math. Program. Ser. B* 129 (2011), pp. 163–195.

[18] B.N. Bhaskar, Gongguo Tang, and B. Recht. "Atomic Norm Denoising With Applications to Line Spectral Estimation". In: *Signal Processing, IEEE Transactions on* 61.23 (Dec. 2013), pp. 5987–5999. ISSN: 1053-587X. DOI: 10.1109/TSP.2013.2273443.

[19] J. M. Bioucas-Dias and M. A. T. Figueiredo. "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration". In: *IEEE Trans. Image Process.* 16.12 (Dec. 2007), pp. 2992–3004.

[20] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford, 1995.

[21] M. Born and E. Wolf. *Principles of Optics*. 7th (expanded) edition. Cambridge Univ. Press, 1999.

[22] M. Born and E. Wolf. "Principles of Optics". In: 7th ed. Cambridge Univ. Press, 2003. Chap. Scattering from inhomogeneous media, pp. 695–734.

[23] M. Born and E. Wolf. "Principles of Optics". In: 7th ed. Cambridge Univ. Press, 2003. Chap. Transition to a Scalar Theory, pp. 430–436.

[24] L. Bottou. "Neural Networks: Tricks of the Trade". In: 2nd ed. Springer, Sept. 2012. Chap. Stochastic Gradient Descent Tricks, pp. 421–437.

[25] L. Bottou, F. E. Curtis, and J. Nocedal. "Optimization Methods for Large-Scale Machine Learning". In: (June 2016). arXiv:1606.04838 [stat.ML].

[26] Matthew B Bouchard et al. "Swept confocally-aligned planar excitation (SCAPE) microscopy for high-speed volumetric imaging of behaving organisms". In: *Nature Photonics* 9.2 (2015), pp. 113–119.

[27] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, 2004.

[28] D. J. Brady et al. "Compressive Holography". In: *Opt. Express* 17.15 (2009), pp. 13040–13049.

[29] K. H. Brenner and A. W. Lohmann. "Wigner distribution function display of complex 1D signals". In: *Optics Communications* 42.5 (1982), pp. 310–314. ISSN: 0030-4018. DOI: `DOI:10.1016/0030-4018(82)90238-3`.

[30] K. H. Brenner, A. W. Lohmann, and J. Ojeda-Castañeda. "The ambiguity function as a polar display of the OTF". In: *Optics Communications* 44.5 (1983), pp. 323–326. ISSN: 0030-4018.

[31] Robert Bridson. "Fast Poisson Disk Sampling in Arbitrary Dimensions". In: *ACM SIGGRAPH 2007 Sketches*. SIGGRAPH '07. San Diego, California: ACM, 2007. ISBN: 978-1-4503-4726-6. DOI: `10.1145/1278780.1278807`.

[32] M. M. Bronstein et al. "Reconstruction in Diffraction Ultrasound Tomography Using Nonuniform FFT". In: *IEEE Trans. Med. Imag.* 21.11 (Nov. 2002), pp. 1395–1401.

[33] Michael Broxton et al. "Wave optics theory and 3-D deconvolution for the light field microscope". In: *Opt. Express* 21.21 (2013), pp. 25418–25439.

[34] F. Bürgel, K. S. Kazimierski, and A. Lechleiter. "A sparsity regularization and total variation based computational framework for the inverse medium problem in scattering". In: *J. Comput. Physics* 339 (2017), pp. 1–30.

[35] E. J. Candès, J. Romberg, and T. Tao. "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information". In: *IEEE Trans. Inf. Theory* 52.2 (Feb. 2006), pp. 489–509.

[36] Emmanuel J Candès and Michael B Wakin. "An introduction to compressive sampling". In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 21–30.

[37] Julie Chang et al. "Variable Aperture Light Field Photography: Overcoming the Diffraction-Limited Spatio-Angular Resolution Tradeoff". In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 3737–3745.

[38] M. Charnotskii. "Extended Huygens-Fresnel principle and optical waves propagation in turbulence: discussion". In: *J. Opt. Soc. Am. A* 32.7 (July 2015), pp. 1357–1365.

[39] P. C. Chaumet and K. Belkebir. "Three-dimensional reconstruction from real data using a conjugate gradient-coupled dipole method". In: *Inv. Probl.* 25.2 (2009), p. 024003.

[40] B. Chen and J. J. Stamnes. "Validity of diffraction tomography based on the first Born and the first Rytov approximations". In: *Appl. Opt.* 37.14 (May 1998), pp. 2996–3006.

