Propagation of partially coherent light in optical systems

Dissertation

for the acquisition of the academic title

Doktor-Ingenieur (Dr.-Ing.)

submitted to the Council of the Faculty of Physics and Astronomy

of the Friedrich-Schiller-Universität Jena

By MSc. Minyi Zhong

born in Gaoyao, China on 30.09.1987

Supervisors:

Major Supervisor: Prof. Dr. Herbert Gross

Associate Supervisor: Prof. Dr. Alois Herkommer

Associate Supervisor: Prof. Dr. Richard Kowarschik

Day of the Disputation: 11 April 2017

Abstract

There are in general three approaches to propagate partially coherent light. They include solving the Helmholtz equations for the correlation function, employing the Wigner function to propagate light in phase space, or using modal expansion. In this work, we investigate the interaction of partially spatially coherent light with optical systems. Our modeling methods are based on the Wigner function and modal expansion.

First, we utilize the thin element approximation in the Wigner function to propagate partially coherent light through discontinuous surfaces. Our examples include the phase step, grating, kinoform lens, axicon and lens array. Phase space provides a vivid picture for understanding the diffraction effects, *e.g.* the diffracted orders formed by a grating, the multiple foci of a kinoform lens, and the beam homogenizing effect of a lens array, *etc*. This approach facilitates the design of diffractive elements and the interpretation of optical effects. Second, we improve the propagation algorithms for the Wigner function. These implementations include removing a parabolic wavefront of a beam, and an efficient propagator based on rotating the phase space. Both algorithms increase the computational speed without losing any physical accuracy.

To overcome certain limitations of our phase space propagators, we investigate the modal expansion method. We use Schell beams as examples, to discuss the advantages and disadvantages of eigenmode and shifted-elementary mode expansions. Afterward, we develop an expansion tool to efficiently propagate partially coherent light inside waveguides. This tool allows quick access to the light fields at any distance inside the waveguide. We thus obtain an accurate modeling of the diffraction effects produced by the waveguide structure.

Our modeling methods pave the path to the experimental measurement of coherence. We discuss two schemes to retrieve the spatial coherence of a beam under test. The first scheme is to solve an inverse problem of the modal expansion. It requires iterative phase retrieval algorithms to recover the complex fields of individual modes. We extend this method to a beam with an arbitrary wavefront and discuss the potential and limits of this technique. The second scheme for coherence retrieval is to solve the inverse problem of the Wigner function, also known as the phase space tomography. To reconstruct a Wigner function requires an inverse Radon transform in a four-dimensional space. We present an overview of the two relevant algorithms, *i.e.* the filtered back projection and the ambiguity function reconstruction. Furthermore, we propose an improved algorithm for the filtered back projection.

This page is left blank intentionally.

Contents

1. Introduction

Partially coherent light sources are widely applied in microscopy[1], lithography[2], optical communications[3], *etc*. This kind of sources is well known for some outstanding optical performance. For example, partial coherence reduces the speckle effect and improves the imaging quality. Compared to completely coherent illumination, partially coherent illumination is more tolerant to surface blemishes. When an optical surface contains dust or scratches, a coherent source may easily generate an image with distinct diffraction patterns. For a partially coherent source, the interference effects are limited within the coherence length. If the distance to the detector is beyond the coherence length, surface blemishes only reduce the image contrast. In differential interference contrast microscopy, partial coherence improves the sectioning performance in a similar manner by preventing the Talbot effect[4]. Furthermore, in microscopy and lithography, it is desired to have a high imaging resolution. Partially coherent light is often employed to provide extended illumination area on the pupil and to increase the numerical aperture and homogeneity. Besides, from an economical point of view, partially coherent sources are sometimes less costly for fabrication and alignment, compared to completely coherent sources. Nowadays, a large variety of laser sources has mixed modes[5]. Multiple modes offer another advantage, which is a significantly high power. These laser sources are thus popular for lighting and display.

Based on the growing demand for partially coherent light in practice, it is of importance to developing efficient modeling tools to propagate such light in optical systems. In general, there are three main strategies to compute the propagation of a partially coherent field.

The first method, also being a conventional one, is to propagate the correlation function of a source[6]. Such a function reveals the correlation of the light fields emitted from different places at different moments in time. It indicates the degree of coherence of the source. As the propagation of a coherent light field obeys the wave equations, the correlation of the fields also follows the rules of wave propagation[7]. Therefore, one can employ the diffraction integrals developed for complete coherence to propagate the coherence function, *e.g.* using the Rayleigh-Sommerfeld integral[8]. However, to solve such an integral takes enormous computational efforts, especially with complicated optical elements. For monochromatic light with two transverse dimensions, the coherence function spans into four dimensions (4D). That makes this approach

computationally expensive and not applicable to practically relevant problems, even with modern computers.

The second method to propagate partially coherent light is to employ the Wigner function. The Wigner function was first proposed by Wigner[9] in 1932 as a quasi-probability distribution to describe quantum mechanics in phase space. Later, the Wigner function was introduced into optics by Dolin and Walther[10-12], to describe an optical signal in phase space. The concept of phase space resembles the angle and position in geometrical optics. We thus can employ the ray transformation matrix for paraxial propagation[13]. That vastly simplifies the propagation operator. In additional, the Wigner function is powerful in visualizing optical effects, not only in geometrical optics but also in wave optics[14]. One disadvantage of this method is that it also contains four axes for the light fields with two transverse dimensions.

The third propagation method for partially coherent light is modal expansion[6,7]. It considers a partially coherent beam as an incoherent sum of multiple modes. Each mode is a coherent beam. Between every two modes, we assume no interference. The propagation of each mode follows the wave equations. Therefore, the interaction of partially coherent light with optical systems can be considered as multiple independent channels. Each channel contains one coherent beam. This approach attracts much interest because it requires one less dimension compared to the other two methods above. However, when the degree of coherence is low, the number of modes can be large.

The question, which propagation method to apply, depends on the prior assumptions of the beam and the system. A suitable propagation operator can significantly improve the computational efficiency and the physical accuracy. Our work contains a comprehensive investigation into the last two methods, *i.e.* the Wigner function and modal expansion.

An additional application of these modeling methods form the basis of the experimental measurement of coherence. In practice, we often encounter a light source with an unknown degree of coherence. While the irradiance of a beam may be easily measured by a camera, the coherence of the beam is not always directly accessible. To retrieve the coherence information from the intensity measurements is equivalent to solving an inverse problem of obtaining the Wigner function from its projection, or an inverse problem of the modal expansion.

For an inverse problem of the modal expansion, we apply the iterative phase retrieval algorithm. The measured intensity is considered as an incoherent sum of multiple modes. The phase

retrieval algorithm recovers of the complex fields of individual modes. After that, the coherence, given by the correlation function, is reconstructed from the modes. Alternatively, to solve the inverse problem of the Wigner function demands a method called phase space tomography. The irradiance at the output plane of a first-order system is considered as a projection of the transformed Wigner function at the input plane. To recover the Wigner function from its projection, the inverse Radon transform is required. These methods of coherence retrieval have their own advantages and disadvantages. Our work includes an extensive exploration of the relavant algorithms.

We arrange this dissertation in the following order. We focus on spatial coherence of monochromatic light and represent light with scalar fields. Chapter 2 is a theoretical introduction. It builds the infrastructure for our research. Chapter 3 is our investigation into the modeling methods for partially coherent light based on the Wigner function and modal expansion. Although the concept of the Wigner function has been developed for decades, few publications describe partially coherent light in phase space with diffractive elements. Section 3.1 concentrates on the interaction of partially coherent light with discontinuous surfaces and the interpretation of optical effects in phase space. Section 3.2 presents our improved propagation algorithms for the Wigner function. In modal expansion, we take Schell beams as examples to compare the different expansions, since Schell beams have some certain symmetry and can be found in many practical approaches. Moreover, we visualize the modes in phase space, to provide a comprehensive understanding of the mode properties. Also, we develop propagation operators for partially coherent light traveling inside a waveguide. Our algorithm allows a fast and accurate modeling of light at any distance along the waveguide. In Chapter 4, we focus on the inverse problems of modal expansion and the Wigner function, to retrieve the coherence from intensity measurements. In mode recovery (Section 4.1), we extend the iterative phase retrieval algorithm to a beam with an arbitrary wavefront. The potential and limit of this approach are discussed. For the phase space tomography in Section 4.2, we propose an efficient implementation of the filtered back projection in the inverse Radon transform to reconstruct a 4D Wigner function.

2. Theory

This chapter is a review of the theory of partially coherent light. It serves as the road map of this dissertation.

The content in this chapter consists of four sections. In Section 2.1 we introduce the fundamentals of partial coherence, including the mathematical and physical properties, also the possibilities to characterize partial coherence. In Section 2.2, we present the Wigner function to describe light in phase space. In Section 2.3, we introduce the theory of modal expansion. Section 2.4 discloses the applications of modal expansion and the Wigner function. They contribute to the coherence measurement in practice. To retrieve the coherence of a beam, it equals to solving the inverse problem of modal expansion or solving an inverse problem of the Wigner function. We describe the major algorithms in this section.

2.1 Partial coherence

The concept of partial coherence includes temporal coherence and spatial coherence. Simply speaking, temporal coherence describes the correlation between fields emitted at different moments in time. A typical setup to realize this detection is the Michelson interferometer. Correspondingly, spatial coherence describes the correlation between fields emitted at different positions in space. A common experiment to detect spatial coherence is Young's interferometer.

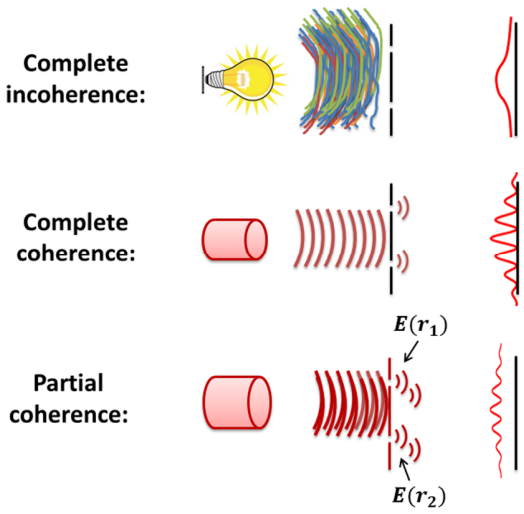


Figure 2-1. Pictorial introduction to spatial coherence in Young's interferometer experiment.

In our work, we concentrate on spatial coherence. Figure 2-1 illustrates a pictorial introduction to spatial coherence in Young's interferometer experiment. A completely incoherent source

yields no intensity fringes at the detector plane. A completely coherent source generates intensity fringes with a maximum contrast. With a partially coherent light source, the contrast of fringes is in between those two extreme cases. The correlation between the fields emitted at the slit positions, denoted by $E(r_1)$ and $E(r_2)$, provides an important quantity to characterize partial coherence.

2.1.1 Mutual coherence function and the degree of coherence

Before introducing the theory of partial coherence, we present a short review of a few terminologies in statistics. These terminologies need to be introduced, as they provide the fundamental basis of partially coherent light. They are called the stationary process, ensemble average, and ergodicity.

A process is called stationary when it fulfills two requirements. First, its probability density function does not depend on time. Second, the time average of this process does not rely on the starting or the ending time, but only on the time interval. A wide-sense stationary process does not fulfill the first requirement strictly. Only its mean and variance, instead of the probability density function, are independent of time.

Assume there is a random process $x(t)$ that fluctuates with time. We define $h(x)$ as a function of this random process. There are two ways to obtain the average of $h(x)$, *i.e.* an ensemble average and a time average. An ensemble average of $h(x)$ is defined as

$$\langle h[x(t=t_0)]\rangle_E = \int_{-\infty}^{\infty} h(x)\, p(x, t=t_0)\, \mathrm{d}x, \tag{2.1}$$

where the symbol $\langle\ \ \rangle_E$ represents the ensemble average and p is the probability density function of x. Equation (2.1) implies that an ensemble average takes the mean of many realizations of the random process, *i.e.* $h(x)$, at a given time. On the contrary, a time average of $h(x)$ is defined as

$$\langle h(x)\rangle_t = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} h[x(t)]\, \mathrm{d}t, \tag{2.2}$$

Where the symbol $\langle\ \ \rangle_t$ represents the time average and T is a time period in the random process. Equation (2.2) means that a time average takes the mean of one realization of the random process over a time interval $-\infty < t < \infty$.

When a process is ergodic, the ensemble average equals to the time average,

$$\langle h[x(t=t_0)]\rangle_E = \langle h(x)\rangle_t\ . \tag{2.3}$$

The advantage of having an ergodic system is that we do not need many realizations of the random process to obtain the ensemble average. Instead, averaging a single realization over a sufficiently large time interval can provide the required information. For a stationary and ergodic process, we exclude the time dependency and have

$$\langle h(x)\rangle_E = \langle h(x)\rangle_t. \tag{2.4}$$

Now we assume a partially coherent source is stationary, at lease in the wide sense, and ergodic. The mutual coherence function J is defined as an ensemble average of the field correlations between two positions (r_1 and r_2) with a time difference τ [15]

$$J(r_1, r_2, \tau) = \langle E^*(r_1, t)E(r_2, t+\tau)\rangle. \tag{2.5}$$

where $r_1 = (x_1, y_1, z_1)$, $r_2 = (x_2, y_2, z_2)$ ($z = 0$ defined at the source plane), E denotes the fluctuating light field, E^* is the complex conjugate of E. From now on, we use $\langle \quad \rangle$ to represent an ensemble average. From the statistical theory above we also have

$$J(r_1, r_2, \tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} E^*(r_1, t)E(r_2, t+\tau)\,\mathrm{d}t, \tag{2.6}$$

where T is the time interval of the random process, *e.g.* the integration time of a CCD camera in practice. Because light oscillates with a high frequency, the integration time of a camera is considered sufficiently large to average out the light fluctuation.

