Compressive Phase Retrieval

by

Lei Tian

B.S., Tsinghua University (2008)
S.M., Massachusetts Institute of Technology (2010)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Author .....
Department of Mechanical Engineering
May 18, 2013

Certified by .....
George Barbastathis
Professor
Thesis Supervisor

Accepted by .....
David E. Hardt
Chairman, Department Committee on Graduate Students

Compressive Phase Retrieval

by

Lei Tian

Submitted to the Department of Mechanical Engineering
on May 18, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Recovering a full description of a wave from limited intensity measurements remains a central problem in optics. Optical waves oscillate too fast for detectors to measure anything but time–averaged intensities. This is unfortunate since the phase can reveal important information about the object. When the light is partially coherent, a complete description of the phase requires knowledge about the statistical correlations for each pair of points in space. Recovery of the correlation function is a much more challenging problem since the number of pairs grows much more rapidly than the number of points.

In this thesis, quantitative phase imaging techniques that works for partially coherent illuminations are investigated. In order to recover the phase information with few measurements, the sparsity in each underly problem and efficient inversion methods are explored under the framework of compressed sensing. In each phase retrieval technique under study, diffraction during spatial propagation is exploited as an effective and convenient mechanism to uniformly distribute the information about the unknown signal into the measurement space.

Holography is useful to record the scattered field from a sparse distribution of particles; the ability of localizing each particles using compressive reconstruction method is studied. When a thin sample is illuminated with partially coherent waves, the transport of intensity phase retrieval method is shown to be effective to recover the optical path length of the sample and remove the effect of the illumination. This technique is particularly suitable for X–ray phase imaging since it does not require a coherent source or any optical components. Compressive tomographic reconstruction, which makes full use of the priors that the sample consists of piecewise constant refractive indices, are demonstrated to make up missing data. The third technique, known as the phase space tomography (PST), addresses the correlation function recovery problem. Implementing the PST involves measuring many intensity images under spatial propagation. Experimental demonstration of a compressive reconstruction method, which finds the sparse solution by decomposing the correlation function into a few mutually uncorrelated coherent modes, is presented to produce accurate reconstruction even when the measurement suffers from the 'missing cone' problem in the Fourier

domain.

Thesis Supervisor: George Barbastathis
Title: Professor

Acknowledgments

First of all, I would like to express my gratitude to my advisor, Prof. George Barbastathis. He has been not only a great advisor but also a sincere friend throughout all these years. The most important thing I learned from him for research is to always look for answers from the basics. I am also grateful for his encouragement when I lack of confidence, and the many great experiences outside of MIT I got thanks to his support, such as Singapore, and many conferences.

I would like to thank my thesis committee members: Prof. Colin Sheppard, Prof. Ramesh Raskar, and Prof. Peter So for taking time to offer valuable advice on my work.

I would also like to thank the members at the 3D Optical Systems group for all their help throughout my graduate school life. I joined the group with almost no idea about optics, the whiteboard drawings and lab demos from Jose A. Dominguez–Caballero and Nick Loomis are the most memorable lessons during my first semester, and I would like to thank for all the helps from them. Laura Waller introduced me the TIE, and we have been collaborators and good friends ever since, which I feel very grateful. I would like to thank Se Baeh Oh, and Zhengyun Zhang for the insightful discussions on partial coherence and phase space and their contributions to the phase space tomography project; Jonathan Petruccelli for all the helps he gave me in various projects we have been working on together; Justin Lee for his helps on compressed sensing and his hospitalities in many occasions; Chih–Hao Chang, Yuan Luo, Baile Zhang, Dipanjan Bhattacharya, and Yongjin Sung for their helps in both research and career advices; Qin Miao for sharing the X–ray imaging data; Jason Ku for his American culture lessons; Nikhil Vadhavkar, Adam Pan, and Kelli Xu for their helps at the last stage before my defense; Seongkeun Cho for proof–reading the thesis. I also enjoyed and feel thankful for the opportunities to learn about the nano–optics research from Nader Shaar, Tony Nichol, Satoshi Takahashi, Martin Deterre, Chee Wee Tan, Se Young Yang, Hyungryul (Johnny) Choi, and Jeong-gil Kim. Furthermore, I had my best memories with my Chinese friends in this group,

Yi Liu, Chih–Hung (Max) Hsieh, Yubo Duan, Yuanhao (Howard) Huang, Hanhong Gao, Yen–Sheng Lu, Xiaogang (Larry) Liu, Wensheng Chen, Zhi Chen and Hongyi (Thomas) Xu, which I would never forget.

Lastly and specially, I would like to thank my family for their love and support.

Contents

List of Figures

Chapter 1

Introduction

Optical waves have amplitude and phase, but light oscillates too fast for electronic detectors to measure anything more than time–averaged intensities (amplitude). This is unfortunate since the phase of an optical wave can reveal important information about the object. Transparent objects, such as most of unstained biological cells and tissues, do not change the intensity of light passing through them, but introduce phase delays due to variations of thicknesses and/or refractive indices. Knowledge about these phase delays allows physical properties of the object, otherwise 'invisible', to be inferred.

When light is partially coherent, a complete description of the phase of the optical wave requires a more detailed treatment. For instance, light from a lightbulb (a partially coherent source) measured at any single point in space will fluctuate randomly on a time scale of femtoseconds. To gain quantitative information from these fluctuations, we can use statistical tools to specify the correlations for each pair of points in an optical wave [70, 49]. The challenge is that the number of pairs grows much more rapidly than the number of points (e.g. 10^3 points have 10^6 possible combinations of pairs). Another useful and intuitive description of partial coherence is via the superposition of mutually uncorrelated coherent modes [70, 49]: fully coherent light contains only a single coherent mode; while partially coherent light consists of a mixture of coherent modes at any single point in space, as illustrated in Fig. 1-1. Partially coherent illumination is widely used in modern imaging and

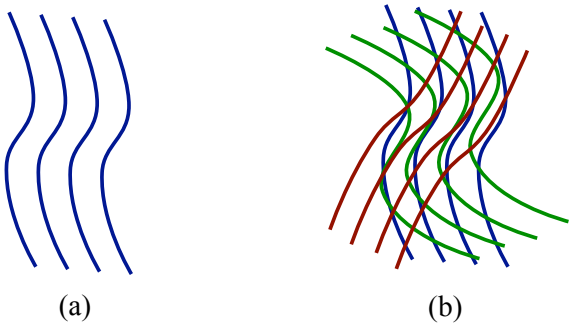


Figure 1-1: (a) A coherent wave has a well–defined amplitude and phase as illustrated by the set of blue curves. (b) Partially coherent light contains a statistical mixture of coherent fields; here, the sets of red, green, and blue curves indicate three different coherent fields that are mixed in space.

manufacturing systems, such as microscopes, telescopes, X–ray imaging systems, and photo–lithography systems, comprehensive characterization of the wave fields may create opportunities for innovation in these areas by exploiting the extra degrees of freedom in partially coherent fields.

1.1 Quantitative phase imaging

Phase contrast was first introduced in the early 1900s [132, 131]. It solves the problem of directly visualizing phase from a single intensity measurement, for which Frits Zernike won the Nobel prize in 1953. However, the greyscale value obtained from a phase contrast image has a nonlinear relationship with the underlying phase, yielding only a qualitative description of the object.

In order to quantitatively recover the phase information, computational methods can be applied to the measured intensity data to account for the measurement process. This idea in principle is applicable to any phase contrast imaging system; in practice, however, a physical measurement that maps the underlying phase to the observed intensity greatly influence the quality of the inversion result. For example, a nonlinear mapping may not be easily inverted, or some spatial frequency components may be lost during the measurement. Here, two types of quantitative phase imaging techniques are investigated: (1) interferometric techniques recover the phase difference

between the unknown field and a known reference field by extracting information from the interference term in a interferogram; (2) propagation based techniques rely on measurements of diffraction during spatial propagation and computational inversion of the propagation operator.

1.1.1 Interferometric/holographic techniques

There are many experimental arrangements for interferometry [117, 95], but the main idea is that an unknown field $f(x,y) = A_f(x,y)\exp[\mathrm{i}\phi(x,y)]$, where A_f is amplitude and ϕ is phase, is interfered with a known reference wave $r(x,y) = A_r(x,y)\exp[\mathrm{i}\phi_0(x,y)]$ and the measured intensity is

$$\begin{aligned} I(x,y) &= |A_f(x,y)\exp(\mathrm{i}\phi(x,y)) + A_r(x,y)\exp(\mathrm{i}\phi_0(x,y))|^2 \\ &= I_f(x,y) + I_r(x,y) + 2A_f(x,y)A_r(x,y)\cos[\phi(x,y) - \phi_0(x,y)], \quad (1.1) \end{aligned}$$

where the first two terms are the intensities of the two fields and the cosine term is the interference term. Depending on the experimental arrangements, ϕ is either directly related to the phase of the underlying object or the wavefront of the propagated field from the object plane. In the first case, ϕ may be recovered by computationally extracting the interference term from I; the phase unwrapping algorithm is often also needed when the phase difference is larger than 2π in the post–processing [117]. For the second case, the recorded intensity is also known as a hologram [95]. The interference term encodes the information about the defocus distance. A computational method that accounts for propagation and other reconstruction artifacts in order to find the 'in–focus' objects is often the concern. A detailed discussion on holographic imaging is in Chapter 2.

When the underlying field is partially coherent, interferometric techniques can also be applied. The measured intensity takes a similar form as the coherent case [Eq. (1.1)], except that the contrast of the interference term is modified by the corre-

lation $\mu(x, y)$ between the two fields, as given by the following expression,

$$I(x, y) = I_f(x, y) + I_r(x, y) + 2\sqrt{I_f(x, y)I_r(x, y)}\mu(x, y). \tag{1.2}$$

Note that often in practical the two fields are originated from different parts of the same field, and in this case μ is the degree of coherence of this field [70, 49]. Equation 1.2 is the principle of interferometric–based coherence state measurement techniques [110, 59, 41, 71].

1.1.2 Propagation based techniques

The propagation based phase retrieval techniques rely on intensity measurements of the unknown field at one or more propagation distances. One such method is the iterative technique, which often uses the image and the far–field diffraction pattern as the input. The unknown phase is estimated by computing the fields at the image and Fourier planes, enforcing constraints based on prior knowledge about the field at each iteration until the algorithm converges to the final solution [48, 44, 74].

An alternative propagation based technique estimates the unknown phase by taking intensity measurements in closely spaced planes. According to energy conservation, the intensity flow in or out of a point in space in the lateral dimension must be compensated by the intensity flow along the longitudinal direction in free space. Under the paraxial approximation, this continuity relation can be written as the transport of intensity equation (TIE) [109]:

$$-\frac{2\pi}{\lambda}\frac{\partial I}{\partial z} = \nabla_\perp \cdot (I\nabla_\perp \phi), \tag{1.3}$$

where $\nabla_\perp$ denotes the gradient in the lateral dimension. The energy flux (the Poynting vector) $\mathbf{F}$ along the transverse direction is related to the gradient of the phase by $\mathbf{F} = I\nabla_\perp\phi$. The phase ϕ is solved given the knowledge about the longitudinal derivative $\partial I/\partial z$, which can be estimated by finite difference method based on the intensity measurements in two or more planes [105, 120]. Extensions to the original

TIE and computational methods to invert the equation will be discussed in detail in Chapters 3–5.

For partially coherent fields, a similar technique known as phase space tomography (PST) allows the reconstruction of the correlation function at a transverse plane from a stack of many propagated intensity measurements [91]. Experiments and reconstruction methods for the PST will be discussed in Chapter 6.

Another method to estimate the energy flux vector is to use an array of lenslets, such as in a Shack–Hartmann sensor [90]. The direction of the vector of the incoming wave averaged over the underlying lenslet's aperture is estimated by the location of the focal spot after the wave propagates to the focal plane of the lens. It turns out that the same setup can be used to estimate the coherence state of partially coherent light. Its properties and limitations will be discussed in detail in Chapter 7.

1.2 Computational imaging and compressed sensing

Quantitative phase imaging falls under the category of 'computational imaging'. Computational imaging refers to the idea that computation is explicitly integrated as a part of the image formation process. As illustrated in Fig. 1-2, comparing to a conventional imaging system [in (a)] in which the desired final image is directly obtained after passing through the optical elements, a computational imaging system [in (b)] first captures a representation of the original object, which is then processed to produce the final image. The goal is no longer to design an optical system to acquire images that replicates the underlying object, but to design a system which optimizes the optics and computation simultaneously to allow information to be transferred more efficiently.

A recent design of computational imaging system is based on the notion that the natural representation of an object is often not the most efficient way to encode information. For example, discrete cosine transform and discrete wavelet transform

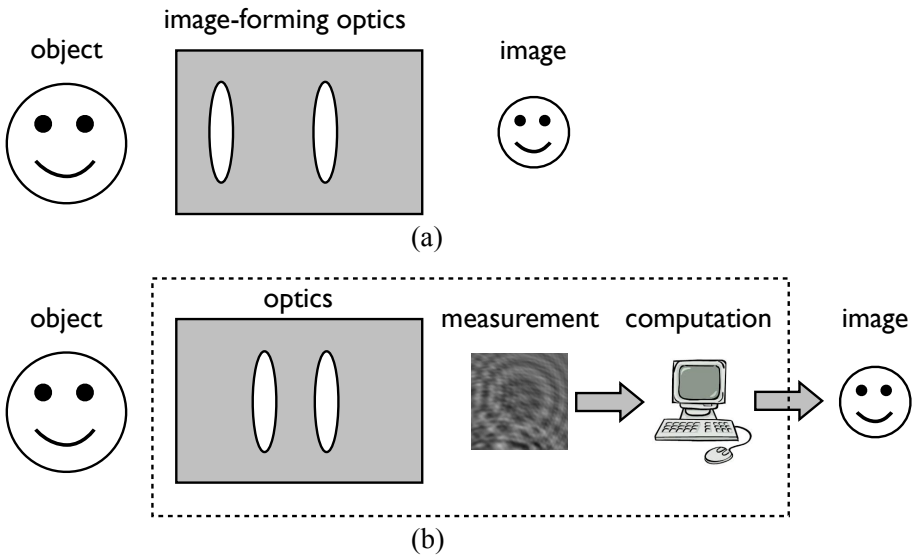


Figure 1-2: Architectures of (a) conventional imaging and (b) computational imaging.

require much fewer coefficients to represent an image of a natural scene, both of which are widely used in modern image compression techniques [108]; human faces can be constructed by just a few 'eigenfaces' [98, 116]. Sensing/imaging models that maximally utilize the knowledge about the 'sparse' representation of an unknown signal are systematically studied by a recent theory, known as compressed sensing or compressive sampling (CS). As contrary to conventional wisdom in data acquisition based upon Nyquist–Shannon theorem, CS theory asserts that one can recover certain signals/images from much fewer measurements without sacrificing reconstruction fidelity [19, 23, 40]. In addition to the 'sparsity' (defined by the number of nonzero coefficients in a specific basis) assumption about the signal of interest, CS also relies on efficient measurement models. The idea is that to make the best use of a measurement, the system should spread out the information about the object evenly. The amount of spreading is measured by a quantity called 'incoherence' parameter (not in the typical sense we assign in Statistical Optics). If the object has a sparse representation in Ψ, an incoherent measurement means that the sensing vectors have an extremely dense representation in Ψ. Under the CS framework, the number of

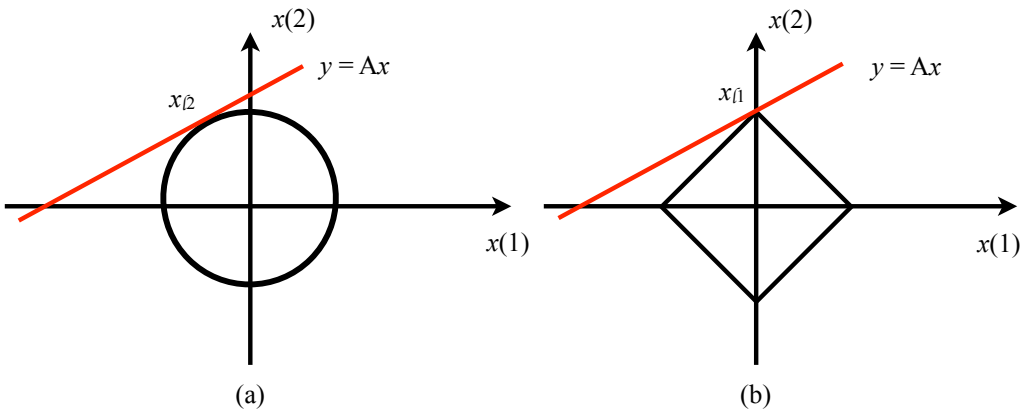


Figure 1-3: Geometry of the (a) ℓ_2 and (b) ℓ_1 recovery for an unknown vector $x = [x(1), x(2)]$. A single measurement $Ax = y$ can be graphically represented by the red line in both figures. Since the cost function for $\|x\|_{\ell_2}$ is a circle, while for $\|x\|_{\ell_1}$ is a diamond, the solution for either problem is found by the intersection between the cost function and the line.

measurements m to ensure successful reconstruction satisfies

$$m \geq C\mu S \log n, \tag{1.4}$$

where n is the length of the unknown object vector, S is the sparsity and measures the compressibility of a signal, μ is the incoherence parameter, and C is a small positive constant. The optimal solution x of a linear system $y = Ax$ with measurement y is found by solving the following ℓ_1 minimization problem

$$\min \|x\|_{\ell_1} \quad \text{subject to} \quad Ax = y, \tag{1.5}$$

where $\|x\|_{\ell_1} = \sum_n |x_n|$. The geometry of ℓ_1 minimization is illustrated by an unknown vector of length two in Fig. 1-3. The solution of ℓ_1 minimization will lie on one of the axes with very high probability, the sparsest solution for this problem is obtained by this method since only one coefficient is nonzero in this case. The application of CS is demonstrated by a simple example in Fig. 1-4. The original unknown signal consists of 10 spikes [in (a)], which means that the signal is sparse in its natural representation. The measurement is carried out by taking samples from the discrete

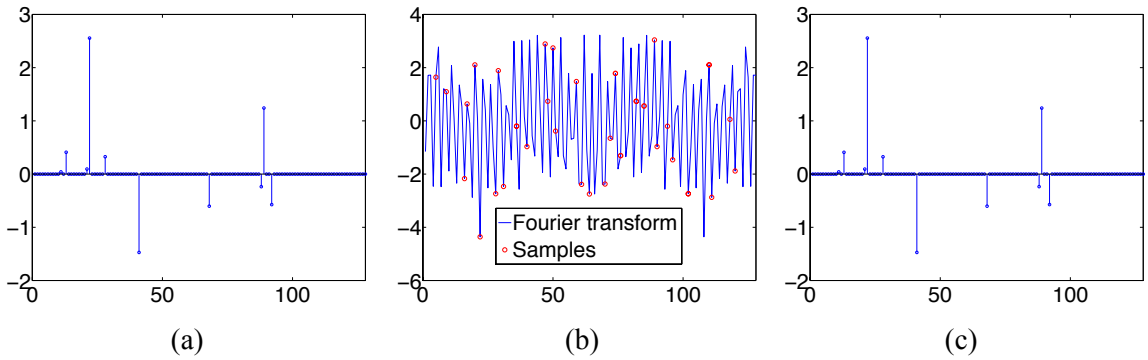


Figure 1-4: Numerical example of compressed sensing. (a) A sparse real–valued signal containing 10 spikes; (b) The Fourier transform (the blue curve) of the signal in (a) is sampled at 40 random locations (as marked by the red circles); (c) The reconstruction by ℓ_1 minimization.

Fourier transform of the signal. 40 samples are taken at random locations [as marked by the red circles in (b)] in the simulation. Taking Fourier transform is an incoherent measurement since a signal contains a single spike in the spatial domain results in an evenly spread–out signal in its Fourier domain. The ℓ_1 minimization accurately recovers the unknown signal, whose result is shown in (c).

A fundamental problem addressed by this thesis is how to best recover phase from a minimal set of measurements. Sparse models and compressive measurement and reconstruction techniques are investigated for different phase retrieval problems. Diffraction during spatial propagation provides a convenient and effective way of evenly mixing the phase information of the object. Efficient compressive inversion methods are designed to recover the phase using much fewer measurements.

1.3 Outline of the thesis

Chapter 2 studies holography under the CS framework. When the complex field scattered from particle–like objects are recorded as an interference pattern, the information about the objects are mixed on the hologram during propagation. The compressive method and its limitation for reconstructing the 3D information about the particles will be discussed.

The TIE is originally derived for a coherent field; the difficulty in extending the TIE from coherent to partially coherent fields arises from the fact that the latter does not have a well–defined phase, as a partially coherent field experiences random fluctuations over time. Chapter 3 presents a new form of the TIE and the experimental demonstration of recovering the optical path length (OPL) of a thin sample under partially coherent illumination.

The phase reconstruction from the TIE often suffers from low–frequency artifacts due to the physics. Chapter 4 describes an iterative reconstruction method, called nonlinear diffusion (NLD) regularization, to mitigate these artifacts under the assumption that the sample is piecewise constant.

Chapter 5 presents the TIE tomography in the X–ray regime using a table–top X–ray microfocus source. A compressive reconstruction method to alleviate the TIE low–frequency artifacts and tomography high–frequency streaking noise from limited sampling angles is demonstrated.

The propagation based correlation function recovery technique, the phase space tomography (PST), for partially coherent fields is described in Chapter 6. Unphysical artifacts often arise in the recovered correlation function using traditional reconstruction method because of inaccessible data in the experiment. A compressive reconstruction method that looks for a sparse solution in the coherent mode superposition representation is demonstrated experimentally for both 1D and 2D fields.

Chapter 7 investigates the use of a lenslet array to measure the coherence state of partially coherent light. Properties and limits of this techniques are studied by numerical examples. Design parameters and constraints are suggested.

Finally, Chapter 8 states conclusions and future work.

Chapter 2

Compressive holography applied to particulate flows

2.1 Introduction

The fundamental problem addressed by holography is that of recording, and later reconstructing, both the amplitude and the phase of a coherent optical wave based on interferometric principle. A hologram is formed by the interference between the scattered field from a coherently illuminated object and a reference wave. Conventional holography is recorded on a photographic film or plate. In the last decades, Digital holography (DH) becomes increasingly popular due to advances in megapixel electronic sensors, e.g. CCD and CMOS, with high spatial resolution and high dynamic range. Compared to traditional holography, DH records holograms on an electronic detector array, which does not have the cumbersome requirement of film positioning, handling, and processing. In addition, with direct access to the hologram in the digital form, various signal processing techniques can be applied to extract useful information from the hologram during reconstruction. DH has proven to be a powerful computational imaging technique in a number of applications. For example, quantitative phase information can be obtained by DH microscopy using the off–axis geometry [73]. The alternative in–line DH geometry, in which both the illumination and the reference waves are served by the same beam, is also widely used due to its

simple optical geometry and better stability. However, since the ghost terms, *e.g.* twin image and halo, cannot be easily removed with a single image, in–line DH has shown to be more suitable for 3D localization applications. For example, it has been used for localizing particulate objects in flow cytometry [31, 96], holographic particle imaging velocimetry (PIV) [38, 85, 61], marine biological imaging [39, 75], and two–phase flows [113, 111]. In these applications, the object under interrogation typically consists of many point–like scatterers, such as particles, blood cells, and air bubbles, sparsely distributed in a 3D volume. The benefit of using the in–line DH is the ability to capture the shape and 3D position information of the scatterers within the 3D volume in a single shot. Since the scatterers are typically some distance away from the detector, the hologram records some defocused information about each scatterer. The task of digital reconstruction is to find the in–focus information.

Traditional reconstruction methods typically consist of two separate steps. First, a focal stack is generated by the back–propagation method (BPM) [94], in which the hologram is convolved with a series of depth dependent free–space propagation point spread functions. An image in this focal stack often contains both in–focus features and diffraction fringes from out–of–focus objects. The goal in the next step is to isolate in–focus objects by using image segmentation techniques. A common approach is to apply an experimentally determined threshold to a focus metric, such as edge sharpness [75], minimum amplitude/intensity [42, 113] and correlation coefficient [130], to each plane in the focal stack. The advantage of this method is that the algorithm can be implemented very efficiently. In practice, however, the values computed from the focus metrics are often sensitive to noise, resulting spurious objects in the segmentation results. Furthermore, the localization accuracy from this method is limited by the Nyquist sampling rate induced by the finite pixel size of the detector [62, 45].

An alternative holographic reconstruction method is to treat the hologram as the output from a linear system with the unknown object as the input signal. Direct inversion will not produce satisfactory results because the equation describing the system is under–determined and it contains infinite numbers of possible solutions.

This solution uniqueness problem can be alleviated by enforcing constraints based on prior knowledge about the objects and then inverted by standard regularization method [9]. It should be noted that the use of focus metric in the previous method can be seen as an approach to utilize priors in order to get an unique interpretation of the object distribution based on the BPM–obtained focus stack; the difference from the method discussed here is that the priors are expressed as a regularization term, and then the reconstruction of a hologram is converted into a constrained optimization problem [100, 101, 36, 13]. Recent studies on the inversion of linear under–determined systems show that it is still possible to obtain a highly accurate solution in such a problem given that the expected solution is known to be sparse in some pre–determined basis; the optimal solution can be found by solving an ℓ_1 minimization problem by standard convex optimization algorithms (compressed sensing) [19, 20, 40]. The application of this theory combined with coherent diffraction in holography is compressive holography [13]. It is shown that the solution to the compressive model is robust to noise; artifacts from the twin–image/halo artifacts are greatly suppressed in the reconstruction [13]. In addition, a recent study on applying compressive holography models to object localization problem has shown orders of magnitude improvement on the localization accuracy in the lateral dimension by finding the sparsity constrained solution [67]. Improvement in the axial localization compared to the BPM–obtained result has also been reported [34].

For the application of imaging particulate flows, an important question is to understand how the quality of the reconstruction is affected by the seeding density of particles. In this chapter, the performance of the compressive and BPM–based methods are compared quantitatively by simulating holograms resulting from scatterings of air bubbles immersed in water with increasing seeding density. The simulation results show that the compressive reconstruction method provides better results over a wide range of seeding densities. The theoretical limit of the seeding density is analyzed in the compressed sensing framework and found in good agreement with the results from the simulation.

2.2 Theory and method

2.2.1 Forward model

The in–line hologram is a record of the interference between a plane reference wave E_{r} of amplitude a and the scattered field E from a 3D object. A schematic diagram of the experimental geometry for the in–line holography is shown in Fig. 2-1. Assuming that the wavefront of the illumination beam is not significantly disturbed due to the presence of the object and $a = 1$ without loss of generality, the intensity g recorded on the detector located at the $z = 0$ plane is

$$\begin{aligned} g(x,y) &= |1 + E(x,y,0)|^2 \\ &= 1 + |E(x,y,0)|^2 + E^*(x,y,0) + E(x,y,0), \end{aligned} \tag{2.1}$$

The object of interest consists of micro–sized dielectric particles sparsely distributed in a transparent medium. If we further assume that the effect due to the multiple scattering between particles is small such that it can be ignored, the total scattered field at the detector plane is

$$E(x,y,0) = \frac{\pi}{\lambda^2} \iiint E_{\mathrm{r}}(x',y',z') f(x',y',z') h(x-x', y-y', 0-z') \mathrm{d}x' \mathrm{d}y' \mathrm{d}z', \tag{2.2}$$

where f is the object function, and h is the depth dependent point spread function for free–space propagation. Under the paraxial approximation,

$$h(x-x', y-y', z-z') = \frac{\exp[\mathrm{i}2\pi(z-z')/\lambda]}{(z-z')} \exp\left\{ \frac{\mathrm{i}\pi}{\lambda(z-z')} [(x-x')^2 + (y-y')^2] \right\}. \tag{2.3}$$

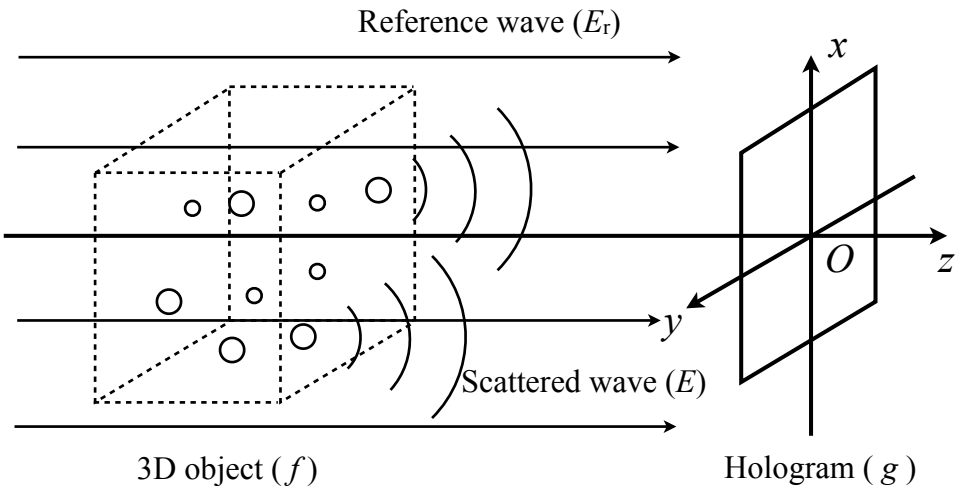


Figure 2-1: Experimental geometry for in–line holography

Together with expression for the plane reference wave $E_{\rm r}(x', y', z') = \exp(\mathrm{i}2\pi z'/\lambda)$, the total scattered field is

$$\begin{aligned} E(x,y,0) &= \frac{\pi}{\lambda^2}\iiint f(x',y',z')\left(-\frac{1}{z'}\right)\exp\left\{-\frac{\mathrm{i}\pi}{\lambda z'}[(x-x')^2+(y-y')^2]\right\}\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'. \\ &= \frac{\mathrm{i}\pi}{\lambda}\iiint\left[\iint f(x',y',z')\exp\{-\mathrm{i}2\pi(ux'+vy')\}\mathrm{d}x'\mathrm{d}y'\right] \\ &\quad \exp\{\mathrm{i}\pi\lambda z'(u^2+v^2)\}\exp\{\mathrm{i}2\pi(ux+vy)\}\mathrm{d}u\mathrm{d}v\mathrm{d}z', \end{aligned} \tag{2.4}$$

where the second equality is due to the Fourier transforming property of convolution integrals. Equation (2.4) is a form of the multi–slice approximation: the integration inside the square brackets with respect to du and dv represents the field contribution from the object at a given depth slice z'; the integration with respect to dz' calculates the superposition of fields from all depth slices.

Next, let us consider the discretization model for Eq. (2.4). Assume the object is discretized into $N_x \times N_y \times N_z$ voxels with lateral spacing Δ, and axial spacing Δ_z. Without loss of generality, assume the number of samples in both lateral dimensions are the same and $N_x = N_y = N$. Equation (2.4) is rewritten as the following

discretized form

$$E_{n_1 n_2} = \frac{\mathrm{i}\pi\Delta_z}{\lambda N^2} \sum_l \left[\sum_p \sum_q \left(\sum_{m_1} \sum_{m_2} f_{m_1 m_2 l} e^{-\mathrm{i}2\pi \frac{pm_1+qm_2}{N}} \right) e^{\mathrm{i}\pi\lambda l\Delta_z(p^2+q^2)\Delta_u^2} e^{\mathrm{i}2\pi \frac{pn_1+qn_2}{N}} \right], \tag{2.5}$$

where the matrix element for a physical variable is denoted by the same letter used in the continuous model with a subscript indicating the index, and the value is determined by $f_{m_1 m_2 l} = f(m_1\Delta, m_2\Delta, l\Delta_z)$ and $E_{n_1 n_2} = E(n_1\Delta, n_2\Delta)$. The spatial frequency sampling pitch is $\Delta_u = 1/(N\Delta)$, with p and q indicating the indices along the two directions.

It is also useful to define a 2D matrix $f^{(l)}$ denoting the l^{th} 2D slice from the 3D object matrix, whose element is defined by $f^{(l)}_{m_1 m_2} = f_{m_1 m_2 l}$. Consider the term enclosed by the square bracket in Eq. (2.5) by rewriting it as the following form

$$\mathbf{E}^{(l)} = H^{(l)} \mathbf{f}^{(l)}, \tag{2.6}$$

where $\mathbf{E}^{(l)}$ is a vector denoting the field contributed from the l^{th} object slice, $\mathbf{f}^{(l)}$ is a vector form of $f^{(l)}$ with its entries defined as $\mathbf{f}^{(l)}_{(N-1)m_1+m_2} = f^{(l)}_{m_1 m_2}$, and

$$H^{(l)} = \mathcal{F}_{\text{2D}}^{-1} Q^{(l)} \mathcal{F}_{\text{2D}} \tag{2.7}$$

is the free–space propagation operator for the l^{th} object slice, which applies the 2D Fourier transform $\mathcal{F}_{\text{2D}}$ of $f^{(l)}$, multiplies a depth dependent quadratic phase function $Q^{(l)} = \exp\{\mathrm{i}\pi\lambda l\Delta_z(p^2+q^2)\Delta_u^2\}$, and then applies the inverse Fourier transform ($\mathcal{F}_{\text{2D}}^{-1}$). The total scattered field is the sum of all the $\mathbf{E}^{(l)}$, and takes the following form

$$\begin{aligned} \mathbf{E} &= \sum_l \mathbf{E}^{(l)} = H\mathbf{f} \\ &\equiv \begin{bmatrix} H^{(1)} & H^{(2)} & \cdots & H^{(N_z)} \end{bmatrix} \begin{bmatrix} \mathbf{f}^{(1)} \\ \mathbf{f}^{(2)} \\ \vdots \\ \mathbf{f}^{(N_z)} \end{bmatrix}, \end{aligned} \tag{2.8}$$

where $\mathbf{E}$ is a vector form of E with its entries defined as $\mathbf{E}_{(N-1)m_1+m_2} = E_{m_1 m_2}$, and H is the propagation operator for all the object slices.

The final form for the discretization model of Eq. (2.1) is

$$\mathbf{g} = \mathbf{1} + |H\mathbf{f}|^2 + H^*\mathbf{f}^* + H\mathbf{f}, \tag{2.9}$$

where $\mathbf{g}$ is a vector containing samples from g, whose entries are defined as $\mathbf{g}_{(N-1)n_1+n_2} = g_{n_1 n_2}$.