[41] Tsai-Wen Chen et al. "Ultrasensitive fluorescent proteins for imaging neuronal activity". In: *Nature* 499.7458 (2013), pp. 295–300.

[42] W. Chen et al. "Empirical concentration bounds for compressive holographic bubble imaging based on a Mie scattering model". In: *Opt. E* 23.4 (2015), February.

[43] W. C. Chew and Y. M. Wang. "Reconstruction of Two-Dimensional Permittivity Distribution Using the Distorted Born Iterative Method". In: *IEEE Trans. Med. Imag.* 9.2 (June 1990), pp. 218–225.

[44] W. Choi et al. "Tomographic Phase Microscopy". In: *Nat. Methods* 4.9 (Sept. 2007), pp. 717–719.

[45] J. M. Cowley and A. F. Moodie. "The scattering of electrons by atoms and crystals. I. A new theoretical approach". In: *Acta Crystallographica* 10.10 (Oct. 1957), pp. 609–619. DOI: 10.1107/S0365110X57002194.

[46] IJ Cox, Colin Sheppard, and Tony Wilson. "Super-resolution by confocal fluorescent microscopy". In: *Optik* 60.4 (1982), pp. 391–396.

[47] L. Crocco, M. D'Urso, and T. Isernia. "Testing the contrast source extended Born inversion method against real data: the TM case". In: *Inv* 21.6 (2005), S33–S51.

[48] M. Cui and C. Yang. "Implementation of a digital optical phase conjugation system and its application to study the robustness of turbidity suppression by phase conjugation". In: *Opt. Express* 18.4 (2010), pp. 3444–3455.

[49] I. Daubechies, M. Defrise, and C. De Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint". In: *Commun. Pure Appl. Math.* 57.11 (Nov. 2004), pp. 1413–1457.

[50] B. J. Davis et al. "Nonparaxial vector-field modeling of optical coherence tomography and interferometric synthetic aperture microscopy". In: *J. Opt. Soc. Am. A* 24.9 (Sept. 2007), pp. 2527–2542.

[51] Winfried Denk, James Strickler, and Watt Webb. "Two-photon laser scanning fluorescence microscopy". In: *Science* 248.4951 (1990), pp. 73–76.

[52] A. J. Devaney. "Inverse-scattering theory within the Rytov approximation". In: *Opt. Lett.* 6.8 (Aug. 1981), pp. 374–376.

[53] D. L. Donoho. "Compressed sensing". In: *IEEE Trans. Inf. Theory* 52.4 (Apr. 2006), pp. 1289–1306.

[54] D. Dragoman. "Phase-space interferences as the source of negative values of the Wigner distribution function". In: *J. Opt. Soc. Am. A* 17.12 (2000), pp. 2481–2485.

[55] A. Dubois, K. Belkebir, and M. Saillard. "Retrieval of inhomogeneous targets from experimental frequency diversity data". In: *Inv. Probl.* 21.6 (2005), S65–S81.

[56] D. V. Dylov, L. Waller, and J. W. Fleischer. "Instability-driven recovery of diffused images". In: *Opt. Lett.* 36.18 (2011), pp. 3711–3713. DOI: 10.1364/OL.36.003711.

[57] Albert C Fannjiang. "Compressive inverse scattering: I. High-frequency SIMO/MISO and MIMO measurements". In: *Inverse Problems* 26.3 (2010), p. 035008.

[58] Sina Farsiu et al. "Statistical detection and imaging of objects hidden in turbid media using ballistic photons". In: *Appl. Opt.* 46.23 (Aug. 2007), pp. 5805–5822. DOI: 10.1364/AO.46.005805.

[59] E. E. Fenimore and T. M. Cannon. "Coded aperture imaging with uniformly redundant arrays". In: *Appl. Opt.* 17.3 (Feb. 1978), pp. 337–347. DOI: 10.1364/AO.17.000337.

[60] O. Féron, B. Duchêne, and A. Mohammad-Djafari. "Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data". In: *Inv. Probl.* 21.6 (2005), S95–S117.

[61] J. R. Fienup. "Phase retrieval algorithms: a comparison". In: *Appl. Opt.* 21.15 (1982), pp. 2758–2769.

[62] J.R. Fienup. "Reconstruction of an object from the modulus of its Fourier transform". In: *Opt. Lett.* 3 (1978).

[63] M. A. T. Figueiredo and R. D. Nowak. "An EM Algorithm for Wavelet-Based Image Restoration". In: *IEEE Trans. Image Process.* 12.8 (Aug. 2003), pp. 906–916.