The complex degree of coherence of the two signals is defined as

$$\gamma(r_1, r_2, \tau) = \frac{J(r_1, r_2, \tau)}{\sqrt{I(r_1, \tau=0)}\sqrt{I(r_2, \tau=0)}}, \tag{2.7}$$

where I represents the irradiance intensity. (For abbreviation, we call it intensity from now on.) Equation (2.7) normalizes the coherence property to a value between 0 to 1. A degree of $\gamma = 1$ means that the two fields are completely coherent. A degree of $\gamma = 0$ represents complete coherence, while $0 < \gamma < 1$ expresses partial coherence.

From Equations (2.5) and (2.6) it is clear that $J(r_1, r_2, \tau)$ is a complex temporal cross-correlation between two fields. Therefore $J(r_1, r_2, \tau)$ is also called as cross-correlation function. The propagation of J in free space obeys the wave equations[16]:

$$\begin{aligned} \nabla_1^2 J(r_1, r_2, \tau) &= \frac{1}{c^2}\frac{\partial^2}{\partial\tau^2} J(r_1, r_2, \tau), \\ \nabla_2^2 J(r_1, r_2, \tau) &= \frac{1}{c^2}\frac{\partial^2}{\partial\tau^2} J(r_1, r_2, \tau). \end{aligned} \tag{2.8}$$

where ∇_1^2 is the Laplacian operator acting on r_1. The correlation of the fields at r_1 and r_2 characterizes spatial coherence. The dependence on τ characterizes temporal coherence. As the

two equations (2.8) couple the dependency of J on all the variables r_1, r_2 and τ, the spatial and temporal coherence properties of light are not independent of each other during propagation[16]. Indeed, according to the Van Cittert-Zernike theorem, a spatially incoherent light source possesses a finite degree of spatial coherence after propagation over a long distance[15]. Mandel[16] provides a qualitative picture to explain this phenomenon. A source with spatially partial coherence has a certain area or volume emitting light. The light fields emitted at different source points can be individually considered as wave trains. These wave trains have a finite duration. That means the light fields have a non-zero bandwidth, which implies finite temporal coherence. As we are interested in spatial coherence in this work, the following section introduces a quantity to separate temporal coherence and spatial coherence.

2.1.2 Cross-spectral density function

The Fourier transform of the mutual coherence function $J(r_1, r_2, \tau)$ with respect to the time difference τ is given by

$$\Gamma(r_1, r_2, \omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} J(r_1, r_2, \tau)\exp(i\omega\tau)\,\mathrm{d}\tau. \tag{2.9}$$

$\Gamma(r_1, r_2, \omega)$ is known as the cross-spectral density function.

The Fourier transform of the fluctuating optical field E with respect to the time variable t is given by

$$E(r, t) = \int_{-\infty}^{\infty} \tilde{E}(r, \omega)\exp(-i\omega t)\,d\omega. \tag{2.10}$$

Keeping Equations (2.9) and (2.10) in mind, we apply Fourier transforms to both the left and right sides of Equation (2.5). It results in

$$\langle \tilde{E}^*(r_1, \omega)\tilde{E}(r_2, \omega')\rangle = \Gamma(r_1, r_2, \omega)\delta(\omega - \omega'). \tag{2.11}$$

Equation (2.11) emphasizes that the cross-spectral density function is a correlation measure of two fluctuating fields at a particular frequency. Therefore, the cross-spectral density function allows us to inspect spatial coherence between r_1 and r_2 at a particular wavelength. As we focus on quasi-monochromatic light sources, from now on we write

$$\Gamma(r_1, r_2, \omega = \omega') =: \Gamma(r_1, r_2) \tag{2.12}$$

The function $\Gamma(r_1, r_2)$ is often called mutual intensity or equivalently, correlation function, when we only consider the spatial coherence of a quasi-monochromatic stationary source. From now on, we address $\Gamma(r_1, r_2)$ as correlation function to avoid confusion. The correlation function satisfies the coupled Helmholtz equations[16],

$$\nabla_1^2 \Gamma(r_1, r_2) + k^2 \Gamma(r_1, r_2) = 0, \\ \nabla_2^2 \Gamma(r_1, r_2) + k^2 \Gamma(r_1, r_2) = 0. \tag{2.13}$$

where $k = \frac{2\pi}{\lambda} = \frac{\omega}{c}$ (λ being the wavelength and c being the speed of light in vacuum).

The propagation of the correlation function in paraxial approximation takes the form [17]

$$\Gamma(r_1, r_2, z) = \left(\frac{k}{2\pi z}\right)^2 \iint_{-\infty}^{\infty} \Gamma^{(0)}(r_1', r_2') \exp\left\{-\frac{ik[(r_1 - r_1')^2 - (r_2 - r_2')^2]}{2z}\right\} \mathrm{d}^2 r_1' d^2 r_2'. \tag{2.14}$$

where $\Gamma^{(0)}$ denotes the correlation function at the source plane.

The correlation function in the far zone, defined as Γ^{∞}, follows a Fourier transform of $\Gamma^{(0)}$,

$$\Gamma^{\infty}(r_1, r_2) = \frac{(2\pi k)^2 \cos\theta_1 \cos\theta_2 \exp[ik(r_2 - r_1)]}{r_1 r_2} \tilde{\Gamma}^{(0)}(-k\hat{r}_{1\perp}, k\hat{r}_{2\perp}), \tag{2.15}$$

where r_1 and r_2 point from the origin at the source plane to the observed points in the far zone, $\hat{\boldsymbol{r}}_{\boldsymbol{1}}$ and $\hat{\boldsymbol{r}}_{\boldsymbol{2}}$ are the unit vectors along $\boldsymbol{r}_{\boldsymbol{1}}$ and $\boldsymbol{r}_{\boldsymbol{2}}$, $\hat{r}_{1\perp}$ and $\hat{r}_{2\perp}$ represent the projections of $\hat{\boldsymbol{r}}_{\boldsymbol{1}}$ and $\hat{\boldsymbol{r}}_{\boldsymbol{2}}$ into the source plane. $\tilde{\Gamma}^{(0)}$ denotes the Fourier transform of $\Gamma^{(0)}$ as

$$\tilde{\Gamma}^{(0)}(k_1, k_2) = \frac{1}{(2\pi)^4} \iint \Gamma^{(0)}(r_1, r_2) \exp[-i(k_1 \cdot r_1 + k_2 \cdot r_2)] \, \mathrm{d}^2 r_1 \mathrm{d}^2 r_2. \tag{2.16}$$

where r_1 and r_2 are vectors from the origin to the arbitrary points at the source plane, and k_1, k_2 represent the spatial frequencies along r_1 and r_2 directions respectively.

By now we know that a correlation function defines the essential properties of a partially coherent beam. It indicates the statistical similarity of the light fields at two spatial positions. In the following, we introduce the important concept of Schell beams, which are often referred to later. They are the Gaussian-Schell beam and the flat-top Schell beam.

2.1.3 Schell beams

Schell beams belong to a common class of partially coherent light. It covers a broad range of light sources in practice, from various lamps, light-emitting diodes, multimode fibers, lasers, to the pulsed broad-area vertical-cavity-surface-emitting-laser[18].

We begin with the definition of a Schell-type source. For simplicity, we confine ourselves to light with one transverse dimension, *i.e.* in coordinates (x, z) where x denotes the transverse coordinate, and z is the optical axis. The correlation function of a Schell-type source is defined as [19]

$$\Gamma(x_1, x_2) = [I(x_1) I(x_2)]^{1/2} \gamma(x_1 - x_2), \tag{2.17}$$

where x_1, x_2 are two arbitrary positions at the source plane, I describes the intensity across the source plane, and γ is the degree of coherence. It is clear from Equation (2.17) that the

correlation of the fields only depends on their relative distance, but not on their specific locations. Such a source has a relatively homogenous property.

Our introduction includes two primary Schell sources. One is called the Gaussian-Schell beam (also known as the Gaussian-Schell model beam). Another is a flat-top Schell beam. Their properties are summarized as follows.

Gaussian-Schell beam

A Gaussian-Schell beam has a cross-spectral density function with Gaussian profiles

$$\Gamma(x_1, x_2) = A \exp\{-(x_1 + x_2)^2/(2{w_o}^2)\} \exp\{-(x_1 - x_2)^2/(2{l_c}^2)\}, \qquad (2.18)$$

where w_0 is the beam waist and l_c represents the coherence length. Here the coherence length means the correlation width at the transverse plane. It differs from the coherence length in the concept of a temporal coherence. When the distance between x_1 and x_2 exceeds the length of l_c, the two fluctuating fields are considered incoherent.

According to Equation (2.14), the propagation of the Gaussian-Schell beam follows

$$\Gamma(x_1, x_2; z) = A \left[\frac{w_0}{w(z)}\right]^2 \exp\left\{-\frac{(x_1+x_2)^2}{2[w(z)]^2}\right\} \exp\left\{-\frac{{w_0}^2(x_1-x_2)^2}{2{l_c}^2[w(z)]^2}\right\} \exp\left\{-i\frac{k({x_1}^2-{x_2}^2)}{2R(z)}\right\}, \quad (2.19)$$

where $w(z) = w_0[1 + (\lambda z/\pi w_0 w_c)^2]^{1/2}$, $R(z) = z[1 + (\pi w_0 w_c/\pi z)^2]$ and $1/{w_c}^2 = 1/{w_0}^2 + 1/{l_c}^2$.

Flat-top Schell beam

The analytical form of a flat-top Schell beam is proposed by Korotkova[20]. The cross-spectral density function at the source plane is defined as

$$\Gamma(x_1, x_2) = \exp\left(-\frac{{x_1}^2+{x_2}^2}{4\sigma^2}\right)\frac{1}{{C_0}^2}\sum_{m=1}^{M}\binom{M}{m}\frac{(-1)^{m-1}}{\sqrt{m}}\exp\left[-\frac{(x_1-x_2)^2}{2m\delta^2}\right], \qquad (2.20)$$

where $\binom{M}{m}$ is a binomial coefficient, $C_0 = \sum_{m=1}^{M}\binom{M}{m}\frac{(-1)^{m-1}}{\sqrt{m}}$, σ represents the rms source width, and δ denotes the rms transverse correlation width. Equation (2.20) implies that a rectangular Schell beam is a spatial superposition of multiple Gaussian-Schell beams. Integer M indicates the number of overlaid Gaussian-Schell beams. A larger integer of M yields a more flat-top intensity profile. With M=1, the intensity approaches Gaussian.

In the far field, we have $r = (x, z)$ and $|r| = \sqrt{x^2 + z^2}$ with one transverse dimension. According to Equation (2.15), the cross spectral density follows

$$\Gamma^{\infty}(\mathrm{r}_1,\mathrm{r}_2)=\frac{(2\pi k)^2\cos\theta_1\cos\theta_2\exp[ikr(\hat{\boldsymbol{r}}_{\boldsymbol{2}}-\hat{\boldsymbol{r}}_{\boldsymbol{1}})]}{{C_0}^2 r_1 r_2}\times \sum_{m=1}^{M}\binom{M}{m}\frac{(-1)^{m-1}a_m}{\sqrt{m}b_m}\exp\left[-c_m\left({\hat{r}_{1x}}^2+{\hat{r}_{2x}}^2\right)-d_m(\hat{r}_{1x}-\hat{r}_{2x})^2\right], \tag{2.21}$$

where $a_m=\sigma\sqrt{\frac{2m\delta^2+4\sigma^2}{m\delta^2+4\sigma^2}}$, $b_m=\sqrt{\frac{1}{m\delta^2}+\frac{1}{\sigma^2}}$, $c_m=\frac{k^2\sigma^2 m\delta^2}{m\delta^2+4\sigma^2}$, $d_m=\frac{2k^2\sigma^4}{m\delta^2+4\sigma^2}$, $C_0=\sum_{m=1}^{M}\frac{(-1)^{m-1}}{\sqrt{m}}$. The radiant intensity in the far field takes the form

$$I^{\infty}(x)=\frac{2k^2\cos^2\theta}{{C_0}^2 r^2}\sum_{m=1}^{M}\binom{M}{m}\frac{(-1)^{m-1}\sqrt{m}a_m}{b_m}\exp[-2c_m x^2]. \tag{2.22}$$

Given by Equation (2.22), the radiant intensity in the far-field exhibits a flat-top profile. That is the reason we call it a flat-top Schell beam.

As Schell beams exist commonly n practice, we employ them in the later chapters as the main illumination sources. After the introduction of the two exemplary Schell beams, in the following, we discuss the efficient propagation of these beams in optical systems.

2.2 Wigner function

In the previous sections, the propagation of a partially coherent beam is performed by solving the correlation function in the Helmholtz equation. In the following, we introduce an efficient alternative approach, the Wigner function, to propagating a partially coherent beam in phase space.