2.2.2 Back–propagation method

In the BPM, a focal stack is generated by convolving the hologram with a series of depth dependent point spread functions for free–space propagation.In its discrete form, the back–propagation operator is simply $H^\dagger$, where † denotes the complex conjugate transpose of a matrix. The focal stack $\mathbf{s}$ generated by the BPM from the hologram $\mathbf{g}$ is given by

$$\begin{aligned} \mathbf{s} &= H^\dagger \mathbf{g} \\ &= H^\dagger \mathbf{1} + H^\dagger |H\mathbf{f}|^2 + H^\dagger (H\mathbf{f})^* + H^\dagger H\mathbf{f}. \end{aligned} \tag{2.10}$$

Since $H^{(m)}H^{(n)} = H^{(m+n)}$, it can be easily verified that the last term in Eq. (2.10) is

$$\begin{aligned} H^\dagger H\mathbf{f} &= \begin{bmatrix} I_N & H^{(1)} & \cdots & H^{(N_z-1)} \\ H^{(-1)} & I_N & \cdots & H^{(N_z-2)} \\ \vdots & \vdots & \ddots & \vdots \\ H^{-(N_z-1)} & H^{-(N_z-2)} & \cdots & I_N \end{bmatrix} \begin{bmatrix} \mathbf{f}^{(1)} \\ \mathbf{f}^{(2)} \\ \vdots \\ \mathbf{f}^{(N_z)} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{f}^{(1)} \\ \mathbf{f}^{(2)} \\ \vdots \\ \mathbf{f}^{(N_z)} \end{bmatrix} + \begin{bmatrix} H^{(1)}\mathbf{f}^{(2)} + H^{(2)}\mathbf{f}^{(3)} + \cdots + H^{(N_z-1)}\mathbf{f}^{(N_z)} \\ H^{(-1)}\mathbf{f}^{(1)} + H^{(1)}\mathbf{f}^{(3)} + \cdots + H^{(N_z-2)}\mathbf{f}^{(N_z)} \\ \vdots \\ H^{-(N_z-1)}\mathbf{f}^{(1)} + H^{-(N_z-2)}\mathbf{f}^{(2)} + \cdots + H^{(-1)}\mathbf{f}^{(N_z-1)} \end{bmatrix}, \end{aligned} \tag{2.11}$$

where the object vector $\mathbf{f}$ is recovered as shown in the first term along with the out–of–focus term in the second vector. In addition, the focal stacks are further corrupted by

the halo [the second term in Eq. (2.10)] and twin–image [the third term in Eq. (2.10)] terms.

2.2.3 Compressive reconstruction method

To construct a linear model for compressive reconstruction, consider an unknown vector $\mathbf{x}$ consisting of the real and imaginary parts of $\mathbf{f}$, Eq. (2.9) is rewritten as

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} 2H_{\mathrm{r}} & -2H_{\mathrm{i}} \end{bmatrix} \begin{bmatrix} \mathbf{f}_{\mathrm{r}} \\ \mathbf{f}_{\mathrm{i}} \end{bmatrix} + |H\mathbf{f}|^2 \\ &\equiv A\mathbf{x} + \mathbf{e}, \end{aligned} \tag{2.12}$$

where $\mathbf{x} = \mathbf{f}_{\mathrm{r}} + \mathrm{i}\mathbf{f}_{\mathrm{i}}$ and $A = H_{\mathrm{r}} + \mathrm{i}H_{\mathrm{i}}$. Equation (2.12) is under–determined because the number of unknowns in $\mathbf{f}$ is $2N_z$ times as the number of measurements in $\mathbf{y}$. As a result, the solution is not unique unless prior information is utilized to constrain the solution space. Compressed sensing tries to find a unique solution by enforcing sparsity constraints.

Successful implementation of compressed sensing is conditioned upon two requirements: incoherence of measurements and sparsity [24]. The first requirement does not mean "incoherence" in the typical sense we assign in statistical optics; rather, it means that the information of the unknown vector must be evenly spread over the set of basis vectors that describe it at the detector. Here, we utilize Fresnel diffraction to generate incoherence at the sensing plane. The mixing produced by the Fresnel propagator is not provably optimal, but it is extremely easy to attain by simple free–space propagation [112]. The second requirement is fulfilled by finding a set of basis functions such that the original unknown function can be expressed by a few non–zero coefficients under this basis. We choose total variation (TV) as our sparsity basis, since it prefers solutions with sharp boundaries and the original object can be modeled as small particles with sharp edges in an otherwise empty space.

The sparse solution under the TV basis is obtained by solving the following min-

imization problem,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathrm{TV}} \text{ such that } \mathbf{y} = A\mathbf{x}, \tag{2.13}$$

where $\|\mathbf{x}\|_{\mathrm{TV}}$ is defined as a function of the magnitude of the gradient of the object function according to [92]

$$\|\mathbf{x}\|_{\mathrm{TV}} = \sum_{l} \sum_{m_1} \sum_{m_2} |\nabla x^{(l)}_{m_1 m_2}|, \tag{2.14}$$

where ∇ denotes the finite difference operator in the lateral dimension. In practice, the following regularization problem is often solved instead to account for imperfect measurement [20, 10]

$$\text{minimize } \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \tau \|\mathbf{x}\|_{\mathrm{TV}}, \tag{2.15}$$

where the TV constraint is adapted as a regularizer, and the regularization parameter τ controls its relative weight as compared to the data fitness term $\|\mathbf{y} - A\mathbf{x}\|^2$.

2.3 Numerical simulations

To study the performance of both reconstruction methods, we simulate holograms of air bubbles immersed in water with increasing seeding density. The bubbles are modeled as perfect spheres, the diameters of which are determined randomly from a uniform distribution ranging from 10μm to 20μm. The bubbles are located at random positions within a volume of $512 \times 512 \times 6000 \mu\mathrm{m}^3$; the center of this volume is 6mm away from the hologram plane. The total scattered field is calculated at the hologram plane, consisting of 256×256 pixels with 2μm pixel pitch. Under a plane wave illumination of 632nm wavelength, the scattered field from each individual bubble is calculated based on the Lorentz–Mie theory. Assuming that the effect of multiple scattering between bubbles are negligible, the total scattered field is the linear superposition of the scattered fields from each bubbles. We also assume that

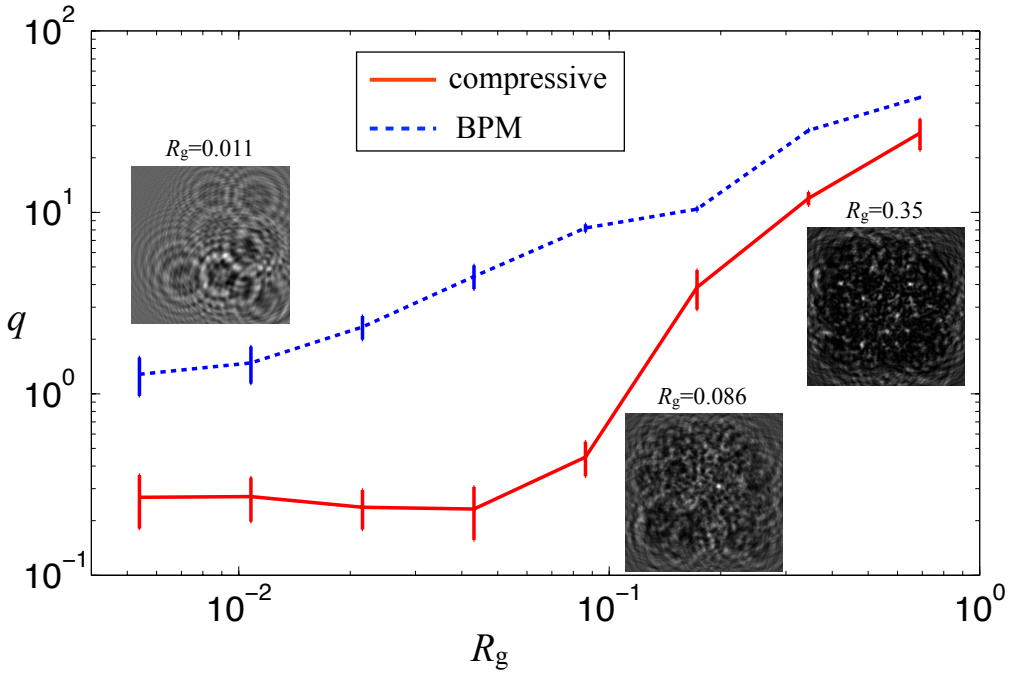


Figure 2-2: The total error q using the compressive reconstruction method (solid red) and BPM-based method (dashed blue) at different seeding densities measured by the geometrical occlusion factor R_g.

the reference wave is an undisturbed plane wave, the hologram is calculated as the interference between the total scattered field and the plane reference wave. Holograms containing 8, 16, 32, 64, 128, 256, 512, 1024 bubbles are calculated. For each data set containing the same number of bubbles, 20 holograms generated from different distributions of bubbles are simulated.

We qualify the seeding density with a dimensionless variable called the geometrical occlusion factor R_g, defined as

$$R_g = \frac{\text{total cross–section area of all bubbles}}{\text{area of the hologram}} \approx \frac{N_p \pi \bar{r}^2}{(N\Delta)^2}, \tag{2.16}$$

where N_p is the number of bubbles, $\bar{r}$ is the mean radius, and R_g measures the fractional area taken up by the parallel projection of all bubbles on the hologram plane.

During the reconstruction using either method, the estimated object function is expressed as a $256 \times 256 \times 30$ matrix corresponding to discretization of the actual

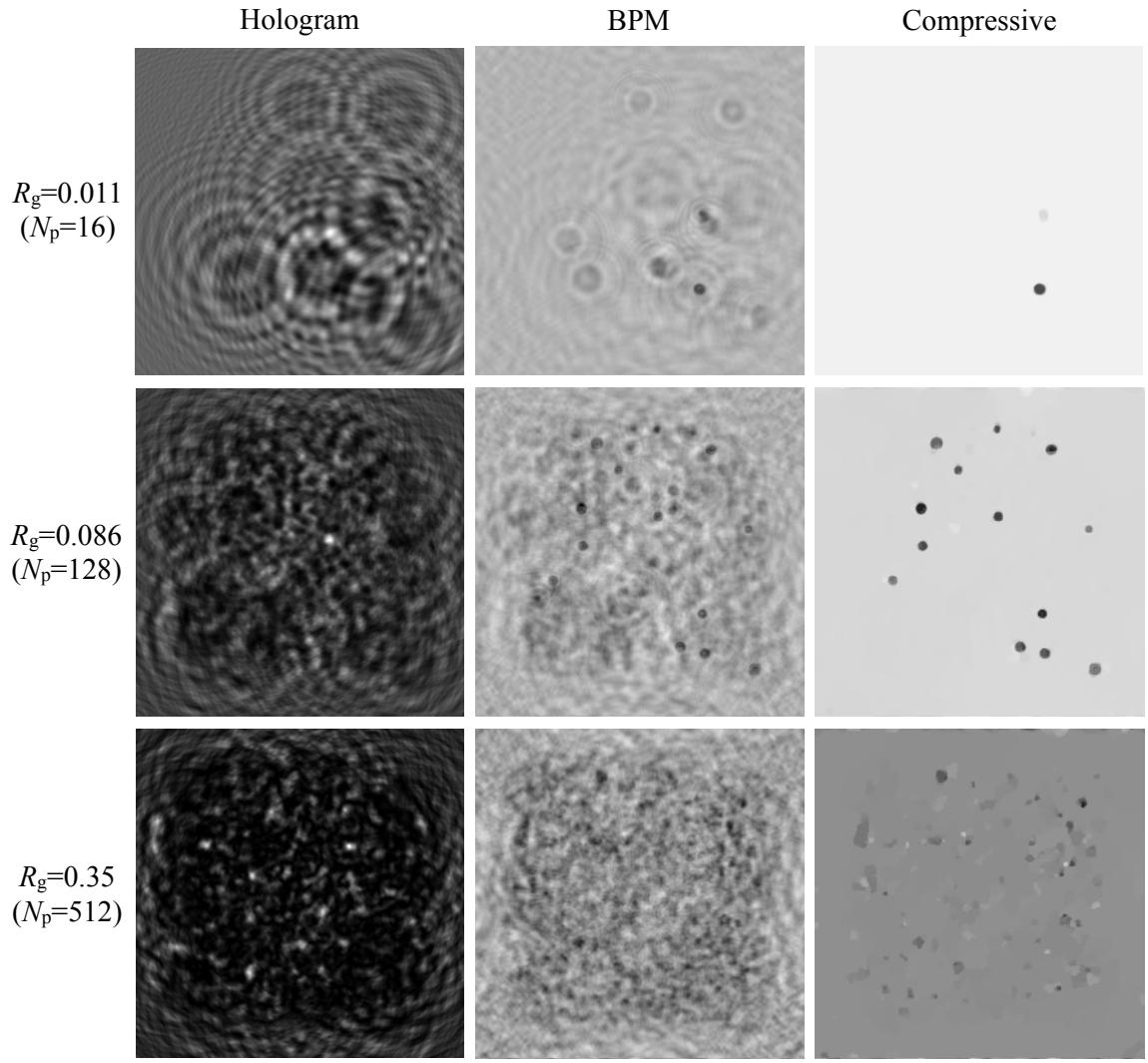


Figure 2-3: Left: Sample holograms; middle: real part of a depth slice from the BPM–obtained focal stack; right: real part of a depth slice from the focal stack estimated by the compressive reconstruction method. First row: $R_g = 0.011$; second row: $R_g = 0.096$; third row: $R_g = 0.35$.

volume with lateral spacing $\Delta = 2\mu$m and axial spacing $\Delta_z = 250\mu$m. The BPM–estimated object is obtained by applying a threshold on the intensity focal stack matrix. Since the diameter of a bubble is much smaller than Δ_z, each bubble is represented as a single disk located at some depth slice in the true object matrix. Note that within the region of a disk where a bubble is located, the reconstructed values from the two methods are different because they have different physical meanings. It is not straightforward to relate the value to some physical quantity, such as the refractive index or the scattering potential, because the Mie scattering is not a linear function of either of these quantities [12]. When comparing the reconstruction results produced by either method with the ground truth, the value within each disk is ignored as long as the disk is correctly detected. The quality of the reconstruction is measured by the number of falsely detected voxels weighted by the value in the original intensity focal stack. The total reconstruction error q is defined as

$$q = \frac{\sum_{\text{all false voxels}} |\text{value of false voxel}|^2}{\sum_{\text{all voxels}} |\text{voxel value of the true object matrix}|^2}; \tag{2.17}$$

The total error q of each reconstruction method is plotted as function of R_{g} in Fig. 2-2. The error of the BPM–based method grows approximately linearly as R_{g} increases when both are plotted in the log–scale. The results from the compressive reconstruction method remain accurate until $R_{\text{g}} > 0.086$ (N_{p}=128). Beyond that value of R_{g}, the error grows rapidly as the seeding density increases. Reconstruction slices from sample holograms with $R_{\text{g}} = 0.011$, $R_{\text{g}} = 0.086$, and $R_{\text{g}} = 0.35$ are shown in Fig. 2-3. A sample hologram at each R_{g} is shown in the first column. It is seen that the quality of the BPM reconstruction degrades as the number of bubbles increases, as shown in the middle column; it is impossible to find a single threshold value to isolate the in–focus bubbles when the degradation is severe. The compressive reconstruction method successfully finds the in–focus bubbles and removes most of the unwanted artifacts in the first two cases, as shown in the third column.

It should be noted that the signal–to–noise ratio (SNR) of a measurement by an in–line hologram decreases as the density of bubbles increases. For an ideal hologram

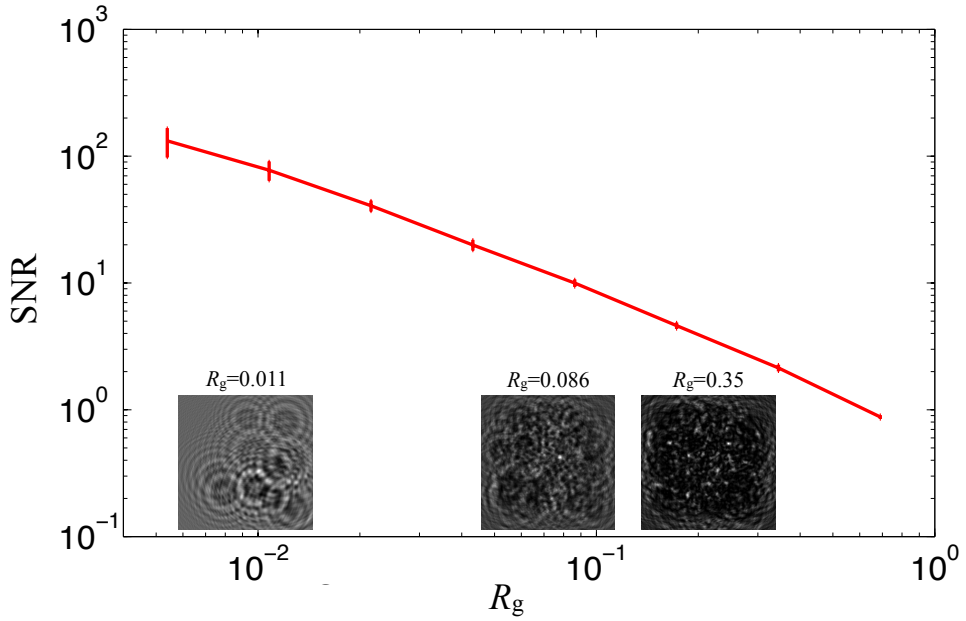


Figure 2-4: The SNR of a hologram decreases as the density of bubbles increases.

with no random measurement noise, the noise is entirely contributed from the halo term, so the SNR is

$$\text{SNR} = \frac{\text{power in the real and twin–image terms}}{\text{power in the halo term}}, \tag{2.18}$$

As R_g increases, the SNR deceases almost linearly when both are plotted in the log–scale), as shown in Fig. 2-4. We expect the qualities of the hologram at high seeding densities are even worse since the effects of multiple scatterings and the deviation from an ideal planar reference wave become severe.

Next, we study the maximum R_g allowed for accurate reconstruction under the compressed sensing framework. Let s denote the "sparsity" of the problem, *i.e.* the number of non–zero coefficients of the unknown expressed in the sparsity basis. Since the TV basis is used in our problem, s can be estimated by the total number of edge voxels for all the bubbles:

$$s \approx N_p 2\pi\bar{r}/\Delta, \tag{2.19}$$

where the average number of edge voxels for a bubble is estimated by the number of voxels on the perimeter of a disk of radius $\bar{r}$. To accurately reconstruct an s–sparse

vector of length N_0, the number of samples M should satisfy [24]

$$M \geq C\mu s \log N_0, \tag{2.20}$$

where C is a small positive constant and μ is another constant called the coherence parameter, which measures the correlation between the measurement matrix and the sparsity basis. In our problem, $M = N^2$, and $N_0 = 2N^2 N_z$. As a result, the number of bubbles N_p that can be successfully reconstructed should satisfy

$$N_\mathrm{p} \leq \frac{N^2 \Delta}{2C\mu\pi\bar{r}\log(2N^2 N_z)}, \tag{2.21}$$

and the corresponding constraint for R_g is

$$R_\mathrm{g} \leq \frac{\bar{r}}{2C\mu\Delta\log(2N^2 N_z)}. \tag{2.22}$$

After substituting the values for all the variables used in the simulation and assuming $C\mu = 1$ [28] in Eq. (2.22), the maximum value of R_g is 0.168.

2.4 Discussion

Our simulation result and theoretical analysis based on compressed sensing both indicate that the maximum geometrical occlusion factor R_g allowed for accurate reconstruction is around 0.086. Equation (2.21) shows that the maximum number of particles that can be accurately reconstructed increases as the number of pixels (space–bandwidth product) of the hologram increases and the size of a particle decreases; the number limit is not very sensitive to the discretization in the axial direction since N_z only affects the results in its log–scale. A nonlinear forward model that properly treats the halo term as part of the useful signal may help to improve the reconstruction.

Chapter 3

Transport of intensity imaging with partially coherent illumination

3.1 Introduction

While interferometric methods, such as holography, are useful for phase measurement for a coherent field, their performance is often limited due to the extreme sensitivity to mechanical stability and environmental conditions and the requirement for a coherent (laser) source and reference beam. In cases where analogous optical elements are difficult to manufacture, such as X–ray [80], electron [58], and neutron imaging [2], the propagation–based (noninterferometric) techniques, which relies on measurements of diffraction patterns of the unknown field upon propagation, are more suitable. The transport of intensity imaging is a non–iterative propagation–based technique, which allows the recovery of the unknown phase by solving a linear equation, *i.e.* the transport of intensity equation (TIE) by measuring intensities in closely spaced planes [109, 105, 120]. It is easy to implement since the measurements can be taken by simply displacing the detector or the sample along the optical axis.

The TIE is originally derived for a coherent field; the difficulty in extending the TIE from coherent to partially coherent fields arises from the fact that the latter does not have a well–defined phase, as a partially coherent field experiences random fluctuations over time. Although it is possible to decompose a partially coherent

field into a sum of fully coherent fields through a process known as coherent mode decomposition [127], it is shown that a weighted average of the phases of all the modes does not match well with the TIE–obtained phase [135]. An alternative interpretation of the TIE–obtained phase is by a scalar phase whose gradient is proportional to the time–averaged transverse component of the Poynting vector [84]. In this chapter, we clarify the TIE–obtained phase in terms of the optical path length (OPL) of a thin sample and show that the OPL can be accurately recovered even with partially coherent illumination from TIE measurement, the effects of the illumination may be removed provided its transverse flux vector is first characterized by another TIE measurement. The remainder of this chapter is structured as follows: the TIE with coherent illumination will be reviewed in Section 3.2; the partially coherent TIE (PC–TIE) for recovering the OPL of a thin sample and the experimental verification of this equation will be presented in Section 3.3.

3.2 TIE with coherent illumination

3.2.1 Theory

Following the original work of Teague [109], TIE for a monochromatic coherent field can be derived directly from the paraxial wave equation. Consider a monochromatic coherent paraxial field of the form

$$U(\mathbf{x}, z) = \sqrt{I_{\mathrm{c}}(\mathbf{x}, z)}\, \exp[\mathrm{i}\phi(\mathbf{x}, z)], \tag{3.1}$$

where I_{c} is the intensity, ϕ is the phase, $\mathbf{x}$ denotes a 2D vector specifying the position in a plane transverse to the optical axis along z, and the subscript c denotes the properties of coherent light. The propagation of the field $U(\mathbf{x}, z)$ satisfies the following paraxial wave equation

$$\frac{\partial U(\mathbf{x}; z)}{\partial z} = \frac{\mathrm{i}}{2k} \nabla_{\perp}^2 U(\mathbf{x}; z), \tag{3.2}$$

where $\nabla_\perp$ is the 2D gradient operator in the transverse plane. To derive the TIE, consider the propagation of intensity by differentiating I_c with respect to z, and since $I_\mathrm{c} = U^*U$,

$$\frac{\partial I_\mathrm{c}(\mathbf{x};z)}{\partial z} = \frac{\mathrm{i}}{2k}\nabla_\perp \cdot [U^*(\mathbf{x};z)\nabla_\perp U(\mathbf{x};z) - U(\mathbf{x};z)\nabla_\perp U^*(\mathbf{x};z)]. \tag{3.3}$$

Equation 3.3 is an expression of energy conservation under paraxial propagation, and the intensity at a point measures the local energy. The local intensity changes as the field propagates along the optical axis, which is compensated by the energy flow in or out of that point from neighboring points; this transverse flow of energy is measured by the transverse energy flux density $\mathbf{F}_\mathrm{c}$:

$$\mathbf{F}_\mathrm{c}(\mathbf{x};z) = -\frac{\mathrm{i}}{2k}[U^*(\mathbf{x};z)\nabla_\perp U(\mathbf{x};z) - U(\mathbf{x};z)\nabla_\perp U^*(\mathbf{x};z)]. \tag{3.4}$$

Substitution of the definition of $U(\mathbf{x};z)$ relates $\mathbf{F}_\mathrm{c}$ to the underlying intensity and phase of the field as

$$\mathbf{F}_\mathrm{c}(\mathbf{x};z) = \frac{1}{k} I_\mathrm{c}(\mathbf{x};z)\nabla_\perp \phi(\mathbf{x};z). \tag{3.5}$$

Finally, the TIE, which relates the phase to variations in the propagation of intensity, is given by

$$\frac{\partial I_\mathrm{c}(\mathbf{x};z)}{\partial z} = -\frac{1}{k}\nabla_\perp \cdot [I_\mathrm{c}(\mathbf{x};z)\nabla_\perp \phi(\mathbf{x};z)]. \tag{3.6}$$

3.2.2 Validity of the TIE

The TIE states that the phase distribution on a plane of constant z can be recovered by solving a partial differential equation (PDE) with the knowledge of the intensity derivative along the optical axis. However, since the intensity derivative cannot be measured directly, finite difference methods are used to approximate the derivative from two or more defocused images in practice [120]. For example, the central difference scheme can be implemented by capturing two defocused images at planes located symmetrically before and after the focal plane, (the focal plane is assumed to be located at $z = 0$ without loss of generality), to estimate the derivative with third

order accuracy:

$$\frac{\partial I_{\mathrm{c}}(\mathbf{x};0)}{\partial z} = \frac{I_{\mathrm{c}}(\mathbf{x},\Delta z) - I_{\mathrm{c}}(\mathbf{x},-\Delta z)}{2\Delta z} + O\left(\frac{\partial^3 I_{\mathrm{c}}(\mathbf{x};0)}{\partial z^3}\right), \tag{3.7}$$

where Δz is some finite defocus distance.

To understand the limits in which TIE is still valid when implemented with finite defocus distance, consider the intensity measured at the plane which is Δz away from the focal plane. The Fourier transform $\tilde{I}_{\mathrm{c}}(\mathbf{u};\Delta z)$ of this intensity can be written as [54]

$$\tilde{I}_{\mathrm{c}}(\mathbf{u};\Delta z) = \int U\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2};0\right) U^*\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2};0\right) \exp(-\mathrm{i}2\pi\mathbf{x}\cdot\mathbf{u})\,\mathrm{d}^2\mathbf{x}, \tag{3.8}$$

where $\mathbf{u}$ is the spatial frequency variable. Assume that the field $U(\mathbf{x};0)$ at the focal plane varies slowly with respect to changes in $\mathbf{u}$, in which case the field can be approximated by the first order Taylor expansion term as

$$U\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2};0\right) = U(\mathbf{x};0) + \frac{1}{2}\lambda\Delta z\mathbf{u}\cdot\nabla_{\perp}U(\mathbf{x};0). \tag{3.9}$$

Substituting Eq. (3.9) into Eq. (3.8) and keeping only the first order terms with respect to Δz yields

$$I_{\mathrm{c}}(\mathbf{x};\Delta z) = I_{\mathrm{c}}(\mathbf{x};0) - \frac{\Delta z}{k}\nabla_{\perp}\cdot[I_{\mathrm{c}}(\mathbf{x};0)\nabla_{\perp}\phi(\mathbf{x};0)]. \tag{3.10}$$

Equation 3.10 can be treated as a forward–difference approximation of the TIE, and it is valid as long as

$$|\lambda\Delta z\mathbf{u}|^2 \ll 1. \tag{3.11}$$

The condition in (3.11) implies that the high spatial frequency components of the field can only be captured with a relatively small defocus distance so that the condition still holds; whereas low spatial frequency components tolerate a much larger defocus distance. When Δz is large, the high spatial frequency components of the field produce diffraction fringes, which violate the linearity assumption inherent in the TIE.

The loss of high spatial frequency information results in a low–pass filtering effect on both the measured data and the recovered phase. A small Δz avoids loss of high frequency information, however, may result in measurement with low signal–to–noise ratio in practice, as the differences between the defocused images may be too small as compared to the noise level. An optimal Δz is object and system dependent, and needs to be found experimentally.

In many practical situations, the concern is to recover the optical properties of a sample rather than the phase of the incident field. Consider a thin sample characterized by the following transmission function

$$g(\mathbf{x}) = \sqrt{T(\mathbf{x})}\exp[\mathrm{i}\psi(\mathbf{x})], \tag{3.12}$$

where $T(\mathbf{x})$ is the transmittance of the sample, which characterizes the total attenuation property of the sample; ψ is the phase of the sample, which is equal to the product of the OPL though the sample with the wavenumber of the illumination, $\psi(\mathbf{x}) = k\mathrm{OPL}$. If the sample consists of homogenous material of known refractive index n, the physical thickness L of the sample is related to the OPL by $L = \mathrm{OPL}/n$. For a sample consisting of multiple materials with different refractive indices, tomographic TIE measurement is required to fully recover the 3D index distribution; experimental results of TIE tomography will be discussed in Chapter 5. Here, the thin sample approximation will be used throughout this chapter.

When a sample is illuminated by a plane wave with constant intensity I_0, the phase of the field at the plane immediately after the sample is entirely due to the sample's phase. The transverse flux vector at this plane is fully characterized by the absorption and phase of the sample:

$$\mathbf{F}_{\mathrm{c}}(\mathbf{x}; z) = \frac{1}{k} I_0 T(\mathbf{x}) \nabla_{\perp} \psi(\mathbf{x}; z). \tag{3.13}$$

A special case of practical interest is when the sample is "pure–phase" as $T(\mathbf{x}) = 1$. In this case, the TIE can be simplified by taking the intensity term I_0 out of the

gradient operator in Eq. (3.6) and rearranged to

$$-k\frac{1}{I_0}\frac{\partial I_{\mathrm{c}}(\mathbf{x};z)}{\partial z} = \nabla_{\perp}^2\psi(\mathbf{x};z). \tag{3.14}$$

Equation 3.14 is a Poisson equation, stating that the change of intensity at small defocus is proportional to the Laplacian of the phase of the sample. When the sample contains both amplitude and phase variations, the full solution of TIE requires to solve two Poisson equations, as described below. First, we introduce an auxiliary function $\Gamma(\mathbf{x};z)$ as the solution of the following equation,

$$\nabla_{\perp}\Gamma(\mathbf{x};z) = I_0T(\mathbf{x})\nabla_{\perp}\psi(\mathbf{x};z). \tag{3.15}$$

Substitution of Eq. (3.15) into Eq. (3.6) gives

$$\nabla_{\perp}^2\Gamma(\mathbf{x};z) = -k\frac{\partial I_{\mathrm{c}}(\mathbf{x};z)}{\partial z}, \tag{3.16}$$

which is a Poisson equation which can be solved for $\Gamma(\mathbf{x};z)$. Next, the second Poisson equation relating the phase can be obtained by rearranging Eq. (3.15) to get

$$\nabla_{\perp}\left[\frac{\nabla_{\perp}\Gamma(\mathbf{x};z)}{I_0T(\mathbf{x})}\right] = \nabla_{\perp}^2\psi(\mathbf{x};z). \tag{3.17}$$

3.3 TIE with partially coherent illumination

3.3.1 Theory

A partially coherent field does not have a well defined phase since the field at any point in the space experiences statistical fluctuations over time. To fully characterize a partially coherent field, second–order correlations between pairs of points need to be used. The correlation function for a stationary quasi–monochromatic spatially partially coherent field on a plane of constant z is the mutual intensity

$$J(\mathbf{x}_1,\mathbf{x}_2;z) = \langle U(\mathbf{x}_1;z)U^*(\mathbf{x}_2;z)\rangle\,, \tag{3.18}$$

where $\langle\cdot\rangle$ denotes the ensemble average over a statistical ensemble of realizations of the field, $\{U(\mathbf{x};z)\}$. The intensity I_{p} of this field is related to J according to

$$I_{\mathrm{p}}(\mathbf{x},z) = J(\mathbf{x},\mathbf{x};z) = \langle U(\mathbf{x};z)U^{*}(\mathbf{x};z)\rangle\,, \tag{3.19}$$

where the subscript p denotes the properties of partially coherent light. Note that J is a complex–valued correlation function and defined over the 4D space spanned by the pair of 2D position vectors, $\{\mathbf{x}_{1,2}\}$. Therefore, although one can still measure a focal stack of intensities and apply the TIE in order to reconstruct a 2D function in this plane, the recovered "phase" clearly does not provide the phase of J.

Since a partially coherent field can be represented in terms of an ensemble of coherent fields, each of which satisfies the paraxial wave equation, the propagation equation of intensity can be obtained by differentiating I_{p} with respect to z, with the substitution of Eqs. (3.2) into (3.19), yielding the PC–TIE:

$$\frac{\partial I_{\mathrm{p}}(\mathbf{x};z)}{\partial z} = -\nabla_{\perp}\cdot\mathbf{F}_{\mathrm{p}}(\mathbf{x};z), \tag{3.20}$$

where $\mathbf{F}_{\mathrm{p}}(\mathbf{x};z)$ is the transverse flux vector of partially coherent field, and $\mathbf{F}_{\mathrm{p}}(\mathbf{x};z)$ is defined as [70]

$$\begin{aligned}\mathbf{F}_{\mathrm{p}}(\mathbf{x};z) &= -\frac{\mathrm{i}}{2k}\left\langle U^{*}(\mathbf{x};z)\nabla_{\perp}U(\mathbf{x};z) - U(\mathbf{x};z)\nabla_{\perp}U^{*}(\mathbf{x};z)\right\rangle \\ &= -\frac{\mathrm{i}}{2k}\nabla_{\mathbf{x}'}J\left(\mathbf{x}+\frac{\mathbf{x}'}{2},\mathbf{x}-\frac{\mathbf{x}'}{2};z\right)\Bigg|_{\mathbf{x}'=0}.\end{aligned} \tag{3.21}$$

Equation 3.20 is the analogous energy conservation equation for partially coherent field under paraxial propagation, which states that the transverse flux can be characterized given the knowledge of intensity derivatives along the propagation direction.

Since $\mathbf{F}_{\mathrm{p}}(\mathbf{x};z)$ is a vector over a transverse plane, it can be uniquely decomposed into a curl–free component and a divergence–free component via the Helmholtz decomposition [84]:

$$\mathbf{F}_{\mathrm{p}}(\mathbf{x};z) = \nabla_{\perp}\phi_{\mathrm{s}} + \nabla_{\perp}\times\mathbf{A}(\mathbf{x}). \tag{3.22}$$

Substitution of Eq. (3.22) into Eq. (3.20) results in a Poisson's equation for ϕ_{s}:

$$\frac{\partial I_{\mathrm{p}}(\mathbf{x};z)}{\partial z} = -\nabla_{\perp}^{2}\phi_{\mathrm{s}}, \tag{3.23}$$

which can be solved for specified boundary conditions, yielding $\mathbf{F}_{\mathrm{p}}$ uniquely up to the addition of $\nabla_{\perp} \times \mathbf{A}(\mathbf{x})$. Note that $\mathbf{F}_{\mathrm{p}}$ retrieved by this procedure has a similar form as the coherent flux in Eq. (3.5), in which the gradient of the scalar "phase" function $\nabla_{\perp}\phi_{\mathrm{s}}$ measures the rate of the transverse energy flow as the field propagates. The coherent case can be treated as a special case of this result where the scalar phase function equals to the phase of the complex field up to a multiplicative constant. While the complex field representation for a coherent light can be fully determined with the knowledge of $\mathbf{F}_{\mathrm{c}}$ and I_{c}; it is not sufficient to characterize the mutual intensity with only the knowledge of $\mathbf{F}_{\mathrm{p}}$ and I_{p} for a partially coherent light expect for trivial cases.