[64] Mark H. Finger and Thomas A. Prince. "Useful Classes of Redundant Arrays for Imaging Applications". In: *Imaging in High Energy Astronomy.* Ed. by L. Bassani and G. Di Cocco. Dordrecht: Springer Netherlands, 1995, pp. 221–226. ISBN: 978-94-011-0407-4.

[65] S. Flewett et al. "Extracting coherent modes from partially coherent wavefields". In: *Opt. Lett.* 34.14 (2009), pp. 2198–2200.

[66] Jack D Gaskill. "Imaging through a randomly inhomogeneous medium by wavefront reconstruction". In: *J. Opt. Soc. Am.* 58.5 (1968), pp. 600–608.

[67] J.-M. Geffrin and P. Sabouroux. "Continuing with the Fresnel database: experimental setup and improvements in 3D scattering measurements". In: *Inv. Probl.* 25.2 (2009), p. 024001.

[68] J.-M. Geffrin, P. Sabouroux, and C. Eyraud. "Free space experimental scattering database continuation: experimental set-up and measurement precision". In: *Inv. Probl.* 21.6 (2005), S117–S130.

[69] S. Gigan et al. "Image transmission through a scattering medium: Inverse problem and sparsity-based imaging". In: *Information Optics (WIO), 2014 13th Workshop on.* July 2014, pp. 1–3. DOI: 10.1109/WIO.2014.6933276.

[70] Marcel J. E. Golay. "Point Arrays Having Compact, Nonredundant Autocorrelations". In: *J. Opt. Soc. Am.* 61.2 (Feb. 1971), pp. 272–273. DOI: 10.1364/JOSA.61.000272.

[71] A. González and Y. Mejía. "Nonredundant array of apertures to measure the spatial coherence in two dimensions with only one interferogram". In: *J. Opt. Soc. Am. A* 28.6 (2011), pp. 1107–1113.

[72] J. W. Goodman. *Introduction to Fourier optics.* Ed. by Crisp, J. & Elliott, B. 3rd ed. Roberts & Company Publishers, 2004.

[73] JW Goodman et al. "Wavefront-Reconstruction Imaging Through Random Media". In: *Applied Physics Letters* 8.12 (1966), pp. 311–313.

[74] D. Gordon. *La Jolla Difference Set Repository.* Apr. 2018. URL: http://www.ccrwest.org/diffsets/index.html.

[75] Daniel M. Gordon. *The prime power conjecture is true for $n < 2,000,000$*. 1994.

[76] Stephen R. Gottesman and E. E. Fenimore. "New family of binary arrays for coded aperture imaging". In: *Appl. Opt.* 28.20 (Oct. 1989), pp. 4344–4352. DOI: `10.1364/AO.28.004344`.

[77] Paul Green et al. "Multi-aperture Photography". In: *ACM Trans. Graph.* 26.3 (July 2007). ISSN: 0730-0301. DOI: `10.1145/1276377.1276462`.

[78] O. Haeberlé et al. "Tomographic diffractive microscopy: basic, techniques, and perspectives". In: *J. Mod. Opt.* 57.9 (May 2010), pp. 686–699.

[79] Eugene Hecht. *Optics*. 4th ed. Addison Wesley, Aug. 2001. ISBN: 0805385665.

[80] Fritjof Helmchen and Winfried Denk. "Deep tissue two-photon microscopy". In: *Nat. Methods* 2.12 (2005), pp. 932–940.

[81] Gabor T Herman. *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media, 2009.

[82] F. J. Herrmann and T. van Leeuwen. "Mitigating local minima in full-waveform inversion by expanding the search space". In: *Geophys. J. Int.* 195.1 (2013), pp. 661–667.

[83] Chia-Lung Hsieh et al. "Imaging through turbid layers by scanning the phase conjugated second harmonic radiation from a nanoparticle". In: *Opt. Express* 18.20 (Sept. 2010), pp. 20723–20731. DOI: `10.1364/OE.18.020723`.

[84] David Huang et al. "Optical coherence tomography". In: *Science* 254.5035 (1991), pp. 1178–1181.

[85] R.L. Hudson. "When is the Wigner quasi-probability density non-negative?" In: *Reports on Mathematical Physics* 6.2 (1974), pp. 249–252. ISSN: 0034-4877. DOI: `10.1016/0034-4877(74)90007-X`.

[86] J. D. Jackson. *Classical Electrodynamics*. 3rd ed. Wiley, 1999.