The term, phase space, expresses a signal in the phase domain and the spatial domain simultaneously. The definition of these domains depends on the property of the signal we want to observe. For an acoustic signal, it is about the temporal frequency against a time variable. Roughly speaking, it provides the possibility to assign a frequency information, which usually is defined globally via a Fourier transform, to a specific moment in time. As a pictorial example, one can image the musical notes of a piano. It plays a tone related to a specific frequency at a given moment in time.

For a spatial signal, it is about the spatial frequency against the spatial coordinate. In this case, the approach allows the assignment of a propagation angle locally to every position in space based on a wave-optical definition. Therefore, this approach can be viewed as an extension of the classical ray optical theory. At every point in space, a bundle of rays could emerge in different directions. Since we investigate spatial coherence in this work, we apply the phase space for a spatial signal.

The Wigner function is defined as a Fourier transform of the correlation function with respect to the correlation width Δx [21],

$$W(x,u) = \int \Gamma\left(x + \frac{\Delta x}{2}, x - \frac{\Delta x}{2}\right) exp\left(-i\frac{2\pi}{\lambda}u\Delta x\right) \mathrm{d}\Delta x, \tag{2.23}$$

where $x = (x_1 + x_2)/2, \Delta x = x_1 - x_2$, variables u and Δx are Fourier conjugates. Integrating $W(x,u)$ over the x axis gives the angular spectrum, while integrating $W(x,u)$ over the u axis yields the intensity distribution in space[22].

There are several remarkable properties of using the Wigner function for analyzing optical systems. Firstly, the idea to describe optical signals simultaneously in spatial frequency and space resembles the concepts of angle and position in geometrical optics. Making use of this property, we can paraxially propagate light in phase space with the ray transfer matrix (also known as the ABCD matrix)[13,23].

The Wigner function passing a first-order optical system follows

$$W_o(x_o, u_o) = W(Ax + Bu, Cx + Du), \tag{2.24}$$

where the subscript $_o$ denotes the signal leaving the system. It is worth noting that the Wigner function is capable of describing higher-order systems [24]. In this case, other models instead of the ABCD matrices are required.

Secondly, the propagation of the Wigner function through a thin element can be further simplified. Assume the thin component can be described by an amplitude or phase modulation $P(x)$. The relation of the light field before and after passing the element is

$$E'(x) = P(x)E(x), \tag{2.25}$$

where the symbol $'$ denotes the transmitted signal behind the thin element. Based on Equation (2.11), we attain

$$\Gamma'(x_1, x_2) = \langle P^*(x_1)P(x_2)E^*(x_1)E(x_2)\rangle = P^*(x_1)P(x_2)\Gamma(x_1, x_2), \tag{2.26}$$

With Equation (2.23), the Wigner function behind the thin element follows

$$W'(x,u) = \int\int \Gamma\left(x + \frac{\Delta x}{2}, x - \frac{\Delta x}{2}\right) P^*\left(x + \frac{\Delta x}{2}\right) P\left(x - \frac{\Delta x}{2}\right) \exp\left(-i\frac{2\pi}{\lambda}u\Delta x\right) d\Delta x. \tag{2.27}$$

With Equation (2.27), we can use the Wigner function to investigate the diffraction effects produced by a thin element.

Figure 2-2 serves as an example to demonstrate the propagation of light in a system by using Wigner functions. Inside the paraxial regions that can be described by ABCD matrices (*e.g.* z_1-z_6), we employ Equation (2.24). In other regions consisting of lenses or stops (*e.g.* d_1-d_4), We

use Equation (2.27). The conversion between Γ and W crossing different regions is carried out by a Fourier transform (Equation (2.23)).

Due to the Hermitian properties of $\Gamma(x_1, x_2)$ (see below in Section 2.3.1), the Wigner function always yields real, but not necessarily positive values. Negative values in the phase space indicate destructive interference[25]. In particular, for partially coherent light, the Wigner function facilitates the interpretation of diffraction effects. These aspects are discussed in detail in Chapter 3.1.

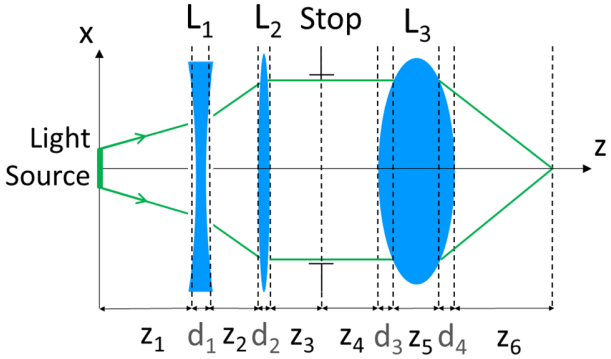


Figure 2-2. Using the Wigner functions to propagate partially coherent light inside a system.

2.3 Modal expansion

Up to now, we have introduced two possibilities to propagate a partially coherent beam. One is to solve the Helmholtz equations given by Equations (2.13). The second option is to employ the Wigner function with the thin element approximation in a ABCD-matrix formalism (Equation (2.27)). However, for a component which cannot be considered as a thin element, Equation (2.27) becomes invalid.

In this section, we introduce another approach to propagate partially coherent light through a wider class of components. We represent the full beam by multiple coherent modes. Each mode is a coherent beam and propagates independently. The effect of the full beam is an incoherent superposition of these modes. This method significantly simplifies the propagation operator for partially coherent light. The corresponding theory is introduced.

2.3.1 Eigenmodes

According to Plancherel's theorem[26], the correlation function $\Gamma(x_1, x_2)$ is a square-integrable function, *i.e.*

$$\iint_{-\infty}^{+\infty} |\Gamma(x_1, x_2)|^2 \, dx_1 dx_2 < \infty. \tag{2.28}$$

From Equations (2.5) and (2.28) we have

$$\Gamma(x_1, x_2) = \Gamma^*(x_2, x_1) \tag{2.29}$$

$$\iint_{-\infty}^{+\infty} \Gamma(x_1, x_2) s^*(x_1) s(x_2)\, dx_1 dx_2 \geq 0, \tag{2.30}$$

for an arbitrary square-integrable function $s(x)$.

Equations (2.28)-(2.30) imply that $\Gamma(x_1, x_2)$ is a Hermitian, non-negative definite Hilbert-Schmidt kernel[27]. According to the Mercer's theorem, there exists a uniformly and absolutely convergent expansion

$$\Gamma(x_1, x_2) = \sum_n \lambda_n \phi_n^*(x_1) \phi_n(x_2) = \sum_n f_n^*(x_1) f_n(x_2), \tag{2.31}$$

where λ_n are the eigenvalues and ϕ_n are the eigenvectors of the Fredholm integral equation,

$$\int_{-\infty}^{\infty} \Gamma(x_1, x_2) \phi_n(x_1) dx_1 = \lambda_n \phi_n(x_2). \tag{2.32}$$

For the convenience of the later discussion, we define $f_n(x) = \sqrt{\lambda_n}\phi_n(x)$ the eigenmodes of $\Gamma(x_1, x_2)$.

Equation (2.31) shows that a partially coherent beam can be equally represented by an incoherent superposition of the eigenmodes. This suggests solving a complicated system by simply propagating each coherent beam separately. There are in general two ways to obtain the eigenmodes of a known $\Gamma(x_1, x_2)$. One is to solve the Equation (2.32) analytically. Another possibility is to sample $\Gamma(x_1, x_2)$ discretely and obtain the eigenmodes from linear algebraic methods, as it is explained in the following.

Analytical eigenmodes

Assume we know the analytical form of a correlation function given by Equation (2.18). The way to derive the analytical solutions for the eigenvalues and eigenmodes is to insert Equation (2.18) into Equation (2.32). We use the Gaussian-Schell beam as an example. Gori[28] and Starikov[29] have derived the following solutions,

$$\phi_n(x) = \left(\frac{2c_0}{\pi}\right)^{\frac{1}{4}} \frac{1}{\sqrt{2^n n!}} H_n\left(x\sqrt{2c_0}\right) \exp(-c_0 x^2), \tag{2.33}$$

$$\lambda_n = A\left(\frac{\pi}{a_0 + b_0 + c_0}\right)^{1/2} \left(\frac{b_0}{a_0 + b_0 + c_0}\right)^n, \tag{2.34}$$

where H_n denotes the n^{th} order Hermite polynomial, $a_0 = 1/w_o^2$, $b_0 = 1/l_c^2$ and $c_0 = ({a_0}^2 + 2a_0 b_0)^{1/2}$. According to Equation (2.31), Equations (2.33) and (2.34) offer us a way to directly represent a Gaussian-Schell beam by the eigenmodes for a given beam waist w_o and the correlation length l_c.

Note that analytical eigenmodes are not always available for all forms of correlation functions. So far, researchers have derived the analytical solutions to partially coherent beams with Gaussian-Schell or Bessel-Schell characters[28-31]. An arbitrary partially coherent light may have a correlation function with a complicated form. In that case, an easier approach to obtaining the eigenmodes is to use the linear algebraic method.

Numerical Eigenmodes

Before starting with the linear algebraic concepts, we firstly assume that our light fields are sampled discretely fulfilling the Nyquist's theorem. Although optical fields are always continuous in nature, a discrete sampling with a sufficient density still provides access to the physical properties of light.

Ozaktas et al.[32,33] have used the linear algebraic theory to characterize partial coherence. They discuss spatial coherence based on the discrete correlation function (*i.e.* correlation matrix). We recall their main ideas as follows.

Assume $\boldsymbol{\Gamma}$ is the correlation matrix of a quasi-monochromatic source. We then have

$$\boldsymbol{\Gamma} = \langle \mathbf{f}\mathbf{f}^{\mathrm{H}} \rangle, \tag{2.35}$$

where $\mathbf{f}$ represents the discretely sampled light field, $\langle \quad \rangle$ denotes ensemble average, and the superscript H represents a Hermitian transpose (*i.e.* $\mathbf{f}^{\mathrm{H}} = (\mathbf{f}^{*})^{\mathrm{t}}$, t being a matrix transpose). As $\boldsymbol{\Gamma}$ is Hermitian symmetric (*i.e.* $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{\mathrm{H}}$) and positive semidefinite (*i.e.* $\mathbf{s}^{\mathrm{H}}\boldsymbol{\Gamma}\mathbf{s} \geq 0$ for all $\mathbf{s} \in R^{n}$), there always exist a group of orthogonal eigenvectors and a group of real and nonnegative eigenvalues for any defined form of $\boldsymbol{\Gamma}$. We can obtain them from a singular value decomposition,

$$\boldsymbol{\Gamma} = \Phi\boldsymbol{\Lambda}\Phi^{\mathrm{H}}, \tag{2.36}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix consisting of eigenvalues λ, and the columns of Φ are the eigenvectors ϕ. We write their relation as

$$\Gamma(m,n) = \sum_{k=1}^{N} \lambda_k \phi_k(m) \phi_k^{*}(n), \tag{2.37}$$

where (m,n) denotes the matrix element in the m^{th} row and in the n^{th} column, and k is the number of the mode. The discrete Γ shows very similar property compared to the continuous Γ in Equation (2.31). Furthermore, we define the complex degree of coherence in a matrix form $\boldsymbol{\gamma}$. Its matrix element follows

$$\gamma(m,n) = \frac{\Gamma(m,n)}{[\Gamma(m,m)\Gamma(n,n)]^{1/2}}. \tag{2.38}$$

In the following, we extract the coherence information from the above-defined matrices and vectors.

In a fully incoherent case, any two distant points are uncorrelated. As a result, the matrices of $\boldsymbol{\Gamma}$ and $\boldsymbol{\gamma}$ are diagonal. Meanwhile $\boldsymbol{\gamma}$ is of full rank ($R = N$) and equals to the identity matrix. The distribution of the eigenvalues λ_n along the matrix diagonal of $\boldsymbol{\Lambda}$ is uniform**.**

In the completely coherent case, every two points are fully correlated to each other. Therefore, every element in $\boldsymbol{\gamma}$ has a magnitude of 1. The rank of $\boldsymbol{\Gamma}$ is 1. Only one eigenvalue is nonzero. That means the distribution of λ_n is as dense as possible.

In the case of partial coherence, the degree of coherence can be measured by the uniformity of the eigenvalues or the spreading of the $\boldsymbol{\gamma}$ matrix between the two extreme cases, *i.e.* full coherence and full incoherence**.**

Definition 1*:* we sort the eigenvalues from the largest (index $n = 1$) to smallest. The measure of the spread is defined as

$$c_1 = \frac{1}{N}\sum_{n=1}^{N}(n-1)^2\lambda_n. \tag{2.39}$$

When all the λ_n are unity, *i.e.* in a complete incoherent case, we have $c_1 = (N-1)(2N-1)/6$. When only one eigen value is nonzero, *i.e.* in a complete coherent case, we have $c_1 = 0$.

***Definition** 2:* an alternative to measuring the eigenvalues is to take the variance,

$$c_2 = \frac{1}{N}\sum_{n=1}^{N}(\lambda_n - 1)^2. \tag{2.40}$$

where we consider 1 as the average of the eigenvalues. In a complete incoherent case, we have $c_2 = N - 1$. In a complete coherent case, we have $c_2 = 0$.

Definition 3*:* we take the concept of entropy to define the maximum order versus disorder,

$$c_3 = -\sum_{n=1}^{N}\frac{\lambda_n}{N}\log\left(\frac{\lambda_n}{N}\right). \tag{2.41}$$

where we have $\sum_{n=1}^{N}(\lambda_n/N) = 1$. In a complete incoherent case, we have $c_3 = \log N$. In a complete coherent case, we have $c_3 = 0$.