In many practical applications, it is only important to characterize the optical properties of a sample rather than the full coherence state of the incident field. It will be shown in the following that it is still possible to recover the optical thickness of a sample by modifying the PC–TIE to account for the partially coherent illumination. A thin sample, described by Eq. (3.12), is illuminated by a partially coherent field with mutual intensity $J_{\mathrm{i}}(\mathbf{x}_1, \mathbf{x}_2; z)$ and intensity $I_{\mathrm{i}}(\mathbf{x}; z)$, where the subscript i denotes the properties of the illumination. At the plane immediately after the sample, the mutual intensity is given by

$$J_{\mathrm{t}}(\mathbf{x}_1, \mathbf{x}_2) = J_{\mathrm{i}}(\mathbf{x}_1, \mathbf{x}_2; z) g(\mathbf{x}_1) g^{*}(\mathbf{x}_2) = \left\langle U_{\mathrm{i}}(\mathbf{x}_1) g(\mathbf{x}_1) U_{\mathrm{i}}^{*}(\mathbf{x}_2) g^{*}(\mathbf{x}_2) \right\rangle, \tag{3.24}$$

where the subscript t denotes the properties of the field immediately after the sample, and the intensity at the same plane is

$$I_{\mathrm{t}}(\mathbf{x}; z) = I_{\mathrm{i}}(\mathbf{x}; z) T(\mathbf{x}). \tag{3.25}$$

The propagation of the intensity $I_{\mathrm{t}}(\mathbf{x}; z)$ follows the PC–TIE in Eq. (3.20); the trans-

verse energy flux vector $\mathbf{F}_\mathrm{t}$ can be obtained by substituting $U_\mathrm{i}(\mathbf{x};z)g(\mathbf{x})$ for $U(\mathbf{x};z)$ in Eq. (3.21), yielding

$$\mathbf{F}_\mathrm{t}(\mathbf{x};z) = T(\mathbf{x})\mathbf{F}_\mathrm{i}(\mathbf{x};z) + \frac{1}{k} I_\mathrm{t}(\mathbf{x};z)\nabla_\perp \psi(\mathbf{x}). \tag{3.26}$$

Equation 3.26 shows that two terms contribute to the total flux $\mathbf{F}_\mathrm{t}$: the first term characterizes the flux of the illumination modified by the absorption of the sample; the second term characterizes the flux due to the presence of the sample, which takes a similar form as the one for the coherent case in Eq. (3.13). Substituting the expression of $\mathbf{F}_\mathrm{t}(\mathbf{x};z)$ into Eq. (3.20) yields the PC–TIE for the field immediately after the thin sample:

$$\frac{\partial I_\mathrm{t}(\mathbf{x};z)}{\partial z} + \nabla_\perp \cdot [T(\mathbf{x})\mathbf{F}_\mathrm{i}(\mathbf{x};z)] = -\frac{1}{k}\nabla_\perp \cdot [I_\mathrm{t}(\mathbf{x};z)\nabla_\perp \psi(\mathbf{x})]. \tag{3.27}$$

Equation 3.27 shows that the phase ψ of a sample can still be recovered by conducting TIE measurements even if the illumination is partially coherent. The only difference from the coherent TIE is that the term $\nabla_\perp \cdot [T(\mathbf{x})\mathbf{F}_\mathrm{i}(\mathbf{x};z)]$ needs to be properly accounted for. To do that, we assume the illumination flux is divergence free so it is fully described by a scalar phase by $\mathbf{F}_\mathrm{i} = \nabla_\perp \phi_\mathrm{s}$. ϕ_s at the sample plane can be calibrated by first taking intensity focal stacks without the sample in place and then recovered by solving Eq. (3.23). Next, the sample's phase can be retrieved by taking the second set of TIE measurements with the sample in place and solving Eq. (3.27).

Note that when $\mathbf{F}_\mathrm{i}(\mathbf{x};z) = 0$, Eq. (3.27) simplifies to

$$\frac{\partial I_\mathrm{t}(\mathbf{x};z)}{\partial z} = -\frac{1}{k}\nabla_\perp \cdot [I_\mathrm{t}(\mathbf{x};z)\nabla_\perp \psi(\mathbf{x})], \tag{3.28}$$

which has the exact same form as the coherent TIE. This implies that the optical thickness of a sample can be retrieved by taking a single set of TIE measurement as long as the illumination does not have transverse flux at the sample plane. This is approximately the case when the sample is illuminated by Köhler illumination in a standard bright–field microscope [105].

Another special case is when the sample is pure–phase, where Eq. (3.27) reduces

to the simpler form

$$\frac{\partial I_{\mathrm{t}}(\mathbf{x};z)}{\partial z} - \frac{\partial I_{\mathrm{i}}(\mathbf{x};z)}{\partial z} = -\frac{1}{k}\nabla_{\perp}\cdot[I_{\mathrm{i}}(\mathbf{x};z)\nabla_{\perp}\psi(\mathbf{x})], \tag{3.29}$$

where Eq. (3.20) is used to express the left–hand–side as the difference between the two intensity derivative measurements taken with and without the sample in place.

The result of Eq. (3.27) assumes that intensity derivative can be measured accurately. In reality, one is limited to estimate the derivative from finite difference measurements. The validity of Eq. (3.27), taking the finite displacement distance along z into account, can be examined as follows. The Fourier transform $\tilde{I}(\mathbf{u};\Delta z)$ of the intensity taken at the plane defocused by Δz from the focal plane (assumed at $z=0$) can be written as [56]

$$\begin{aligned}\tilde{I}(\mathbf{u};\Delta z) &= \int J_{\mathrm{i}}\left(\mathbf{x}+\frac{\lambda\Delta z\mathbf{u}}{2},\mathbf{x}-\frac{\lambda\Delta z\mathbf{u}}{2}\right)\sqrt{T\left(\mathbf{x}+\frac{\lambda\Delta z\mathbf{u}}{2}\right)T\left(\mathbf{x}-\frac{\lambda\Delta z\mathbf{u}}{2}\right)} \\ &\quad\times\exp\left\{\mathrm{i}\left[\phi\left(\mathbf{x}+\frac{\lambda\Delta z\mathbf{u}}{2}\right)-\phi\left(\mathbf{x}-\frac{\lambda\Delta z\mathbf{u}}{2}\right)\right]\right\} \\ &\quad\times\exp(-\mathrm{i}2\pi\mathbf{x}\cdot\mathbf{u})\,\mathrm{d}^2\mathbf{x},\end{aligned} \tag{3.30}$$

where $\mathbf{u}$ is the spatial frequency variable. Assume that all functions of $\lambda\Delta z\mathbf{u}$ vary slowly with respect to changes in $\mathbf{u}$ and

$$|\lambda\Delta z\mathbf{u}|^2 \ll 1, \tag{3.31}$$

Eq. (3.30) can be simplified by taking up to the first order terms with respect to $\lambda\Delta z\mathbf{u}$, yielding

$$\begin{aligned}\tilde{I}(\mathbf{u};\Delta z) &= \int[T(\mathbf{x})I_{\mathrm{i}}(\mathbf{x};0)+\mathrm{i}2\pi\Delta z T(\mathbf{x})\mathbf{u}\cdot\mathbf{F}_{\mathrm{i}}(\mathbf{x};0)+\mathrm{i}\lambda\Delta z I_{\mathrm{t}}(\mathbf{x};0)\mathbf{u}\cdot\nabla_{\perp}\psi(\mathbf{x})] \\ &\quad\times\exp(-\mathrm{i}2\pi\mathbf{x}\cdot\mathbf{u})\,\mathrm{d}^2\mathbf{x}.\end{aligned} \tag{3.32}$$

Taking inverse Fourier transform on both sides of Eq. (3.32) yields:

$$I(\mathbf{x};\Delta z) = T(\mathbf{x})I_{\mathrm{i}}(\mathbf{x};0) - \Delta z\nabla_{\perp}\cdot[T(\mathbf{x})\mathbf{F}_{\mathrm{i}}(\mathbf{x};0)] - \frac{\Delta z}{k}\nabla_{\perp}\cdot[I_{\mathrm{t}}(\mathbf{x};0)\nabla_{\perp}\psi(\mathbf{x})]. \quad (3.33)$$

Equation 3.33 is a good approximation to Eq. (3.27) as long as the condition in (3.31) is satisfied. The spatial frequency dependence in this condition implies that the validity of PC–TIE with finite difference implementation depends not only on spatial features in the sample, but also on the coherence state of the incident light.

3.3.2 Experimental verification of the PC–TIE

The experimental arrangement used to test the PC–TIE is illustrated in Fig. 3-1. An LED with central wavelength $\bar{\lambda}$ =620nm and bandwidth 20nm is placed before a pinhole with a diameter $d = 500\mu$m; a diffuser is placed in between to create a uniform angular distribution of light. We use a Köhler illumination system, by placing a condenser lens at one focal length (f =50mm) away from the pinhole, to create a uniform intensity distribution after the lens. If we assume that the lens is perfect and neglect the effect of finite apertures, the field after the lens consists of plane waves propagating at different directions. If we further assume the primary source is incoherent with uniform intensity I_0, which is a good approximation in our case, then the mutual intensity at any arbitrary plane z after the lens can be obtained according to the van–Cittert Zernike theorem as

$$J_{\mathrm{i}}(\mathbf{x}_1,\mathbf{x}_2,z) = \frac{\pi I_0 d}{k|\mathbf{x}_2-\mathbf{x}_1|}J_1\left(\frac{kd|\mathbf{x}_2-\mathbf{x}_1|}{2f}\right), \quad (3.34)$$

where $k = 2\pi/\bar{\lambda}$. Note that $\mathbf{F}_{\mathrm{i}} = 0$ due to the symmetry in this illumination. In order to produce a partially coherent field with non–zero flux, a field modulation mask is placed in the beam prior to the sample being imaged. The field modulation mask is an MIT logo, and the phase object being imaged is an MIT beaver. Both samples were made by etching into a glass microscope slide ($n \approx 1.5$) to a depth of approximately 100 nm. The object is relayed onto the detector (Aptina MT9P031)

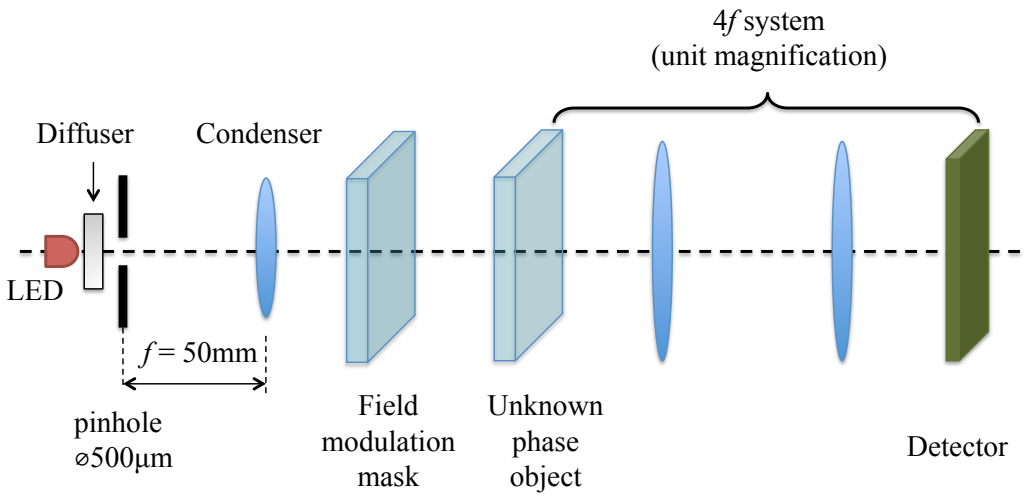


Figure 3-1: Imaging system used for the partially coherent TIE

by a 4f system with unit magnification. The detector is placed on a linear motion stage (Zaber T–LSM025A), which can be moved along the optical axis. For each set of TIE measurements, an under–focused, over–focused and in–focus images were taken with axial displacement in steps of 250μm. For display, all intensities shown in Fig. 3-2 are normalized so that the maximum value is unity. The TIE was solved by the Fourier domain method with the adoption of Tikhonov regularization to suppress the low–frequency artifacts in the reconstruction [83].

To provide a baseline measurement for comparison, measurements were first taken with only the phase object in place with quasi–uniform illumination, as shown in Figs. 3-2(a–b). The thickness of the sample is recovered by solving the simplified PC–TIE in Eq. (3.28), and the result are shown in Fig. 3-2(c).

In order to test the PC–TIE when the object is illuminated by non–uniform partially coherent light, measurements were taken with both the field modulation mask and the object in place, as shown in Figs. 3-2(g–h). As already explained in the theory, to correctly reconstruct the phase of the object in this case, the transverse flux vector $\mathbf{F}_\mathrm{i}$ of the illumination needs to be calibrated alone by another TIE measurement. After taking the intensity measurements shown in Figs. 3-2(d–e), the scalar phase ϕ_i shown in Fig. 3-2(f) is reconstructed by solving Eq. (3.23); $\mathbf{F}_\mathrm{i}$ is determined by $\nabla_\perp \phi_\mathrm{s}$, assuming zero contribution from the divergence–free component. The result

of applying the coherent TIE, without accounting for the illumination, is shown in Fig. 3-2(i), which clearly mixes illumination flux with sample's phase. The sample's thickness reconstructed from Eq. (3.29), assuming a pure–phase sample, is shown in Fig. 3-2(j), while the result of applying Eq. (3.27) is shown in Fig. 3-2(k). The difference between Figs. 3-2(j) and (k) is minimal, due to the fact that the in–focus intensity of the sample exhibits minimal variation, *i.e.* it is nearly pure–phase.

3.4 Conclusion

The PC–TIE thats allows recovery of the OPL of a thin sample and accounts for the effect of partially coherent illumination is demonstrated. For illuminations with transverse symmetries, this equation reduces to the same form as the coherent TIE. Otherwise, two sets of TIE measurements are required in order to account for the nonuniform illumination flux.

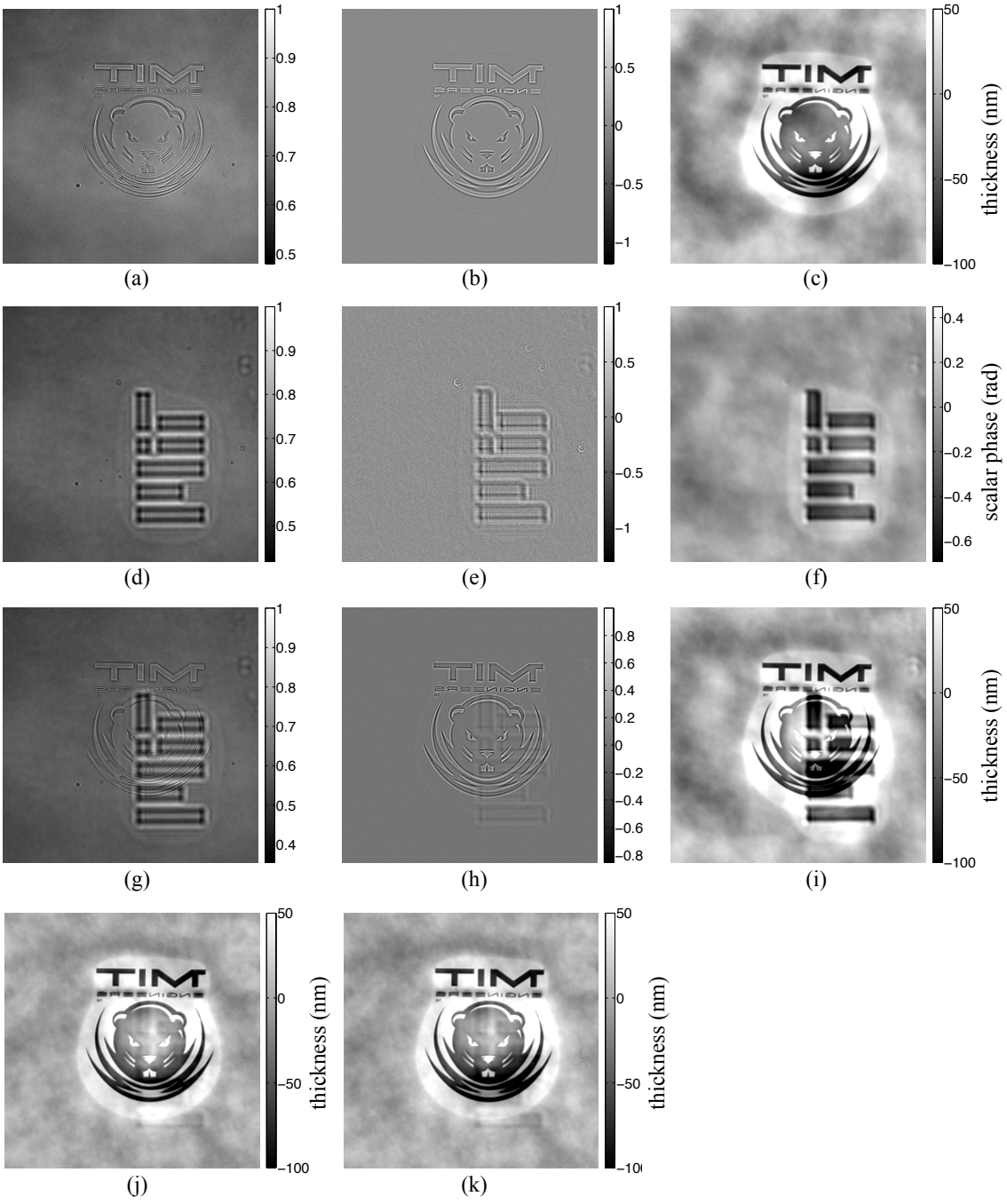


Figure 3-2: Experimental measurements and results for (a–c) the sample with uniform illumination, (d–f) the illumination modulation mask alone, and (g–k) both sample and modulation mask in place. (a,d,g) In–focus intensity measurements. (b,e,h) Intesnity difference between defocused measurements. (f) The scalar phase reconstruction for the illumination with only the field modulation mask in place. Sample thickness reconstructed from (c) Eq. (3.28) with uniform illumination, (i) Eq. (3.28) with both sample and modulation mask in place, (j) Eq. (3.29), assuming a pure–phase sample, and (k) Eq. (3.27), assuming a sample containing both amplitude and phase variations.

Chapter 4

Transport of intensity phase imaging with nonlinear diffusion regularization

4.1 Introduction

The transport of intensity equation (TIE) follows from the wave equation under paraxial propagation and in the simplest case relates the first derivative of intensity along the optical axis to the Laplacian of the phase in the transverse direction. In practice, recovery of phase from the TIE involves estimating the first intensity derivative from finite differences taken between two or more intensity images at different propagation distances [120, 121, 112] followed by numerical inversion of the Laplacian operator. Phase reconstructed from noisy images using the TIE often contains unwanted low–frequency artifacts [83], which existing denoising methods often do not remove effectively [58, 83]. On the other hand, proper adaptation of prior knowledge of the object can significantly improve recovery, as has been shown in numerous other inverse problems. In this chapter, we consider the prior, often encountered in practice, of the objects consisting of regions of piecewise constant phase and demonstrate how to incorporate this prior into the TIE solution as nonlinear diffusion (NLD)

regularization.

4.2 Theory and method

Consider a pure phase object oriented perpendicular to the optical axis, z, whose optical thickness is specified by $\phi(x, y)$. If this sample is illuminated by a normally incident plane wave of intensity I_0, the intensity immediately after the object, $I(x, y, z)$, satisfies a simplified form of the TIE

$$-\frac{k}{I_0}\frac{\partial I(x, y, z)}{\partial z} = \nabla^2\phi(x, y), \tag{4.1}$$

where $k = 2\pi/\lambda$ is the wave number. This has the form of Poisson's equation in 2D. If $\partial I/\partial z$ is known along with suitable boundary conditions for the phase, the phase ϕ can be uniquely computed from Eq. (4.1). For a more general phase and amplitude object, the full solution of the TIE can be obtained by solving two Poisson's equations [109], the reconstructed phase has been shown to be unique up to phase singularities [55].

To understand low–frequency artifacts in the TIE solution, it is instructive to consider the TIE in the Fourier domain. Consider a single Fourier component of the phase, which has spatial variation proportional to $\sin(u_0 x)$. From Eq. (4.1) it is clear that the change in intensity upon propagation, $\partial I/\partial z$, is proportional to u_0^2 (a more rigorous analysis for a bright–field microscope may be found in [3, 97]). Therefore, low spatial frequency components yield a much weaker signal than those with high spatial frequency. The inversion of the TIE to recover phase involves multiplication in the Fourier domain by the transfer function, $H = -1/(u^2 + v^2)$, where u and v are the spatial frequency variables along x and y, respectively. This has the effect of significantly amplifying low frequency noise in the phase reconstruction. Traditional linear denoising utilizes a Tikhonov regularized transfer function, $H_{\mathrm{TR}} = (u^2 + v^2) \,/\, \left[(u^2 + v^2)^2 + \gamma\right]$, where γ is a regularization parameter. H_{TR} behaves like a high–pass filtered version of H and γ controls the lower cutoff frequency.

An alternative nonlinear method is to write the TIE in terms of a linear model

$A = \mathcal{F}^* H \mathcal{F}$ that relates the measured intensity derivative f to the phase, where $\mathcal{F}$ is the Fourier transform operator; then solve it as a minimization problem where the quadratic error is balanced against a denoising (regularization) term by a weight τ as

$$\text{minimize} \;\; \frac{1}{2} \left\| f - A\phi \right\|^2 + \tau \iint \Psi \left(|\nabla \phi| \right) \mathrm{d}x \mathrm{d}y. \tag{4.2}$$

The argument of Ψ is deliberately chosen as the magnitude of the phase gradient, so that a proper choice of Ψ can successfully penalize low–frequency artifacts.

In a pure denoising context, A in Eq. (4.2) is simply the identity matrix and the method of steepest descent for minimizing Eq. (4.2) results in a partial differential equation for NLD denoising [93, 88, 125]:

$$\frac{\partial \phi(x, y; t)}{\partial t} = \tau \nabla \cdot \left(\frac{\Psi'(|\nabla \phi|)}{|\nabla \phi|} \nabla \phi \right). \tag{4.3}$$

The diffused image at $t = T$ is generally computed iteratively by finite differences. At each step in t, the local flux points along the direction of the gradient, $\nabla \phi$ with magnitude given by $F = \Psi'(|\nabla \phi|)$. For denoising TIE, Ψ should be such that F is preferentially large for small values of the phase gradient and nearly zero above some cutoff value. This would result in denoising through diffusion for low–frequency artifacts while preventing diffusion across sharp boundaries.

A widely used Ψ for recovering piecewise constant signals is the total variation (TV) function [106], $\Psi_{\mathrm{TV}} = |\nabla \phi|$, $F_{\mathrm{TV}} = 1$. Because F is constant, TV regularization behaves like an unbiased diffusion process so that low–frequency artifacts are not preferentially removed. In order to preserve sharp features while diffusing low–frequency regions, Weickert instead introduced an NLD flux of the form $F_{\mathrm{W}} = 1.67 s \left[1 - \exp(-3.86/s^{12}) \right]$ (where $s = |\nabla \phi| / k_0$ is the normalized magnitude of the gradient) [124]. Here k_0 is the critical gradient that controls the position of the deflection point in both the curve of Ψ_{W} and F_{W} in Fig. 4-1(a) and (b), respectively. The Weickert function has small flux at high gradient, and thus preserves sharp features. However, it also has linearly decreasing flux below k_0 and therefore does not remove low–frequency artifacts efficiently. To meet that goal as well, we removed the

"1.67s" term from the Weikert function, resulting instead in the "hybrid" NLD flux

$$F_{\mathrm{H}}(s) = 1 - \exp(-3.86/s^{12}). \tag{4.4}$$

This hybrid NLD regularization function and the magnitude of the flux are shown in Fig. 4-1(a) and (b), respectively. Below the critical gradient, F_{H} behaves like the TV function; and above the critical gradient, it behaves as the Weickert function.

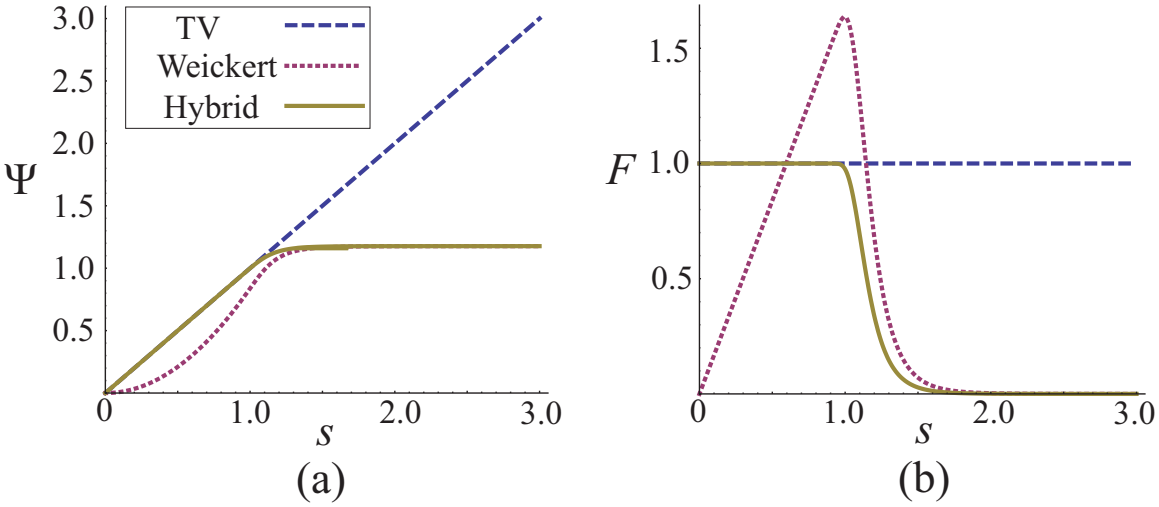


Figure 4-1: (a) The NLD regularizing function and (b) the magnitude of the flux as a function of the normalized magnitude of the gradient of the phase.

Pseudocode to implement the optimization in Eq. (4.2) based on the iterative shrinkage/thresholding(IST) algorithm [10] is as follows:

1. Initialize $\hat{\phi}_{\mathrm{b}}^{(0)} = A^* f$;

2. For $i = 1, 2, \ldots,$

(2.a) apply gradient descent with a step size α for the data fidelity term: $\hat{\phi}_{\mathrm{a}}^{(i)} = \hat{\phi}_{\mathrm{b}}^{(i-1)} + \alpha A^*(f - A\hat{\phi}_{\mathrm{b}}^{(i-1)})$;

(2.b) apply NLD with time t such that $t/\alpha = \tau$: $\hat{\phi}_{\mathrm{b}}^{(i)} = \hat{\phi}_{\mathrm{a}}^{(i)} + t\nabla\cdot\left(\Psi'(|\nabla\hat{\phi}_{\mathrm{a}}^{(i)}|)\nabla\hat{\phi}_{\mathrm{a}}^{(i)}/|\nabla\hat{\phi}_{\mathrm{a}}^{(i)}|\right)$;

3. Stop and yield the final estimate $\hat{\phi} = \hat{\phi}_{\mathrm{b}}^{(i-1)}$ when the change in the minimizing functional in Eq. (4.2) between consecutive steps falls below some threshold.

For large t, we implement the NLD with n time–marching steps of step size dt ($t = n\mathrm{d}t$) using the additive operator splitting (AOS) scheme [124]. Additionally, before computing the flux at each step, we smoothen the gradient by convolution

with a Gaussian to reduce potential staircasing artifacts [125].

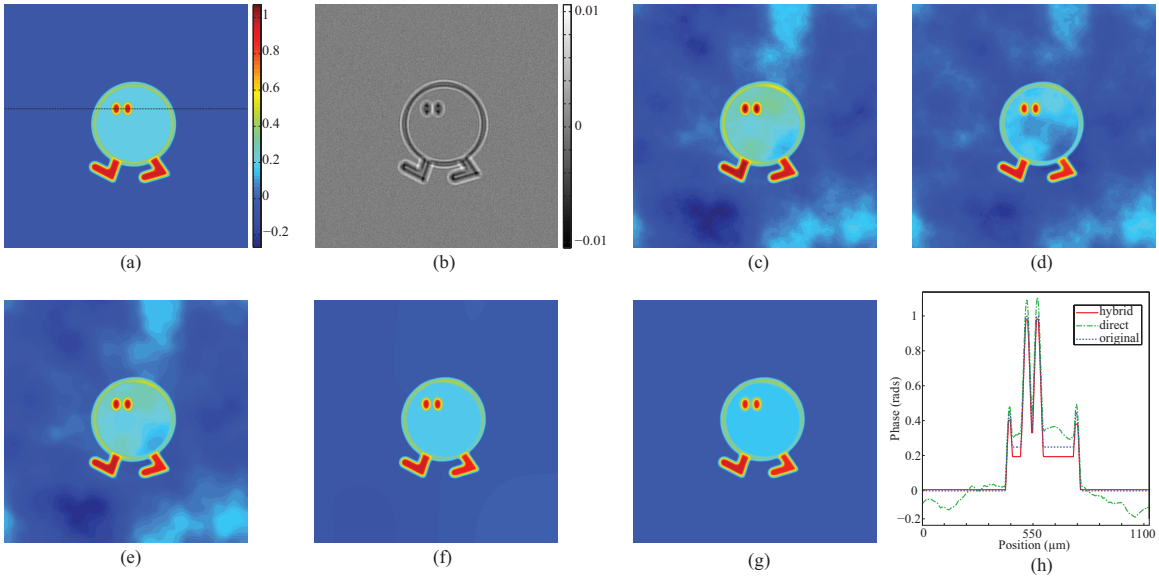


Figure 4-2: (a) Original phase, (b) intensity derivative, (c) direct, (d) Tikhonov–regularized, (e) TV–regularized, (f) Weickert function regularized and (g) the hybrid function regularized solutions. (f) Phase cross–sections taken along the dashed–line in (a).

4.3 Simulations

We demonstrate our method by a numerical example: a pure phase object with the phase distribution given in Fig. 4-2(a) is illuminated by a normally incident plane wave. The intensity derivative in Fig. 4-2(b) was generated by simulating over– and under–focused images, each defocused from the in–focus plane by a normalized distance $z_{\text{norm}} = 8$ ($z_{\text{norm}} = z \cdot \text{NA}^2/\lambda$, where z is the physical distance and NA=0.14 is the numerical aperture). Each defocused image was corrupted by Poisson noise such that the signal–to–noise ratio(SNR) in this intensity derivative data is 0dB. Low frequency artifacts are clearly seen in the directly inverted result in Fig. 4-2(c). The effect of Tikhonov regularization with the regularizer $\gamma = 10^{-12}$ (chosen to minimize the average per–pixel root–mean–square (RMS) error) is shown in Fig. 4-2(d). We implemented the NLD regularization with the TV, Weickert and hybrid functions and the results are shown in Fig. 4-2(e), (f), and (g), respectively. The TV function

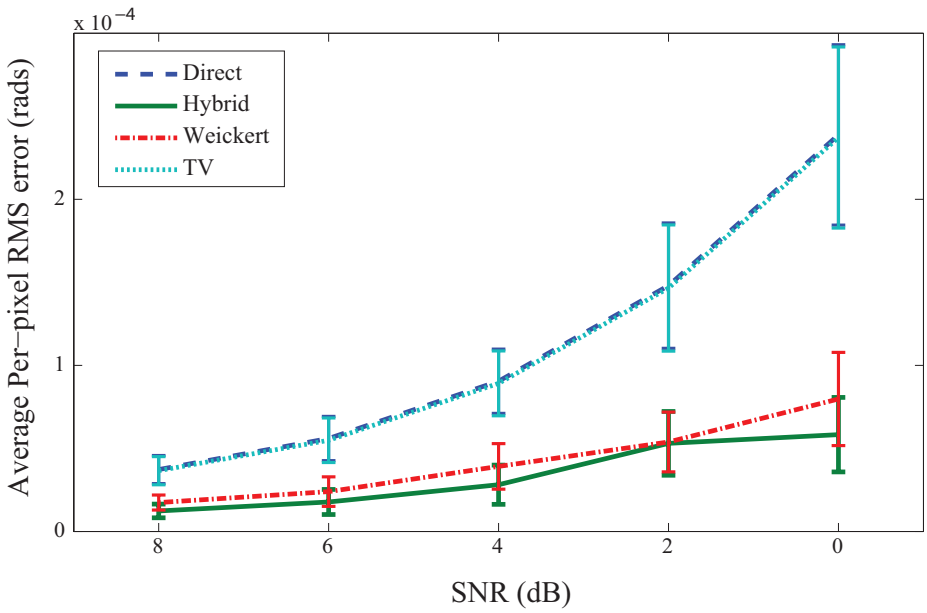


Figure 4-3: Average per-pixel RMS error versus SNR

smooths the result. Both the Weickert and hybrid functions preserve the edge features and diffuse the low frequency artifacts into piecewise constant patches; however, the Weickert function requires approximately 5 times more steps than the hybrid function to converge. Figure 4-2(h) compares the phase distributions using NLD diffusion along the dashed line indicated in Fig. 4-2(a) to phase recovered without regularization. The noise performance of the NLD method is studied in Fig. 4-3. For each SNR level, we repeat the simulation 100 times. In Fig. 4-3, the data points are the average per-pixel RMS errors in the recovered phase and the error bars are the standard deviations. In terms of RMS after convergence, the hybrid and Weickert functions perform similarly but, as mentioned previously, the latter converges much slower.

4.4 Discussion

We have demonstrated that the use of the piecewise constant priors in the TIE by NLD regularization to account for low–frequency artifacts while preserving sharp edges and

introduced a new function combining the desired properties of TV and Weickert NLD. It should be noted that, aside from the TV case, the NLD regularization explored here is not a provably convex problem; therefore, the recovered phase is not necessarily the global minimizer of Eq. (4.2). This problem also occurs in edge-preserving NLD since the regularization functions, Ψ that preserve edges and smooth low–frequency artifacts are not convex. However, in NLD it is generally observed that discretization due to finite differences (or Gaussian smoothing of the gradient prior to performing the NLD) greatly stabilizes the results [125]. Based on our observations, this appears to be the case for the technique presented here as well. For phase maps that are not piecewise constant, but have strong sparsity priors in other domains, *e.g.* the phase is piecewise continuous or consist of a small set of Zernike modes, similar techniques may prove useful for removing low–frequency artifacts.

Chapter 5

Compressive X–ray phase tomography based on transport of intensity

5.1 Introduction

Traditional tomography with hard X–rays recovers the attenuation of an object. Attenuation does not always provide good contrast when imaging objects made of materials with low electron density, *e.g.* soft tissues. In these cases, richer information is often contained in the phase, *i.e.* the optical thickness of the sample [76, 89, 129, 35], since the X–ray phase shift is almost 10^3 times larger than the X–ray attenuation for low–Z elements [107], such as hydrogen, carbon, nitrogen and oxygen. X–ray interferometry based on Mach–Zehnder geometry using Bragg crystals [76] provides X–ray phase information; however, it suffers from extreme sensitivity to mechanical stability and alignment, the coherent synchrotron source is necessary for good fringe contrast. The diffraction enhanced X–ray phase imaging technique measures the refraction angle after an X–ray beam passing through a sample using crystal analyzers. Enhancement of the X–ray phase contrast has been demonstrated using this technique [29], however, the experimental setup is also very complicated and relies on a

synchrotron source. Recently, a grating based X–ray interferometer is demonstrated using a table–top X–ray source [77, 89]; the spatial coherence requirement is satisfied by the introduction of a narrow pitch grating at the source plane. However, the measurement process is cumbersome since it requires scanning a analyzer grating in front of the detector. Propagation based techniques are particularly suitable for X–ray phase imaging because they allow phase to be recovered from intensity images taken at multiple propagation distances without the need for optical elements [54, 81]. Here we adopt the transport of intensity equation (TIE) which relates the measured intensity to the Laplacian of the phase under a weakly–attenuating sample approximation. Implementing TIE at many angles while rotating the object allows tomographic reconstruction of the refractive index distribution [16, 57, 51].

TIE tomographic reconstruction consists of two tasks: (1) retrieving the phase projections by solving the TIE at each angle and (2) applying a tomographic inversion algorithm to the phase projections to reconstruct the refractive index distribution. The two steps are often treated as separate problems [17]. First, the TIE is solved by an off–the–shelf Poisson equation solver. This step typically requires regularization, e.g. Tikhonov [51, 17] to mitigate amplification of low spatial frequency noise by the inverse Laplacian operator. For the second step, standard tomographic reconstruction is carried out, e.g. using the filtered back–projection (FBP) method. This requires additional regularization to mitigate the well known streaking (high–frequency) artifacts due to under–sampling in the Fourier domain. In this chapter, we design a forward model that combines TIE and tomography operations in a single, discretized linear operator and develop a compressive reconstruction method which allows suppression of both low and high–frequency artifacts.