[87] J. Christopher James, Gisele Welch, and William T. Rhodes. "Nonuniform coherence sampling for fields produced by incoherent three-dimensional sources". In: *J. Opt. Soc. Am. A* 20.4 (Apr. 2003), pp. 668–677. DOI: `10.1364/JOSAA.20.000668`.

[88] Na Ji, Daniel E Milkie, and Eric Betzig. "Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues". In: *Nat. Methods* 7.2 (Feb. 2010), pp. 141–147.

[89] D. Jin et al. "Tomographic Phase Microscopy: Principles and Applications in Bioimaging". In: *J. Opt. Soc. Am. B* 34.5 (2017), B64–B77.

[90] H. M. Jol, ed. *Ground Penetrating Radar: Theory and Applications*. Amsterdam: Elsevier, 2009.

[91] Eric Jonas et al. "Occupy the Cloud: Distributed Computing for the 99%". In: *CoRR* abs/1702.04024 (2017). arXiv: `1702.04024`. URL: `http://arxiv.org/abs/1702.04024`.

[92] Dieter Jungnickel and Bernhard Schmidt. "Difference Sets: An Update". In: (1997). URL: `http://www.ntu.edu.sg/home/bernhard/Publications/pub/update1.pdf`.

[93] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE, 1988.

[94] U. S. Kamilov. "A Parallel Proximal Algorithm for Anisotropic Total Variation Minimization". In: *IEEE Trans. Image Process.* 26.2 (Feb. 2017), pp. 539–548.

[95] U. S. Kamilov et al. "A Recursive Born Approach to Nonlinear Inverse Scattering". In: *IEEE Signal Process. Lett.* 23.8 (Aug. 2016), pp. 1052–1056.

[96] U. S. Kamilov et al. "Learning Approach to Optical Tomography". In: *Optica* 2.6 (June 2015), pp. 517–522.

[97] U. S. Kamilov et al. "Optical Tomographic Image Reconstruction Based on Beam Propagation and Sparse Regularization". In: *IEEE Trans. Comp. Imag.* 2.1 (Mar. 2016), pp. 59–70.

[98] Gergely Katona et al. "Fast two-photon in vivo imaging with three-dimensional random-access scanning in large tissue volumes". In: *Nature Methods* 9.2 (Feb. 2012), pp. 201–208.

[99] Ori Katz, Eran Small, and Yaron Silberberg. "Looking around corners and through thin turbid layers in real time with scattered incoherent light". In: *Nature Photonics* 6.8 (Aug. 2012), pp. 549–553. DOI: `10.1038/nphoton.2012.150`.

[100] Ori Katz et al. "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations". In: *Nature Photonics* 8.10 (2014), pp. 784–790.

[101] Philipp Keller et al. "Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy". In: *Science* 322.5904 (2008), pp. 1065–1069.

[102] Jaewon Kim et al. "Descattering Transmission via Angular Filtering". In: *Proceedings of the 11th European Conference on Computer Vision: Part I*. ECCV'10. Heraklion, Crete, Greece: Springer-Verlag, 2010, pp. 86–99. ISBN: 3-642-15548-0, 978-3-642-15548-2.

[103] Moonseok Kim et al. "Maximal energy transport through disordered media with the implementation of transmission eigenchannels". In: *Nature Photonics* 6.9 (July 2012), pp. 583–587. ISSN: 1749-4885. DOI: `10.1038/nphoton.2012.159`.

[104] T. Kim et al. "White-light diffraction tomography of unlabelled live cells". In: *Nat. Photonics* 8 (Mar. 2014), pp. 256–263.

[105] A. I. Kirkland, W. O. Saxton, and G. Chand. "Multi beam tilt microscopy for super resolved imaging". In: *J. Electron Microsc. (Tokyo)* 46.11 (Jan. 1997), pp. 11–22.

[106] R. E. Kleinman, G. F. Roach, and P. M. van den Berg. "Convergent Born series for large refractive indices". In: *J. Opt. Soc. Am. A* 7.5 (May 1990), pp. 890–897.

[107] H Kogelnik and KS Pennington. "Holographic imaging through a random medium". In: *J. Opt. Soc. Am.* 58.2 (1968), pp. 273–274.

[108] V. Lauer. "New approach to optical diffraction tomography yielding a vector equation of diffraction tomography and a novel tomographic microscope". In: *J. Microsc.* 205.2 (2002), pp. 165–176.

[109] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning". In: *Nature* 521 (May 2015), pp. 436–444.