Definition 4*:* we define the spatial variance of the matrix $\boldsymbol{\gamma}$ as

$$c_4 = \sum_{n=1}^{N}\sum_{m=1}^{N}(m-n)^2|\gamma(m,n)|^2. \tag{2.42}$$

When $\boldsymbol{\gamma}$ is a unity matrix (*i.e.* for completely incoherent light), we have $c_4 = 0$. When every element in $\boldsymbol{\gamma}$ is 1 (*i.e.* for completely coherent light), we have $c_4 = 2\sum_{l=1}^{N}(N-l)l^2$.

Definition 5*:* we measure the energy of the matrix $\boldsymbol{\gamma}$ as

$$c_5 = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}|\gamma(m,n)|^2. \quad (2.43)$$

When $\boldsymbol{\gamma}$ is a unity matrix (*i.e.* for completely incoherent light), we have $c_5 = 1/N$. When every element in $\boldsymbol{\gamma}$ is 1 (*i.e.* for completely coherent light), we have $c_5 = 1$.

Above are the five measures of the coherence degree based on a given correlation matrix. So far we have discussed the expansion of orthogonal modes. In the coming section, we introduce another way of modal expansion to provide more freedom in simulation work.

2.3.2 Shifted-elementary modes

The shifted-elementary mode is another type of modal expansion for a particular class of partially coherent light. It assumes the angular correlation function of the source, *i.e.* the far-field correlation function, obeys a Schell-model type. The Schell beams we have discussed above also fall into this category. This theory is originally proposed by Vahimaa and Turunen[18]. We recall their ideas in the following.

We first define a cross-spectral density function at the source plane ($z = 0$) as

$$\Gamma^{(0)}(r_1, r_2) = [I(r_1)I(r_2)]^{1/2}\gamma(r_1, r_2). \quad (2.44)$$

where $r = (x, y, z = 0)$. Its angular correlation function at the source plane reads

$$A^{(0)}(k_{1\perp}, k_{2\perp}) = \iint_{-\infty}^{\infty}\Gamma^{(0)}(r_1, r_2)\exp[-i(k_{1\perp}\cdot r_1 - k_{2\perp}\cdot r_2)]\,d^2r_1 d^2r_2. \quad (2.45)$$

where $A^{(0)}$ denotes the correlation in the angular spectrum at the source plane.

In the far zone, $r = (x, y, z > 0)$ the correlation function follows

$$\Gamma^{\infty}(r_1\hat{\boldsymbol{r}}_1, r_2\hat{\boldsymbol{r}}_2) = (2\pi k)^2\cos\theta_1\cos\theta_2\, A^{(0)}(\hat{\boldsymbol{r}}_{1\perp}, \hat{\boldsymbol{r}}_{2\perp})\frac{\exp[ik(r_1 - r_2)]}{r_1 r_2}, \quad (2.46)$$

where $k = 2\pi/\lambda$ is the wave number (λ being the wavelength), r_1 and r_2 denote the distances from the origin at the source plane to the observed points in the far zone, $\hat{\boldsymbol{r}}_1$ and $\hat{\boldsymbol{r}}_2$ are the unit vectors along r_1 and r_2, $\hat{\boldsymbol{r}}_{1\perp}$ and $\hat{\boldsymbol{r}}_{2\perp}$ represent the projections of $\hat{\boldsymbol{r}}_1$ and $\hat{\boldsymbol{r}}_2$ into the source plane, θ_1 (or θ_2) represents the angle between $\hat{\boldsymbol{r}}_1$ (or $\hat{\boldsymbol{r}}_2$) and the optical axis.

The essential assumption in the shifted elementary modal expansion is that the correlation function in the far-zone obeys a Schell-model form

$$A^{(0)}(k_{1\perp}, k_{2\perp}) = [I^{\infty}(k_{1\perp})I^{\infty}(k_{2\perp})]^{\frac{1}{2}}\gamma_A(\Delta k_{\perp}), \quad (2.47)$$

where $\Delta k_{\perp} = k_{1\perp} - k_{2\perp}$, $\gamma_A(\Delta k_{\perp})$ is the angular degree of coherence, and I^{∞} is a positive intensity function related to the far-zone radiant intensity in this form

$$I^{\infty}(\hat{\boldsymbol{r}}) = r^2\Gamma^{\infty}(k\hat{\boldsymbol{r}}, k\hat{\boldsymbol{r}}) = (2\pi k)^2\cos\theta^2\, I^{\infty}(k\hat{\boldsymbol{r}}_{\perp}). \quad (2.48)$$

Based on Equation (2.47), the correlation function at the source plane can be written as

$$\Gamma^{(0)}(r_1, r_2) = \int_{-\infty}^{\infty} \tilde{\gamma}_A(r') f^*(r_1 - r') f(r_2 - r')\, d^2 r'. \tag{2.49}$$

where $\tilde{\gamma}_A(r')$ is the Fourier transform of the angular correlation function $\gamma_A(\Delta k_\perp)$, and $f(r)$ represents the elementary field. Equation (2.49) implies an incoherent superposition of the coherent modes. Each mode has an identical field profile given by f centered at $r = r'$.

In case of a Schell beam, the correlation function at the source plane (assuming one transverse dimension for simplicity) is

$$\Gamma^{(0)}(r_1, r_2) = I(\bar{r})\gamma(\Delta r), \tag{2.50}$$

where $\bar{r} = (r_1 + r_2)/2$ and $\Delta r = r_1 - r_2$. The angular correlation function at the source plane is

$$A^{(0)}(k_{1\perp}, k_{2\perp}) = \tilde{I}(\Delta k_\perp)\tilde{\gamma}\left(\bar{k}_\perp\right), \tag{2.51}$$

where $\bar{k}_\perp = (k_{1\perp} + k_{2\perp})/2$, $\tilde{I}(\Delta k_\perp)$ is the Fourier transform of the intensity function $I(\bar{x})$, and $\tilde{\gamma}\left(\bar{k}_\perp\right)$ is the Fourier transform of the complex coherence degree $\gamma(\Delta x)$. The elementary function can be derived from

$$\left|\tilde{f}(k_\perp)\right|^2 = \left|\tilde{\gamma}\left(\bar{k}_\perp\right)\right|. \tag{2.52}$$

If the support of I is larger than f and γ, we simplify Equation (2.49) into the form,

$$\Gamma^{(0)}(r_1, r_2) = I(\bar{r}) \int_{-\infty}^{\infty} f^*(r_1 - r') f(r_2 - r')\, d^2 r'. \tag{2.53}$$

In numerical simulation, the integration in Equation (2.49) can be treated as a sum,

$$\Gamma^{(0)}(r_1, r_2) \approx I(\bar{r}) \sum_{m=1}^{M} f^*(r_1 - r_m) f(r_2 - r_m). \tag{2.54}$$

where r_m denotes the spatial shift between two adjacent modes.

Equation (2.54) declares that any partially coherent light that shows a Schell-type correlation function in the far-zone can be considered, at the source plane, as an incoherent sum of shifted elementary modes.

Unlike the eigenmodes forming an orthogonal expansion, shifted-elementary modes are non-orthogonal. In Chapter 4, we investigate further the properties of these two types of expansion with specific examples.

2.4 Coherence retrieval

Up to now, we discuss light propagation based on the assumption that the correlation function of a source is known. However, this is not always the case in practice. Instead, we often need to

firstly characterize the coherence of an unknown source in experiments, before employing it for any further realization. How to regain the coherence information of a given source becomes the main topic of this section. That is, in fact, an inverse problem of the modal expansion, or an inverse problem of the Wigner function.

For simplicity, we restrict ourselves on monochromatic sources and retrieving spatial coherence. There are various methods for recovering spatial coherence. We classify them into the following four categories: (a) interferometric method based on the visibility of interference fringes, *e.g.* Young's experiment[34]; (b) wavefront measurement, *e.g.* parameter retrieval by using a Shack-Hartmann sensor[35,36]; (c) mode recovery followed by iterative phase retrieval algorithms[37,38]; (d) phase space tomography[39-41]. A proper choice among these methods dependents on the degrees of freedom that tested beam offers. No matter which approach we employ, the common goal is to reconstruct the correlation function from pure intensity measurements.

In our work, we solve this inverse problem based on the modal expansion and the Wigner function. That corresponds to methods of (c) and (d). There are two reasons for our choice. First, unlike methods of (a) and (b), systems based on (c) and (d) require simpler measurements. Only a few elementary optical components (*e.g.* a few cylindrical lenses and a camera) are sufficient. Second, the techniques of (c) and (d) provide a full picture of the correlation function, while the measurements of (a) and (b) characterize the beam point by point, or area by area, which can be more time-consuming.

2.4.1 Mode recovery

As discussed above, we can always replace the correlation function Γ by an incoherent sum of multiple coherent beams f_n,

$$\Gamma(r_1, r_2, z) = \sum_n f_n^*(r_1, z) f_n(r_2, z), \tag{2.55}$$

where $r = (x, y)$, n is the index of the coherent mode. Here we may use either orthogonal or non-orthogonal mode expansion. The intensity of the full beam is

$$I(r, z) = \sum_n |f_n(r, z)|^2, \tag{2.56}$$

The essence of this method is to measure the intensity of the full beam at several transverse planes, denoted as $I^{ref}(r, z)$ for $z = z_1, z_2, \dots, z_k$. Afterwards, we reconstruct the fields of individual modes, *i.e.* $f_n(r, z)$, by using iterative phase retrieval algorithms[42]. Then the correlation function Γ can be reconstructed by using Equation (2.55).

In what follows, we outline a representative algorithm proposed by Rydberg[37]. We start from an initial guess for the amplitudes of each mode, g_n^0, on the plane of $z = z_1$. The intensities on this plane based on the guessed amplitudes are calculated by

$$I(r, z_1) = \sum_n |g_n^o(r, z_1)|^2. \tag{2.57}$$

A correction term is defined as

$$cr(r, z_1) = \sqrt{I^{ref}(r, z_1)/I(r, z_1)}. \tag{2.58}$$

The amplitude of each mode on this transverse plane is individually modified by the correction term:

$$g_n^+(r, z_1) = cr(r, z_1) \cdot g_n^o(r, z_1). \tag{2.59}$$

We then propagate each mode of $g_n^+(r, z_1)$ from the local position $z = z_1$ to the next transverse plane $z = z_2$ by

$$g_n(r, z_2) = \boldsymbol{P}\{g_n^+(r, z_1)\}, \tag{2.60}$$

where $\boldsymbol{P}$ is the propagation operator in free space. Here we employ the angular spectrum method for propagation,

$$g_n(r, z_2) = \boldsymbol{FT}_{k_x,k_y}{}^{-1}\left\{\boldsymbol{FT}_{x,y}[g_n^+(r, z_1)] \cdot \exp\left[i\frac{2\pi}{\lambda}(z_2 - z_1)\sqrt{1 - (\lambda k_x)^2 - \left(\lambda k_y\right)^2}\right]\right\}, \tag{2.61}$$

where the coordinates (k_x, k_y) indicate the spatial frequencies along (x, y) directions.

The amplitudes $g_n(r, z_2)$ obtained from Equation (2.60) are considered as the initial amplitudes on the next plane $z = z_2$. We repeat the same process of Equations (2.57-2.60) for $z = z_2, z_3, \dots, z_k$. Afterwards, we propagate the fields $g_n^+(r, z_k)$ back to z_1, and repeat the iterations again. The iteration stops when the composed intensity given by Equation (2.57) approaches the measured intensity I^{ref}. Then the correlation function follows

$$\Gamma(r_1, r_2, z_i) = \sum_i g_n^*(r_1, z_i) g_n(r_2, z_i). \tag{2.62}$$

where z_i is one of the transverse planes.

The above algorithm is, in fact, the Gerchberg-Saxton algorithm in an adapted form to partially coherent light. We investigate this method for a beam with phase aberration in Chapter 5.

Moreover, using the phase retrieval algorithms for the coherence recovery brings up the following indispensable question. Are the recovered fields g_n and thus the recovered correlation function Γ always unique? In other words, can multiple transverse intensities uniquely define

the spatial coherence of a beam? The answer is nontrivial. Related research[37,43,44] has been carried on for years. For a beam with one transverse dimension, the propagated intensities in the entire space uniquely determine the spatial coherence[43]. However, for a beam with two transverse dimensions, the propagated intensities in space do not determine a unique correlation function. A typical example is the twisted Gaussian-Schell beam[45]. Its propagated intensities in space can be identical to a Gaussian-Schell beam. However, their individual correlation functions have distinct difference. That means, in order to retrieve the coherence information with two transverse dimensions, a simple measurement of the intensities in free space is not sufficient. We need to introduce one more degree of freedom into the system, in order to remove the azimuthal ambiguity. That can be achieved by inserting a rotationally asymmetric component into the beam path. A common and simple choice can be a cylindrical lens.

In Chapter 5 we extend this approach to a partially coherent beam with an arbitrary wavefront and discuss the potential and limit of this method.

2.4.2 Phase space tomography

Phase space tomography is an alternative approach to the coherence reconstruction. This approach reconstructs the Wigner function from intensity measurements. As the Wigner function is a Fourier conjugate of the correlation function, by recovering one of them, we have direct access to the other.