The remainder of this chapter is structured as follows: Section 5.2 describes the forward model that is used to relate the unknown 3D refractive index distribution to the intensities measured at different rotational angles; the compressive reconstruction method to invert the model is developed in Section 5.3; experimental results from X–ray TIE tomography are presented in Sections 5.4; Section 5.5 concludes this chapter with some discussions.

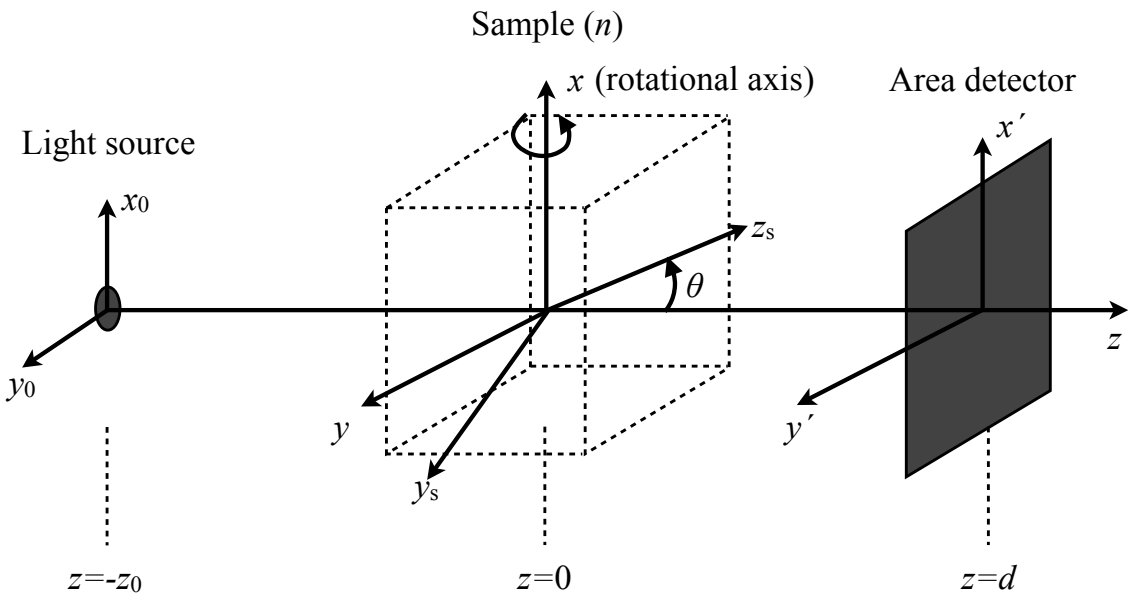


Figure 5-1: Imaging geometry for TIE tomography

5.2 Forward model

A schematic diagram of the imaging geometry for X–ray TIE tomography is shown in Fig. 5-1. A quasi–monochromatic source with mean wavelength λ is located at the plane of $z = -z_0$ with coordinates $\mathbf{x}_0 = (x_0, y_0)$. Here, we consider a planar incoherent source with intensity distribution I_0, which is a good approximation for a table–top X–ray source. The mutual intensity of the incident field J_i at the plane of $z = 0$ with coordinates $\mathbf{x} = (x, y)$, under the paraxial approximation, is [70]:

$$J_\mathrm{i}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\lambda^2 z_0^2} \exp\left\{\frac{\mathrm{i}k}{2z_0}\left(|\mathbf{x}_1|^2 - |\mathbf{x}_2|^2\right)\right\} H_0\left(\frac{\mathbf{x}_1 - \mathbf{x}_2}{\lambda z_0}\right) \tag{5.1}$$

where the exponential term measures the wavefront curvature in the illumination; the term H_0 is source dependent:

$$H_0\left(\frac{\mathbf{x}_1 - \mathbf{x}_2}{\lambda z_0}\right) = \iint I_0(\mathbf{x}_0) \exp\left\{\mathrm{i}2\pi \frac{\mathbf{x}_1 - \mathbf{x}_2}{\lambda z_0} \cdot \mathbf{x}_0\right\} \mathrm{d}^2\mathbf{x}_0, \tag{5.2}$$

which relates the mutual intensity of the partially coherent illumination to the Fourier transform of the intensity distribution of the primary incoherent source.

The sample is characterized by the 3D distribution of the complex refractive index, $n(\mathbf{x}, z) = n_{\mathrm{r}}(\mathbf{x}, z) + \mathrm{i}n_{\mathrm{i}}(\mathbf{x}, z)$, where the origin of the cartesian coordinates is located within the object. We further assume that the interaction between the sample and the field can be treated using the projection approximation. Under this assumption, the intensity attenuation and phase delay imparted upon the field passing through the sample can be characterized by a simple transmission function, $t(\mathbf{x}) = \sqrt{T(\mathbf{x})}\exp[\mathrm{i}\phi(\mathbf{x})]$, where the transmittance function T is defined by the integral $T(\mathbf{x}) = \exp\left(\int_L \alpha(\mathbf{x}, z)dl\right)$, and L defines a ray path of a straight line in the (y, z) plane, and the absorption coefficient $\alpha = kn_{\mathrm{i}}$; the phase delay ϕ measures the total optical path length (OPL) along the same ray path L: $\phi(\mathbf{x}) = \int_L kn_r(\mathbf{x}, z)dl$. By adopting this approximation, we assume that the size of the sample is small compared to the curvature of the illumination at the object and the wavelength of the incident light is much smaller than the size of the finest structure in the sample. Under these assumptions, the mutual intensity J' immediately after the sample is given by

$$J'(\mathbf{x}_1, \mathbf{x}_2) = J_{\mathrm{i}}(\mathbf{x}_1, \mathbf{x}_2)t(\mathbf{x}_1)t^*(\mathbf{x}_2). \tag{5.3}$$

The intensity after light passing through the object is recorded by an area detector located at the plane of $z = d$ with coordinates $\mathbf{x}' = (x', y')$. Under the paraxial approximation, the intensity I on the detector plane is given by the following propagation integrals:

$$I(\mathbf{x}') = \frac{1}{\lambda^2 d^2}\iiiint J'(\mathbf{x}_1, \mathbf{x}_2)\exp\left[\frac{\mathrm{i}k}{2d}\left(|\mathbf{x}' - \mathbf{x}_1|^2 - |\mathbf{x}' - \mathbf{x}_2|^2\right)\right]\mathrm{d}^2\mathbf{x}_1\mathrm{d}^2\mathbf{x}_2. \tag{5.4}$$

Substitute Eqs. (5.1–5.3) into Eq. (5.4), and then take the Fourier transform on both side of Eq. (5.4), yielding

$$\begin{aligned}\tilde{I}\left(\frac{\mathbf{u}}{M_{\mathrm{s}}}\right) = \frac{1}{\lambda^2 z_0^2}H_0\left(-\frac{\mathbf{u}}{M_{\mathrm{i}}}\right)\iint\sqrt{T\left(\mathbf{x} - \frac{\lambda d'\mathbf{u}}{2}\right)T\left(\mathbf{x} + \frac{\lambda d'\mathbf{u}}{2}\right)} \\ \exp\left\{\mathrm{i}\left[\phi\left(\mathbf{x} - \frac{\lambda d'\mathbf{u}}{2}\right) - \phi\left(\mathbf{x} + \frac{\lambda d'\mathbf{u}}{2}\right)\right]\right\}\exp(-\mathrm{i}2\pi\mathbf{x}\cdot\mathbf{u})\,\mathrm{d}^2\mathbf{x}\end{aligned} \tag{5.5}$$

where $\tilde{}$ denotes the Fourier transform, $\mathbf{u} = (u, v)$ is the spatial frequency variable, $\mathbf{x} = (\mathbf{x}_1 + \mathbf{x}_2)/2$, $M_\mathrm{s} = (z_0 + d)/z_0$ and $M_\mathrm{i} = (z_0 + d)/d$ are the geometrical magnification factors for the sample and the source, respectively, and $d' = z_0 d/(z_0 + d)$ is effective propagation distance. Assuming that $|\lambda d' \mathbf{u}|$ is small compared to the variations in the transmission and phase functions, Eq. (5.5) can be expanded to the first order in $\lambda d' \mathbf{u}$, yielding

$$\tilde{I}\left(\frac{\mathbf{u}}{M_\mathrm{s}}\right) = \frac{1}{\lambda^2 z_0^2} H_0\left(-\frac{\mathbf{u}}{M_\mathrm{i}}\right)\left\{\tilde{T}(\mathbf{u}) - \int [\mathrm{i}\lambda d' T(\mathbf{x})\mathbf{u}\cdot\nabla_{xy}\phi(\mathbf{x})]\exp(-\mathrm{i}2\pi\mathbf{x}\cdot\mathbf{u})\,\mathrm{d}^2\mathbf{x}\right\}. \tag{5.6}$$

By taking the inverse Fourier transform on both sides, Eq. (5.6) can be expressed as

$$g(\mathbf{x}) \equiv -\frac{k}{d'}[M_\mathrm{s}^2 I(M_\mathrm{s}\mathbf{x}) - I_\mathrm{d}(\mathbf{x})] = \frac{M_\mathrm{i}^2}{\lambda^2 z_0^2} I_0(M_\mathrm{i}\mathbf{x}) * \nabla_{xy}\cdot[T(\mathbf{x})\nabla_{xy}\phi(\mathbf{x})], \tag{5.7}$$

where $*$ denotes the convolution operator. The in–focus intensity measurement at the plane immediately after the sample is $I_\mathrm{d}(\mathbf{x}) = \frac{M_\mathrm{i}^2}{\lambda^2 z_0^2} I_0(M_\mathrm{i}\mathbf{x}) * T(\mathbf{x})$. Equation 5.7 is a modified TIE that includes the filtering effect due to the finite size of an incoherent source. Since the total intensity can be treated as an incoherent superposition of intensities generated from every points on the source, each of which illuminates the sample with a slightly different displacement with respect to the center, the overall effect on the final intensity is the convolution with intensity distribution of the source.

When the sample is weakly attenuating such that $T \approx 1$, I_d is entirely determined by the intensity profile of the source. If we further ignore the filtering effect due to the finite source size, Eq. (5.7) is simplified to

$$g(\mathbf{x}) \equiv -\frac{k}{d'}[I_\mathrm{s}(M_\mathrm{s}\mathbf{x}) - I_\mathrm{i}(M_\mathrm{s}\mathbf{x})] = I_\mathrm{i}(M_\mathrm{s}\mathbf{x})\nabla_{xy}^2\phi(\mathbf{x}), \tag{5.8}$$

where I_i is the background image taken without the sample in place. Equation 5.7 is a form of the TIE that uses two images taken with and without the sample in place and no detector motion in order to recover the phase projection of a weakly attenuating sample. The measurement process is illustrated in Fig. 5-2.

In a tomographic measurement, the sample is rotated about the x axis and the

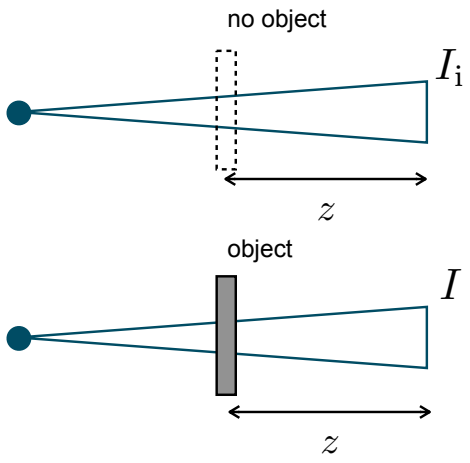


Figure 5-2: Experimental process for a pure phase sample. A background image I_{i} is first taken without the object in place; images I with the sample in place are taken next at the same plane.

phase projection ϕ for a given rotation angle θ can be expressed as:

$$\phi(\mathbf{x};\theta) = \iint n_{\mathrm{r}}(x, y_{\mathrm{s}}, z_{\mathrm{s}})\delta(y - y_{\mathrm{s}}\cos\theta - z_{\mathrm{s}}\sin\theta)\,\mathrm{d}y_{\mathrm{s}}\mathrm{d}z_{\mathrm{s}}, \tag{5.9}$$

The set of measurements g obtained from many angles of rotation can be arranged into a vector $\mathbf{g}$. The forward model that relates the unknown n to the data takes the simple form

$$\mathbf{g} = PR\,n \equiv An. \tag{5.10}$$

where the TIE operator P is performed on each phase projection and the forward operator $A = PR$ takes the projection of n by R at each angle.

5.3 Compressive reconstruction method

Since n generally contains more unknowns than measurements in $\mathbf{g}$, inversion is ill–posed; assuming n can be expressed as sparse, *i.e.* it contains only a small number of nonzero coefficients in some specified basis, an effective and elegant approach to invert it is to adapt compressed sensing theory [19]. Since the sample of interest often consists of regions with constant refractive indices, we choose the following

compressive reconstruction model

$$\hat{n}_{\mathrm{r}} = \arg\min_{n} \|n\|_{\mathrm{TV}} \text{ such that } \mathbf{g} = An, \tag{5.11}$$

where the total variation (TV) function $\|n\|_{\mathrm{TV}}$ is our sparsity basis, defined as

$$\|n\|_{\mathrm{TV}} = \sum \sqrt{(\nabla_x n)^2 + (\nabla_y n)^2 + (\nabla_z n)^2}, \tag{5.12}$$

and ∇_x, ∇_y, and ∇_z are the finite difference operators in the three spatial dimensions. The solution of the optimization in Eq. (5.11) is found by solving the following regularization problem instead to account for imperfect measurement [20, 10]

$$\text{minimize } \frac{1}{2}\|\mathbf{g} - An_{\mathrm{r}}\|^2 + \tau\|n_{\mathrm{r}}\|_{\mathrm{TV}}, \tag{5.13}$$

where the regularization parameter τ controls the relative weight of the TV constraint as compared to the data fidelity term $\|g - A\ n_{\mathrm{r}}\|^2$. We adapt the two–step iterative shrinkage/thresholding algorithm (TwIST) [10] to solve this minimization problem.

When implementing the algorithm, n_{r} is discretized into a $N \times N \times N$ matrix with voxel length Δ. The operator P is performed in the Fourier domain:

$$\begin{aligned} P &\equiv I_0 * \nabla_{xy} \cdot T\nabla_{xy} \\ &= \mathcal{F}_{xy}^{-1} H_{\mathrm{i}} \circ H_x \circ \mathcal{F}_{xy} T \circ \mathcal{F}_{xy}^{-1} H_x \circ \mathcal{F}_{xy} \\ &\quad + \mathcal{F}_{xy}^{-1} H_{\mathrm{i}} \circ H_y \circ \mathcal{F}_{xy} T \circ \mathcal{F}_{xy}^{-1} H_y \circ \mathcal{F}_{xy} \end{aligned} \tag{5.14}$$

where the gradient operator ∇_{xy} is replaced by $(\mathcal{F}_{xy}^{-1} H_x \circ \mathcal{F}_{xy}, \mathcal{F}_{xy}^{-1} H_y \circ \mathcal{F}_{xy})$ and the entries of H_x and H_y at (p, q) are given by $H_{x,pq} = \mathrm{i}2\pi p/(N\Delta)$ and $H_{y,pq} = \mathrm{i}2\pi q/(N\Delta)$, respectively; $\mathcal{F}_{xy}$ and $\mathcal{F}_{xy}^{-1}$ denote the discrete Fourier and inverse Fourier transform, respectively, with respect to the (x, y) variables; the transfer function matrix due to the incoherent source is considered by H_{i}, whose entry is determined by $H_{\mathrm{i},pq} = H_0(-p/(M_{\mathrm{i}}N\Delta), -q/(M_{\mathrm{i}}N\Delta))/\lambda^2 z_0^2$; and $\circ$ denotes entry–wise multiplication. For a weakly attenuating sample illuminated by an ideal point source, the TIE

operator P_{ph} takes a much simpler form:

$$P_{\text{ph}} = I_{\text{i}} \circ \mathcal{F}_{xy}^{-1} H \circ \mathcal{F}_{xy}, \tag{5.15}$$

where H is the transfer function matrix with entries defined as $H_{pq} = -4\pi^2(p^2 + q^2)/(N\Delta)^2$.

The operator R can be implemented in either the spatial domain [50] or the Fourier domain [43]. Generally, it is observed that the performance of the Fourier domain method is more robust to discretization and noise, which has been demonstrated in both X–ray CT [82] and MRI [68]. We adopt the Fourier domain method to write R as

$$R = \mathcal{F}_{y}^{-1} S \mathcal{F}_{yz,x} \tag{5.16}$$

where $\mathcal{F}_{yz,x}$ denotes the discrete 2D Fourier transform taken for the cross section at a constant x, S takes radial slices in the Fourier domain, and $\mathcal{F}_{y}^{-1} = [\mathcal{F}_{y,\theta_1}^{-1}, \mathcal{F}_{y,\theta_2}^{-1}, \cdots, \mathcal{F}_{y,\theta_N}^{-1}]$, where $\mathcal{F}_{y,\theta_{\text{m}}}^{-1}$ denotes the 1D inverse Fourier transform of a radial slice along the angle θ_{m}. Since the radial slices need to be taken from points equally spaced along the radial direction, while the result of $\mathcal{F}_{yz,x}$ is on a Cartesian grid, S requires interpolation and then resampling (gridding) [43]. We implement S together with $\mathcal{F}_{yz,x}$ by adapting the non-uniform FFT (NUFFT) algorithm [43].

The TIE and projection operators may be combined and simplified as

$$\begin{aligned} A \;=\; & (\mathcal{F}_{xy}^{-1} H_{\text{i}} \circ H_x \circ \mathcal{F}_{xy} T \circ \mathcal{F}_{xy}^{-1} H_x \circ \mathcal{F}_x \\ & + \mathcal{F}_{xy}^{-1} H_{\text{i}} \circ H_y \circ \mathcal{F}_{xy} T \circ \mathcal{F}_{xy}^{-1} H_y \circ \mathcal{F}_x) \mathcal{F}_{\text{NU}}, \end{aligned} \tag{5.17}$$

where $\mathcal{F}_{\text{NU}}$ is the NUFFT operator. For a weakly attenuating sample illuminated by an ideal point source, A is further simplified as

$$A = \mathcal{F}_{xy}^{-1} H \circ \mathcal{F}_x \mathcal{F}_{\text{NU}}. \tag{5.18}$$

5.4 X–ray phase tomography

The experimental setup for the X–ray TIE tomography is shown in Fig. 5-3. A microfocus source (Hamamatsu, L8121–03) with a circular focal spot size of 5μm in diameter, located at $z_0 = 0.765$, was operated at 20kVp to produce a diverging X–ray beam with central wavelength $\lambda = 0.062$nm. For a beetle sample, intensity images were obtained with an X–ray detector consisting of a CsI:Tl scintillator (RMD Inc.) coupled to an EMCCD (Andor, ixon, 512×512 pixels, 16μm pixel size) by an imaging system with 6× demagnification; the effective pixel size of an image directly seen at the scintillator is 96μm. The scintillator was placed at $d = 1.711$m. We assume the beetle sample is weakly attenuating. During the tomographic measurement, a single image was taken at every 5 degrees with 6.7 seconds exposure time. The intensity of the incident beam I_{i} was calibrated by taking a single background image without the sample in place.

The coherence width x_{c} of the illumination at the sample plane is estimated by calculating the width of the main lobe of the Fourier transform according to Eq. (5.2), yielding $x_{\mathrm{c}} = 19\mu$m. Since x_{c} is much smaller than the finest feature size of the sample, the low–pass filtering effect due to the finite source was ignored for the inversion. We have also verified that blurring due to the finite spectral bandwidth of the source is orders of magnitude smaller than x_{c}, so our model based on monochromaticity assumption applies in this experiment.

During the reconstruction, we first compute g for each angle, four examples of which are shown in Fig. 5-4. Reconstruction results using three different methods are compared in Fig. 5-6. In (a), the Fourier domain TIE solver with Tikhonov regularization chosen to provide optimal results is used to compute phase projections at each angle (the reconstructed phase projections for the measurements in Fig. 5-4 are shown in Fig. 5-5), and then the FBP method with a Ram–Lak filter is applied for the tomographic inversion. In (b), the same phase projections are inverted using the Fourier domain tomographic reconstruction method. The results from both methods suffer from severe streaking artifacts due to missing samples between slices in the

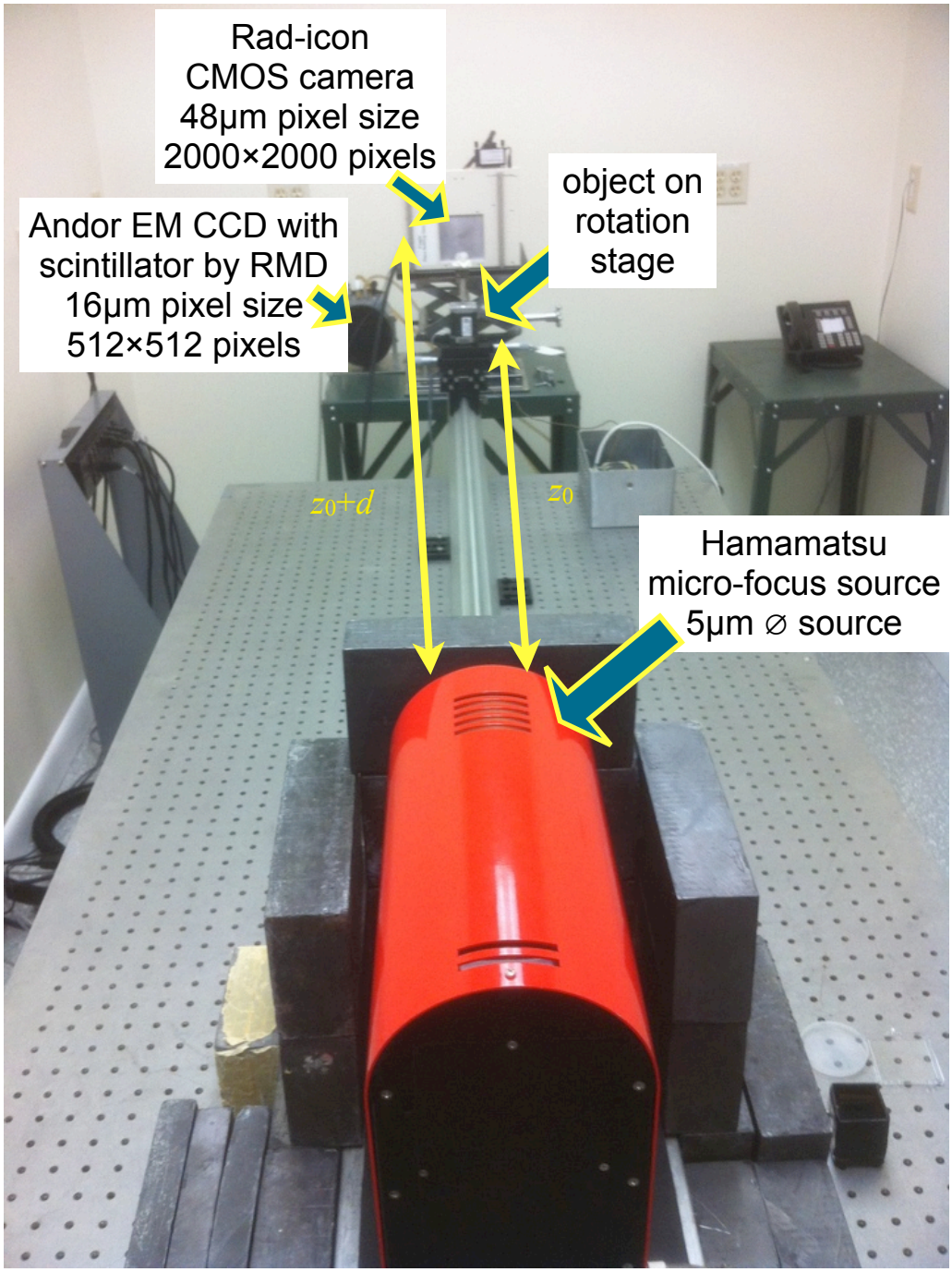


Figure 5-3: X–ray TIE tomography experimental setup

Fourier domain. Low–frequency artifacts (blurring) around edges are also observable but are less severe as compared to a single phase projection reconstruction. This is likely due to denser sampling around the origin resulting from the intersection of the Fourier slices. Both artifacts can be greatly suppressed using compressive reconstruction with TV minimization, Eq. (5.11), whose results are shown in (c), since TV minimization favors large structures with sharp edges. A 3D rendering of the refractive index reconstruction by the compressive method is shown in Fig. 5-7.

5.5 Conclusion and future work

A forward model that relates the refractive index distribution of the underlying sample to the intensity measurements of the straight–through X–ray beam has been developed. A compressive reconstruction method that simultaneously suppresses both high and low–frequency reconstruction artifacts is proposed and demonstrated experimentally for a weakly attenuating object in a microfocus X–ray system.

The current forward model shows that the intensity distribution of the primary incoherent source acts like a filter in the Fourier domain. The fringe contrast of the measured intensity data is controlled by the source size. When the blurring due to low–pass filtering from a medium–sized source is not severe, deconvolution techniques might help to improve the phase reconstruction. From an opposite perspective, the intensity distribution of the primary source provides extra degrees of freedom for designing the phase imaging system.

Compressed sensing techniques work best when the measurement is "incoherent," *i.e.* the sparse information in the unknown is evenly spread out in the measurement [19]. This is achieved in our model by the projection operator R. The measurement could be made more incoherent through the use of source coding or coded apertures [69].

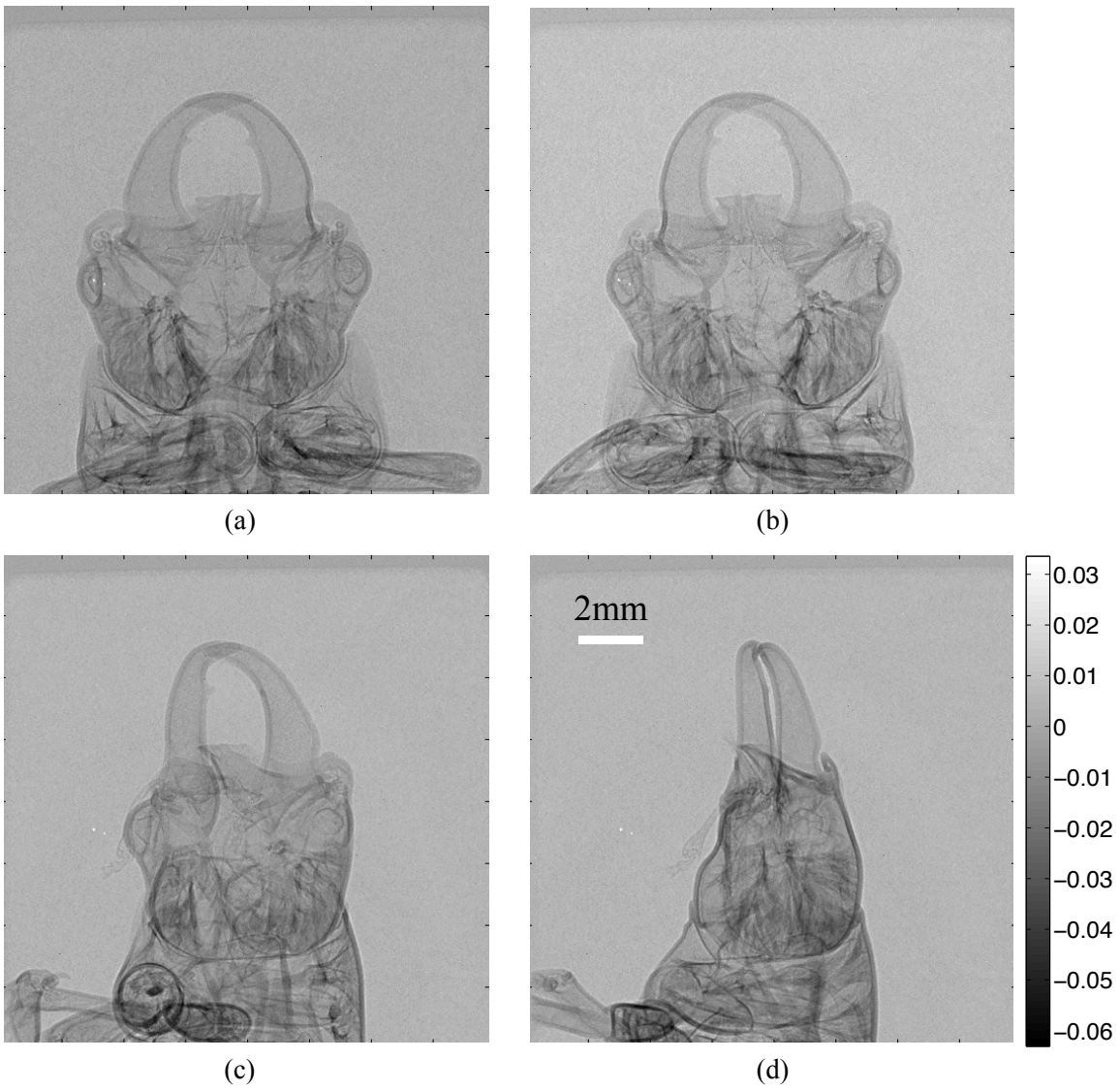


Figure 5-4: Normalized intensity variation of a beetle sample measured at (a) $\theta = 0°$, (b) $\theta = 30°$, (c) $\theta = 60°$, (d) $\theta = 90°$.

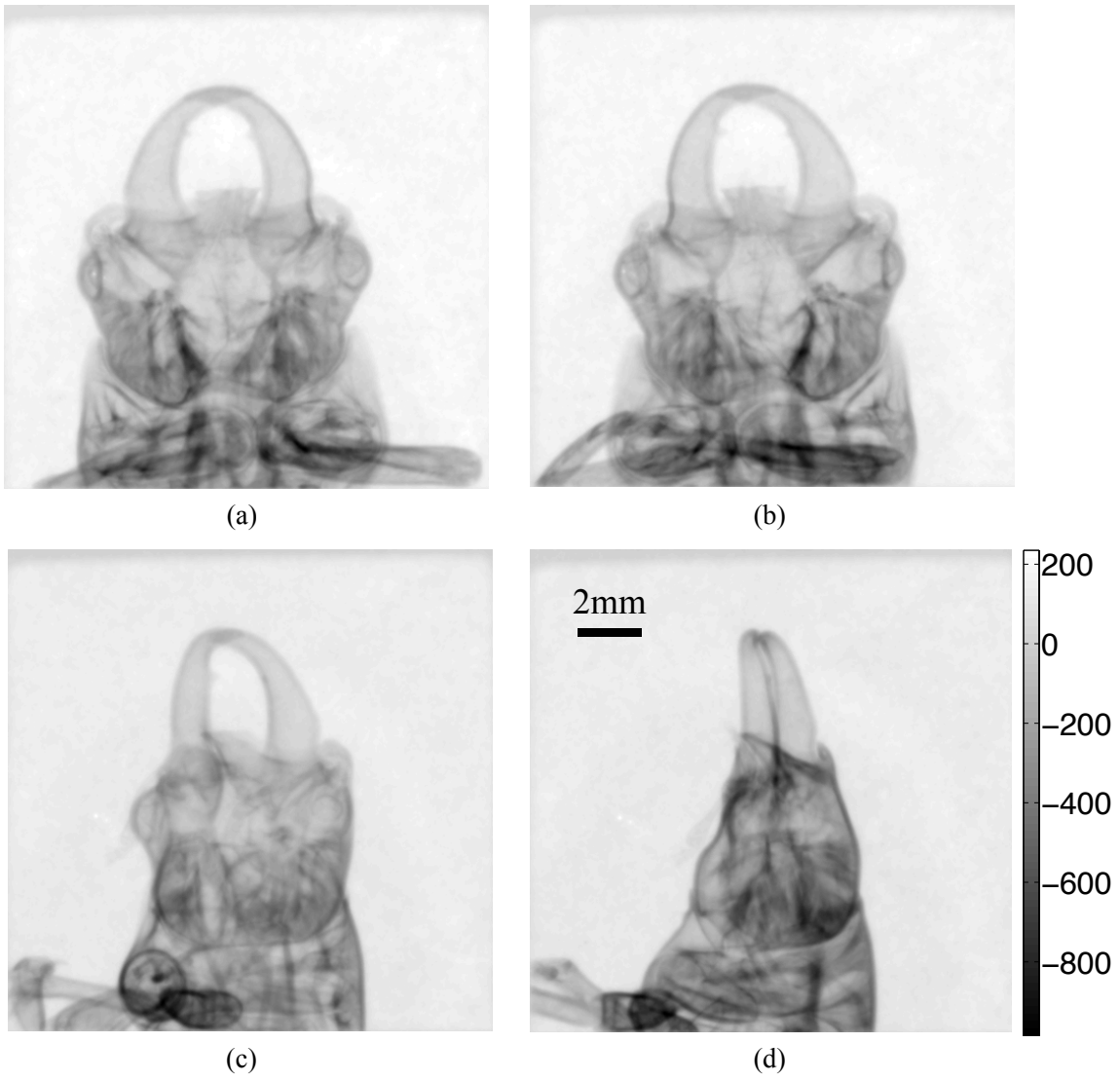
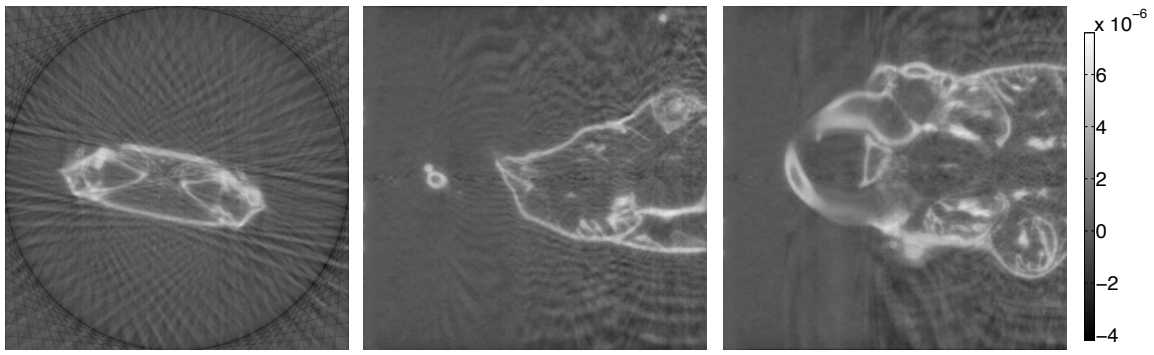
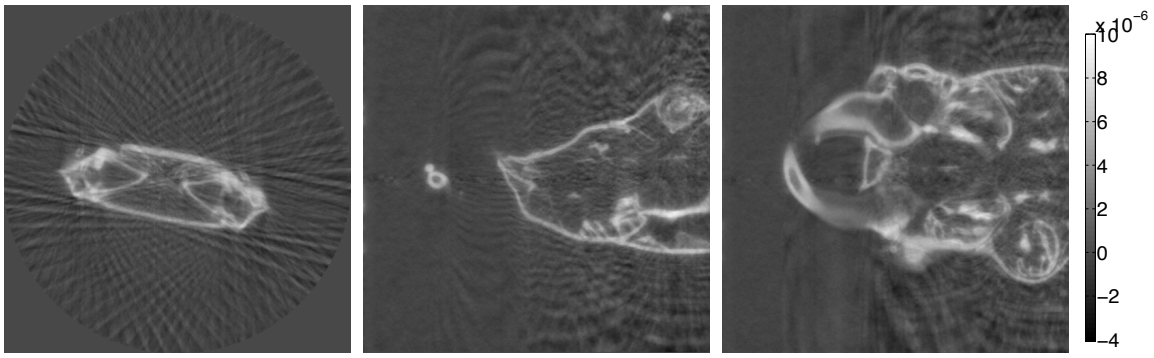


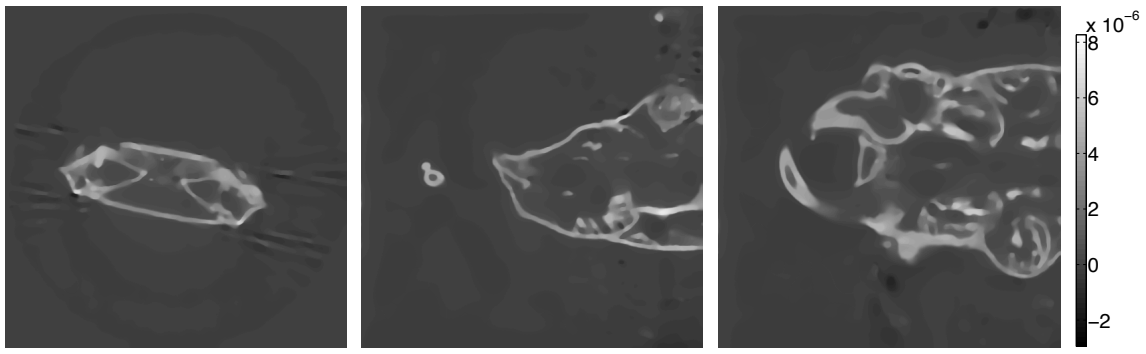
Figure 5-5: Phase projection reconstructions of the beetle sample at (a) $\theta = 0°$, (b) $\theta = 30°$, (c) $\theta = 60°$, (d) $\theta = 90°$.