[110] Daniel D. Lee and H. Sebastian Seung. "Algorithms for Non-negative Matrix Factorization". In: *Advances in Neural Information Processing Systems 13*. Ed. by T.K. Leen, T.G. Dietterich, and V. Tresp. MIT Press, 2001, pp. 556–562.

[111] Ok Kyun Lee et al. "A non-iterative method for the electrical impedance tomography based on joint sparse recovery". In: *Inverse Problems* 31.7 (2015), p. 075002.

[112] Okkyun Lee and Jong Chul Ye. "Joint sparsity-driven non-iterative simultaneous reconstruction of absorption and scattering in diffuse optical tomography". In: *Opt. Express* 21.22 (Nov. 2013), pp. 26589–26604. DOI: 10.1364/OE.21.026589.

[113] O. Lee et al. "Compressive Diffuse Optical Tomography: Noniterative Exact Reconstruction Using Joint Sparsity". In: *IEEE Transactions on Medical Imaging* 30.5 (May 2011), pp. 1129–1142. ISSN: 0278-0062. DOI: 10.1109/TMI.2011.2125983.

[114] M. Leigsnering et al. "Multipath Exploitation in Through-the-wall Radar Imaging Using Sparse Reconstruction". In: *IEEE Trans. Aerosp. Electron. Syst.* 50.2 (Apr. 2014), pp. 920–939.

[115] Anat Levin et al. "Image and Depth from a Conventional Camera with a Coded Aperture". In: *ACM Trans. Graph.* 26.3 (July 2007). ISSN: 0730-0301. DOI: 10.1145/1276377.1276464.

[116] M. Levoy. "Light Fields and Computational Imaging". In: *Computer* 39.8 (2006), pp. 46–55. ISSN: 0018-9162. DOI: 10.1109/MC.2006.270.

[117] Marc Levoy et al. "Light Field Microscopy". In: *ACM Trans. Graph.* 25.3 (July 2006), pp. 924–934. ISSN: 0730-0301. DOI: 10.1145/1141911.1141976.

[118] C. Liang et al. "Programmable Aperture Photography: Multiplexed Light Field Acquisition". In: *Proceedings of ACM SIGGRAPH*. Los Angeles, California: ACM, 2008, 55:1–55:10. ISBN: 978-1-4503-0112-1. DOI: 10.1145/1399504.1360654.

[119] J. W. Lim et al. "Comparative study of iterative reconstruction algorithms for missing cone problems in optical diffraction tomography". In: *Opt. Express* 23.13 (June 2015), pp. 16933–16948.

[120] D. Liu, U. S. Kamilov, and P. T. Boufounos. "Compressive Tomographic Radar Imaging with Total Variation Regularization". In: *Proc. IEEE 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar, and Remote Sensing (CoSeRa 2016)*. Aachen, Germany, Sept. 2016, pp. 120–123.

[121] Hsiou-Yuan Liu, Jingshan Zhong, and Laura Waller. "4D phase-space multiplexing for fluorescent microscopy". In: *Proc. SPIE* 9720 (2016), DOI: `10.1117/12.2213610`.

[122] Hsiou-Yuan Liu et al. "3D imaging in volumetric scattering media using phase-space measurements". In: *Opt. Express* 23.11 (June 2015), pp. 14461–14471. DOI: `10.1364/OE.23.014461`.

[123] H.-Y. Liu et al. "Compressive Imaging with Iterative Forward Models". In: *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP 2017)*. New Orleans, LA, USA, Mar. 2017, pp. 6025–6029.

[124] Miles E Lopes. "Estimating unknown sparsity in compressed sensing". In: *ArXiv preprint arXiv:1204.4227* (2012).

[125] M. Lustig, D. Donoho, and J.M. Pauly. "Sparse MRI: The application of compressed sensing for rapid MR imaging". In: *Magnetic Resonance in Medicine* 58.6 (2007), pp. 1182–1195.

[126] Richard MacKenzie. "Path integral methods and applications". In: *ArXiv e-prints* (2000). arXiv: `quant-ph/0004090 [quant-ph]`.

[127] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg. "Ptychographic transmission microscopy in three dimensions using a multi-slice approach". In: *J. Opt. Soc. Am. A* 29.8 (Aug. 2012), pp. 1606–1614. DOI: `10.1364/JOSAA.29.001606`.

[128] J. Mairal, F. Bach, and J. Ponce. "Sparse Modeling for Image and Vision Processing". In: *Foundations and Trends in Machine Learning* 8.2-3 (2014), pp. 1–199.