For a beam with two transverse dimensions, the Wigner function spans into four dimensions (4D), *i.e.* $W(x, y, u, v)$, where (x, y) denotes a spatial position on the transverse plane, u and v are the propagation angles with respect to the optical axis. Similar to the 2D Wigner function (Section 2.2), an integration of the 4D Wigner function over the angular axes, *i.e.* $\int W \, \mathrm{d}u\mathrm{d}v$, yields a 2D intensity distribution in space. An integration of the Wigner function over the spatial axes, *i.e.* $\int W \, \mathrm{d}x\mathrm{d}y$, corresponds to the 2D angular spectrum. A propagation through a first-order system can be expressed in the Wigner function by a 4 by 4 ABCD matrix.

The goal of phase space tomography is to reconstruct the 4D Wigner function at the input plane of a first-order system. The idea is as follows. The parameters of the first-order system define the ABCD matrix. The Wigner function at the input and output planes are directly linked by this ABCD matrix. We use a camera to capture the 2D intensity at the output plane. The measured intensity is an integration of the output Wigner function over the angular axes. Since the ABCD matrix of the system is given, the connection between the measured intensity and the input

Wigner function is accessible. We then change the parameters of the system and obtain varied intensities at the output plane. The input Wigner function can be reconstructed from these intensities. This approach is similar to recovering data from certain projections. That is the reason we call this approach as phase space tomography.

The essence of this method concentrates on the matrix transformation, especially the projecting relation between the input Wigner function and the measured intensity at the output plane. Schäfer[46], Bastiaans and Alieva (chapters 1 and 3 in the reference[13]) have given a conductive summary of the corresponding theory. We recall it here as a theoretical introduction.

We begin with the definition of a symplectic matrix[13]. Let

$$\boldsymbol{J} = i\begin{bmatrix} 0 & -\boldsymbol{I} \\ \boldsymbol{I} & 0 \end{bmatrix} = \boldsymbol{J}^{-1} = -\boldsymbol{J}^{t} = \boldsymbol{J}^{H}, \tag{2.63}$$

where $\boldsymbol{I}$ denotes a 2 x 2 identity matrix, $\boldsymbol{J}^{-1}$, $\boldsymbol{J}^{t}$ and $\boldsymbol{J}^{H}$ represent the inverse, transpose and Hermitian transpose of $\boldsymbol{J}$. If there is a 4 by 4 matrix $\boldsymbol{T}$ satisfying

$$\boldsymbol{T}^{-1} = \boldsymbol{J}\boldsymbol{T}^{t}\boldsymbol{J}, \tag{2.64}$$

then we call $\boldsymbol{T}$ as a symplectic matrix. Note that, a first-order ray transformation matrix (*i.e.* the 4 by 4 ABCD matrix) is symplectic,

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}, \tag{2.65}$$

where each symbol of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{D}$ represents a 2 by 2 submatrix.

$$\boldsymbol{S}^{-1} = \boldsymbol{J}\boldsymbol{S}^{t}\boldsymbol{J}. \tag{2.66}$$

We then have

$$W_o(\vec{x}, \vec{u}) = W_i(\boldsymbol{D}^{t}\vec{x} - \boldsymbol{B}^{t}\vec{u}, -\boldsymbol{C}^{t}\vec{x} + \boldsymbol{A}^{t}\vec{u}). \tag{2.67}$$

where W_i and W_o denote the Wigner functions at the input and output planes of the first-order system, $\vec{x} = (x, y)$, $\vec{u} = (u, v)$.

The measured intensity at the output plane follows

$$I_o(\vec{x}) = \int W_o(\vec{x}, \vec{u})\, \mathrm{d}\vec{u}, \tag{2.68}$$

We use the Dirac's delta function to express Equation (2.68),

$$I_o(\vec{x}) = \iint W_o(\vec{x}_o, \vec{u}_o)\delta(\vec{x}_o - \vec{x})\mathrm{d}\vec{x}_o\mathrm{d}\vec{u}_o, \tag{2.69}$$

Note that, $(\vec{x}_o, \vec{u}_o)$ are the phase space coordinates at the output plane. Additionally, we define the phase space coordinates at the input plane as $(\vec{x}_i, \vec{u}_i)$. We then have, $\vec{x}_o = \boldsymbol{A}\vec{x}_i + \mathbf{B}\vec{u}_i$, $\vec{u}_o = \boldsymbol{C}\vec{x}_i + \mathbf{D}\vec{u}_i$, $\mathrm{d}\vec{x}_o \sim \mathrm{d}\vec{x}_i$ and $\mathrm{d}\vec{u}_o \sim \mathrm{d}\vec{u}_i$ (symbol $\sim$ denoting a proportional relation). Equation (2.69) becomes

$$I_o(\vec{x}) = \iint W_i(\vec{x}_i, \vec{u}_i)\delta(\boldsymbol{A}\vec{x}_i + \mathbf{B}\vec{u}_i - \vec{x})\mathrm{d}\vec{x}_i\mathrm{d}\vec{u}_i. \tag{2.70}$$

Equation (2.70) paves the path to recovering W_i by knowing I_o. The 2D intensity I_o is a projection of the 4D function W_i. The projection direction is defined by the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The reconstruction of the 4D matrix W_i is an inverse Radon transform. We focus on the following two algorithms to realize this method. One is the filtered back projection. The other algorithm is the Fourier domain reconstruction. These two algorithms are mathematically equivalent. However, in practice, they involve different numerical techniques. We summarize these two algorithms as follows.

Fourier domain reconstruction

A Fourier transform of Equation (2.70) follows,

$$\begin{aligned}\int I_o(\vec{x}) \exp\left(i\vec{k}^t \cdot \vec{x}\right) \mathrm{d}\vec{x} &= \iiint W_i(\vec{x}_i, \vec{u}_i)\delta(\boldsymbol{A}\vec{x}_i + \mathbf{B}\vec{u}_i - \vec{x}) \exp\left(i\vec{k}^t \cdot \vec{x}\right) \mathrm{d}\vec{x}_i\mathrm{d}\vec{u}_i\mathrm{d}\vec{x}, \\ &= \iint W_i(\vec{x}_i, \vec{u}_i) \exp\left(i\vec{k}^t\boldsymbol{A}\vec{x}_i\right) \exp\left(i\vec{k}^t\boldsymbol{B}\vec{u}_i\right) \mathrm{d}\vec{x}_i\mathrm{d}\vec{u}_i.\end{aligned} \tag{2.71}$$

where $\vec{k} = \left(k_x, k_y\right)$ represents the spatial frequencies along x, y directions respectively. The last integral in Equation (2.71) is in a similar form of the ambiguity function. The definition of an ambiguity function Z is a 4D Fourier transform of the Wigner function W [47],

$$Z(\vec{q}_x, \vec{q}_u) = \iint \iint W_i(\vec{x}_i, \vec{u}_i) \exp(i\vec{q}_x^t \cdot \vec{x}_i) \exp(i\vec{q}_u^t \cdot \vec{u}_i) \mathrm{d}\vec{x}^i\mathrm{d}\vec{u}^i, \tag{2.72}$$

where $\vec{q}_x = \left(q_x, q_y\right)$, $\vec{q}_u = (q_u, q_v)$.

By comparing Equations (2.71) and (2.72), we have

$$Z(\vec{q}_x, \vec{q}_u) = \int I_o(\vec{x}) \exp\left(i\vec{k}^t \cdot \vec{x}\right) \mathrm{d}\vec{x} = \widetilde{I_o}\left(\vec{k}\right). \tag{2.73}$$

with

$$\vec{q}_x = \boldsymbol{A}^t\vec{k}, \quad \vec{q}_u = \boldsymbol{B}^t\vec{k}. \tag{2.74}$$

Equation (2.73) and (2.74) indicates the key steps in the 4D Fourier domain reconstruction. The value of the ambiguity function $Z(\vec{q}_x, \vec{q}_u)$ is equal to the Fourier transform of the intensity distribution, *i.e.* $\widetilde{I_o}\left(\vec{k}\right)$. The reconstruction process is merely a coordinate transform. To display an intuitive relation, we rewrite Equation (2.74) as

$$\begin{bmatrix} q_x \\ q_y \end{bmatrix} = \boldsymbol{A}^t \begin{bmatrix} k_x \\ k_y \end{bmatrix}, \ \begin{bmatrix} q_u \\ q_v \end{bmatrix} = \boldsymbol{B}^t \begin{bmatrix} k_x \\ k_y \end{bmatrix}. \tag{2.75}$$

Equation (2.75) reveals that each coordinate at the Fourier transform of the intensity corresponds to another coordinate at the ambiguity function. This coordinate relation is determined by the

parameters in the matrices $\boldsymbol{A}^t$ and $\boldsymbol{B}^t$. Therefore, the Fourier domain reconstruction is an ambiguity function reconstruction.

We write the system matrix in a complete way,

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} A_x & 0 & B_x & 0 \\ 0 & A_y & 0 & B_y \\ C_x & 0 & D_x & 0 \\ 0 & C_y & 0 & D_y \end{bmatrix}. \tag{2.76}$$

Equation (2.75) then becomes

$$\begin{bmatrix} q_x \\ q_y \end{bmatrix} = \begin{bmatrix} A_x k_x \\ A_y k_y \end{bmatrix}, \begin{bmatrix} q_u \\ q_v \end{bmatrix} = \begin{bmatrix} B_x k_x \\ B_y k_y \end{bmatrix}. \tag{2.77}$$

If the optical system is stigmatic (*i.e.* rotationally symmetric), it yields $A_x = A_y$ and $B_x = B_y$. In that case, only the points with the following coordinates can be recovered,

$$q_x/q_u = q_y/q_v. \tag{2.78}$$

Equation (2.78) corresponds to a 3D space inside the 4D ambiguity function. This is not a full reconstruction of the entire ambiguity function. Therefore, a stigmatic system is incapable of measuring the coherence of an arbitrary source. It only works for a limited class of sources. For example, a separable beam has an ambiguity function $Z(q_x, q_y, q_u, q_v) = Z_x(q_x, q_u)Z_y(q_y, q_v)$. In that case, we can use a stigmatic system to measure Z_x and Z_y separately. The same argument exists in the filtered back projection.

Therefore, for a general coherence measurement, we need an astigmatic system to fulfill $A_x \neq A_y$ and $B_x \neq B_y$. The values of these parameters are determined in the following way. Assume we want to obtain the value of a specific point with coordinates (q_x, q_y, q_u, q_v) in the ambiguity function. According to Equation (2.77), we have

$$k_x = q_x/A_x = q_u/B_x \text{ , } k_y = q_y/A_y = q_v/B_y. \tag{2.79}$$

Equation (2.79) yields this relation,

$$B_x/A_x = q_u/q_x \text{ , } B_y/A_y = q_v/q_y. \tag{2.80}$$

We then vary the system parameters to fulfill Equation (2.80). The value of the point (q_x, q_y, q_u, q_v) in the ambiguity function equals to the Fourier transform of the intensity at the output plane of the system, *i.e.* $\widetilde{I_o}(\vec{k})$, with a spatial frequency coordinate

$$\begin{bmatrix} k_x \\ k_y \end{bmatrix} = \begin{bmatrix} q_x/A_x \\ q_y/A_y \end{bmatrix} = \begin{bmatrix} q_u/B_x \\ q_v/B_y \end{bmatrix}. \tag{2.81}$$

After introducing the key steps of the ambiguity function reconstruction, we present the filtered back projection as follows.

Phase space rotator and filtered back projection

Before starting the algorithm of the filtered back projection, we first summarize the theory of the phase space rotator. It serves as a prerequisite for the filtered back projection.

According to the modified Iwasawa decomposition[48], a normalized symplectic ray transformation matrix can be decomposed into three fundamental matrices,

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{G} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{K} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ -\boldsymbol{Y} & \boldsymbol{X} \end{bmatrix} = \boldsymbol{T_L T_M T_O}, \tag{2.82}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ are each a 2 by 2 matrix, and $\boldsymbol{I}$ represents a unitary matrix. The matrix $\boldsymbol{T}_L$ corresponds to a lens transformation, where $\boldsymbol{G}$ is a symmetric matrix,

$$\boldsymbol{G} = -(\boldsymbol{C}\boldsymbol{A}^t + \boldsymbol{D}\boldsymbol{B}^t)(\boldsymbol{A}\boldsymbol{A}^t + \boldsymbol{B}\boldsymbol{B}^t)^{-1} = \boldsymbol{G}^t. \tag{2.83}$$

The physical meaning of $\boldsymbol{T}_L$ equals to second-order phase modulation. The second matrix $\boldsymbol{T}_M$ represents a magnifier, where $\boldsymbol{K}$ is a symmetric matrix,

$$\boldsymbol{K} = (\boldsymbol{A}\boldsymbol{A}^t + \boldsymbol{B}\boldsymbol{B}^t)^{1/2} = \boldsymbol{K}^t. \tag{2.84}$$

Therefore, $\boldsymbol{T}_M$ simply introduces a magnification factor into the system.