(a)



(b)



(c)

Figure 5-6: Reconstruction results for the real part of the refractive index. (a) Fourier based TIE solver + FBP; (b) Fourier based TIE solver + Fourier domain tomographic inversion; (c) Compressive reconstruction by solving Eq. (5.13). The three cross–sections are taken from the three orthogonal planes going through the center of the sample.

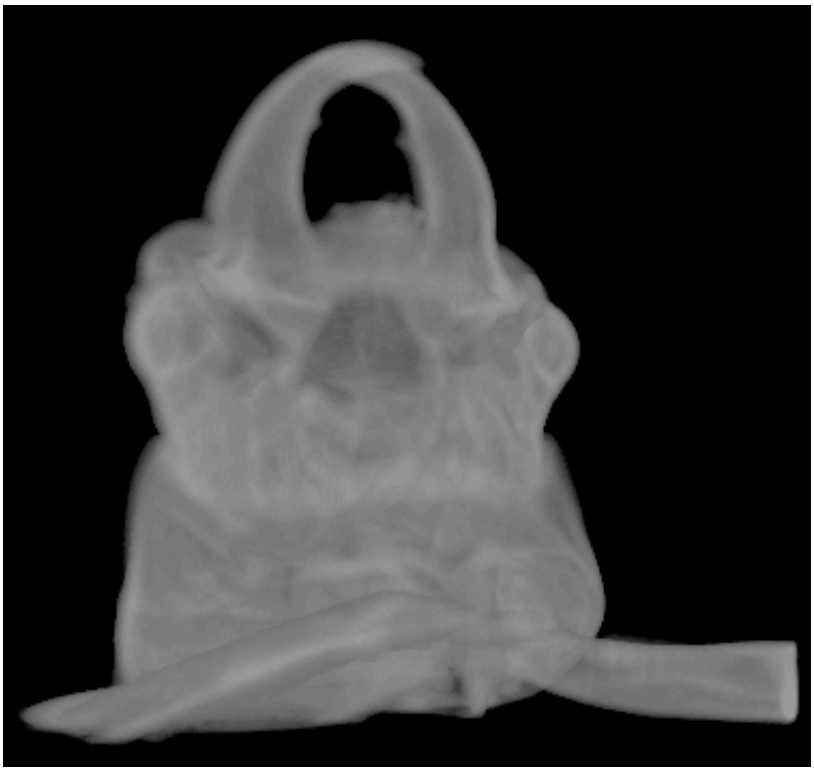


Figure 5-7: 3D rendering of the refractive index reconstruction by the compressive method.

Chapter 6

Compressive Phase Space Tomography

6.1 Introduction

In the previous three chapters, the TIE method is shown to be able to recover the phase based on measurements of propagated intensities. When the light is partially coherent, the TIE technique cannot recover the full statistical correlation information about the field. In this chapter, a technique, called phase space tomography (PST), which also relies on taking an axial stack of intensity images while the beam propagates, will be described. By using a different method to process the data, PST allows the recovery of the full correlation function of the underlying partially coherent field.

Correlation functions provide complete characterization of wave fields in several branches of physics, e.g. the mutual intensity of stationary quasi–monochromatic partially coherent light [70], and the density matrix of conservative quantum systems (*i.e.*, those with a time–independent Hamiltonian) [11]. Classical mutual intensity expresses the joint statistics between two points on a wavefront, and it is traditionally measured using interferometry: two sheared versions of a field are overlapped in a Young, Mach–Zehnder, or rotational shear [59, 71] arrangement, and two–point ensemble statistics are estimated as time averages by a slow detector under the assumption of ergodicity [70, 49].

As an alternative to interferometry, PST is an elegant method to measure correlation functions. In classical optics, PST involves measuring the intensity under spatial propagation [79, 91, 114] or time evolution [8]. In quantum mechanics, analogous techniques apply [118, 99, 64, 63]. However, the large dimensionality of the unknown state makes tomography difficult. In order to recover the correlation matrix corresponding to just n points in space, a standard implementation would require at least n^2 data points.

Compressed sensing [19, 20, 40] exploits sparsity priors to recover missing data with high confidence from a few measurements derived from a linear operator. Here, sparsity means that the unknown vector contains only a small number of nonzero entries in some specified basis. Low–rank matrix recovery (LRMR) [25, 27] is a generalization of compressed sensing from vectors to matrices: one attempts to reconstruct a high–fidelity and low–rank description of the unknown matrix from very few linear measurements.

In this chapter, the experimental measurement and verification of the correlation function of a classical partially coherent field using LRMR will be presented. It is worth noting that LRMR came about in the context of compressive quantum state tomography (QST) [53], which utilizes different physics to attain the same end goal of reconstructing the quantum state. In PST, one performs tomographic projection measurements, rotating the Wigner space between successive projections by evolving the wave function [79, 91]. This is directly analogous to the classical optical experiment we are presenting here, where we perform intensity measurements (*i.e.*, tomographic projections in Wigner space) and utilize propagation along the optical axis to rotate the Wigner space between projections. The difference lies in the fact that in QST the state is recovered via successive applications of the Pauli dimensionality–reducing operator, and there is no need to evolve the state. Nevertheless, both approaches lead to the same Hermitian LRMR problem, as long as the assumption of a quasi–pure unknown state is satisfied. In [52], it was shown that estimation of a low–rank matrix of dimension n and rank r requires only $O(rn \ln n)$ to $O(rn \ln^2 n)$ data points.

The low–rank assumption for classical partially coherent light anticipates a source

composed of a small number of mutually incoherent effective sources, *i.e.* coherent modes [127], to describe measurements. This is essentially equivalent to the low entropy assumption [53], *i.e.* a nearly pure quantum state in the quantum analogue. This assumption is valid for lasers, synchrotron and table–top X–ray sources [87], and Köhler illumination in optical microscopes [70]. An additional requirement for LRMR to succeed is that measurements are "incoherent" with respect to the eigenvectors of the matrix, *i.e.* the measured energy is approximately evenly spread between modes [52, 26]. Diffraction certainly mixes the coherent modes of the source rapidly, so we expect LRMR to perform well for classical PST. The same expectation for QST has already been established [53].

6.2 Theory and method

The two–point correlation function of a stationary quasi–monochromatic partially spatially coherent field is the mutual intensity [70]

$$J(\mathbf{x}_1, \mathbf{x}_2) = \langle g^*(\mathbf{x}_1) g(\mathbf{x}_2) \rangle \,, \tag{6.1}$$

where $\langle \cdot \rangle$ denotes the expectation value over a statistical ensemble of realizations of the field $g(\mathbf{x})$.

PST recovers the mutual intensity by successive measurement of intensities at multiple propagation distances. For simplicity, let us first consider a 1D case, as shown in Fig. 6-1. The intensity after propagation by distance z is [70]

$$I(x_\mathrm{o}; z) = \iint J(x_1, x_2) \exp\left(-\frac{\mathrm{i}\pi}{\lambda z}(x_1^2 - x_2^2)\right) \exp\left(\mathrm{i}2\pi \frac{x_1 - x_2}{\lambda z} x_\mathrm{o}\right) \mathrm{d}x_1 \mathrm{d}x_2. \tag{6.2}$$

This can be expressed in operator form as

$$I = \mathrm{tr}(P_{x_\mathrm{o}} J), \tag{6.3}$$

where P denotes the propagation operator that combines both the quadratic phase

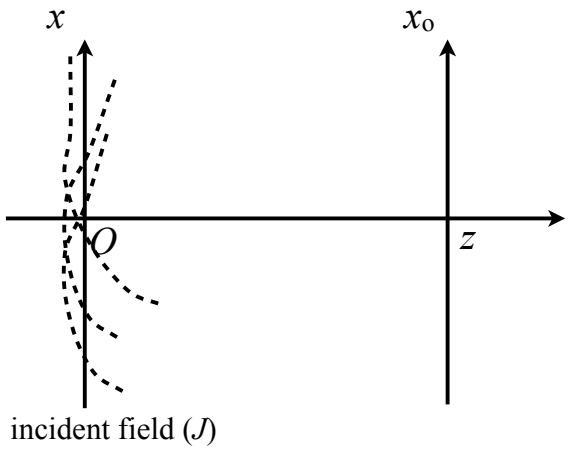


Figure 6-1: Experimental arrangement of PST for recovering the 2D mutual intensity J at the plane of incidence O.

and Fourier transform operations in Eq. (6.2), $\mathrm{tr}(\cdot)$ computes the trace, and x_o denotes the lateral coordinate at the observation plane. By changing variables $x = (x_1 + x_2)/2$, $x' = x_1 - x_2$ and Fourier transforming the mutual intensity with respect to x we obtain the Ambiguity Function (AF) [14, 15, 115]

$$\mathcal{A}(u', x') = \int J\left(x + \frac{x'}{2}, x - \frac{x'}{2}\right) \exp\left(-\mathrm{i}2\pi u' x\right) \mathrm{d}x. \tag{6.4}$$

Eq. (6.2) can be written as [79, 91, 114, 14, 115],

$$\tilde{I}(u'; z) = \mathcal{A}(u', \lambda z u'), \tag{6.5}$$

where $\tilde{I}$ is the Fourier transform of the vector of measured intensities with respect to x_o. Thus, radial slices of the AF may be obtained from Fourier transforming the vectors of intensities measured at corresponding propagation distances, and from the AF the mutual intensity of a 1D field can be recovered by an additional inverse Fourier transform, subject to sufficient sampling.

The extension of Eq. (6.5) for a 2D partially coherent field [described by the 4D

mutual intensity $J(x_1, y_1, x_2, y_2)]$ is given by

$$\tilde{I}(u', v'; z_x, z_y) = \mathcal{A}\left(u', v', \lambda z_x u', \lambda z_y v'\right). \tag{6.6}$$

It is worth noting that measurements by simple free space propagation arrangement can only access specific hyper–planes in the 4D Ambiguity space due to the constraint $z_x = z_y = z$. To break this symmetry in the x and y dimensions, astigmatic optics can be used in the measurement [91, 72]. Consider the geometry in Fig. 6-2, a pair of cylindrical lenses oriented perpendicularly are placed in between the plane of incidence and the detection plane. Without loss of generality, we define the plane of incidence at $z = 0$, the detection plane is at $z = z_{\mathrm{o}}$, whereas the cylindrical lenses oriented along x and y are located at $z = z_1$ and $z = z_2$, respectively. The intensity measured at point $(x_{\mathrm{o}}, y_{\mathrm{o}})$ is

$$\begin{aligned} I(x_{\mathrm{o}}, y_{\mathrm{o}}; z_{\mathrm{o}}) = \iiiint J(x_1, y_1, x_2, y_2) \exp\left[\mathrm{i}\pi\left(\frac{x_1^2 - x_2^2}{\lambda z_{xa}} + \frac{y_1^2 - y_2^2}{\lambda z_{ya}}\right)\right] \\ \exp\left[-\mathrm{i}2\pi\left(\frac{x_1 - x_2}{\lambda z_{xb}} x_{\mathrm{o}} + \frac{y_1 - y_2}{\lambda z_{yb}} y_{\mathrm{o}}\right)\right], \end{aligned} \tag{6.7}$$

where

$$\begin{aligned} z_{xa} &= z_1 \frac{1/z_1 + 1/(z_{\mathrm{o}} - z_1) - 1/f_1}{1/(z_{\mathrm{o}} - z_1) - 1/f_1}, \;\; z_{xb} = z_{\mathrm{o}} - \frac{z_1(z_{\mathrm{o}} - z_1)}{f_1}, \\ z_{ya} &= z_2 \frac{1/z_2 + 1/(z_{\mathrm{o}} - z_2) - 1/f_2}{1/(z_{\mathrm{o}} - z_2) - 1/f_2}, \;\; z_{yb} = z_{\mathrm{o}} - \frac{z_2(z_{\mathrm{o}} - z_2)}{f_2}, \end{aligned} \tag{6.8}$$

and f_1 and f_2 are the focal length of the cylindrical lenses oriented along x and y, respectively. The 2D Fourier transform of the measured intensity is related to the AF by

$$\tilde{I}(u, v) = \mathcal{A}\left(-\frac{z_{xb}}{z_{xa}} u, -\lambda z_{xb} u, -\frac{z_{yb}}{z_{ya}} v, -\lambda z_{yb} v\right), \tag{6.9}$$

which shows that the effective propagation distances in x and y are z_{xa} and z_{ya}, respectively. By changing the distances (z_1, z_2, and z_{o}) and/or the focal lengths (f_1 and f_2), z_{xa} and z_{ya} can be varied independently.

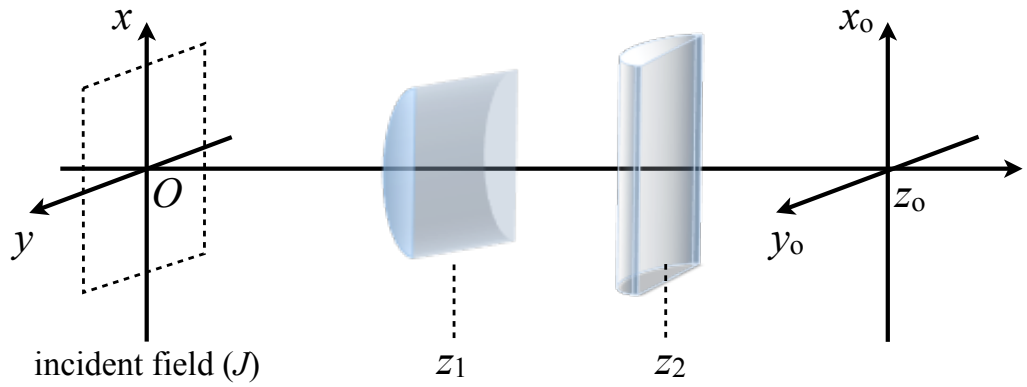


Figure 6-2: Experimental arrangement of PST for recovering the 4D mutual intensity J at the plane of incidence O. A pair of cylindrical lenses oriented perpendicularly are used to introduce astigmatism to the measurement. Intensities are measured at planes with axial coordinate z_o.

In both cases, a linear model for compressive PST can formulated by first arranging the measured intensity data in the Ambiguity space. The mutual intensity is defined as the "sparse" unknown to solve for. To relate the unknowns (mutual intensity) to the measurements (AF), the center–difference coordinate–transform is first applied, expressed as a linear transformation T upon the mutual intensity J, followed by Fourier transform $\mathcal{F}$, and adding measurement noise e as

$$\mathcal{A} = \mathcal{F} \cdot T \cdot J + e. \tag{6.10}$$

The mutual intensity propagation operator is unitary and Hermitian, since it preserves energy. We use eigenvalue decomposition to determine the basis where the measurement is sparse. The resulting basis, *i.e.* the set of eigenvectors, is also known as coherent modes in optical coherence theory, whereas the whole process is known as coherent mode decomposition [127]. The goal of the LRMR method is to minimize the number of coherent modes to describe measurements. By doing LRMR, we impose two physically meaningful priors: *(1)* existence of the coherent modes [127], and *(2)* sparse representation of the partially coherent field in terms of coherent modes.

Mathematically, if we define all the eigenvalues λ_i and the estimated mutual in-

tensity as $\hat{J}$, the method can be written as

$$\begin{aligned} \text{minimize} \quad & \text{rank}(\hat{J}) \\ \text{subject to} \quad & \mathcal{A} = \mathcal{F} \cdot T \cdot \hat{J}, \\ & \lambda_i \geq 0, \text{and } \sum_i \lambda_i = 1. \end{aligned} \tag{6.11}$$

Direct rank minimization is NP–hard; however, it can be accomplished by solving instead a proxy problem: convex minimization of the "nuclear norm" (ℓ_1 norm) of the matrix J [25, 22]. The corresponding problem is stated as

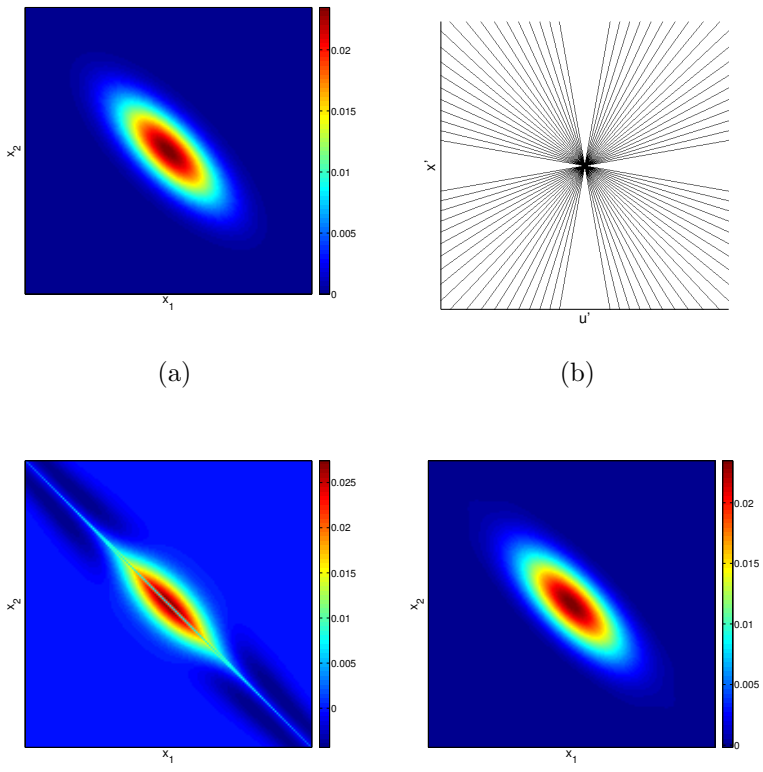
$$\begin{aligned} \text{minimize} \quad & \| \hat{J} \|_* \\ \text{subject to} \quad & \mathcal{A} = \mathcal{F} \cdot T \cdot \hat{J}, \\ & \lambda_i \geq 0, \text{and } \sum_i \lambda_i = 1, \end{aligned} \tag{6.12}$$

where the nuclear norm is the sum of the singular values $\sigma_i = |\lambda_i|$, $\| \hat{J} \|_* = \sum_i \sigma_i$. This problem is convex and a number of numerical solvers can be applied to solve it. In our implementation, we used the singular value thresholding (SVT) method [18]. The output estimate after each iteration of SVT typically has a sub-normalized total energy, *i.e.* $\sum_i \lambda_i < 1$; we compensated for this by renormalizing at the end of each iteration [53].

6.3 Numerical simulations

First, we demonstrate the LRMR method with a numerical example using a 1D Gaussian-Schell model source (GSMS). Both the intensity distribution and the degree of coherence of GSMS follow a Gaussian distribution [103]

$$J(x_1, x_2) = [I(x_1)]^{1/2} [I(x_2)]^{1/2} \mu(x_1 - x_2), \tag{6.13}$$



(a) (b)

(c) (d)

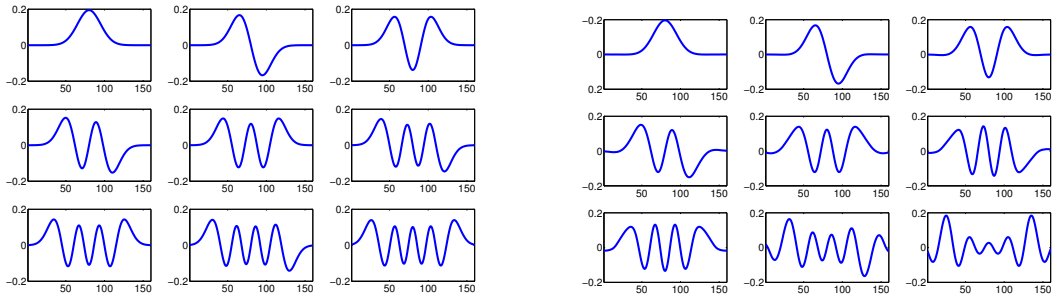
Figure 6-3: (a) Input mutual intensity of a GSMS with paramters $\sigma_I = 17$ and $\sigma_c = 13$, (b) data point locations in the Ambiguity space, mutual intensities estimated by (c) FBP and (d) LRMR methods.

and

$$I(x) = \exp\left(\frac{x^2}{2\sigma_I^2}\right),$$
$$\mu(x_1 - x_2) = \exp\left(\frac{(x_1 - x_2)^2}{2\sigma_c^2}\right), \tag{6.14}$$

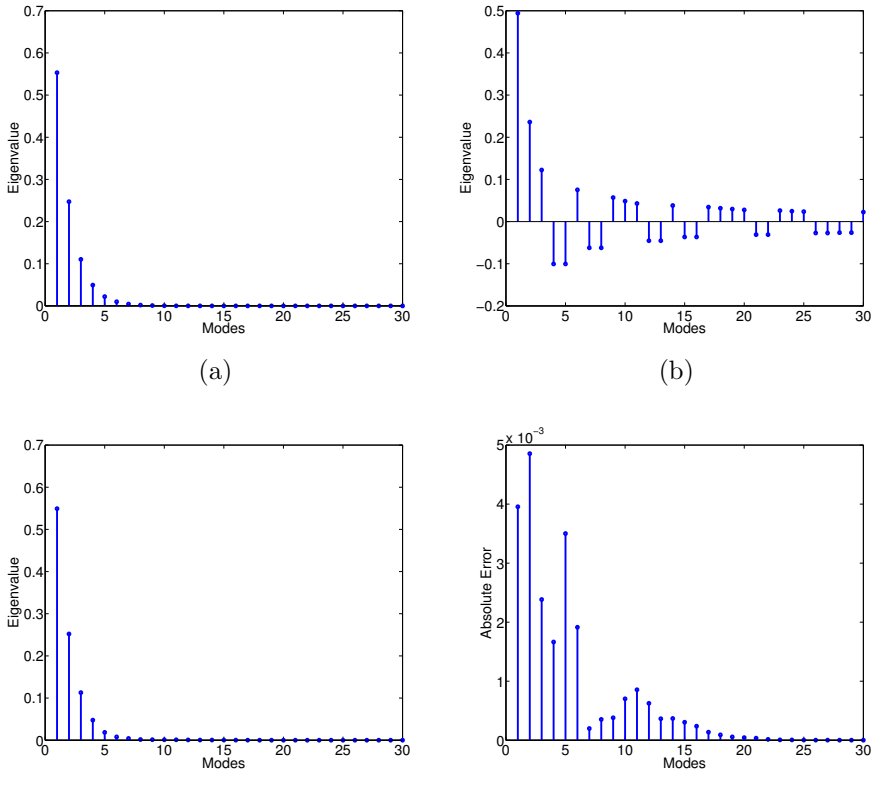
where σ_I determines the spatial extent of the source, and σ_c is proportional to the coherence length and determines the number of coherent modes in the input source. The eigenvalues of GSMS are never zero (analytical solution given in [103]). We defined the number of modes (rank of the source) r as the first r modes containing the 99% of the total energy.

One example is shown in Fig. 6-3(a). The parameters in this example are $\sigma_I = 17$



(a) (b)

Figure 6-4: The first nine coherent modes of the mutual intensity in Fig. 6-3(a). (a) Theoretical modes, and (b) LRMR estimates.



(a) (b)

(c) (d)

Figure 6-5: Eigenvalues of the mutual intensity in Fig. 6-3(a). (a) Theoretical values, (b) FBP estimates, (c) LRMR estimates, and (d) absolute errors in the LRMR estimates versus mode index.

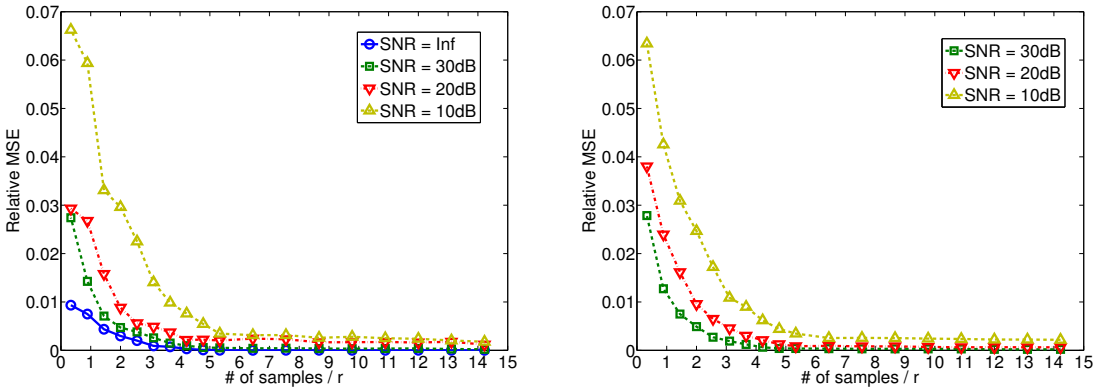
and $\sigma_c = 13$ (rank $r = 6$). Intensities are calculated at 40 different axial distances and the coverage in Ambiguity space is shown in Fig. 6-3(b). We simulate the case

where data from both the near field and the far field are missing due to the finite range of camera scanning motion allowed in the actual experiment. The missing cone around the u'-axis is due to missing data from near field, while the data missing from far field results in the missing cone around the x'-axis. Both cones have an apex angle of 20 degrees.

For comparison, the data are first processed using the traditional filtered-backprojection (FBP) method [60]. Applying the Fourier-slice theorem to Eq. (6.5) implies that the 1D Fourier transform of a radial slice in the Ambiguity space (an intensity measurement) is related to a projection in the AF's 2D Fourier space (the Wigner space [5, 7]). The Wigner distribution function (WDF) is related to the mutual intensity by

$$\mathcal{W}(x,u) = \int J\left(x + \frac{x'}{2}, x - \frac{x'}{2}\right) \exp\left(-i2\pi u x'\right) dx'. \tag{6.15}$$

To implement the FBP method, each intensity projection is first filtered by a Ram–Lak kernel apodized by a Hamming window; the estimated WDF is obtained by back-projecting all the filtered intensities, and then an inverse Fourier transform is applied to produce the mutual intensity. Figure 6-3(c) shows the reconstructed mutual intensity following this procedure. Three types of artifacts can be seen in this reconstruction. First, the reconstructed mutual intensity has lower values along the diagonal of the matrix due to the missing cones. However, this is unphysical because a correlation function should always have maximum value at zero separation. Second, the estimated degree of coherence is lower than the original field. The third artifact is the high frequency noise around the diagonal of the matrix, which is due to undersampling between the radial slices. All these artifacts have been greatly suppressed or completely removed by LRMR, whose reconstruction result is shown in Fig. 6-3(d). The disappearance of the correlation peak along the diagonal (*i.e.*, the intensity) when we use FBP for the reconstruction can be best explained with the help of Figure 6-3(b). Going from the Ambiguity space to the mutual intensity space involves Fourier transforming along horizontal lines, parallel to the u' axis. The diagonal in particular corresponds to the line $x' = 0$. It can be easily seen that, due



(a) Gaussian noise

(b) Poisson noise

Figure 6-6: Oversampling rate versus relative MSE of LRMR estimates. The input field is a GSMS with parameters $\sigma_I = 36$ and $\sigma_c = 18$. The noisy data is generated with different SNR from (a) an additive random Gaussian noise model, and (b) a Poisson noise model.

to the missing cone, pretty much all the data are missing from that line, except near the origin; thus resulting in a low–pass filtering effect. The fact that the compressive reconstruction method manages to restore the physically correct values of the correlation along the diagonal corroborates that the missing cone is successfully retrieved in our LRMR reconstruction. The FBP reconstruction may also be compared quantitatively to the compressive reconstruction in terms of the global degree of coherence parameter $\bar{\mu} = \frac{\sqrt{\sum_i \lambda_i^2}}{\sum_i |\lambda_i|}$ [102, 6], which was found as 0.150 and 0.617, respectively; the true state has $\bar{\mu} = 0.618$.

The coherent modes for a GSMS are Hermite–Gaussian sources [103]. The theoretical and LRMR estimated first nine coherent modes in this example are shown in Fig. 6-4(a) and 6-4(b), respectively. The theoretical eigenvalues are shown in Fig. 6-5(a). The FBP and LRMR estimated eigenvalues are compared in Fig. 6-5(b) and 6-5(c), respectively. The FBP estimates have several negative values, which does not satisfy the positive energy constraint. The absolute errors in LRMR estimates are plotted in Fig. 6-5(d).

Next, we study the noise performance of the LRMR method with a numerical example. In this simulation, the dimension of the input GSMS is 256×256 with

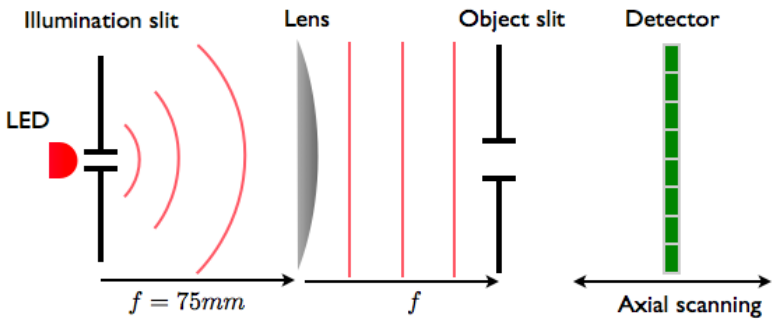


Figure 6-7: Experimental arrangement for 2D PST. A 1D rectangular slit is illuminated by a partially coherent light. The coherence of the illumination is controlled by the size of the slit placed at the front focal plane of the condenser. Free space diffraction patterns after the object slit are recorded at multiple axial locations.

parameters $\sigma_I = 36$ and $\sigma_c = 18$ (rank $r = 9$). We generate noisy data with different signal-to-noise ratio (SNR) from both an additive random Gaussian noise model and a Poisson noise model. However, we emphasize that the reconstruction algorithm does not make use of the noise statistics. For each SNR level, we repeat the simulation 100 times with different random noise terms, and then record the average relative mean-square-error (MSE) from the LRMR reconstruction. The ratio between the number of samples taken from the intensity measurements and the rank r of the input mutual intensity matrix determines the oversampling rate [21]. This rate is plotted versus relative MSE for different SNR cases in Fig. 6-6. For good performance, the required oversampling rate is at least 5–6 (the theoretical oversampling rate is on the order of $\ln(256) = 5.5$ according to [52]). Furthermore, the LRMR method is robust to noise in the sense that the reconstruction degrades gracefully as the SNR decreases.

6.4 Experimental results for 2D PST

The experimental arrangement for 2D PST is illustrated in Fig. 6-7. The illumination is generated by an LED with 620nm central wavelength and 20nm bandwidth. To generate partially coherent illumination, a single slit of width 355.6μm (0.014″) is placed immediately after the LED and one focal length (75 mm) to the left of a cylindrical lens. One focal length to the right of the lens, we place the second single

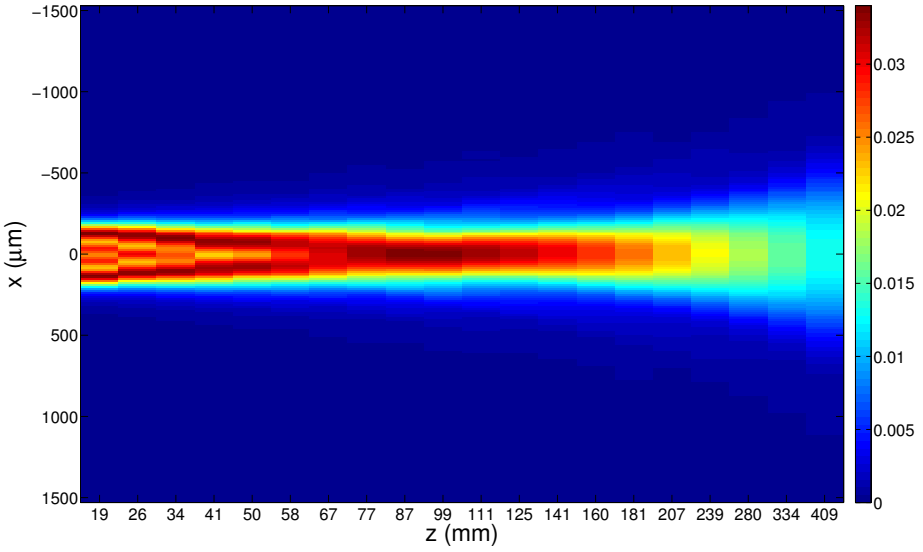
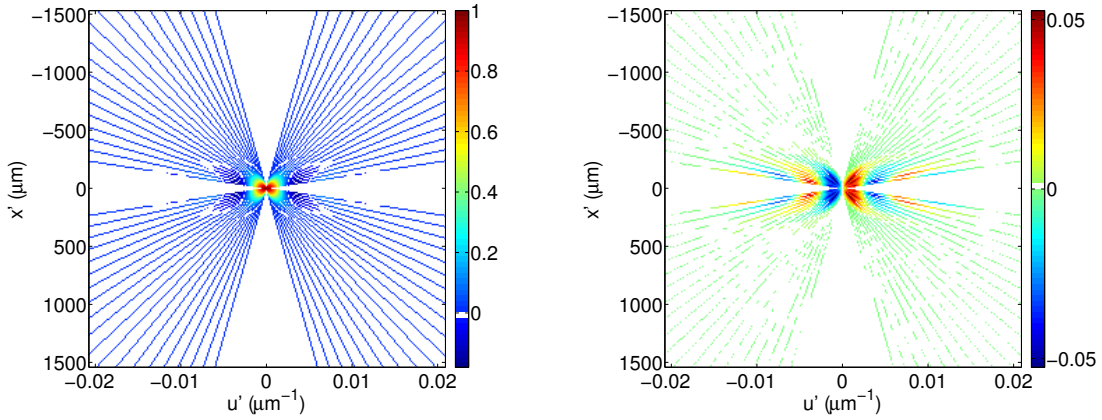


Figure 6-8: Intensity measurements at several unequally spaced propagation distances.



(a) (b)

Figure 6-9: (a) Real and (b) imaginary parts of the radial slices in Ambiguity space from Fourier transforming the vectors of intensities measured at corresponding propagation distances.

slit of width 457.2μm (0.018″), which is used as a one–dimensional (1D) object.