[129] Joseph N. Mait, Ravindra Anant Athale, and Gary W. Euliss. "Two Decades of Computational Imaging: Still serving old wine in new bottles?" In: *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*. Optical Society of America, 2017, CM2B.1. DOI: `10.1364/COSI.2017.CM2B.1`.

[130] Daniel L. Marks, Ronald A. Stack, and David J. Brady. "Three-dimensional coherence imaging in the Fresnel domain". In: *Appl. Opt.* 38.8 (Mar. 1999), pp. 1332–1342. DOI: `10.1364/AO.38.001332`.

[131] D. Marks et al. "Visible cone-beam tomography with a lensless interferometric camera". In: *Science* 284 (1999), pp. 2164–2166.

[132] Emily J. McDowell et al. "Turbidity suppression from the ballistic to the diffusive regime in biological tissues using optical phase conjugation". In: *Journal of Biomedical Optics* 15.2 (2010), DOI: `10.1117/1.3381188`.

[133] Hanno Meyer et al. "Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissal cortex". In: *Cerebral Cortex* 20.10 (2010), pp. 2277–2286.

[134] Inkyu Moon and Bahram Javidi. "Three-dimensional visualization of objects in scattering medium by use of computational integral imaging". In: *Opt. Express* 16.17 (2008), pp. 13080–13089.

[135] E. Mudry et al. "Electromagnetic wave imaging of three-dimensional targets using a hybrid iterative inversion method". In: *Inv. Probl.* 28.6 (Apr. 2012), p. 065007.

[136] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course.* Kluwer Academic Publishers, 2004.

[137] Y. E. Nesterov. "A method for solving the convex programming problem with convergence rate $O(1/k^2)$". In: *Dokl. Akad. Nauk SSSR* 269 (1983). (in Russian), pp. 543–547.

[138] Andrew Y. Ng. "Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance". In: *Proceedings of the Twenty-first International Conference on Machine Learning.* ICML '04. Banff, Alberta, Canada: ACM, 2004, pp. 78–. ISBN: 1-58113-838-5. DOI: 10.1145/1015330.1015435.

[139] R. Ng. "Fourier slice photography". In: *SIGGRAPH '05.* Los Angeles, California, 2005. DOI: 10.1145/1186822.1073256.

[140] R. Ng et al. *Light Field Photography with a Hand-held Plenoptic Camera.* Tech. rep. CTSR 2005-02. Stanford, 2005.

[141] N. K. Nikolova et al. "Sensitivity analysis of network parameters with electromagnetic frequency-domain simulators". In: *IEEE Transactions on Microwave Theory and Techniques* 54.2 (Feb. 2006), pp. 670–681. ISSN: 0018-9480. DOI: 10.1109/TMTT.2005.862663.

[142] J. Nocedal and S. J. Wright. *Numerical Optimization.* 2nd ed. Springer, 2006.

[143] V. Ntziachristos. "Going deeper than microscopy: the optical imaging frontier in biology". In: *Nat. Methods* 7.8 (Aug. 2010), pp. 603–614.

[144] Travis E Oliphant. *A guide to NumPy.* Vol. 1. Trelgol Publishing USA, 2006.

[145] Sri Rama Prasanna Pavani et al. "Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function". In: *Proc. Natl. Acad. Sci. USA* 106.9 (2009), pp. 2995–2999.

[146] Nicolas C Pégard et al. "Compressive light-field microscopy for 3D neural activity recording". In: *Optica* 3.5 (2016), pp. 517–524.

[147] Carl Petersen, Amiram Grinvald, and Bert Sakmann. "Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured in vivo by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions". In: *The Journal of Neuroscience* 23.4 (2003), pp. 1298–1309.

[148] Thomas Planchon et al. "Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination". In: *Nat. Methods* 8.5 (2011), pp. 417–423.

[149] Eftychios A. Pnevmatikakis et al. "Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data". In: *Neuron* 89.2 (2016), pp. 285–299. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2015.11.037.

[150] S. M. Popoff et al. "Measuring the Transmission Matrix in Optics: An Approach to the Study and Control of Light Propagation in Disordered Media". In: *Phys. Rev. Lett.* 104 (10 Mar. 2010), p. 100601. DOI: 10.1103/PhysRevLett.104.100601.

[151] Robert Prevedel et al. "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy". In: *Nature Methods* (2014).

[152] T. S. Ralston et al. "Inverse scattering for optical coherence tomography". In: *J. Opt. Soc. Am. A* 23.5 (May 2006), pp. 1027–1037.