The third matrix $\boldsymbol{T}_O$ is an orthogonal and symplectic matrix. It yields the most crucial contribution to the signal transformation. According to Wolf[48], any orthogonal symplectic matrix can be decomposed into three basic systems,

$$\boldsymbol{T}_O = \boldsymbol{T}_r(\vartheta)\boldsymbol{T}_f(\gamma_x, \gamma_y)\boldsymbol{T}_r(\epsilon), \tag{2.85}$$

To understand these matrices in an intuitive way, we rewrite $\boldsymbol{T}_O$ as a 2 by 2 matrix $\boldsymbol{U}_O$,

$$\boldsymbol{U}_O = \boldsymbol{X} + i\boldsymbol{Y} = (\boldsymbol{A}\boldsymbol{A}^t + \boldsymbol{B}\boldsymbol{B}^t)^{-1/2}(\boldsymbol{A} + i\boldsymbol{B}). \tag{2.86}$$

We then have

$$\boldsymbol{U}_O = \boldsymbol{U}_r(\vartheta)\boldsymbol{U}_f(\gamma_x, \gamma_y)\boldsymbol{U}_r(\epsilon). \tag{2.87}$$

The form of $\boldsymbol{U}_r$ and $\boldsymbol{U}_f$ is as follows:

$$\boldsymbol{U}_r(\vartheta) = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix}, \tag{2.88}$$

where $\boldsymbol{U}_r(\vartheta)$ is a clockwise rotation in the xy and uv planes by an angle ϑ, and

$$\boldsymbol{U}_f(\gamma_x, \gamma_y) = \begin{bmatrix} \exp(i\gamma_x) & 0 \\ 0 & \exp(i\gamma_y) \end{bmatrix}, \tag{2.89}$$

where $\boldsymbol{U}_f$ represents a rotation in the xu plane by an angle γ_x, and another rotation in the yv plane by an angle γ_y. Therefore, $\boldsymbol{T}_O$ in Equation (2.85) can be interpreted as a fractional Fourier transform $\boldsymbol{T}_f$ that is embedded between two rotators of $\boldsymbol{T}_r$.

The above theory indicates that we can use phase space rotations to express for any first order system. When we change the parameters of the systems, the rotation angles vary accordingly. Each intensity distribution captured by the camera is a projection of the rotated Wigner function. Based on this idea, we can reconstruct the Wigner function by using the filtered back projection. A back projection is an inverse implementation of the Radon transform. A forward Radon transform projects a 2D distribution, denoted as $W(x,u)$, along a radial angle γ. It equals to an integration of $W(x,u)$ along parallel lines that form an angle γ with the y-axis. This yields a 1D projected distribution, defined as $S(\mu)$, for each specific angle γ, μ being the spatial axis of the projected distribution. In the inverse case, we are given the projections of $W(x,u)$ along varying angles γ, *i.e.* $S(\mu,\gamma)$. The goal is to recover the original distribution of $W(x,u)$ from $S(\mu,\gamma)$.

We also call the 2D projected distribution $S(\mu,\gamma)$ as a sinogram for the back projection, where μ is the spatial axis, and γ being the angular axis. Before processing the back projection, we apply a filter to the sinogram. There are various filters available. One of the common ways is to take the derivative of the sinogram $S(\mu,\gamma)$ along the μ axis, defined as $\partial S(\mu,\gamma)/\partial\mu$. Afterwards, we back project each line of $\partial S(\mu,\gamma)/\partial\mu$ with a specific angle γ to fill $W(x,u)$.

A common and simple setup to realize this method is the fractional Fourier transform system. One can build the setup by using two cylindrical lenses and one camera[39]. The degrees of freedom in the system are the distances between the components. By changing the distances, different angles of the phase space rotation are achieved. Alternatively, one can fix the distances and only rotate the cylindrical lenses[41]. That prevents axially shifting the components, and thus improves the measurement accuracy. Furthermore, Camara[49] proposed using two spatial light modulators to replace the cylindrical lenses[49]. Each spatial light modulator combines the phase of two rotating cylindrical lenses. The degrees of freedom are the rotating angle and the power

of each lens. This setup significantly accelerates the coherence measurement by a digital synchronization.

If we compared the above two algorithms, *i.e.* Fourier domain reconstruction and the filtered back projection, they are mathematically equivalent and connected by a Fourier transform (Equation (2.71)). However, they differ in numerical procedures. In the Fourier domain reconstruction, each value of $\widetilde{I_o}(\vec{k})$ is mapped only once into the 4D ambiguity function $Z(\vec{q}_x, \vec{q}_u)$. In contrast, the filtered back projection maps each value multiple times into the 4D Wigner function. Therefore, the computational effort of the Fourier space reconstruction is less. However, given by Equation (2.74), using direct regridding interpolation is inevitable in reconstructing the ambiguity function. Its numerical scheme crucially decides the accuracy of the recovered Wigner function.

To improve the numerical accuracy, in Chapter 5.2, we present an efficient implementation of the 2D filter back projection. Based on that, we propose an algorithm for reconstructing the 4D Wigner function.

3 Interaction of partially coherent light with optical systems

This chapter contains our results and discussion of modeling partially coherent light in optical systems. Our modeling methods include the Wigner function and the modal expansion. These two techniques represent partially coherent light in different perspectives.

Phase space expresses an optical signal with a specific position and angle. If we consider each point in phase space as a delta function, a light field is the overall contribution of all the delta functions. This representation is independent of the coherence degree. Therefore, it is a particularly convenient tool for a beam with a low degree of coherence. However, in simulation work, propagating the full beam involves processing individual delta functions. That requires much computational time and effort. Especially when we extend our investigation into two transverse dimensions in space, the Wigner function spans four degrees of freedom.

In contrast, modal expansion models a partially coherent beam as an incoherent sum of multiple modes. The amount of computational effort relies on the coherence degree. This approach is particularly beneficial for a beam with a high degree of coherence. A finite amount of mode coefficients is sufficient to describe the full beam propagation. Figure 3-1 depicts a qualitative comparison of the computational effort between these two approaches.

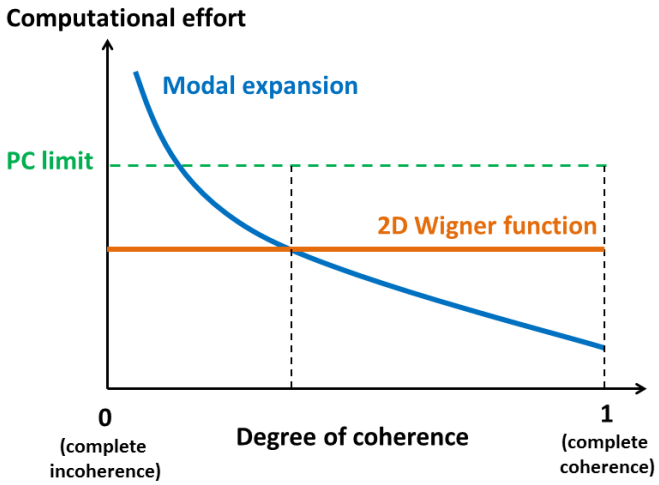


Figure 3-1 Qualitative comparison of the computational effort between the Wigner function and the modal expansion.

In this chapter, we explore individual properties of these two modeling methods and develop propagation operators. Using these operators, we interpret the optical effects of partially coherent light in different systems. The content of this chapter is arranged in the following order.

Section 3.1 focuses on the phase space representation. We divide this into two parts. First, we examine the diffraction effects generated by various discontinuous surfaces (Section 3.1.1). Second, we present our improved propagation algorithms for the Wigner function (Section 3.1.2). Section 3.2 presents our investigation into the modal expansion. We use Schell beams as examples to compare different expansion approaches (Section 3.2.1). Later, we propose an efficient algorithm to compute the parameters of the shifted elementary modes (Section 3.2.2). Furthermore, we develop an expansion tool to propagate partially coherent light inside waveguides, to accurately model the diffraction effect generated by the waveguide structure (Section 3.2.3 and 3.2.4).

3.1 Wigner function in phase space representation

Our first modeling tool for partially coherent light is the Wigner function. In particular, we use phase space to analyze the diffraction effects generated by discontinuous surfaces. These surfaces are treated as thin elements and classified into two types, those with a discontinuity in space and those with a discontinuity in slope. The corresponding optical effects are discussed in the Wigner function. This approach explains the performance of segmented elements during the transition from the refractive into the diffractive regime.

3.1.1 Discontinuous surfaces in Wigner functions

We first discuss the diffraction effects generated by a single discontinuity (*e.g.* a phase step and a linear axicon). Later on, we consider surfaces with periodic discontinuities, *e.g.* gratings, to study the formation of multiple diffracted orders. A kinoform lens is given as an example to visualize the change from pure refraction to diffraction. Moreover, we present the beam homogenizing effect in phase space generated by the lens arrays.

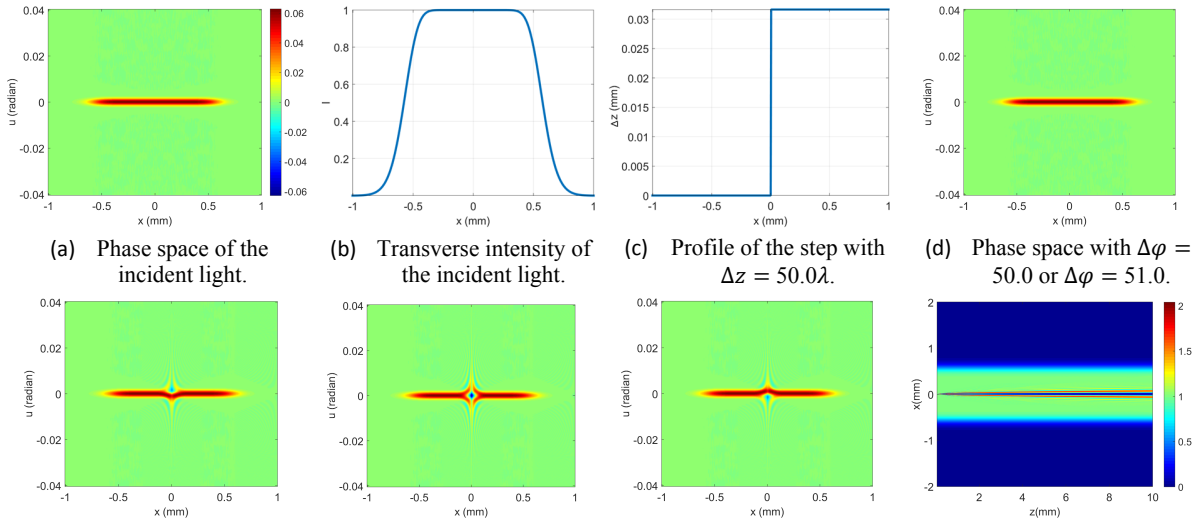
Before starting the surface examples, we introduce an important parameter $\Delta\varphi$ to denote the optical path difference produced by a phase step. A ray of light in the wavelength λ traveling through a refractive material with a height of Δz accumulates the phase term $\exp(i2\pi n_{rf}\Delta z/\lambda)$, where n_{rf} is the refractive index of the material. Meanwhile, another ray of light next to the material traveling through the air of the same thickness Δz carries another phase term $\exp(i2\pi\Delta z/\lambda)$. We define the normalized phase difference between the two different rays as $\Delta\varphi = (n_{rf} - 1)\Delta z/\lambda$, *i.e.* the phase difference divided by 2π. In the special blazed condition, $\Delta\varphi$ is an integer. It means the optical path difference is a multiple of the wavelength. Thus it

results in no interference effects. In the following discussions, we concentrate on cases where $\Delta\varphi$ is a non-integer, to examine the diffraction effects.

Phase step

We employ a flat-top Schell beam[20] to be the source, to ensure a uniform illumination on the diffractive element. The source is characterized by three near-field parameters according to Equation (2.20): RMS beam width σ, rms correlation width δ and the mode count M [20]. The degree of coherence in our model decreases when the value of σ increases. A larger integer of M produces a more flat-top profile in the far-field intensity.

We define the refractive index $n_{rf} = 2.0$ for all optical elements and the wavelength $\lambda = 0.6328$ μm. In Fig. 3-2 the step heights are defined as $\Delta z = 50.0\lambda$, 50.2λ, 50.5λ, 50.8λ and 51.0λ, yielding $\Delta\varphi = 50.0$, 50.2, 50.5, 50.8 and 51.0 respectively (Fig. 3-2 d-g).



(a) Phase space of the incident light.
(b) Transverse intensity of the incident light.
(c) Profile of the step with $\Delta z = 50.0\lambda$.
(d) Phase space with $\Delta\varphi = 50.0$ or $\Delta\varphi = 51.0$.
(e) Phase space with $\Delta\varphi = 50.2$.
(f) Phase space with $\Delta\varphi = 50.5$.
(g) Phase space with $\Delta\varphi = 50.8$.
(h) Intensity along propagation with $\Delta\varphi = 50.5$.

Fig. 3-2. Phase steps of various heights. The parameters for the incident light source are σ= 1 μm, δ = 0.5 μm and M= 40.

With $\Delta\varphi = 50.0$ or $\Delta\varphi = 51.0$ there is no change in the phase space between the incoming beam and the outgoing beam (compare Fig. 3-2a, d). That is because the light experiences no phase jump after passing the step.