The goal is to retrieve the mutual intensity immediately to the right of the object from a sequence of intensity measurements at varying z–distances downstream from the object, as described in the theory. We measured the intensities at 20 z–distances, ranging from 18.2mm to 467.2mm, to the right of the object. The data are given

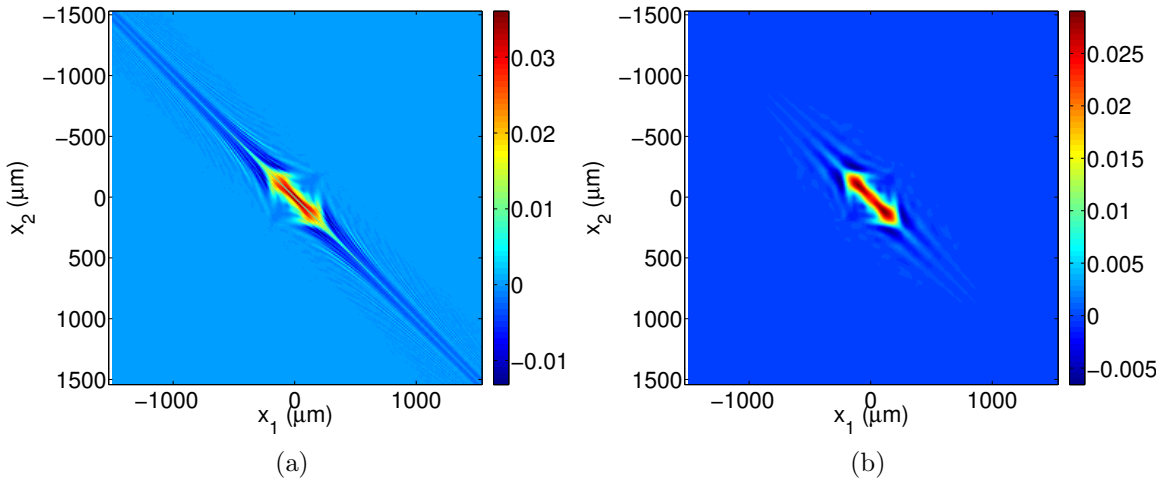
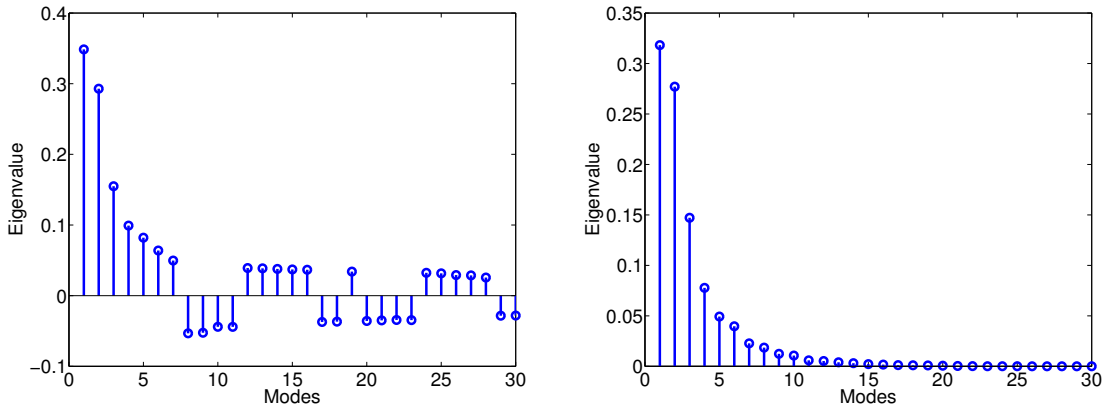


Figure 6-10: Real part of the reconstructed mutual intensity from (a) FBP; (b) LRMR method.

in Fig. 6-8. Each 1D intensity measurement consists of 512 samples, captured by a CMOS sensor with 12μm pixel size. The dimension of the unknown mutual intensity matrix to be recovered is 512×512. Since only intensities at positive z, *i.e.* downstream from the object, are accessible, we can only fill up the top right and bottom left quadrants of Ambiguity space. The other two quadrants are filled symmetrically, *i.e.* assuming that if the field propagating to the right of the object were phase conjugated with respect to the axial variable z, it would yield the correct field to the left of the object, *i.e.* negative z [114, 63]. Under this assumption, a total of 40 radial slices are sampled in Ambiguity space, as shown in Fig. 6-9. The apex angle of the missing cone around the u'–axis is approximately 17.4 degrees, and the one around the x'–axis is approximately 28.6 degrees. The number of measurements is only 7.8% of the total number of entries in the unknown mutual intensity matrix.

The reconstructions from the FBP and LRMR methods are compared in Fig. 6-10(a) and 6-10(b), respectively. The FBP reconstruction suffers from the same artifacts detailed in the numerical simulations section. All these artifacts are greatly suppressed or completely removed in the LRMR reconstruction. In the real part of the reconstruction, the width of the square at the center is approximately 456μm (38 pixels), which agrees with the actual width of the slit. The imaginary part is orders



(a) (b)

Figure 6-11: Eigenvalues estimated by (a) FBP, and (b) LRMR method.

of magnitude smaller than the real part.

FBP estimated eigenvalues contain several negative values, and are shown in Fig. 6-11(a). This does not satisfy the positive energy constraint. LRMR estimated eigenvalues are compared in Fig. 6-11(b), and all eigenvalues are positive.

We further validated our compressive estimates by measuring the field intensity immediately to the right of the illumination slit [Fig. 6-12(a)]. Assuming that the illumination is spatially incoherent (a good assumption in the LED case), the mutual intensity of the field immediately to the left of the object is the Fourier transform of the measured intensity, according to the van Cittert–Zernike theorem [70, 49]. This calculated mutual intensity, based on the measurement of Fig. 6-12(a) and screened by the object slit, is shown in Fig. 6-12(b). The eigenvalues computed by coherent mode decomposition are shown in Fig. 6-12(c) and are in good agreement with the LRMR estimates, as compared in Fig. 6-12(d). It is seen that 99% of the energy is contained in the first 13 modes, which confirms our low–rank assumption. The FBP reconstruction may also be compared quantitatively to the compressive reconstruction in terms of the global degree of coherence parameters, which were experimentally found as 0.12 and 0.46, respectively; whereas the estimate yielded by the van Cittert–Zernike theorem is 0.49. The first nine eigenvectors of each individual mode are shown and compared in Fig. 6-13. Small errors in the compressive estimate are because the missing cone

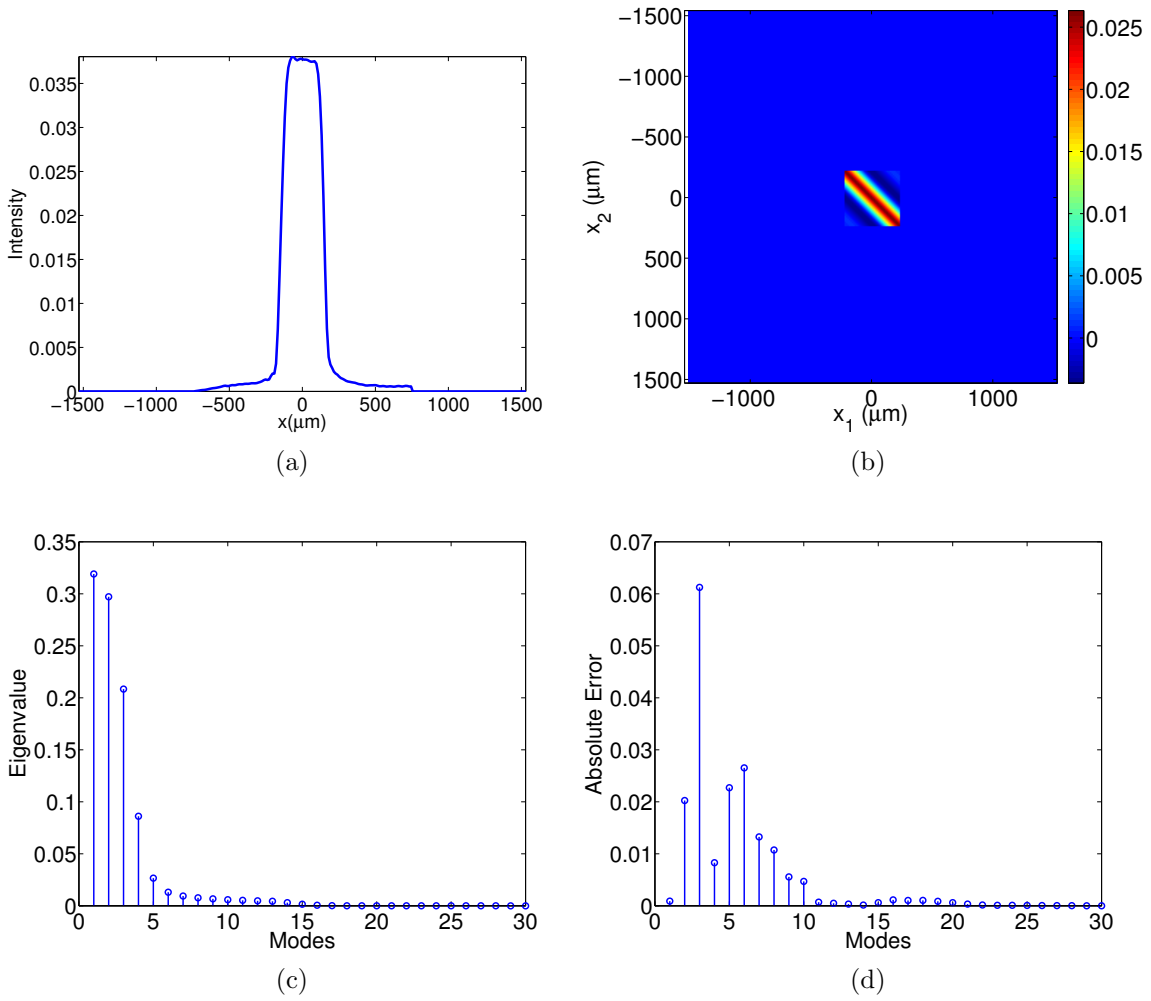
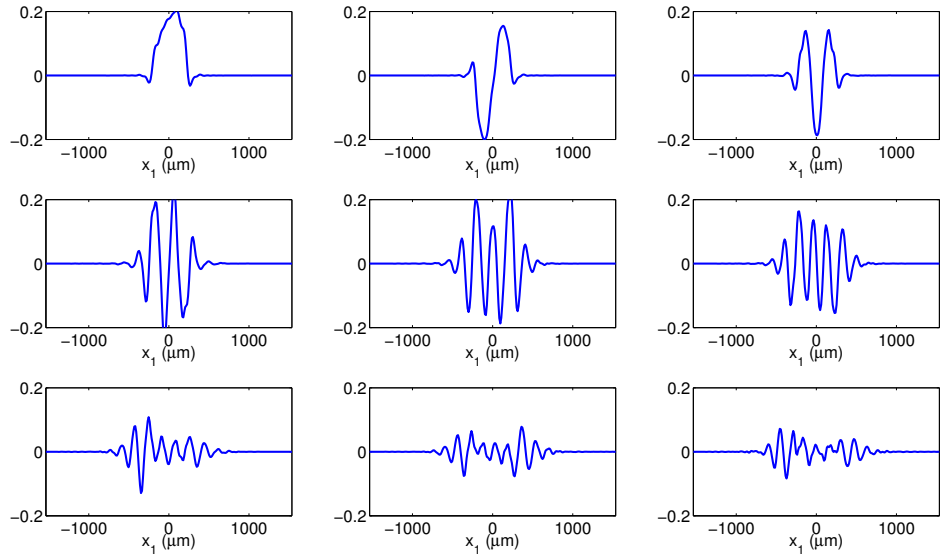
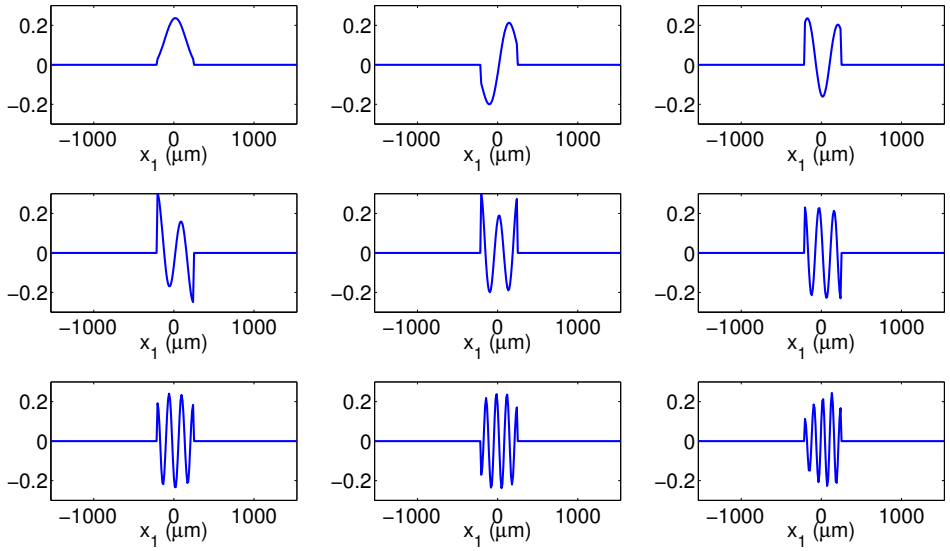


Figure 6-12: (a) Intensity measured immediately to the right of the illumination slit; (b) real part of van Cittert–Zernike theorem estimated mutual intensity immediately to the right of the object slit; (c) eigenvalues of the mutual intensity in (b); (d) absolute error between the eigenvalues in Fig. 6-11(b) and 6-12(c) versus mode index.



(a)



(b)

Figure 6-13: (a) LRMR estimated coherent modes of the mutual intensities in Fig. 6-10(b), and (b) coherent modes of the mutual intensities in Fig. 6-12(b), calculated via use of the van Cittert–Zernike theorem, and assumption of incoherent illumination.

is still not perfectly compensated by the compressive approach, and because of other experimental imperfections.

6.5 Experimental results for 4D PST

The experimental arrangement for the 4D PST is shown in Fig. 6-14. The illumination is generated by an LED with 620nm central wavelength and 20nm bandwidth. A diffuser is used to generate a quasi–uniform intensity distribution. A rectangular aperture (100μm$\times$ 300μm) is placed immediately after the diffuser and one focal length (100 mm) to the left of a spherical lens to control the coherence of the illumination. The goal is to retrieve the mutual intensity immediately to the right of the circular iris object ($\varnothing$640μm). A pair of cylindrical lenses are used to intentionally introduce astigmatism to the system; Lens 1 is oriented along x and Lens 2 is along y, both of which have focal length 100mm. Lens 2 and the camera are placed on separate linear motion stages to allow them moving independently along the optical axis z. For each lens figuration, an axial stack containing 23 intensity images is recorded by displacing the camera between the back focal plane of Lens 1 (BFP1) and the image plane of lens 2 (IP2). A total of 12 axial stacks are taken by displacing Lens 2 from the location immediately to the right of Lens 1 to BFP1. Each image contains 512$\times$512 pixels with pixel size 2.2μm.

To process the data, each image is down–sampled to 32$\times$32 due to memory limitation for the computation. The Fourier transforms of the down–sampled intensity images are arranged into the Ambiguity space according to Eq. (6.9). Due to the scanning range limitation, only a portion of the Ambiguity space is accessible, as shown by the white region in Fig. 6-15; the missing data in the black region need to be compensated by the post–processing. The LRMR and FBP reconstructed mutual intensities are compared with the theoretical prediction based on the van Cittert–Zernike theorem in Fig. 6-16. Similar to the 2D case, LRMR reconstruction alleviates most of the artifacts in the FBP reconstruction. The reconstructed eigenvalues are shown in Fig. 6-17. The LRMR reconstructed eigenvalues are very similar to the theoretical

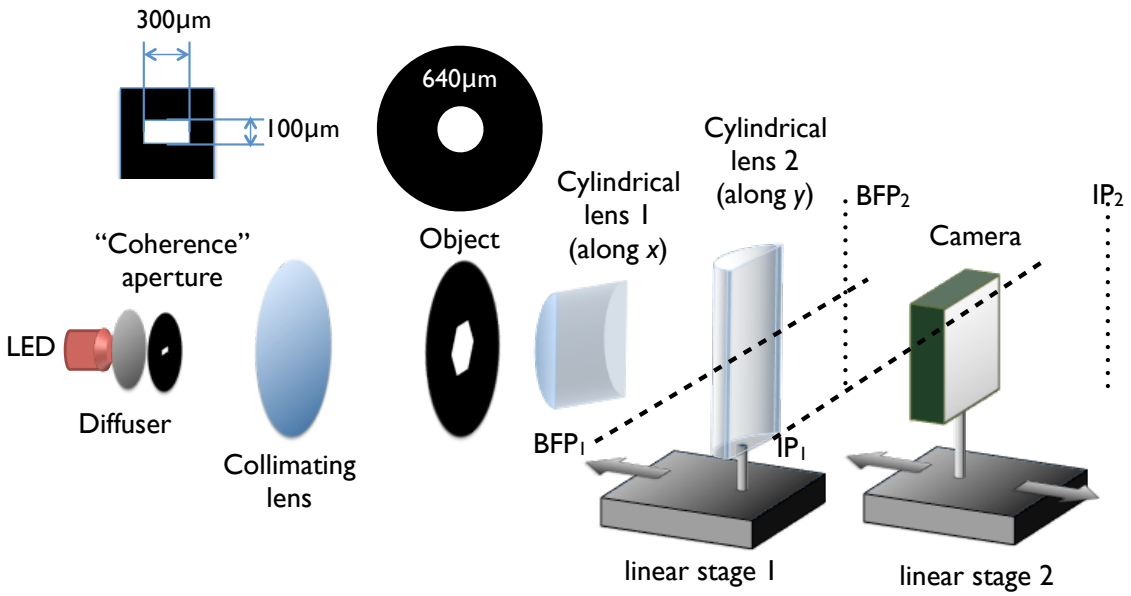


Figure 6-14: Experimental arrangement for 4D PST. A circular object is illuminated by a partially coherent illumination, which is generated by placing a rectangular aperture at the front focal plane of a spherical lens. A pair of cylindrical lenses are used to intentionally introduced astigmatism to the system. During the measurement, both the second cylindrical lens oriented along y and the camera is scanned axially.

prediction, while the FBP result contains unphysical negative eigenvalues. The first 9 coherent modes estimated from the LRMR reconstruction and the van Cittert–Zernike theorem are compared in Fig. 6-18. Since the coherence aperture is wider along the horizontal direction, the field is less coherent in this direction, resulting in more oscillatory modal structures in the coherent mode representation. The LRMR reconstruction clearly matches this prediction.

6.6 Discussion

In conclusion, we experimentally demonstrated compressive reconstruction of the mutual intensity function of a classical partially coherent source using phase space tomography. By exploiting the physically justifiable assumption of a quasi–pure source, both measurement and post–processing dimensionality are greatly reduced. We used the van Cittert–Zernike theorem to estimate the true mutual intensity function as a way to cross–validate the compressive reconstruction, and found indeed good agree-

ment.

In this classical experiment, we have the benefit that direct observation of the object is available; thus, we were able to carry out quantitative analysis of the accuracy of the compressive estimate. In the quantum analogue of measuring a complete quantum state, direct observation would have of course not been possible, but the accuracy attained through the compressive estimate should be comparable, provided the low entropy assumption holds [53].

Here we followed a much simplified version of the approach described in [26] which showed that the complex operators describing the measurements should be uniformly distributed in the n–dimensional unit sphere, whereas we simply utilized free space propagation. The phase masks described in [26] to implement optimal sampling are outside the scope of the present work.

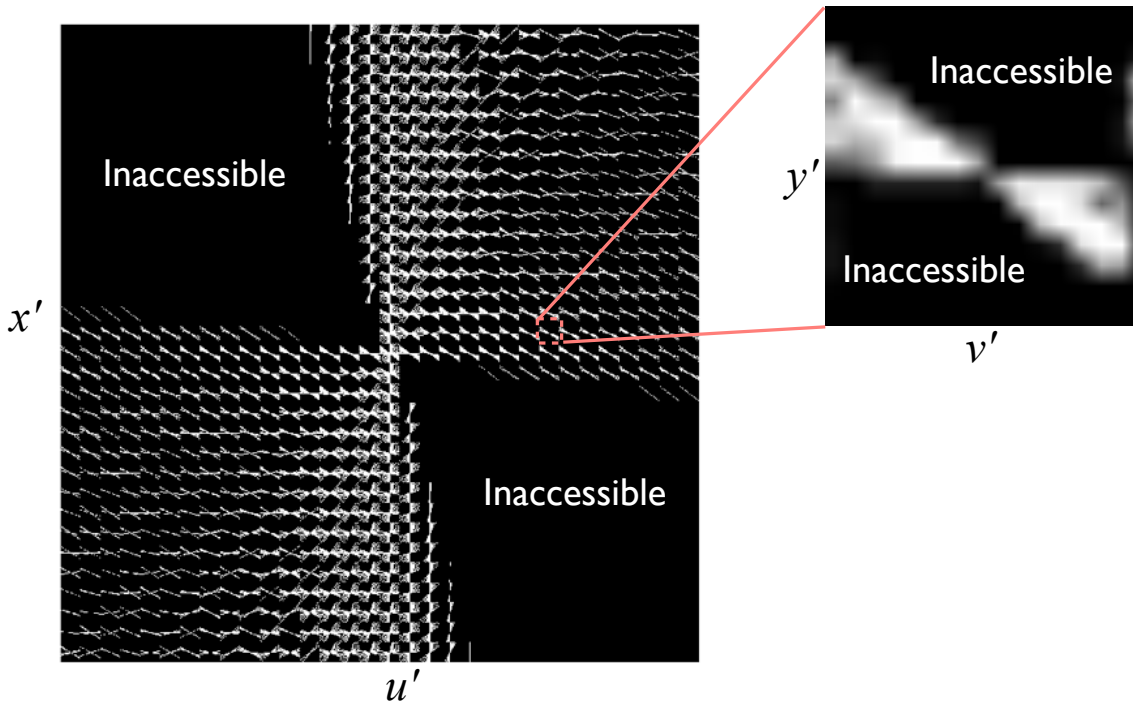


Figure 6-15: Missing data problem in the 4D PST. The white region indicates the locations of the measured data in the Ambiguity space, and the black region the inaccessible points due to the limitation of the experimental arrangement.

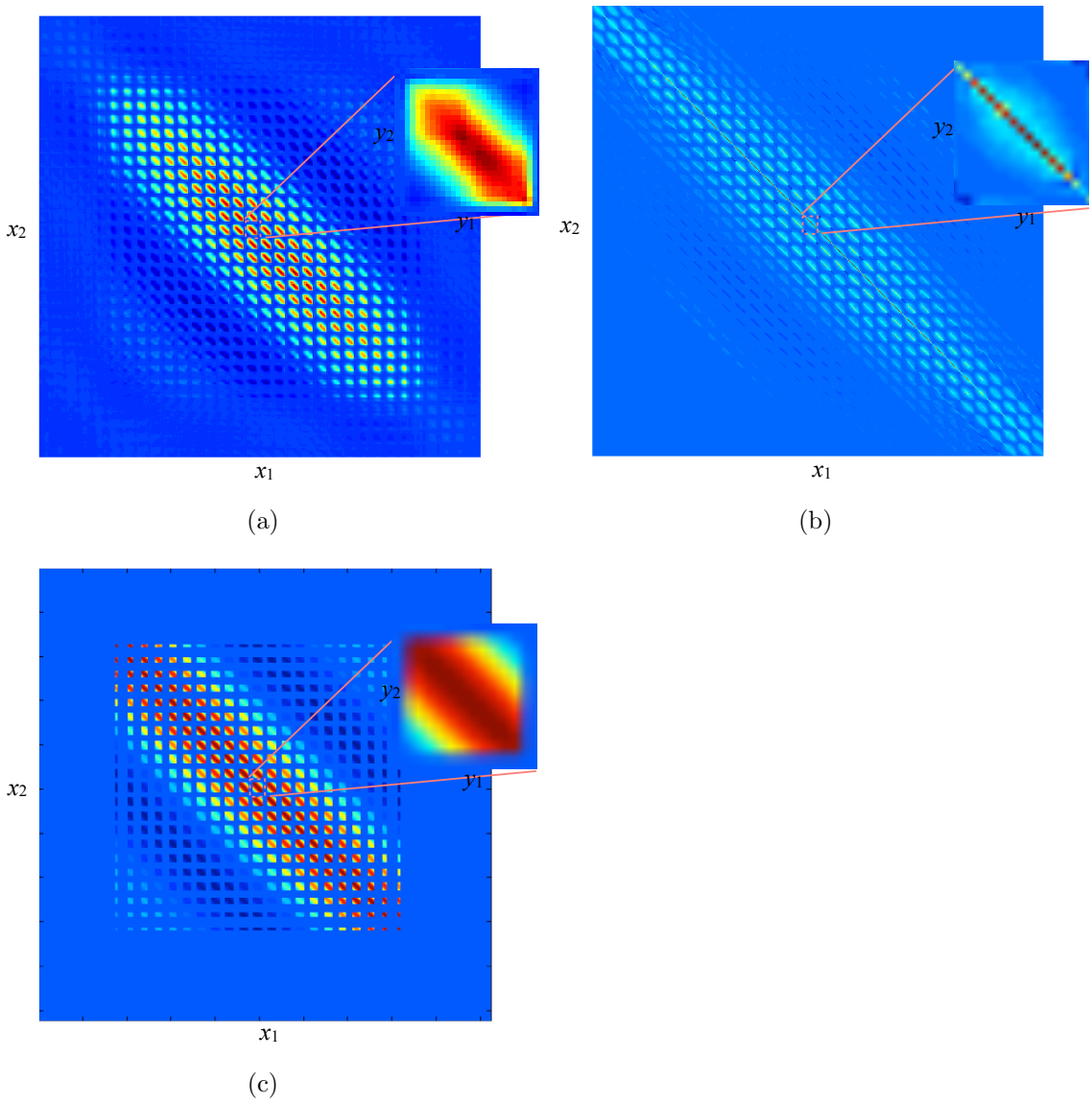


Figure 6-16: Real part of the mutual intensity from (a) LRMR, (b) FBP reconstructions, and (c) the van Cittert–Zernike theorem estimation.

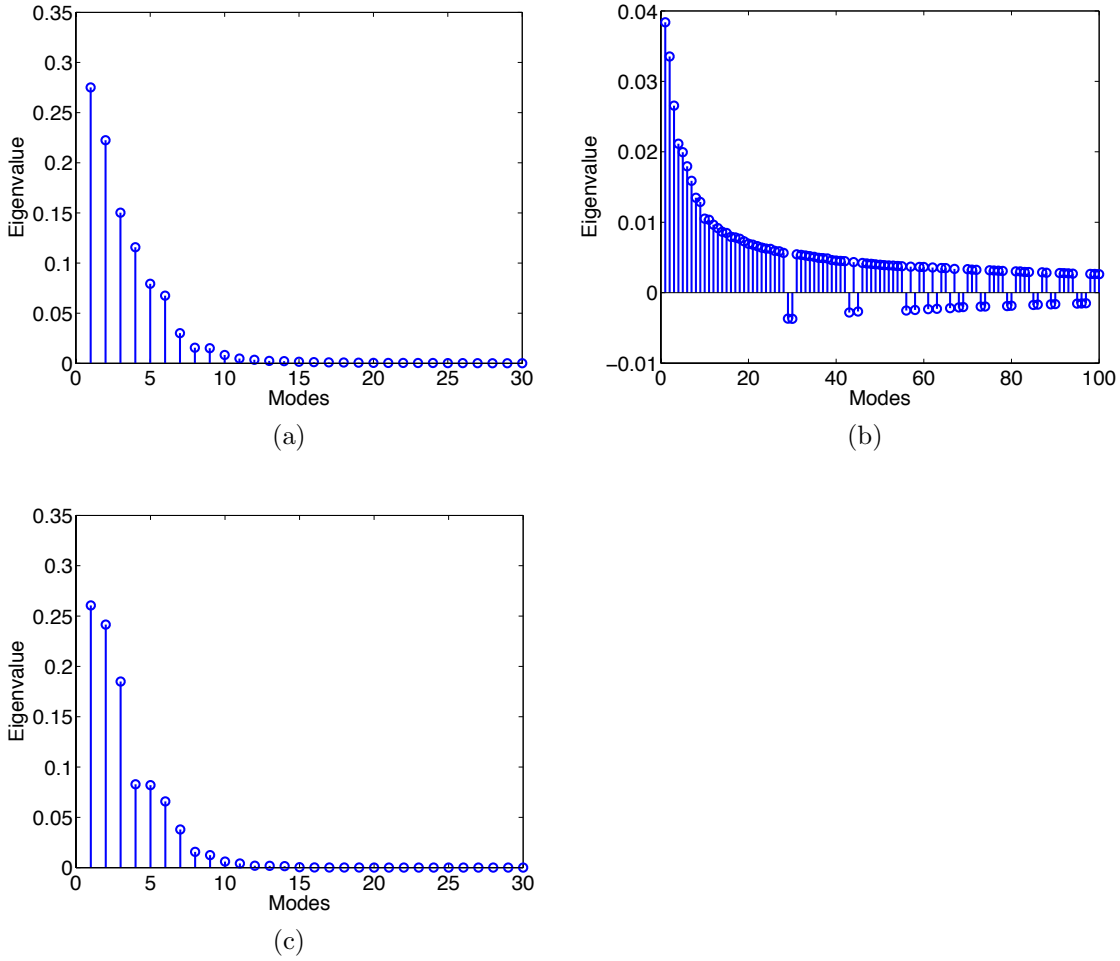
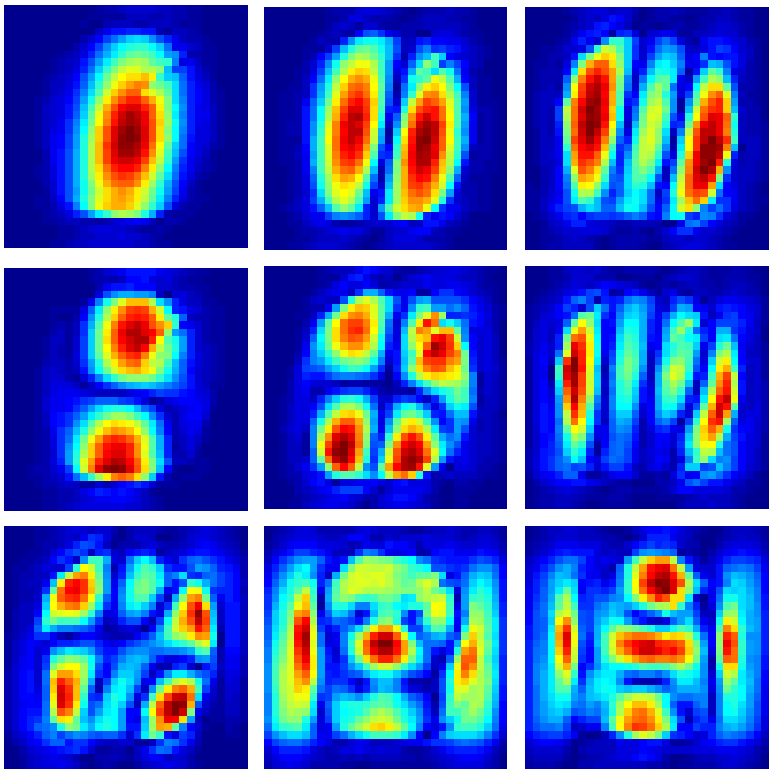
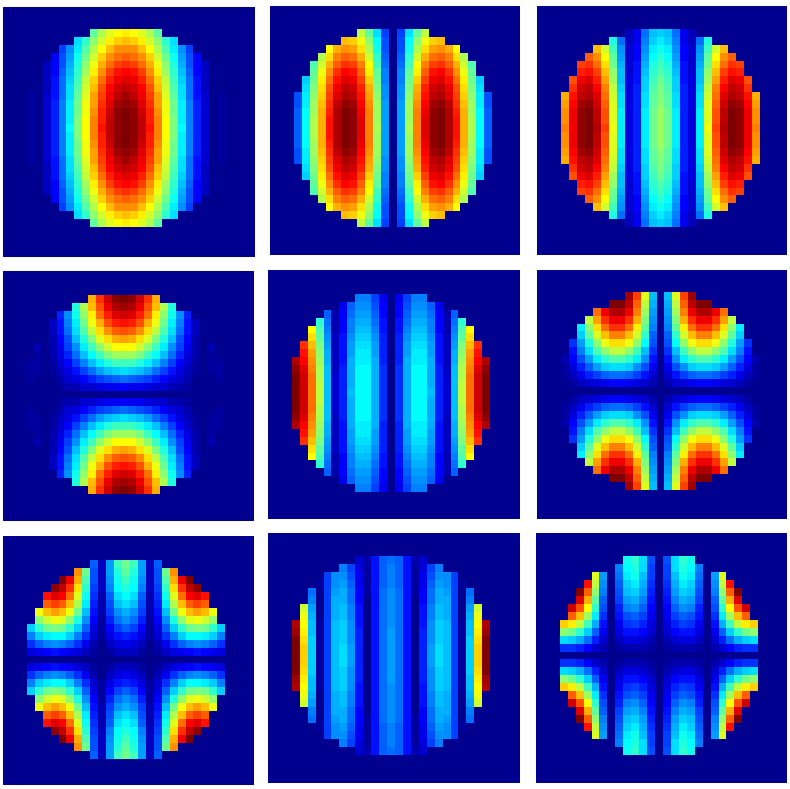


Figure 6-17: Eigenvalues estimated from (a) LRMR, (b) FBP reconstructions, and (c) the van Cittert–Zernike theorem.



(a)



(b)

Figure 6-18: The first 9 coherent modes estimated from (a) LRMR, and (b) the van Cittert–Zernike theorem.

Chapter 7

Wigner function measurement using a lenslet array

7.1 Introduction

In this chapter, another non–interferometric coherence recovery technique based on the measurement of the radiance will be described. The radiance is a function that characterizes the distribution of the power of an optical field over position and direction. Let the radiance at some plane z be denoted by $B(\mathbf{r}, \mathbf{p}; z)$ where $\mathbf{p} = (p_x, p_y)$ is the transverse component of the unit direction vector. Propagation of radiance is based on geometric optics and is simple to calculate:

$$B(\mathbf{r}, \mathbf{p}; z) = B(\mathbf{r} - z\mathbf{p}, \mathbf{p}; 0). \tag{7.1}$$

The intensity at any point is given by the integral of B over all directions

$$I(\mathbf{r}; z) = \int B(\mathbf{r}, \mathbf{p}; z) \mathrm{d}^2\mathbf{p} = \int B(\mathbf{r} - z\mathbf{p}, \mathbf{p}; 0) \mathrm{d}^2\mathbf{p}. \tag{7.2}$$

Use of the radiance predates the wave theory of light, and it was initially described by assigning non-negative values to all trajectories coming from source points. Such descriptions are insufficient to model wave effects since these trajectories contain

no information about constructive or destructive interference. However, in certain situations, these wave effects can be safely ignored, and a lenslet array can then be used to obtain an estimate for the radiance since they allow joint measurement of the spatial and directional distribution of light, as in Shack-Hartmann sensors [90], integral imaging systems [66, 104, 86] and light field cameras [1, 78].