[153] M. Raymer, M. Beck, and D. McAlister. "Complex wave-field reconstruction using phase-space tomography". In: *Phys. Rev. Lett.* 72.8 (1994), pp. 1137–1140. DOI: 10.1103/PhysRevLett.72.1137.

[154] Joseph Rosen and David Abookasis. "Seeing through biological tissues using the fly eye principle". In: *Opt. Express* 11.26 (2003), pp. 3605–3611.

[155] L. I. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D* 60.1–4 (Nov. 1992), pp. 259–268.

[156] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D: Nonlinear Phenomena* 60.1–4 (1992), pp. 259–268. ISSN: 0167-2789. DOI: 10.1016/0167-2789(92)90242-F.

[157] Yoav Y. Schechner, Rafael Piestun, and Joseph Shamir. "Wave propagation with rotating intensity distributions". In: *Phys. Rev. E* 54 (1 July 1996), R50–R53. DOI: 10.1103/PhysRevE.54.R50.

[158] Tina Schrödel et al. "Brain-wide 3D imaging of neuronal activity in Caenorhabditis elegans with sculpted light". In: *Nature Methods* 10.10 (2013), pp. 1013–1020.

[159] J. Sharpe et al. "Optical Projection Tomography as a Tool for 3D Microscopy and Gene Expression Studies". In: *Science* 296.5567 (Apr. 2002), pp. 541–545.

[160] Yoav Shechtman et al. "Optimal Point Spread Function Design for 3D Imaging". In: *Phys. Rev. Lett.* 113 (13 Sept. 2014), p. 133902. DOI: 10.1103/PhysRevLett.113.133902.

[161] C. J. R. Sheppard and C. J. Cogswell. "Three-dimensional image formation in confocal microscopy". In: *J. Microsc.* 159.2 (Aug. 1990), pp. 179–194.

[162] S. Silver. "Microwave Aperture Antennas and Diffraction Theory". In: *J. Opt. Soc. Am.* 52.2 (Feb. 1962), pp. 131–139.

[163] James Singer. "A Theorem in Finite Projective Geometry and Some Applications to Number Theory". In: *Transactions of the American Mathematical Society* 43.3 (1938), pp. 377–385. ISSN: 00029947.

[164] A. K. Singh et al. "Exploiting scattering media for exploring 3D objects". In: *Light Sci. Appl.* 6.2 (Feb. 2017), pp. 1–7.

[165] Alok Kumar Singh et al. "Looking through a diffuser and around an opaque surface: A holographic approach". In: *Opt. Express* 22.7 (Apr. 2014), pp. 7694–7701. DOI: 10.1364/OE.22.007694.

[166] G.K. Skinner. "Imaging with coded-aperture masks". In: *Nuclear Instruments and Methods in Physics Research* 221.1 (1984). Proceedings of the International Workshop on X- and γ-Ray Imaging Techniques, pp. 33–40. ISSN: 0167-5087. DOI: 10.1016/0167-5087(84)90174-1.

[167] Eran Small, Ori Katz, and Yaron Silberberg. "Spatiotemporal focusing through a thin scattering layer". In: *Opt. Express* 20.5 (Feb. 2012), pp. 5189–5195. DOI: 10.1364/OE.20.005189.

[168] Y. Sung and R. R. Dasari. "Deterministic regularization of three-dimensional optical diffraction tomography". In: *J. Opt. Soc. Am. A* 28.8 (Aug. 2011), pp. 1554–1561.

[169] Y. Sung et al. "Optical Diffraction Tomography for high resolution live cell imaging". In: *Opt. Express* 17.1 (Dec. 2009), pp. 266–277.

[170] Kevin T. Takasaki and Jason W. Fleischer. "Phase-space measurement for depth-resolved memory-effect imaging". In: *Opt. Express* 22.25 (Dec. 2014), pp. 31426–31433. DOI: 10.1364/OE.22.031426.

[171] Gongguo Tang, B.N. Bhaskar, and B. Recht. "Sparse recovery over continuous dictionaries-just discretize". In: *Signals, Systems and Computers, 2013 Asilomar Conference on.* Nov. 2013, pp. 1043–1047. DOI: 10.1109/ACSSC.2013.6810450.

[172] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation.* SIAM, 2005.

[173] M. Testorf, B. Hennelly, and J. Ojeda-Castañeda. *Phase-Space Optics: Fundamentals and Applications.* McGraw-Hill Professional, 2009. ISBN: 0-07-159798-0.

[174] Lei Tian and Laura Waller. "3D intensity and phase imaging from light field measurements in an LED array microscope". In: *Optica* 2.2 (Feb. 2015), pp. 104–111. DOI: 10.1364/OPTICA.2.000104.