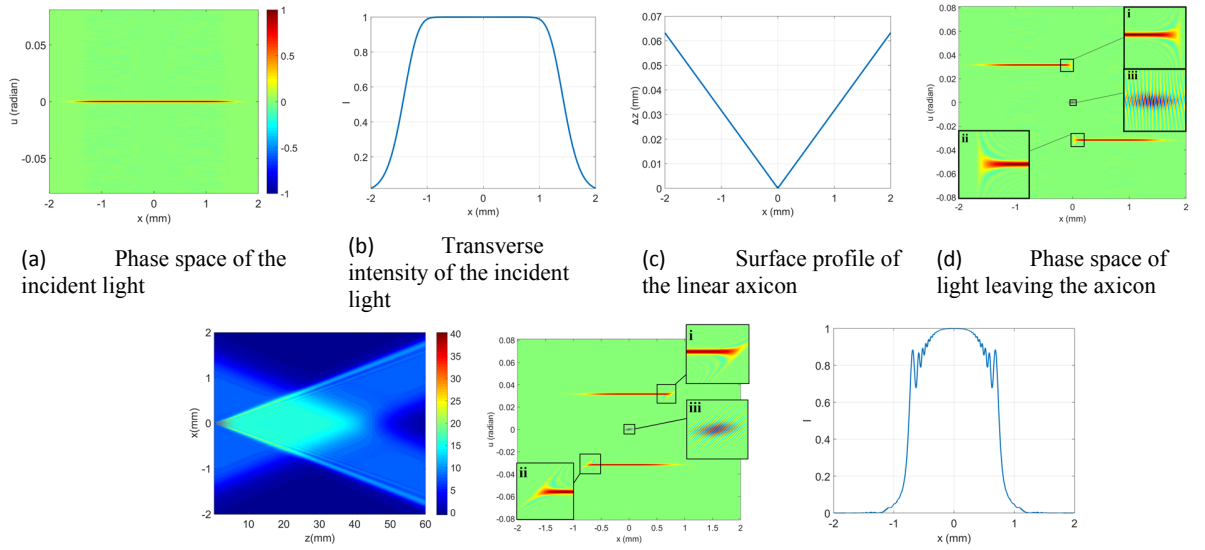
In the cases of $\Delta\varphi$ being a non-integer (Fig. 3-2 e-g), ripples occur in phase space. They are the visualization of diffraction effects centered at the step location. These diffraction ripples are an indication of the Fourier series of the Heaviside step function. In the case of $\Delta\varphi = 50.5$, the

diffraction ripples are most pronounced. The ripple distribution is symmetric about the origin (*i.e.* $x = 0$ and $u = 0$). In the intensity along the propagated beam (Fig. 3-2h) the destructive interference produces a minimum at the transverse intensity at $x = 0$. Besides, the phase space of $\Delta\varphi = 50.2$ (Fig. 3-2e) is a 180° rotation of the phase space of $\Delta\varphi = 50.8$ (Fig. 3-2g). The same applies to pairs of $\Delta\varphi = 50.3$ and 50.7, $\Delta\varphi = 50.2$ and 50.8, $\Delta\varphi = 50.1$ and 50.9.

It is worth mentioning that when the light source has a larger coherence length, the diffraction ripples are more pronounced. This is because a more coherent beam generates more noticeable interference effects. Conversely, as the coherence length of the light source decreases, the diffraction ripples become fainter.

Axicon

We take a linear axicon as another example to discuss the single discontinuity in phase. Compared to the phase step, a linear axicon is continuous in space but discontinuous in slope. Diffraction effects always happen when there is a slope discontinuity on the surface, even though there is no abrupt change in the step height. In this case $\Delta\varphi$ is no longer a criterion to evaluate the destructive interference.



(a) Phase space of the incident light

(b) Transverse intensity of the incident light

(c) Surface profile of the linear axicon

(d) Phase space of light leaving the axicon

(e) Intensity along the propagated beam

(f) Phase space of light at z= 25 mm

(g) Transverse intensity at z= 25 mm

Fig. 3-3. Phase space of a quasi-collimated Schell-model beam passing a linear axicon. The parameters for the incident light source are σ = 0.5 μm, δ = 0.2 μm and M= 40.

Corresponding results are shown in Fig. 3-3. The axicon splits a quasi-collimated incident beam (Fig. 3-3a) into two beamlets with different angles. This is indicated in the phase space (Fig. 3-3d) as two horizontal lines located at opposite angles. A split happens where the surface slope changes, *i.e.* $x = 0$.

Also, diffraction ripples appear due to the phase wedge of the axicon. The ripples in Fig. 3-3 d(i) and (ii) are connected to the two horizontal lines. They resemble the diffraction ripples generated by the phase step in Section 3.2.1. These ripples propagate with the beamlets and produce oscillations in the transverse intensity near the outer edges of the beamlets (Fig. 3-3 g). Meanwhile the ripples in Fig. 3-3d (iii) near the origin are separate from the two lines. They result from interference between the two beamlets. The ripple frequency increases with the angle between the two beamlets[50]. When we integrate these ripples over the angular axes, they produce oscillations in the transverse intensity. For a longer propagated distance, the ripples are sheared, and their effect on the transverse intensity diminishes. The axial range where oscillations can still be observed ($z < 3\ mm$) in the transverse intensity is proportional to the coherence length of the incident beam.

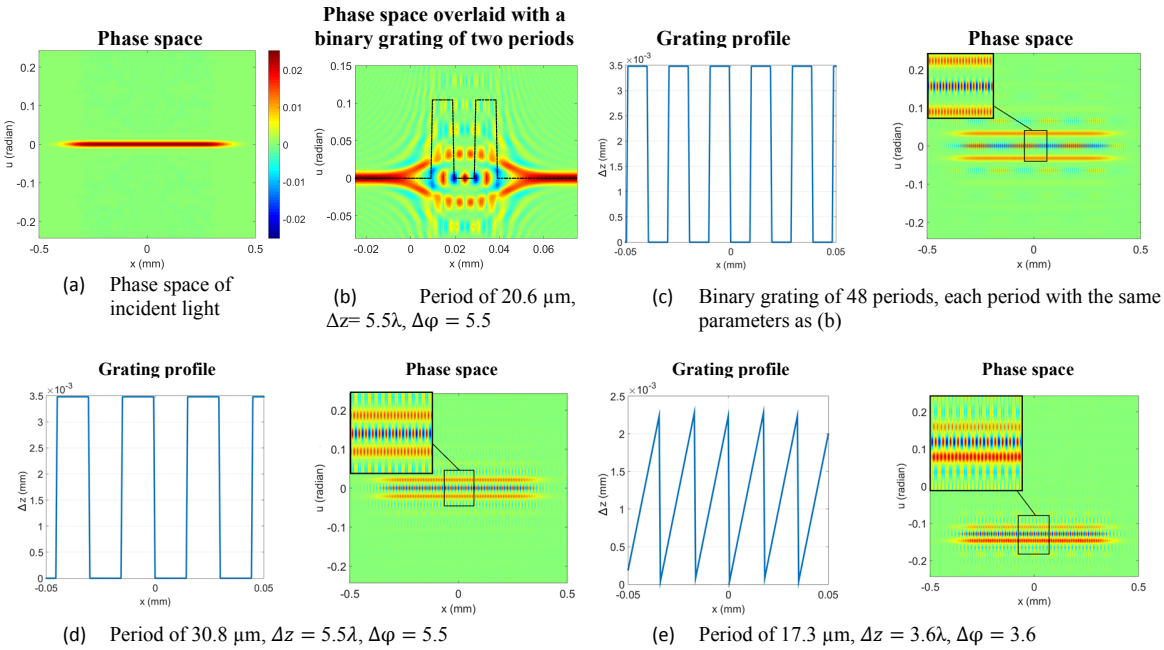
Grating

In the following, we show several examples of phase space corresponding to various phase gratings.

A binary grating can be considered as a periodic superposition of Heaviside steps. The phase space of a single Heaviside step is discussed in the previous section (Fig. 3-4 d-g). Here we start with a grating composed of four Heaviside steps. The grating profile is overlaid on top of the phase space in Fig. 3-4b. On the line with $u = 0$ in phase space, negative values occur at four separate x positions. These positions coincide with the phase discontinuity of the surface. Another notable feature in Fig. 3-4b is the positive peaks on the line through $u = 0.033\ radian$. The three horizontal lines of signals along $u = 0$ and $u = \pm 0.03$ radian form the 0 and ±1 diffracted orders, respectively. The angles of diffracted orders follow the formula $\sin u = m_g \lambda / d_g$, where u is the diffracted angle, m_g being the diffracted order, λ representing the wavelength and d_g denoting the grating period.

In Fig. 3-4c the grating has the same period as the one in Fig. 3-4b, but covering a larger spatial extent. If we derive the angular spectrum from the phase space, there are sharper peaks

representing the diffracted orders (not shown in the figures). However, the width of the peaks in the angular spectrum is also influenced by the angular extent of the partially coherent source. Fig. 3-4d shows a binary grating with a larger period. According to the diffraction formula, a larger value of d results in a smaller diffracted angle u. The ±1 diffracted orders are at $u = \pm 0.023\ radian$. Fig. 3-4e illustrates a blazed grating (also called echelette grating). Due to the asymmetric profile of this grating, both refraction and diffraction take place. All the diffracted orders move along the angular axis. The 0 order is repositioned at the angle $u = -0.12\ radian$.



(a) Phase space of incident light

(b) Period of 20.6 μm, Δz= 5.5λ, Δφ = 5.5

(c) Binary grating of 48 periods, each period with the same parameters as (b)

(d) Period of 30.8 μm, $\Delta z = 5.5\lambda$, Δφ = 5.5

(e) Period of 17.3 μm, $\Delta z = 3.6\lambda$, Δφ = 3.6

Fig. 3-4. Phase gratings. The total spatial extent of the gratings in (c), (d) and (e) is from -0.5 mm to 0.5 mm along the x axis. We only show the grating profiles along the x axis from -0.05 mm to 0.05 mm for a clear visibility of the periods. The parameters for the incident light source are σ= 5 μm, δ= 0.8 μm and M=40.

Moreover, the step height of the blazed grating defines the value of the normalized phase difference $\Delta\varphi$. With $\Delta\varphi = 3.6$, the -1 order contains more energy than the +1 order (Fig. 3-4e). For a binary grating, the energies in the +1 and -1 orders are always equal.

There is another interesting feature about the phase space of a grating. The periodic signals in the Wigner function (Fig. 3-4 b-e) lead to a self-imaging of the grating. We explain this by propagating the diffracted light further in free space. That is expressed as a shearing of the Wigner function. As a result, the diffracted orders in the Wigner function are shifted horizontally

away from each other. Because each diffracted order consists of periodic structures, their relative shift makes the negative and positive signals pass each other periodically. At a certain propagation distance, the vertical integration of all signals gives rise to a periodic intensity distribution. It regenerates the spatial structure of the grating. This is also called as a Talbot effect (Chapter 9.10 in the reference[13]).

Kinoform lens

A kinoform lens can be interpreted as a grating with locally varying periods. When the periods can be approximated by a linear blazed grating, one may use the method proposed by Sinzinger[51] to analyze the diffraction effects based on Fourier analysis. In our work, we express the profile of the kinoform lens as a conical surface with $z = c_R x^2 / \left[1 + (1 - (1+\kappa) {c_R}^2 x^2)^{1/2}\right]$, where c_R denotes the surface curvature (*i.e.* the reciprocal of the surface radius), and κ is the conic constant. We cut this surface into slices with an equal height Δz, and leave out all the coplanar parts. This cutting yields a kinoform lens with quadratically-decreasing zone widths from the inner to the outer region (Fig. 3-5a). The profiles of individual grooves remain conical. We discuss the diffraction effects generated by such a kinoform lens in Fig. 3-5, with $c_R = 0.25\ mm^{-1}$, $\kappa = -4$ and various values of Δz.

When the kinoform lens works in the blazed condition we call it a Fresnel lens (Fig. 3-5c, $\Delta\varphi$ = 2.0). In this case, the transmitted light is dominated by refraction. Ray optics is sufficient to describe its optical effects in the paraxial regime. In phase space, the focusing effect of the Fresnel lens is expressed by a shearing of signals along the angular axis. The propagation of light is performed by an additional shearing along the spatial axis. At the focus position, the signals in phase space form a vertical line. As there are no diffraction ripples in phase space, the intensity of the propagated beam is symmetric about the focal plane ($z = 4\ mm$).

When the kinoform lens does not fulfill the blazed condition, *i.e.* $\Delta\varphi$ being a non-integer, the phase mismatch generated by the groove height can contribute to significant diffractive phenomena. Multiple diffracted orders appear when the zone width on the kinoform lens is in the range of several wavelengths.

In Fig. 3-5c the phase space is composed of several stripes crossing each other at the origin. Each stripe represents one diffracted order. The central stripe is the 0 order indicating the refracted beam path. Thus the 0 order is always at the same location as the phase space signals

in Fig. 3-5b. Any other higher orders (±1, ±2, *etc.*) are the outcomes of interference effects. The fine structures in phase space result from a superposition of all the interference effects generated by individual grooves.