More recently, generalized radiance functions have been proposed that describe interference while satisfying Eqs. (7.1) and (7.2); B is allowed to become negative to account for destructive interference, and it is also allowed to have non–zero values outside of the source [46]. The generalized radiance and mutual intensity both contain equivalent information: when defined over a plane, they are 4D functions that fully describe the intensity of an optical field (including wave effects) over a region of free space. In fact, for a quasi–monochromatic paraxial field of mean wavelength λ, the two descriptions are related by a simple Fourier transform relationship [126, 37, 122]:

$$\frac{1}{\lambda^2} B\left(\mathbf{r}, \mathbf{u}\lambda\right) = \mathcal{W}(\mathbf{r}, \mathbf{u}) = \int J(\mathbf{r} + \mathbf{r}'/2, \mathbf{r} - \mathbf{r}'/2) e^{-\mathrm{i}2\pi \mathbf{u}\cdot\mathbf{r}'} \mathrm{d}^2\mathbf{r}', \tag{7.3}$$

where $\mathcal{W}$ is known as the Wigner distribution function (WDF); it is related to the radiance by a change of variables from transverse direction vector $\mathbf{p}$ to spatial frequency $\mathbf{u} = (u_x, u_y) = \mathbf{p}/\lambda$. From Eq. (7.3), it is evident that if $\mathcal{W}$ is known, an inverse Fourier transform can be used to obtain the mutual intensity.

It turns out that not only can lenslet arrays measure radiance in the geometric optics sense, they can also provide information about the WDF when wave optics is considered. In order to understand how a lenslet array measurement relates to the WDF, we first consider a single lenslet with an aperture function p centered at position $\mathbf{r}$ and placed one focal length f away from a detector [133]. If this lens is illuminated by a field with WDF $\mathcal{W}_\mathrm{i}$, then the intensity in the detector plane is a function of both the incident WDF $\mathcal{W}_\mathrm{i}$ and the WDF $\mathcal{W}_\mathrm{p}$ of a plane wave illuminating the aperture [4, 123, 30, 133, 119]:

$$I(\mathbf{r} + \lambda f \mathbf{u}) = \frac{1}{\lambda f} S[\mathcal{W}_\mathrm{i}, \mathcal{W}_\mathrm{p}](\mathbf{r}, \mathbf{u}), \tag{7.4}$$

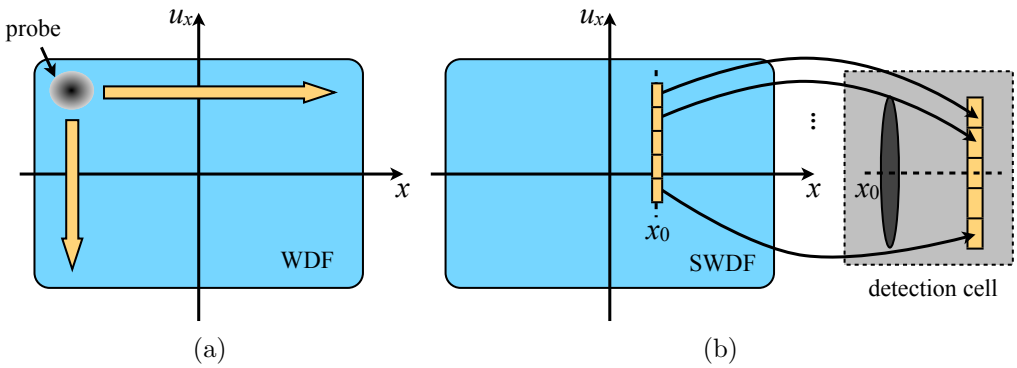


Figure 7-1: Illustration of a lens measurement of the SWDF in 1D. (a) A probe scans across the WDF forms an SWDF measurement according to the convolution in Eq. (7.5). (b) Points at x_0 and along u_x in the SWDF domain are mapped to detector pixels at $x_0 + \lambda f u_x$ at the lens' Fourier plane according to Eq. (7.4).

where $S[\mathcal{W}, P]$ denotes the smoothed WDF, defined as a convolution between two WDFs:

$$S[\mathcal{W}, P](\mathbf{r}, \mathbf{u}) = \int \mathcal{W}(\mathbf{r}', \mathbf{u}') P(\mathbf{r}' - \mathbf{r}, \mathbf{u}' - \mathbf{u}) \mathrm{d}^2\mathbf{r}' \mathrm{d}^2\mathbf{u}'. \tag{7.5}$$

The SWDF can be thought of as measuring $\mathcal{W}$ at each point by taking a weighted average of a neighborhood around the point, with weighting function P, as illustrated in Fig. 7-1(a). Equation (7.4) also implies that the detector pixel at $\mathbf{r}+\lambda f\mathbf{u}$ is mapped to the unique point at $(\mathbf{r}, \mathbf{u})$ of the SWDF, illustrated in Fig. 7-1(b). In order to fully measure the SWDF, one could scan the lenslet and take multiple images [123, 119]. If it were possible to use an aperture for which $P = \delta(\mathbf{r})\delta(\mathbf{u})$, where δ is the Dirac delta function, the WDF could be measured with perfect resolution. However it may be easily verified from Eq. (7.3) that the existence of this WDF is physically impossible; uncertainty relationships place limits on the product of the widths of P in space and spatial frequency. The finite width of P limits the resolution in space and spatial frequency of the measurement of $\mathcal{W}_\mathrm{i}$. Despite these limitations, in many cases the SWDF itself is useful to provide a direct estimate of the WDF [123, 119]. However, if the exact WDF is required, Eq. (7.5) shows that the recovery of the WDF from the SWDF generally requires deconvolution [30].

A periodic array of lenslets enables measurements of different $\mathbf{r}$ simultaneously

and removes the need to scan. The advantage of using lenslet arrays is that the SWDF may be measured in a single snapshot. However, if light passing through different lenslets overlaps at the detector, the unique mapping implied by Eq. (7.4) no longer holds. In this chapter, we conduct a more rigorous analysis, considering the mapping between pixels on a detector and point samples of the SWDF in the presence of cross–talk. For simplicity, we consider only scalar fields in one spatial dimension. We show that the intensity at a detector pixel in general contains contributions from multiple points of the SWDF due to cross–talk. In Section 2, we develop a rigorous mathematical theory and show that both fully incoherent and fully coherent cases have large amounts of cross–talk; it is caused by the large angular spread in the former and high number of interference terms in the latter. In Section 3, we illustrate tradeoffs between coherence and fidelity using a numerical example, showing that there exists an optimal "Goldilocks" regime for array pitch, given the the coherence width of the input light, such that cross–talk is reduced to a minimum without the need for additional barriers to block light between lenslets. It is in this optimal regime that each detector pixel corresponds to a single point in the SWDF domain, allowing lenslet array systems to measure the SWDF with high accuracy.

7.2 Theory

We now present a rigorous analysis of a 1D field passing through a 1D lenslet array. This analysis can be easily extended to a 2D rectangular array; other configurations, such as a 2D hexagonal array, require straightforward modifications. Consider a quasi–monochromatic paraxial field with wavelength λ incident upon an ideal lenslet array with 100% fill factor, as illustrated in Fig. 7-2. A detector is placed at the back focal plane of the lenslet array, and we will refer to the region directly behind each lenslet on the detector as that lenslet's detector cell. We assume an array of $2N + 1$ identical unaberrated thin lenses, each of width w and focal length f. The

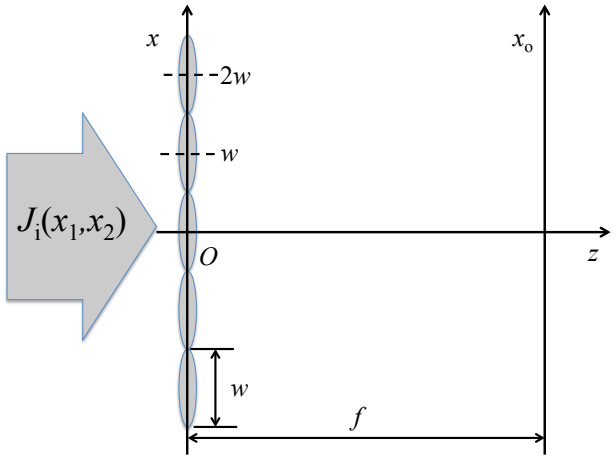


Figure 7-2: Lenslet array geometry.

transmittance function of such an array is given by

$$T(x) = \sum_{l=-N}^{N} \text{rect}\left(\frac{x - lw}{w}\right) \exp\left[-\frac{\mathrm{i}\pi}{\lambda f}(x - lw)^2\right], \tag{7.6}$$

where rect$(\cdot)$ denotes a rectangular function. In this configuration, we assign an integer index l to each lenslet, with the center lenslet having index $l = 0$; the N lenslets above and N lenslets below the center lenslet take on positive and negative values of l, respectively. Thus, the center of each lenslet is located at $x = lw$. We have assumed an odd number of lenslets to simplify notation, although the results we obtain can easily be extended to an even number of lenslets.

The mutual intensity immediately to the right of the lenslet array is [70]

$$J_1\left(\mathbf{x} + \frac{x'}{2}, \mathbf{x} - \frac{x'}{2}\right) = J_\mathrm{i}\left(\mathbf{x} + \frac{x'}{2}, \mathbf{x} - \frac{x'}{2}\right) T\left(\mathbf{x} + \frac{x'}{2}\right) T^*\left(\mathbf{x} - \frac{x'}{2}\right), \tag{7.7}$$

where $\mathbf{x}$ and x' are the center and difference coordinates, respectively, J_i is the mutual intensity of the illumination immediately before the lenslet array; the subscript i indicates that its associated function describes properties of the incident field at the input plane, and we will use this notation through the rest of the chapter.

As a stepping stone to the full relationship between the incident field and the observed intensity behind the lenslet array, we will first consider a simpler system

wherein we scan through the lenslets. That is, instead of letting light pass simultaneously through all the lenslets while recording the intensity image, we only let light pass through one lenslet at a time, cycling through all the lenslets while still recording a single image. This removes the effect of cross–lenslet interference, whose derivation we will consider later.

According to Eq. (7.4), each measurement samples the SWDF over spatial frequency with position fixed at the lenslet's center, $x = lw$. The aperture of each lenslet is a rect function of width w, and thus the weighting WDF is given by

$$\mathcal{W}_{\mathrm{r}}(x,u) = \frac{\sin[2\pi u(w-2|x|)]}{\pi u}\mathrm{rect}\left(\frac{x}{w}\right) \tag{7.8}$$

The total intensity at the detector plane is given by

$$I_0(x_{\mathrm{o}}) = \frac{1}{\lambda f}\sum_{l=-N}^{N} S[\mathcal{W}_{\mathrm{i}},\mathcal{W}_{\mathrm{r}}]\left(lw, \frac{x_{\mathrm{o}}-lw}{\lambda f}\right). \tag{7.9}$$

where $\frac{1}{\lambda f}S[\mathcal{W}_{\mathrm{i}},\mathcal{W}_{\mathrm{r}}]\left(lw, \frac{x_{\mathrm{o}}-lw}{\lambda f}\right)$ is the contribution of light through a single lenslet.

It is clear from this equation that the SWDF is sampled spatially at intervals of w, the spacing of the lenslet centers. The sampling rate along the spatial frequency axis in the SWDF is determined by both the detector pixel size and the linear mapping $u = (x_{\mathrm{o}} - lw)/\lambda f$ between detector coordinate x_{o} and spatial frequency coordinate u. The mapping can be explained by the fact that $(x_{\mathrm{o}} - lw)/f$ equals to the angle between the ray reaching the detector pixel at x_{o} and the optical axis of the l^{th} lenslet under a small angle approximation. Note that if the angular spread of the SWDF is large enough, each detector cell will include contributions to intensity not only from the SWDF associated with its lenslet, but also from neighboring lenslets. This can be prevented by increasing the size of the lenslets or by decreasing the angular spread of the incident field by placing either a main lens with finite numerical aperture in front of the array [78] or physical barriers between lenslets [33]. If we assume that each detector cell measures only light from its associated lenslet, then we would have

a detected intensity of the following form

$$I_{\mathrm{SWDF}}(x_{\mathrm{o}}) = \frac{1}{\lambda f} \sum_{l=-N}^{N} S[\mathcal{W}_{\mathrm{i}}, \mathcal{W}_{\mathrm{r}}] \left(lw, \frac{x_{\mathrm{o}} - lw}{\lambda f} \right) \mathrm{rect} \left(\frac{x_{\mathrm{o}} - lw}{w} \right). \tag{7.10}$$

We refer to this expression as I_{SWDF}, because the intensity measured at x_{o} maps uniquely to the point $[\hat{l}w, (x_{\mathrm{o}} - \hat{l}w)/\lambda f]$ in the SWDF, where $\hat{l}$ is x_{o}/w rounded to the nearest integer.

The difference between Eq. (7.9) and Eq. (7.10) yields what we will call the 0^{th} order cross–talk term $I_{\mathrm{c}}^{(0)}$:

$$I_{\mathrm{c}}^{(0)}(x_{\mathrm{o}}) = \frac{1}{\lambda f} \sum_{l=-N}^{N} S[\mathcal{W}_{\mathrm{i}}, \mathcal{W}_{\mathrm{r}}] \left(lw, \frac{x_{\mathrm{o}} - lw}{\lambda f} \right) \left[1 - \mathrm{rect} \left(\frac{x_{\mathrm{o}} - lw}{w} \right) \right]. \tag{7.11}$$

This term describes the additional light from other lenslets if the previous assumption were to not hold. In this situation multiple points in the SWDF can contribute to the same point on the detector.

To demonstrate the sampling described by Eqs. (7.9–7.11), an array containing three lenslets (centered at $-w, 0, w$) is shown in Fig. 7-3. According to Eq. (7.9), three lines sampled at spatial coordinates $-w, 0, w$ parallel to the u–axis from the SWDF are mapped to the detector plane (marked by different colors in Fig. 7-3). To ensure one–to–one mapping, the maximum spatial frequency u_{m} of the l^{th} line sample cannot exceed $w/2\lambda f$, as shown in case (a); otherwise, points at (lw, u_{m}) and $[(l+1)w, u_{\mathrm{m}} - w/(\lambda f)]$ from the SWDF domain will be measured by the same detector pixel at $x_{\mathrm{o}} = lw + \lambda f u_{\mathrm{m}}$, as shown in case (b).

So far, we have only considered the incoherent superposition of light from all of the lenslets, whereas light passing through all of the lenslets simultaneously should create additional interference terms. Since light from lenslets separated by a distance greater than the incident field's coherence width will not create appreciable interference when mixed, it is useful to enumerate these cross–talk terms with an index n proportional to the lenslet separation. All possible pairs of lenslets with indices l' and l'' such that

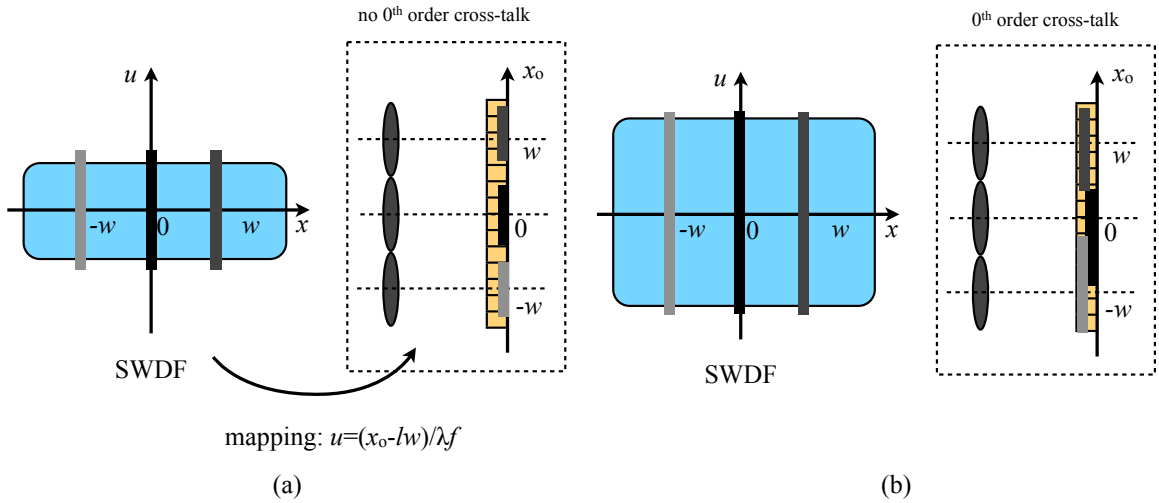


Figure 7-3: Sampling of the SWDF using an array of three lenslets. (a) One–to–one mapping from the SWDF to the detector coordinate according to $u = (x_\mathrm{o} - lw)/(\lambda f)$ as the angular spread of the SWDF is narrower than the numerical aperture of a lenslet. (b) Multiple points in the SWDF domain contribute to detector pixels in the cross–talk region as the angular spread of the incident field is wider than the numerical aperture of a lenslet, which produces the 0^{th} order cross–talk.

$|l' - l''| = n > 0$ contribute to the n^{th} order cross–talk term $I_\mathrm{c}^{(n)}$, given by

$$I_\mathrm{c}^{(n)}(x_\mathrm{o}) = \frac{2}{\lambda f} \sum_{l=-N+\frac{n}{2}}^{N-\frac{n}{2}} \iint \mathcal{W}_\mathrm{i}(\mathbf{x}, u) \mathcal{W}_\mathrm{r}\left(\mathbf{x} - lw, u - \frac{x_\mathrm{o} - lw}{\lambda f}\right) \cos\left[2\pi\left(\frac{\mathbf{x} - x_\mathrm{o}}{\lambda f} + u\right) nw\right] \mathrm{d}\mathbf{x}\mathrm{d}u. \tag{7.12}$$

Note that when n is odd, l takes a value halfway between two integers, and thus $\mathcal{W}_\mathrm{r}$ is centered at the edge between the $(l - 1/2)^{\mathrm{th}}$ and $(l + 1/2)^{\mathrm{th}}$ lenslets; when n is even, l takes every integer value, thus $\mathcal{W}_\mathrm{r}$ is centered at the l^{th} lenslet. We expect the $n = 1$ term to be significant even in highly incoherent fields, since some points near the boundary between two neighboring lenslets are expected to be within the coherence width of the field.

The total output intensity, considering all of the discussed effects, can be written as the sum of three components

$$I(x_\mathrm{o}) = I_\mathrm{SWDF}(x_\mathrm{o}) + I_\mathrm{c}^{(0)}(x_\mathrm{o}) + \sum_{n=1}^{2N} I_\mathrm{c}^{(n)}(x_\mathrm{o}). \tag{7.13}$$

A detailed derivation of this result, obtained by performing Fresnel propagation integrals on Eq. (7.7), is given in Appendix A. Equation (7.13) demonstrates that if all orders of the cross–talk could be made small, then the measured intensity would be an accurate representation of the SWDF. In order for the cross–talk to be negligible, both the angular spread of the SWDF should be small [for $I_{\mathrm{c}}^{(0)}(x_{\mathrm{o}})$] and the coherence width should be less than the width of a single lenslet [for $I_{\mathrm{c}}^{(n)}(x_{\mathrm{o}})$]. In order to optimally measure the SWDF, the angular and coherence widths of the SWDF should be balanced so that as much of each lenslet's detector cell is utilized as possible while minimizing cross–talk. It should also be noted that even with minimal cross-talk, the measurement yields only the SWDF; recovery of the mutual intensity (or WDF) of the field still requires deconvolution of the SWDF with the aperture WDF.

7.3 Numerical Example

We study the effect of coherence width on the quality of the resulting measurement by studying the following example. Let us consider a spatially homogeneous Gaussian–correlated Schell–model incident field, which can be described by the following mutual intensity,

$$J_{\mathrm{i}}(x_1, x_2) = \exp\left[-\frac{(x_2 - x_1)^2}{2\sigma_{\mathrm{c}}^2}\right]. \tag{7.14}$$

The coherence width is proportional to the standard deviation σ_{c} of the coherence term. The WDF of the incident field is

$$\mathcal{W}_{\mathrm{i}}(\mathbf{x}, u) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{u}}} \exp\left[-\frac{u^2}{2\sigma_{\mathrm{u}}^2}\right], \tag{7.15}$$

where $\sigma_{\mathrm{u}} = 1/(2\pi\sigma_{\mathrm{c}})$ quantifies the spatial frequency bandwidth of the WDF and is proportional to the angular spread of the field. The SWDF resulting from the convolution between the WDF of the input field and that of a rectangular aperture is

$$S[\mathcal{W}_{\mathrm{i}}, \mathcal{W}_{\mathrm{r}}](\mathbf{x}, u) = \frac{w^2}{\sqrt{2\pi}\sigma_{\mathrm{u}}} \int \exp\left[-\frac{(u - u')^2}{2\sigma_{\mathrm{u}}^2}\right] \left[\frac{\sin(\pi w u')}{\pi w u'}\right]^2 \mathrm{d}u'. \tag{7.16}$$

Therefore the SWDF term in Eq. (7.13) is

$$I_{\mathrm{SWDF}}(x_{\mathrm{o}}) = \frac{w^2}{\sqrt{2\pi}\sigma_{\mathrm{u}}\lambda f}\sum_{l=-N}^{N}\mathrm{rect}\left(\frac{x_{\mathrm{o}}-lw}{w}\right)\int\exp\left\{-\frac{[(x_{\mathrm{o}}-lw)/\lambda f-u']^2}{2\sigma_{\mathrm{u}}^2}\right\}\left[\frac{\sin(\pi w u')}{\pi w u'}\right]^2\mathrm{d}u', \tag{7.17}$$

and the 0^{th} order cross–talk term is

$$\begin{aligned} I_{\mathrm{c}}^{(0)}(x_{\mathrm{o}}) &= \frac{w^2}{\sqrt{2\pi}\sigma_{\mathrm{u}}\lambda f}\sum_{l=-N}^{N}\left[1-\mathrm{rect}\left(\frac{x_{\mathrm{o}}-lw}{w}\right)\right] \\ &\quad \int\exp\left\{-\frac{[(x_{\mathrm{o}}-lw)/\lambda f-u']^2}{2\sigma_{\mathrm{u}}^2}\right\}\left[\frac{\sin(\pi w u')}{\pi w u'}\right]^2\mathrm{d}u'. \end{aligned} \tag{7.18}$$

The n^{th} order cross–talk term by carrying out the integration in Eq. (7.12) is

$$\begin{aligned} I_{\mathrm{c}}^{(n)}(x_{\mathrm{o}}) &= \frac{2w^2}{\sqrt{2\pi}\sigma_{\mathrm{u}}\lambda f}\sum_{l=-N+\frac{n}{2}}^{N-\frac{n}{2}}\int\exp\left\{-\frac{[(x_{\mathrm{o}}-lw)/\lambda f-u']^2}{2\sigma_{\mathrm{u}}^2}\right\}\cos(2\pi n w u') \\ &\quad \frac{\sin[\pi w(nu_0+2u')/2]}{\pi w(nu_0+2u')/2}\frac{\sin[\pi w(nu_0-2u')/2]}{\pi w(nu_0-2u')/2}\mathrm{d}u', \end{aligned} \tag{7.19}$$

where $u_0 = w/\lambda f$ is the spatial frequency support of a single lenslet.

Three different cases with varying coherence widths are simulated based on the results in Eqs. (7.16–7.19). In the simulation, the wavelength of the incident field is 500nm. Five lenslets are used in the array, each having width $w = 330\mu$m and focal length $f = 5$mm, yielding spatial frequency support of $u_0 = 0.132\mu\mathrm{m}^{-1}$. The simulation results are shown in Fig. 7-4. For all three cases, the total output intensity in row (a) is composed of the SWDF term in row (b) and the total contribution of cross–talk in row (c). The total cross–talk is further analyzed by decomposing it as the 0^{th} order term in row (d) and the total of higher order terms in row (e). Simulations on arrays with larger numbers of lenslets were also conducted; results are not shown here because they are very similar to the ones in Fig. 7-4.

In the highly incoherent case shown in the left column ($\sigma_{\mathrm{c}} = 0.01w$), higher order cross–talk is minimal. However, due to the large angular spread in the incident field, the measurement is corrupted by 0^{th} order cross–talk. The opposite is the highly

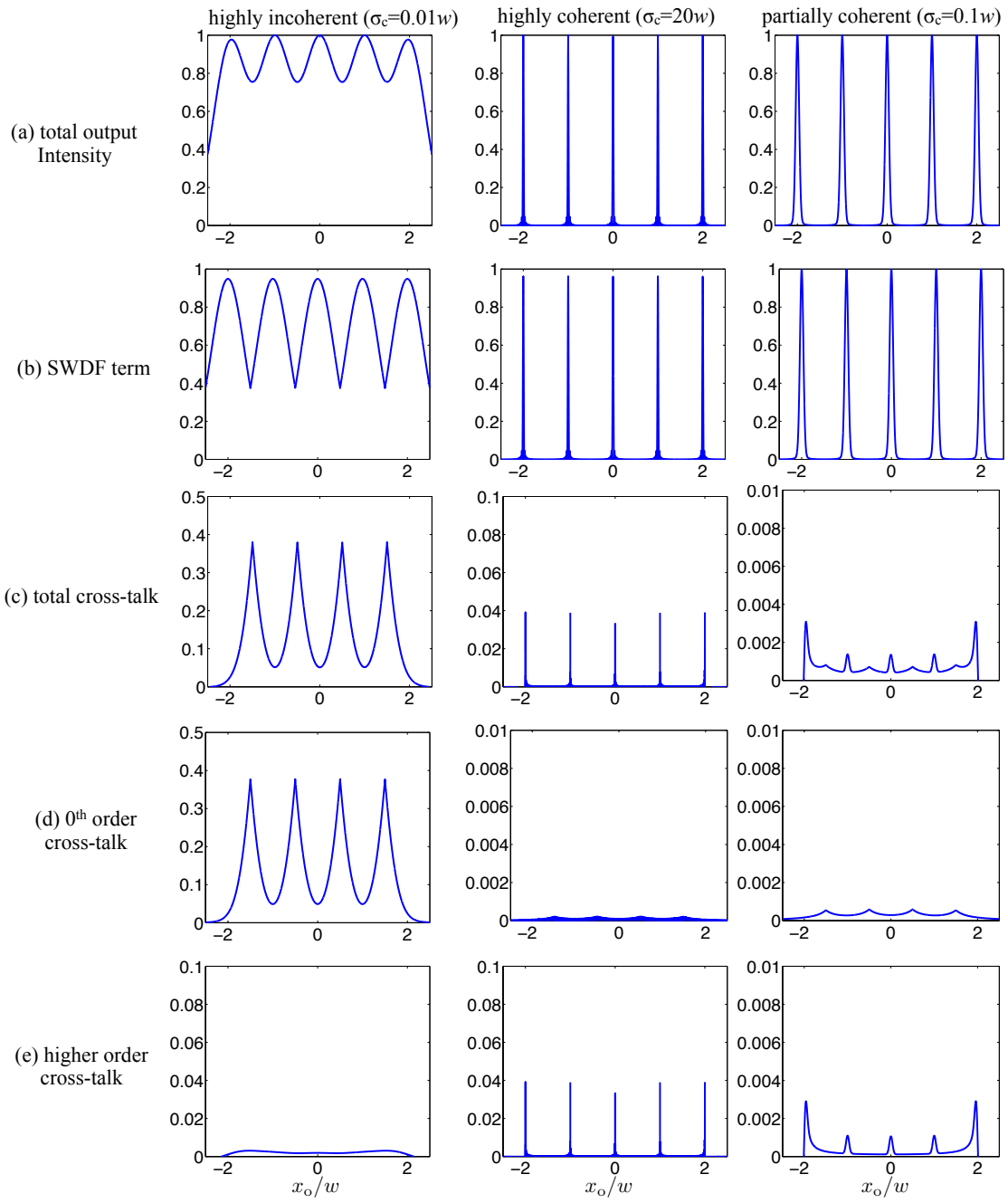


Figure 7-4: Left: highly incoherent; middle: highly coherent; and right: partially coherent case. (a) Total output intensity is composed of (b) SWDF term and (c) total contribution from cross–talk terms. The total cross–talk is composed of (d) 0^{th} order cross–talk and (e) total of higher order cross–talk. All the intensities are normalized to the maximum value in the total output. The horizontal axis is the spatial coordinate normalized by the width of a lenslet.

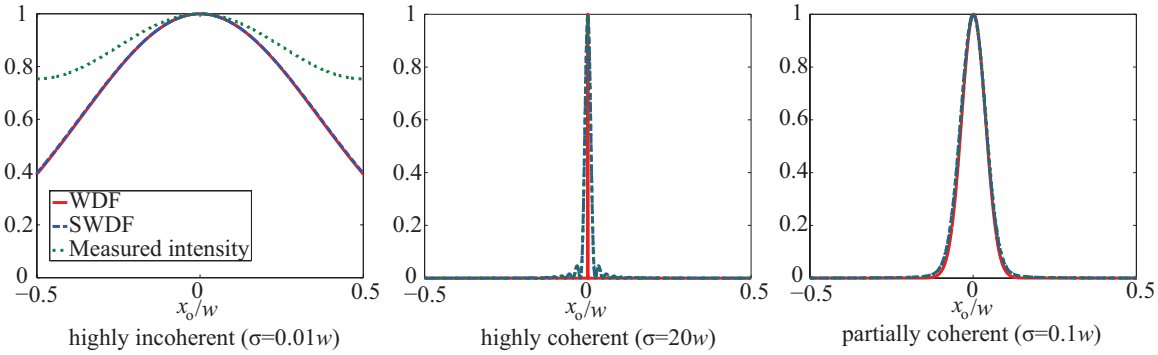


Figure 7-5: Comparison of WDF (solid red line), SWDF (dashed blue lines) and measured intensity (dotted green lines) for (a) highly incoherent ($\sigma_c = 0.01w$), (b) highly coherent ($\sigma_c = 20w$), and (c) partially coherent ($\sigma_c = 0.1w$) incident light.

coherent case, shown in the middle column ($\sigma_c = 20w$). Here, most of the cross–talk comes from higher order terms. The results for a partially coherent field ($\sigma_c = 0.1w$) is shown in the right column; cross–talk contributes minimally to the final intensity, although both 0^{th} order and higher order terms are present.

The effect of the convolution with the aperture on the accuracy of WDF estimation is studied as follows. In Fig. 7-5 we compare the measured intensity to the actual SWDF and WDF for each of the three different fields. Since the fields are spatially homogeneous, we only look at slices along u of the WDF (solid red lines) and SWDF (dashed blue lines), mapped to the x_o coordinate. The intensity distribution behind the central lenslet is shown using dotted green lines. In both the highly incoherent and partially coherent cases, the SWDF and WDF are very similar, since the WDF of the aperture is much smaller than any variations in the incident WDFs. In the highly coherent case, the incident WDF is narrower in u than the aperture WDF, and therefore the SWDF is significantly broadened by the convolution. In order to recover the WDF from the measured intensity, deconvolution is necessary [30].

We define a total error metric R_{error} by measuring the difference between the output intensity and the original WDF as

$$R_{\text{error}} = \frac{|\text{output intensity} - \text{sampled WDF}|^2}{\text{total output power}}. \tag{7.20}$$

Note that the "sampled WDF" refers to the values of the WDF over a single line in $x-u$ space corresponding to the line of the SWDF mapped to the output intensity. To quantify the cross–talk corruption in the output, define the cross–talk power fraction $R_{\text{cross–talk}}$ as

$$R_{\text{cross–talk}} = \frac{\text{total power in all terms of cross–talk}}{\text{total output power}}. \tag{7.21}$$

The signal broadening caused by the convolution in the SWDF is also considered by qualifying the difference between the SWDF and the WDF by the signal broadening metric R_{conv} defined as

$$R_{\text{conv}} = \frac{|\text{sampled SWDF} - \text{sampled WDF}|^2}{\text{total output power}}. \tag{7.22}$$

Here, both the SWDF and WDF are sampled over the same line.

All these variables are plotted as functions of the coherence of incident light (measured by the ratio of σ_{c} to w) in Fig. 7-6. As seen in the dashed green curve, the contribution from cross–talk increases quickly as the field becomes less coherent. When the field becomes more coherent, the contribution from cross–talk also increases until it saturates to the point in which the field is coherent within the whole array. There exists a partially coherent regime where the SWDF can be measured with minimal cross–talk corruption. Depending on accuracy requirements, this regime may provide acceptable measurements. For example, if less than 1% of cross–talk can be tolerated, then the coherence width should be such that $0.02w < \sigma_{\text{c}} < w$. On the other hand, signal broadening increases as the field becomes more coherent, making the SWDF a less accurate estimate of the WDF in these situations. The total error metric, which considers artifacts from both cross–talk and signal broadening, has a similar shape to the cross–talk curve. The measurement deviates from the original WDF except in a partially coherent region. If error needs to be at most 1%, then we would need $0.02w < \sigma_{\text{c}} < 0.4w$.

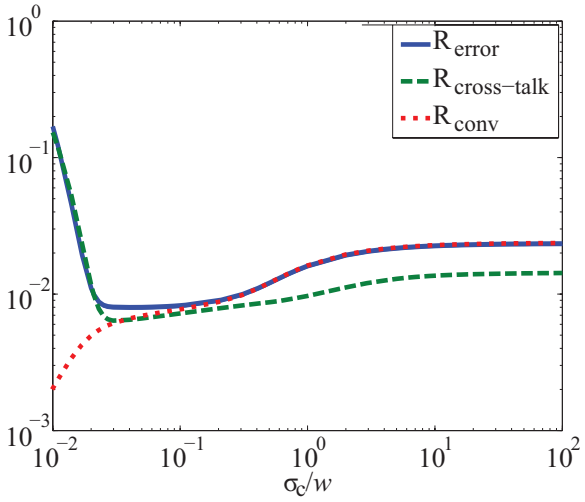


Figure 7-6: Error metric R_{error} in solid blue curve, cross–talk power fraction $R_{\text{cross–talk}}$ in dashed green curve, and signal broadening metric R_{conv} in red dotted curve as functions of the normalized coherence length of incident light σ_c/w.

7.4 Concluding Remarks

Although the numerical example was chosen explicitly to consider the effect of coherence width on the measurement of the SWDF using a lenslet array, this simple model can also provide useful insights for a much broader class of fields whose intensity varies slowly across the field, with features much wider than the coherence width. As a rule of thumb, higher order (coherent) cross-talk can be reduced by ensuring that the lenslet apertures are at least one coherence width in size. This makes intuitive sense, since an aperture larger than the coherence width will not cause the incident beam to diffract significantly, and any light that is diffracted from the aperture will not interfere with that from neighboring lenslets. Both 0^{th} and higher order crosstalk can be reduced by ensuring the incident illumination's angular spread is such that each lenslet primarily illuminates only the pixels lying within its detector cell, such that there is a nearly one–to–one mapping from SWDF space to each detector pixel.

It should also be noted that we have derived these results under the paraxial approximation and that both the 0^{th} and higher order cross–talk can include contributions for which the light propagates highly non–paraxially from one lenslet to its neighbors. In these cases, we expect that a similar analysis can be performed using

non–paraxial versions of the Wigner function [128, 32], although this is outside the scope of our current work.

As was discussed while analyzing the example, there are cases where the SWDF is not an accurate estimate of the WDF. Performing deconvolution to recover the WDF may benefit from techniques such as coded apertures [65, 47] and compressed sensing [112, 134].