[175] Lei Tian et al. "Multiplexed coded illumination for Fourier Ptychography with an LED array microscope". In: *Biomed. Opt. Express* 5.7 (July 2014), pp. 2376–2389. DOI: 10.1364/BOE.5.002376.

[176] Lei Tian et al. "Wigner function measurement using a lenslet array". In: *Opt. Express* 21.9 (2013), pp. 10511–10525.

[177] L. Tian et al. "Quantitative measurement of size and three-dimensional position of fast-moving bubbles in air-water mixture flows using digital holography". In: *Appl. Opt.* 49.9 (Mar. 2010), pp. 1549–1554.

[178] Robert Tibshirani. "Regression Shrinkage and Selection Via the Lasso". In: *Journal of the Royal Statistical Society, Series B* 58 (1994), pp. 267–288.

[179] A. G. Tijhuis. "Born-type reconstruction of material parameters of an inhomogeneous, lossy dielectric slab from reflected-field data". In: *Wave Motion* 11.2 (May 1989), pp. 151–173.

[180] Stephen A. Vavasis. "On the complexity of nonnegative matrix factorization". In: *CoRR* abs/0708.4149 (2007). URL: http://arxiv.org/abs/0708.4149.

[181] IM Vellekoop and AP Mosk. "Focusing coherent light through opaque strongly scattering media". In: *Opt. Lett.* 32.16 (2007), pp. 2309–2311.

[182] J. Virieux and S. Operto. "An overview of full-waveform inversion in exploration geophysics". In: *Geophyscs* 74.6 (Nov. 2009), WCC127–WCC152.

[183] L. Waller, G. Situ, and J.W. Fleischer. "Phase-space measurement and coherence synthesis of optical beams". In: *Nat. Photonics* 6 (2012), pp. 474–479.

[184] A. Walther. "Radiometry and coherence". In: *J. Opt. Soc. Am.* 58.9 (1968), pp. 1256–1259.

[185] L Wang et al. "Ballistic 2-D imaging through scattering walls using an ultrafast optical Kerr gate". In: *Science* 253.5021 (1991), pp. 769–771.

[186] Y. M. Wang and W. C. Chew. "An Iterative Solution of the Two-Dimensional Electromagnetic Inverse Scattering Problem". In: *Int. J. Imag. Syst Tech.* 1 (1989), pp. 100–108.

[187] Emil Wolf. "Three-dimensional structure determination of semi-transparent objects from holographic data". In: *Opt. Commun.* 1.4 (Sept. 1969), pp. 153–156.

[188] James K Wood et al. "Using shadows to measure spatial coherence". In: *Opt. Lett.* 39.16 (2014), pp. 4927–4930.

[189] S.J. Wright, R.D. Nowak, and M.A.T. Figueiredo. "Sparse Reconstruction by Separable Approximation". In: *Signal Processing, IEEE Transactions on* 57.7 (July 2009), pp. 2479–2493. ISSN: 1053-587X. DOI: 10.1109/TSP.2009.2016892.

[190] Tong Chul Ye and Su Yeon Lee. "Non-iterative exact inverse scattering using simultaneous orthogonal matching pursuit (S-OMP)". In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. Mar. 2008, pp. 2457–2460. DOI: 10.1109/ICASSP.2008.4518145.

[191] Li-Hao Yeh et al. "Experimental robustness of Fourier ptychography phase retrieval algorithms". In: *Opt. Express* 23.26 (Dec. 2015), pp. 33214–33240. DOI: 10.1364/OE.23.033214.

[192] Jaejun Yoo et al. "A Joint Sparse Recovery Framework for Accurate Reconstruction of Inclusions in Elastic Media". In: *SIAM Journal on Imaging Sciences* 10.3 (2017), pp. 1104–1138. DOI: 10.1137/16M110318X.

[193] T. Zhang et al. "Far-field diffraction microscopy at $\lambda/10$ resolution". In: *Optica* 3.6 (June 2016), pp. 609–612.

[194] Zhengyun Zhang and M. Levoy. "Wigner distributions and how they relate to the light field". In: *Proceedings of IEEE International Conference on Computational Photography*. IEEE, Apr. 2009, pp. 1–10. DOI: 10.1109/ICCPHOT.2009.5559007.

[195] Guoan Zheng, Roarke Horstmeyer, and Changhuei Yang. "Wide-field, high-resolution Fourier ptychographic microscopy". In: *Nature Photonics* 7.9 (2013), pp. 739–745.