Chapter 8

Conclusion and future work

This thesis has explored various imaging methods and computational algorithms for quantitative phase retrieval. A fundamental problem that the author hope to address is the recovery of a full description of wave fields (either fully coherent or partially coherent) with a small number of measurements. Compressed sensing provides an intriguing recipe that allows the recovery of an unknown signal with the number of measurements scaled by the 'information rate' (sparsity) of the signal. A central idea that has been explored throughout is that free space propagation provides easily accessible phase information. For instance, by recording the propagated wavefront through holography, 3D information can be efficiently encoded on a 2D plane; the transport of intensity relies on defocus to measure the curvature of the phase; the coherence state of light can be recovered by phase space tomography using a stack of intensity measurements during propagation.

The main contributions of the thesis include:

- Proposed and numerically verified a model to analyze the maximum number of particles that can be reconstructed from an in–line hologram based on the compressed sensing framework.

- Developed and experimentally verified a form of the TIE that can be used to recover the OPL of a thin sample that works for partially coherent illumination.

- Developed an iterative algorithm based nonlinear diffusion regularization to invert the TIE to recover piecewise constant phase signals.

- Developed a compressive reconstruction model for X–ray phase tomography based on the TIE measurements. The experimental results demonstrated that both high and low–frequency artifacts are mitigated by this method.

- Proposed and experimentally verified a compressive phase space tomography model based on coherent mode decomposition to reconstruct the correlation function of a partially coherent field.

- Developed a rigorous mathematical theory to analyze the measurement from a lenslet array for partially coherent light. An optimal design allowing accurate sampling of the Wigner distribution function.

Beyond this work, there are many more unexplored problems in the application of novel sensing method to quantitative phase recovery problem. Venue to extend the current work includes:

- In holography, it is well known that the depth localization capability is much worse than the one in the lateral dimension due to the missing spatial frequency samples in the longitudinal direction. Compressive reconstruction might help to improve the result. A rigorous study on the limit of axial localization improvement by using compressive reconstruction methods may be interesting for future study.

- The effect of the partially coherent illumination on the TIE is treated as artifacts which needs to be compensated in the current study. However, active control over illumination could also create opportunities on obtaining phase information with better contrast or making the inverse problem better conditioned. Coded source and/or combination of measurements with different source figurations are interesting directions for future work.

- Phase space tomography takes projection measurements of the Wigner distribution function while rotating the phase space. It may be possible to tailor the sampling trajectories in the Wigner space beyond the tomographic approach. To do that, optical elements, such as volume holograms, coded apertures or nonlinear materials, may be introduced in the path between the unknown wave and detector. Most light sources are not fully coherent; since partially coherent waves contain many more degrees of freedom than coherent waves, they are more difficult to deal with, yet they offer more opportunity for manipulation of light in ways that coherent light cannot. This will lead to cases where information is encoded in or retrieved from the wave, with applications in optical information science, lithography, X-ray sciences, medical imaging, and biological microscopy. It is also interesting to note that an analogous problem exists in quantum mechanics for the recovery of quantum correlations (the density matrix). The understanding of this problem might have further implications to quantum information theory.

Appendix A

Proof of Eq. (7.13)

Appendix A: Proof of Eq. (7.13)

Assume the mutual intensity of the incident field is

$$J_{\mathrm{i}}(x_1, x_2) = \langle U_{\mathrm{i}}(x_1) U_{\mathrm{i}}^*(x_2) \rangle \tag{A.1}$$

where $\langle \cdot \rangle$ denotes the expectation value over a statistical ensemble of realizations of the field ψ_{i}, x_1 and x_2 denote positions at the input plane immediately to the left of the lenslet array. According to Equation (7.3), the WDF of the 1D input field is related to the mutual intensity by

$$\mathcal{W}_{\mathrm{i}}(\mathbf{x}, u) = \int J_{\mathrm{i}} \left(\mathbf{x} + \frac{x'}{2}, \mathbf{x} - \frac{x'}{2} \right) \exp(-\mathrm{i}2\pi u x') \mathrm{d}x', \tag{A.2}$$

where

$$\mathbf{x} = \frac{x_1 + x_2}{2}, \; x' = x_1 - x_2, \tag{A.3}$$

are the center and difference coordinates; u is the local spatial frequency variable. The mutual intensity immediately to the right of the lenslet array is [70]

$$J_1(x_1, x_2) = J_{\mathrm{i}}(x_1, x_2) T(x_1) T^*(x_2), \tag{A.4}$$

where T is the transmittance function of the lenslet array

$$T(x) = \sum_{l=-N}^{N} \text{rect}\left(\frac{x - lw}{w}\right) \exp\left[-\frac{\mathrm{i}\pi}{\lambda f}(x - lw)^2\right]. \tag{A.5}$$

The intensity at the output plane x_o at one focal length to the right of the lenslet array is related to J_1 by double Fresnel integrals under the paraxial approximation,

$$I(x_\mathrm{o}) = \frac{1}{\lambda f} \iint J_1(x_1, x_2) \exp\left\{\frac{\mathrm{i}\pi}{\lambda f}\left[(x_\mathrm{o} - x_1)^2 - (x_\mathrm{o} - x_2)^2\right]\right\} \mathrm{d}x_1 \mathrm{d}x_2. \tag{A.6}$$

Substitutions of Eqs. (A.5) and (A.4) into Eq. (A.6) shows that the calculation of $I(x_\mathrm{o})$ requires in general carrying out double summations with respect to different lenslet indices l_1 and l_2. After some simplification, the output intensity is rewritten as

$$\begin{aligned} I(x_\mathrm{o}) = \frac{1}{\lambda f} \sum_{l_1=-N}^{N} \sum_{l_2=-N}^{N} \iint J_\mathrm{i}(x_1, x_2) \text{rect}\left(\frac{x_1 - l_1 w}{w}\right) \text{rect}\left(\frac{x_2 - l_2 w}{w}\right) \\ \exp\left\{-\mathrm{i}\frac{2\pi}{\lambda f}\left[(l_1^2 - l_2^2)\frac{w^2}{2} + (x_1 - x_2)x_\mathrm{o} - l_1 w x_1 + l_2 w x_2\right]\right\} \mathrm{d}x_1 \mathrm{d}x_2. \end{aligned} \tag{A.7}$$

The double summations are simply due to the fact that the intensity at the region directly behind a particular lenslet results from both the field passing through the local lenslet and cross–talk from the field passing through neighboring lenslets. We proceed with the derivation by the change of variables in Eq. (A.3) and

$$m = l_1 + l_2, \;\; n = l_1 - l_2. \tag{A.8}$$

The consequence of the change of variables in Eq. (A.8) can be studied as follows.

a) When n is even, we can write

$$n = 2q, \;\; \text{where } q = -N, -N+1, ..., N-1, N; \tag{A.9}$$

and since $n = l_1 - l_2$, we can assume

$$l_1 = l + q, l_2 = l - q, \tag{A.10}$$

and

$$m = 2l, \text{ where } l = -N + |q|, -N + |q| + 1, ..., N - |q| - 1, N - |q|. \tag{A.11}$$

b) When n is odd, we can again write

$$n = 2q, \text{ where } q = -N + \frac{1}{2}, -N + \frac{3}{2}, ..., N - \frac{3}{2}, N - \frac{1}{2}, \tag{A.12}$$

and

$$l_1 = l + q, l_2 = l - q, \tag{A.13}$$

with

$$m = 2l, \text{ where } l = -N + |q|, -N + |q| + 1, ..., N - |q| - 1, N - |q|. \tag{A.14}$$

The substitution of the change of variables leads

$$\begin{aligned} I(x_{\mathrm{o}}) &= \frac{1}{\lambda f} \sum_{\substack{n=-2N \\ \text{even}}}^{2N} \sum_{l=-N+\frac{|n|}{2}}^{N-\frac{|n|}{2}} \iint \operatorname{rect}\left(\frac{(\mathbf{x} - lw) + (x' - nw)/2}{w}\right) \operatorname{rect}\left(\frac{(\mathbf{x} - lw) - (x' - nw)/2}{w}\right) \\ &\quad J_{\mathrm{i}}\left(\mathbf{x} + \frac{x'}{2}, \mathbf{x} - \frac{x'}{2}\right) \exp\left[-\mathrm{i}\frac{2\pi}{\lambda f}(x_{\mathrm{o}} - lw)\, x' + \mathrm{i}\frac{2\pi}{\lambda f}(\mathbf{x} - lw)\, nw\right] \mathrm{d}\mathbf{x}\mathrm{d}x' \\ &+ \frac{1}{\lambda f} \sum_{\substack{n=-2N+1 \\ \text{odd}}}^{2N-1} \sum_{l=-N+\frac{|n|}{2}}^{N-\frac{|n|}{2}} \iint \operatorname{rect}\left(\frac{(\mathbf{x} - lw) + (x' - nw)/2}{w}\right) \operatorname{rect}\left(\frac{(\mathbf{x} - lw) - (x' - nw)/2}{w}\right) \\ &\quad J_{\mathrm{i}}\left(\mathbf{x} + \frac{x'}{2}, \mathbf{x} - \frac{x'}{2}\right) \exp\left[-\mathrm{i}\frac{2\pi}{\lambda f}(x_{\mathrm{o}} - lw)\, x' + \mathrm{i}\frac{2\pi}{\lambda f}(\mathbf{x} - lw)\, nw\right] \mathrm{d}\mathbf{x}\mathrm{d}x'. \end{aligned} \tag{A.15}$$

Notice that a term of fixed n contributes a non–zero value to $I(x_{\mathrm{o}})$ only if the two rect–functions overlap. This implies that the separation x' between the pair of correlating points on the incident field can only take certain values, as determined by the following

inequalities

$$|\mathbf{x} - lw| < w/4, \tag{A.16}$$

$$(n-1)w + 2|\mathbf{x} - lw| < x' < (n+1)w - 2|\mathbf{x} - lw|. \tag{A.17}$$

Eq. (A.17) implies that x' is bounded to a region of width $2w - 4|\mathbf{x} - lw|$ centered at nw. Also recall that the magnitude of mutual intensity is significantly larger than zero at large separation distance x' only if the field is highly coherent. This implies that more terms in the summation over n need to be considered if the field is more coherent. To simplify Eq. (A.15), we relate $I(x_\mathrm{o})$ to the WDF of the incident field and the WDF of a rectangular aperture of width w

$$\mathcal{W}_\mathrm{r}(\mathbf{x}, u) = \int \mathrm{rect}\left(\frac{\mathbf{x} + x'/2}{w}\right) \mathrm{rect}\left(\frac{\mathbf{x} - x'/2}{w}\right) \exp(-\mathrm{i}2\pi u \cdot x')\mathrm{d}x', \tag{A.18}$$

by completing the integration with respect to x' to yield

$$\begin{aligned} I(x_\mathrm{o}) = & \\ & + \frac{1}{\lambda f} \sum_{\substack{n=-2N+1 \\ \text{odd}}}^{2N-1} \sum_{l=-N+\frac{|n|}{2}}^{N-\frac{|n|}{2}} \iint \mathcal{W}_\mathrm{i}(\mathbf{x}, u) \mathcal{W}_\mathrm{r}\left(\mathbf{x} - lw, u - \frac{x_\mathrm{o} - lw}{\lambda f}\right) \exp\left[\mathrm{i}2\pi\left(\frac{\mathbf{x} - x_\mathrm{o}}{\lambda f} + u\right) nw\right] \mathrm{d}\mathbf{x}\mathrm{d}u \\ & + \frac{1}{\lambda f} \sum_{\substack{n=-2N \\ \text{even}}}^{2N} \sum_{l=-N+\frac{|n|}{2}}^{N-\frac{|n|}{2}} \iint \mathcal{W}_\mathrm{i}(\mathbf{x}, u) \mathcal{W}_\mathrm{r}\left(\mathbf{x} - lw, u - \frac{x_\mathrm{o} - lw}{\lambda f}\right) \exp\left[\mathrm{i}2\pi\left(\frac{\mathbf{x} - x_\mathrm{o}}{\lambda f} + u\right) nw\right] \mathrm{d}\mathbf{x}\mathrm{d}u. \end{aligned} \tag{A.19}$$

Finally, by combining the complex conjugate terms in n, we arrive at Eq. (7.13)

$$
\begin{aligned}
I(x_{\mathrm{o}}) = & \frac{1}{\lambda f} \sum_{l=-N}^{N} \iint \mathcal{W}_{\mathrm{i}}(\mathbf{x}, u) \mathcal{W}_{\mathrm{r}} \left(\mathbf{x} - lw, u - \frac{x_{\mathrm{o}} - lw}{\lambda f} \right) \mathrm{d}\mathbf{x}\mathrm{d}u \\
& + \frac{2}{\lambda f} \sum_{\substack{n=1 \\ \text{odd}}}^{2N-1} \sum_{l=-N+\frac{n}{2}}^{N-\frac{n}{2}} \iint \mathcal{W}_{\mathrm{i}}(\mathbf{x}, u) \mathcal{W}_{\mathrm{r}} \left(\mathbf{x} - lw, u - \frac{x_{\mathrm{o}} - lw}{\lambda f} \right) \cos \left[2\pi \left(\frac{\mathbf{x} - x_{\mathrm{o}}}{\lambda f} + u \right) nw \right] \mathrm{d}\mathbf{x}\mathrm{d}u \\
& + \frac{2}{\lambda f} \sum_{\substack{n=2 \\ \text{even}}}^{2N} \sum_{l=-N+\frac{n}{2}}^{N-\frac{n}{2}} \iint \mathcal{W}_{\mathrm{i}}(\mathbf{x}, u) \mathcal{W}_{\mathrm{r}} \left(\mathbf{x} - lw, u - \frac{x_{\mathrm{o}} - lw}{\lambda f} \right) \cos \left[2\pi \left(\frac{\mathbf{x} - x_{\mathrm{o}}}{\lambda f} + u \right) nw \right] \mathrm{d}\mathbf{x}\mathrm{d}u.
\end{aligned}
$$

Bibliography

[1] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):99–106, 1992.

[2] B. Allman, P. McMahon, K. Nugent, D. Paganin, D. Jacobson, M. Arif, and S. Werner. Imaging: phase radiography with neutrons. *Nature*, 408:158–159, 2000.

[3] E. D. Barone-Nugent, A. Barty, and K. A. Nugent. Quantitative phase-amplitude microscopy I: optical microscopy. *Journal of Microscopy*, 206(3):194–203, 2002.

[4] H.O. Bartelt, K.-H. Brenner, and A.W. Lohmann. The Wigner distribution function and its optical production. *Optics Communications*, 32(1):32 – 38, 1980.

[5] M. J. Bastiaans. The Wigner distribution function applied to optical signals and systems. *Optics Communications*, 25:26–30, April 1978.

[6] Martin J. Bastiaans. New class of uncertainty relations for partially coherent light. *J. Opt. Soc. Am. A*, 1(7):711–715, Jul 1984.

[7] Martin J. Bastiaans. Application of the Wigner distribution function to partially coherent light. *J. Opt. Soc. Am. A*, 3(8):1227–1238, Aug 1986.

[8] M. Beck, M. G. Raymer, I. A. Walmsley, and V. Wong. Chronocyclic tomography for measuring the amplitude and phase structure of optical pulses. *Opt. Lett.*, 18(23):2041–2043, Dec 1993.

[9] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging*. Taylor & Francis, 1998.

[10] J. M. Bioucas-Dias and M. A. T. Figueiredo. A New TwIST: Two-Step Iterative Shrinkage/Thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16:2992–3004, December 2007.

[11] K. Blum. *Density matrix theory and applications*. Plenum Press, 1981.

[12] M. Born and E. Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 7 edition, October 1999.

[13] David J. Brady, Kerkil Choi, Daniel L. Marks, Ryoichi Horisaki, and Sehoon Lim. Compressive holography. *Opt. Express*, 17(15):13040–13049, Jul 2009.

[14] K.-H. Brenner, A.W. Lohmann, and J. Ojeda-Castañeda. The ambiguity function as a polar display of the OTF. *Opt. Commun.*, 44(5):323 – 326, 1983.

[15] K.-H. Brenner and J. Ojeda-Castañeda. Ambiguity function and Wigner distribution function applied to partially coherent imagery. *Opt. Acta.*, 31:213–223, February 1984.

[16] Andrei V. Bronnikov. Theory of quantitative phase-contrast computed tomography. *J. Opt. Soc. Am. A*, 19(3):472–480, Mar 2002.

[17] Anna Burvall, Ulf Lundström, Per A. C. Takman, Daniel H. Larsson, and Hans M. Hertz. Phase retrieval in x-ray phase-contrast imaging suitable for tomography. *Opt. Express*, 19(11):10359–10376, May 2011.

[18] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *ArXiv: 0810.3286*, October 2008.

[19] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[20] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, August 2006.

[21] E. J. Candès, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *ArXiv: 1109.0573*, September 2011.

[22] E. J. Candès and Y. Plan. Matrix completion with noise. *ArXiv: 0903.3131*, March 2009.

[23] E.J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, Dec. 2006.

[24] Emmanuel J. Candès and Yaniv Plan. A probabilistic and ripless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.

[25] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[26] Emmanuel J. Candès, Thomas Strohmer, and Vladislav Voroninski. Phaselift: exact and stable signal recovery from magnitude measurements via convex programming. *ArXiv: 1109.4499v1*, 09 2011.

[27] Emmanuel J. Candès and Terence Tao. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56:2053–2080, May 2010.

[28] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[29] D. Chapman, W. Thomlinson, RE Johnston, D. Washburn, E. Pisano, N. Gmür, Z. Zhong, R. Menk, F. Arfelli, and D. Sayers. Diffraction enhanced x-ray imaging. *Physics in medicine and biology*, 42(11):2015, 1999.

[30] Henry N. Chapman. Phase-retrieval X-ray microscopy by Wigner–distribution deconvolution. *Ultramicroscopy*, 66(3-4):153 – 172, 1996.

[31] F. C. Cheong, B. Sun, R. Dreyfus, J. Amato-Grill, K. Xiao, L. Dixon, and D. G. Grier. Flow visualization and flow cytometry with holographic video microscopy. *Opt. Express*, 17(15):13071–13079, 2009.

[32] S. Cho, JC Petruccelli, and MA Alonso. Wigner functions for paraxial and nonparaxial fields. *Journal of Modern Optics*, 56(17):1843–1852, 2009.

[33] Heejin Choi, Sung-Wook Min, Sungyong Jung, Jae-Hyeung Park, and Byoungho Lee. Multiple-viewing-zone integral imaging using a dynamic barrier array for three-dimensional displays. *Opt. Express*, 11(8):927–932, Apr 2003.

[34] Christy Fernandez Cull, David A. Wikner, Joseph N. Mait, Michael Mattheiss, and David J. Brady. Millimeter-wave compressive holography. *Appl. Opt.*, 49(19):E67–E82, Jul 2010.

[35] TJ Davis, D Gao, TE Gureyev, AW Stevenson, and SW Wilkins. Phase-contrast imaging of weakly absorbing materials using hard x-rays. *Nature*, 373(6515):595–598, 1995.

[36] Loïc Denis, Dirk Lorenz, Eric Thiébaut, Corinne Fournier, and Dennis Trede. Inline hologram reconstruction with sparsity constraints. *Opt. Lett.*, 34(22):3475–3477, Nov 2009.

[37] LS Dolin. Beam description of weakly-inhomogeneous wave fields. *Izv. Vyssh. Uchebn. Zaved. Radiofiz*, 7:559–563, 1964.

[38] J. A. Dominguez-Caballero and G. Barbastathis. Stability of inversion in digital holographic particle imaging: Theory and experimental validation. In *Frontiers in Optics*, page FThV4. Optical Society of America, 2008.

[39] Jose A. Dominguez-Caballero, Nick Loomis, Weichang Li, Qiao Hu, Jerome Milgram, George Barbastathis, and Cabell Davis. Advances in plankton imaging using digital holography. In *Computational Optical Sensing and Imaging*, page DMB5. Optical Society of America, 2007.

[40] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.

[41] Thomas Dresel, Gerd Häusler, and Holger Venzke. Three-dimensional sensing of rough surfaces by coherence radar. *Applied Optics*, 31(7):919–925, 1992.

[42] F. Dubois, C. Schockaert, N. Callens, and C. Yourassowsky. Focus plane detection criteria in digital holography microscopy by amplitude analysis. *Opt. Express*, 14(13):5895–5908, June 2006.

[43] J.A. Fessler and B.P. Sutton. Nonuniform fast fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560–574, 2003.

[44] J. R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21(15):2758–2769, 1982.

[45] Corinne Fournier, Loïc Denis, and Thierry Fournel. On the single point resolution of on-axis digital holography. *J. Opt. Soc. Am. A*, 27(8):1856–1862, Aug 2010.

[46] A. T. Friberg. On the existence of a radiance function for finite planar sources of arbitrary states of coherence. *J. Opt. Soc. Am.*, 69(1):192–198, 1979.

[47] Michael E. Gehm, Scott T. McCain, Nikos P. Pitsianis, David J. Brady, Prasant Potuluri, and Michael E. Sullivan. Static two-dimensional aperture coding for multimodal, multiplex spectroscopy. *Appl. Opt.*, 45(13):2965–2974, May 2006.

[48] RW Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237, 1972.

[49] J. W. Goodman. *Statistical Optics*. Wiley-Interscience, July 2000.

[50] R. Gordon, R. Bender, and G.T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. *Journal of theoretical Biology*, 29(3):471, 1970.

[51] A. Groso, R. Abela, and M. Stampanoni. Implementation of a fast method for high resolution phase contrast tomography. *Opt. Express*, 14(18):8103–8110, Sep 2006.

[52] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory*, 57(3):1548 –1566, march 2011.

[53] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105(15):150401, Oct 2010.

[54] Jean Pierre Guigay, Max Langer, Renaud Boistel, and Peter Cloetens. Mixed transfer function and transport of intensity approach for phase retrieval in the fresnel region. *Opt. Lett.*, 32(12):1617–1619, Jun 2007.

[55] T. E. Gureyev, A. Roberts, and K. A. Nugent. Partially coherent fields, the transport-of-intensity equation, and phase uniqueness. *J. Opt. Soc. Am. A*, 12(9):1942–1946, Sep 1995.

[56] T.E. Gureyev, Ya.I. Nesterets, D.M. Paganin, A. Pogany, and S.W. Wilkins. Linear algorithms for phase retrieval in the fresnel region. 2. partially coherent illumination. *Optics Communications*, 259(2):569 – 580, 2006.

[57] TE Gureyev, DM Paganin, GR Myers, Y.I. Nesterets, and SW Wilkins. Phase-and-amplitude computer tomography. *Applied physics letters*, 89(3):034102–034102, 2006.

[58] Kazuo Ishizuka and Brendan Allman. Phase measurement of atomic resolution image using transport of intensity equation. *Journal of Electron Microscopy*, 54(3):191–197, June 2005.

[59] Kazuyoshi Itoh and Yoshihiro Ohtsuka. Fourier-transform spectral imaging: retrieval of source information from three-dimensional spatial coherence. *J. Opt. Soc. Am. A*, 3(1):94–100, Jan 1986.

[60] A. C. Kak and M. Slaney. *Principle of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001.

[61] J. Katz and J. Sheng. Applications of holography in fluid mechanics and particle dynamics. *Annual Review of Fluid Mechanics*, 42(1):531–555, 2010.

[62] D.P. Kelly, B.M. Hennelly, N. Pandey, T.J. Naughton, and W.T. Rhodes. Resolution limits in practical digital holographic systems. *Optical Engineering*, 48(9):095801–095801, 2009.

[63] C. Kurtsiefer, T. Pfau, and J. Mlynek. Measurement of the Wigner function of an ensemble of Helium atoms. *Nature*, 386:150–153, March 1997.

[64] Ulf Leonhardt. Quantum–state tomography and discrete Wigner function. *Phys. Rev. Lett.*, 74(21):4101–4105, May 1995.

[65] Norbert Lindlein, Johannes Pfund, and Johannes Schwider. Algorithm for expanding the dynamic range of a shack-hartmann sensor by using a spatial light modulator array. *Optical Engineering*, 40(5):837–840, 2001.

[66] G. Lippmann. La photographie integrale. *Comptes-Rendus, Academie des Sciences*, 146(446-551), 1908.

[67] Yi Liu, Lei Tian, Justin W. Lee, Howard Y. H. Huang, Michael S. Triantafyllou, and George Barbastathis. Scanning-free compressive holography for object localization with subpixel accuracy. *Opt. Lett.*, 37(16):3357–3359, Aug 2012.

[68] M. Lustig, D. Donoho, and J.M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[69] Kenneth MacCabe, Kalyani Krishnamurthy, Amarpreet Chawla, Daniel Marks, Ehsan Samei, and David Brady. Pencil beam coded aperture x-ray scatter imaging. *Opt. Express*, 20(15):16310–16320, Jul 2012.

[70] L. Mandel and E. Wolf. *Optical coherence and quantum optics*. Cambridge University Press, 1995.

[71] Daniel L. Marks, Ronald A. Stack, and David J. Brady. Three-dimensional coherence imaging in the Fresnel domain. *Appl. Opt.*, 38(8):1332–1342, Mar 1999.

[72] Daniel M. Marks, Ronald A. Stack, and David J. Brady. Astigmatic coherence sensor for digital imaging. *Opt. Lett.*, 25(23):1726–1728, Dec 2000.

[73] Pierre Marquet, Benjamin Rappaz, Pierre J. Magistretti, Etienne Cuche, Yves Emery, Tristan Colomb, and Christian Depeursinge. Digital holographic microscopy: a noninvasive contrastimaging technique allowing quantitative visualization ofliving cells with subwavelength axial accuracy. *Opt. Lett.*, 30(5):468–470, Mar 2005.

[74] Jianwei Miao, Pambos Charalambous, Janos Kirz, and David Sayre. Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(6742):342–344, 1999.

[75] J. H. Milgram and W. Li. Computational reconstruction of images from holograms. *Appl. Opt.*, 41:853–864, February 2002.

[76] A. Momose, T. Takeda, Y. Itai, and K. Hirano. Phase–contrast x–ray computed tomography for observing biological soft tissues. *Nature medicine*, 2(4):473–475, 1996.

[77] Atsushi Momose, Wataru Yashiro, Yoshihiro Takeda, Yoshio Suzuki, and Tadashi Hattori. Phase tomography by x-ray talbot interferometry for biological imaging. *Japanese journal of applied physics*, 45:5254, 2006.

[78] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. Technical Report CTSR 2005-02, Stanford, 2005.

[79] K. A. Nugent. Wave field determination using three-dimensional intensity information. *Phys. Rev. Lett.*, 68(15):2261–2264, Apr 1992.

[80] K. A. Nugent, T. E. Gureyev, D. F. Cookson, D. Paganin, and Z. Barnea. Quantitative phase imaging using hard x rays. *Phys. Rev. Lett.*, 77:2961–2964, Sep 1996.

[81] Keith A Nugent. Coherent methods in the x-ray sciences. *Advances in Physics*, 59(1):1–99, 2010.

[82] J O'Sullivan. A fast sinc function gridding algorithm for fourier inversion in computer tomography. *IEEE Transactions on Medical Imaging*, 4(4):200–207, 1985.

[83] D. Paganin, A. Barty, P. J. McMahon, and K. A. Nugent. Quantitative phase-amplitude microscopy. III. The effects of noise. *Journal of Microscopy*, 214(1):51–61, 2004.

[84] D. Paganin and K. A. Nugent. Noninterferometric phase imaging with partially coherent light. *Physical Review Letters*, 80(12):2586–2589, 1998.

[85] G. Pan and H. Meng. Digital holography of particle fields: reconstruction by use of complex amplitude. *Appl. Opt.*, 42(5):827–833, February 2003.

[86] Jae-Hyeung Park, Keehoon Hong, and Byoungho Lee. Recent progress in three-dimensional information processing based on integral imaging. *Appl. Opt.*, 48(34):H77–H94, 2009.

[87] Daniele Pelliccia, Andrei Y. Nikulin, Herbert O. Moser, and Keith A. Nugent. Experimental characterization of the coherence properties of hard x-ray sources. *Opt. Express*, 19(9):8073–8078, Apr 2011.

[88] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629 –639, July 1990.

[89] Franz Pfeiffer, Timm Weitkamp, Oliver Bunk, and Christian David. Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources. *Nature Physics*, 2(4):258–261, 2006.

[90] B. C. Platt and R. Shack. History and principles of Shack-Hartmann wavefront sensing. *Journal of Refractive Surgery*, 17, 2001.

[91] M. G. Raymer, M. Beck, and D. McAlister. Complex wave-field reconstruction using phase-space tomography. *Phys. Rev. Lett.*, 72(8):1137–1140, Feb 1994.

[92] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259 – 268, 1992.

[93] Otmar Scherzer and Joachim Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12(1):43–63, 2000.

[94] U. Schnars and W. P. Juptner. Direct recording of holograms by a CCD target and numerical reconstruction. *Appl. Opt.*, 33:179–181, January 1994.

[95] Ulf Schnars and Werner Jueptner. *Digital holography: digital hologram recording, numerical reconstruction, and related techniques*. Springer, 2004.

[96] Sungkyu Seo, Ting-Wei Su, Derek K. Tseng, Anthony Erlinger, and Aydogan Ozcan. Lensfree holographic imaging for on-chip cytometry and diagnostics. *Lab Chip*, 9:777–787, 2009.

[97] Colin J. R. Sheppard. Defocused transfer function for a partially coherent microscope and application to phase retrieval. *J. Opt. Soc. Am. A*, 21(5):828–831, May 2004.

[98] Lawrence Sirovich and Michael Kirby. Low-dimensional procedure for the characterization of human faces. *JOSA A*, 4(3):519–524, 1987.

[99] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: application to squeezed states and the vacuum. *Phys. Rev. Lett.*, 70(9):1244–1247, Mar 1993.

[100] Saowapak Sotthivirat and Jeffrey A. Fessler. Penalized-likelihood image reconstruction for digital holography. *J. Opt. Soc. Am. A*, 21(5):737–750, May 2004.

[101] Ferréol Soulez, Loïc Denis, Corinne Fournier, Éric Thiébaut, and Charles Goepfert. Inverse-problem approach for particle digital holography: accurate location based on local optimization. *J. Opt. Soc. Am. A*, 24(4):1164–1171, Apr 2007.

[102] A. Starikov. Effective number of degrees of freedom of partially coherent sources. *J. Opt. Soc. Am.*, 72(11):1538–1544, Nov 1982.

[103] A. Starikov and E. Wolf. Coherent-mode representation of Gaussian Schell-model sources and of their radiation fields. *J. Opt. Soc. Am.*, 72(7):923–928, Jul 1982.

[104] Adrian Stern and Bahram Javidi. Three-dimensional image sensing and reconstruction with time-division multiplexed computational integral imaging. *Appl. Opt.*, 42(35):7036–7042, Dec 2003.

[105] N. Streibl. Phase imaging by the transport equation of intensity. *Opt. Commun.*, 49(1):6 – 10, 1984.

[106] D.M. Strong and T.F. Chan. Exact solutions to total variation regularization problems. *Math Department CAM Report*, 9641, 1996.

[107] Tohoru Takeda, Atsushi Momose, Yuji Itai, Wu Jin, and Keiichi Hirano. Phase-contrast imaging with synchrotron x-rays for detecting cancer lesions. *Academic Radiology*, 2(9):799 – 803, 1995.

[108] David S Taubman, Michael W Marcellin, and Majid Rabbani. Jpeg2000: Image compression fundamentals, standards and practice. *Journal of Electronic Imaging*, 11(2):286–287, 2002.

[109] M. R. Teague. Deterministic phase retrieval - A Green's function solution. *J. Opt. Soc. Am.*, 73:1434–1441, November 1983.

[110] BJ Thompson and E Wolf. Two-beam interference with partially coherent light. *J. Opt. Soc. Am.*, 47(10):895, 1957.

[111] Lei Tian and George Barbastathis. Digital holography applied to quantitative measurement of oil–drop in oil–water two-phase flows. In *Digital Holography and Three-Dimensional Imaging*, page DMC4. Optical Society of America, 2010.

[112] Lei Tian, Justin Lee, Se Baek Oh, and George Barbastathis. Experimental compressive phase space tomography. *Opt. Express*, 20(8):8296–8308, Apr 2012.

[113] Lei Tian, Nick Loomis, José A. Domínguez-Caballero, and George Barbastathis. Quantitative measurement of size and three-dimensional position of fast-moving bubbles in air-water mixture flows using digital holography. *Appl. Opt.*, 49(9):1549–1554, Mar 2010.

[114] Chanh Q. Tran, Andrew G. Peele, Ann Roberts, Keith A. Nugent, David Paterson, and Ian McNulty. X-ray imaging: a generalized approach using phase-space tomography. *J. Opt. Soc. Am. A*, 22(8):1691–1700, Aug 2005.

[115] Jinhong Tu. Wave field determination using tomography of the ambiguity function. *Physical Review E*, 55(2):1946–1949, 1997.

[116] Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

[117] Charles M Vest. Holographic interferometry. *New York, John Wiley and Sons, Inc., 1979. 476 p.*, 1, 1979.

[118] K. Vogel and H. Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A*, 40:2847–2849, September 1989.

[119] L. Waller, G. Situ, and J.W. Fleischer. Phase–space measurement and coherence synthesis of optical beams. *Nature Photonics*, 6:474–479, June 2012.

[120] Laura Waller, Lei Tian, and George Barbastathis. Transport of intensity phase-amplitude imaging with higher order intensity derivatives. *Opt. Express*, 18(12):12552–12561, Jun 2010.

[121] Laura Waller, Mankei Tsang, Sameera Ponda, Se Young Yang, and George Barbastathis. Phase and amplitude imaging from noisy images by kalman filtering. *Opt. Express*, 19(3):2805–2814, Jan 2011.

[122] A. Walther. Radiometry and coherence. *J. Opt. Soc. Am.*, 58(9):1256–1259, 1968.

[123] A. Wax and J. E. Thomas. Optical heterodyne imaging and Wigner phase space distributions. *Opt. Lett.*, 21(18):1427–1429, 1996.

[124] J. Weickert, B.M.T.H. Romeny, and M.A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *Image Processing, IEEE Transactions on*, 7(3):398 –410, mar 1998.

[125] Joachim Weickert. A review of nonlinear diffusion filtering. In *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 1–28. Springer Berlin / Heidelberg, 1997.

[126] E. Wigner. On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40(5):0749–0759, June 1932.

[127] Emil Wolf. New theory of partial coherence in the space-frequency domain. Part I: spectra and cross spectra of steady-state sources. *J. Opt. Soc. Am.*, 72(3):343–351, Mar 1982.

[128] Kurt Bernardo Wolf, Miguel Angel Alonso, and Gregory W. Forbes. Wigner functions for helmholtz wave fields. *J. Opt. Soc. Am. A*, 16(10):2476–2487, Oct 1999.

[129] Xizeng Wu and Hong Liu. Clinical implementation of x-ray phase-contrast imaging: Theoretical foundations and design considerations. *Medical physics*, 30:2169, 2003.

[130] Y. Yang, B. Kang, and Y. Choo. Application of the correlation coefficient method for determination of the focal plane to digital particle holography. *Appl. Opt.*, 47(6):817–824, 2008.

[131] Frits Zernike. Phase contrast, a new method for the microscopic observation of transparent objects. *Physica*, 9:686–698, 1942.

[132] Frits Zernike. How I discovered phase contrast. *Science*, 121(3141):345, 1955.

[133] Z. Zhang and M. Levoy. Wigner distributions and how they relate to the light field. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2009.

[134] Zhengyun Zhang, Zhi Chen, Shakil Rehman, and George Barbastathis. Factored form descent: a practical algorithm for coherence retrieval. *Opt. Express*, 21(5):5759–5780, Mar 2013.

[135] Adam M. Zysk, Robert W. Schoonover, P. Scott Carney, and Mark A. Anastasio. Transport of intensity and spectrum for partially coherent fields. *Opt. Lett.*, 35(13):2239–2241, Jul 2010.