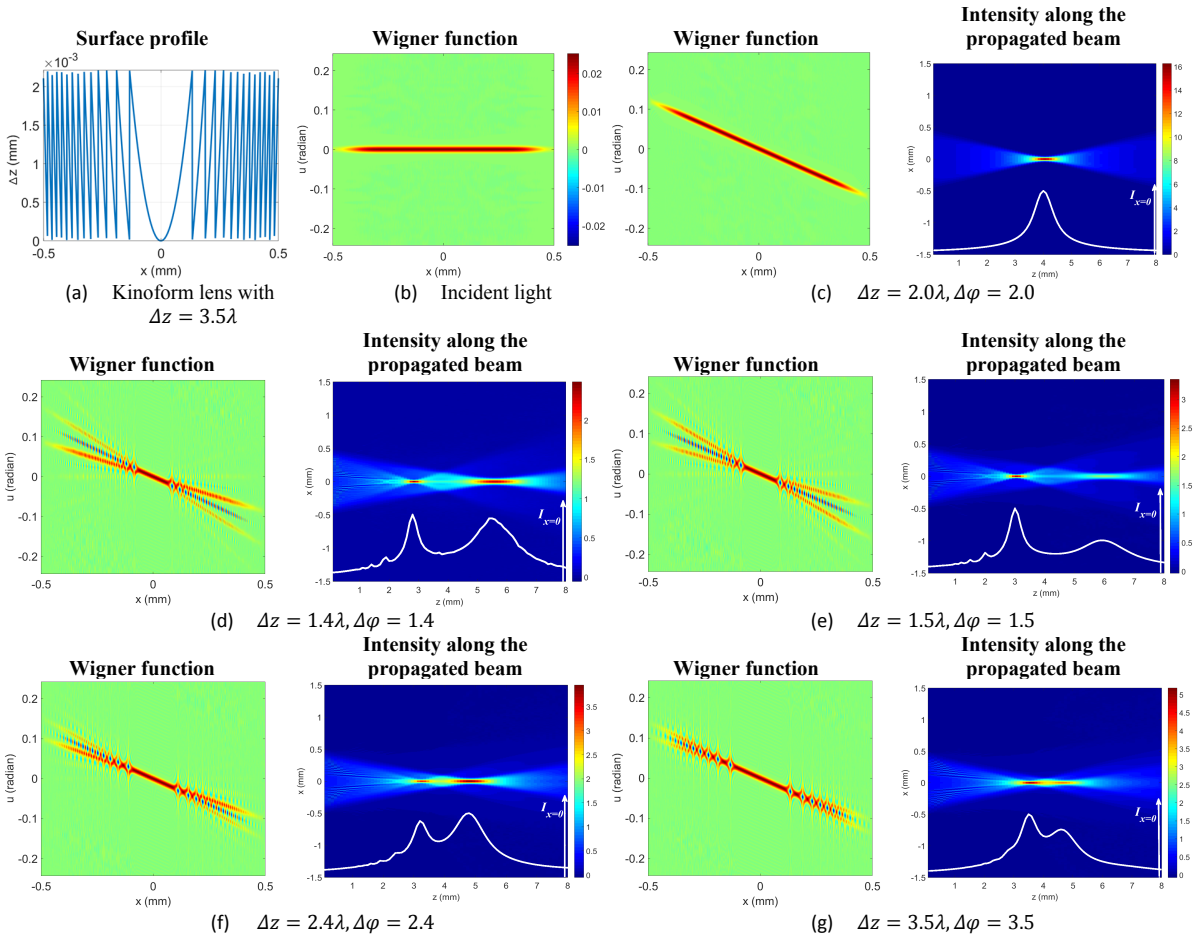


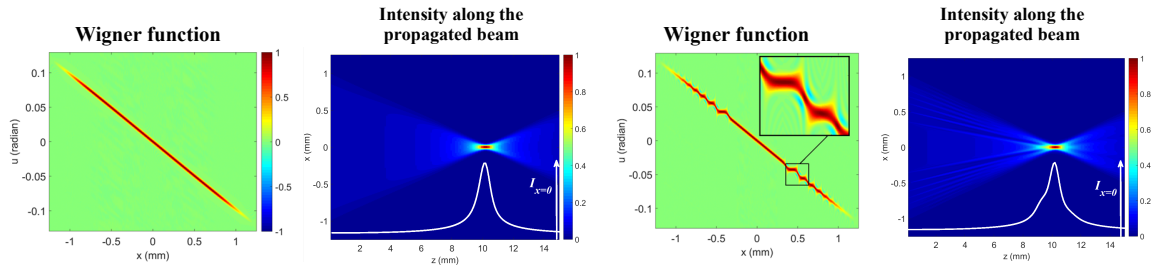
Fig. 3-5. Kinoform lenses with various groove heights. The lens profile in (a) does not have perfectly uniform groove heights due to finite samplings. The zone widths vary from 10.4 to 41.8 μm in (c), from 7.8 to 34.6 μm in (d), from 7.8 to 36.6 μm in (e), from 11.7 to 47 μm in (f), and from 17.6 to 55.4 μm in (g). Figures of phase space share the color bar in (b). The parameters for the incident light source are σ= 5 μm, δ= 0.7 μm and M= 40.

With $\Delta\varphi$ between 1.4 and 2.4, the diffracted orders are separated from each other by a large distance. In free-space propagation, each diffracted order produces its axial focus. The energy ratio of the foci depends on the energy distribution among the diffracted orders. With $\Delta\varphi = 1.5$, both +1 and -1 orders contain equal energy in phase space. Correspondingly the foci formed by the +1 and -1 orders share the same amount of energy. However, due to different numerical apertures, their peak heights are not equal. Meanwhile, there is no intensity peak at $z = 4\ mm$

for $\Delta\varphi = 1.5$ in Fig. 3-5d. That is because the integration of the phase space signals in the 0 order returns zero.

The transition from diffraction to pure refraction happens when the zone width grows considerably larger than the wavelength (λ =0.6328 μm). When the height of the kinoform lens is increased to $\Delta z = 3.5\lambda$, *i.e.* $\Delta\varphi = 3.5$, the zone width is above 17 μm. The diffracted orders are no longer distinguishable in the Wigner function (Fig. 3-5f). All the diffracted orders start to merge. Individual foci in the propagated beam path are not separable anymore. If the zone width is increased, even more, diffraction effects gradually fade. Eventually, this leads to one single focus as in ray optics.

Due to manufacturing limitations, a Fresnel lens is commonly produced with linear saw-tooth grooves. In Fig. 3-6 we compare this type of Fresnel lens ($\Delta\varphi = 10.0$) and a lens with conical grooves. Because $\Delta\varphi$ is an integer, the Fresnel lens with conical grooves gives the same performance as a continuous conical lens. The axial intensity distribution along propagation is symmetric about the focus position (Fig. 3-6a). Correspondingly, a Fresnel lens with linear saw-tooth grooves is equivalent to a piecewise linear approximation of a conical surface. This approximation creates discontinuous slopes on the surface, although the groove height yields no phase mismatch due to $\Delta\varphi$ being an integer. The corresponding diffraction effects are seen in phase space as additional ripples. Furthermore, the linear profile of each groove bends the light into slightly non-focusing angles. Under the influence of ray effects and wave effects, an asymmetric axial focus is found in the propagated beam path (Fig. 3-6b).

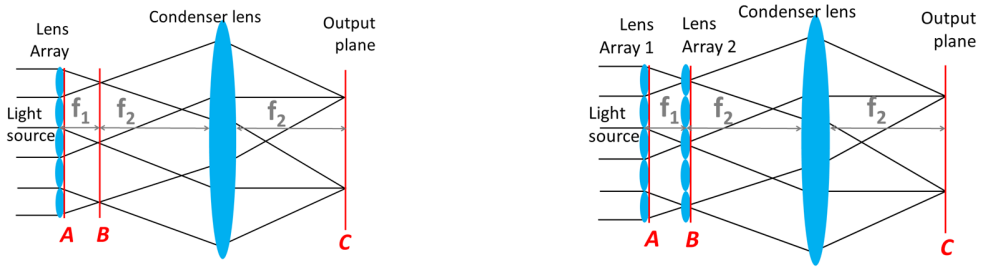


(a) Conical grooves with $\Delta\varphi = 10.0$ (b) Linear saw-tooth grooves with $\Delta\varphi = 10.0$

Fig. 3-6. Fresnel lens with conical grooves (a) and linear saw-tooth grooves (b), both with a focal length of 10 mm. Figures with the same titles share the same color bars. The parameters for the incident light source are σ= 2 μm, δ= 0.3 μm and M= 40.

Lens array

The lens array is a common component in optical systems for beam homogenizing. Fig. 3-7 shows two setups for this application.



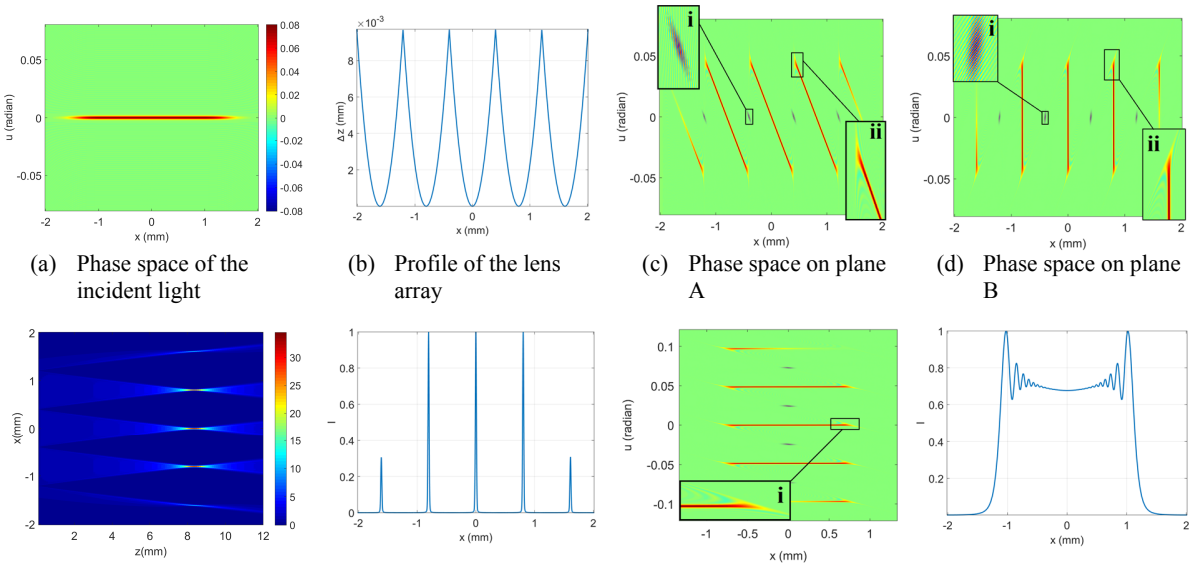
(a) One-array system (b) Two-array system

Fig. 3-7. Setups for beam homogenizing. The focal length of the lens array is f_1. The focal length of the condenser lens is f_2. Plane A marks the plane right behind the array. Plane B lies on the back focal plane of the array. Plane C is the output plane for a homogenized beam.

Fig. 3-8 shows the results based on the setup in Fig. 3-7 a. The light leaving the array is distributed into five channels defined by the individual lenslets. This is indicated by five parallel tilted lines in phase space on the plane A (Fig. 3-8c). Between the adjacent lines, there are additional diffraction ripples caused by the discontinuous slope of the array. Similar to the case of an axicon, ripples occur near the zero angle regions and at each end of the five lines (Fig. 3-8c (i) and (ii)). As the plane B is at the back focal plane of the lens array, five tilted lines in phase space on the plane A are sheared into a vertical orientation (Fig. 3-8d). By integrating the signals in Fig. 3-8 d vertically we derive five narrow peaks in the transverse intensity on the plane B (Fig. 3-8f). They indicate five focused beamlets at the back focal plane of the lens array. The weaker intensity peaks in Fig. 3-8f originate from the non-uniform illumination. The ripples in Fig. 3-8d (i) give no contribution to the transverse intensity after the integration.

Since the plane B and the output plane C form a pair of Fourier conjugates, the phase space on the plane C is a 90^o rotation of the phase space on the plane B. The ripples in Fig. 3-8 g(i) produce oscillations in the transverse intensity at the edges of the beam (Fig. 3-8 h).

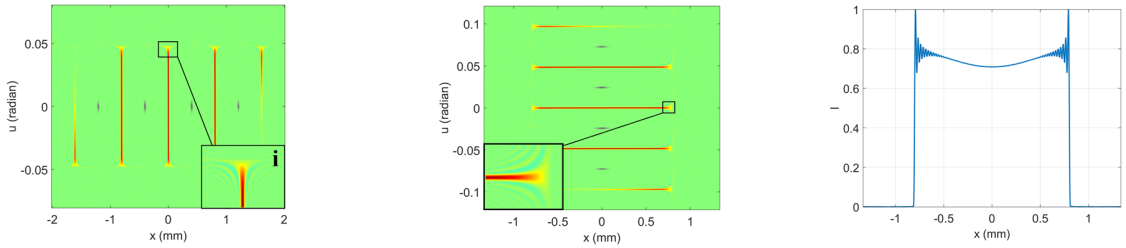
The only difference between the two setups in Fig. 3-7 is the second lens array. It is placed at the back focal plane of the first array, acting as a field lens. This field lens introduces an extra focusing effect at the pupil. In phase space, this is expressed as a vertical shearing of signals. Thus the ripples in Fig. 3-9 a(i) are sheared into a symmetric distribution, compared with Fig. 3-8 d. Physically it means the field lens captures light with large angles and bends them into the acceptable numerical aperture of the condenser lens. Thus the light falls into a better-defined spatial region on the output plane C. Fig. 3-9 c shows a more homogenized intensity profile than Fig. 3-8 h.



(a) Phase space of the incident light (b) Profile of the lens array (c) Phase space on plane A (d) Phase space on plane B

(e) Intensity of the propagated beam (f) Transverse intensity on plane B (g) Phase space on plane C (h) Transverse intensity on plane C

Fig. 3-8. Phase space results of the one-array system for beam homogenizing given by Fig. 3-7 a. Figures of phase space share the color bar in (a). The parameters for the incident light source are σ= 1 μm, δ= 0.18 μm and M= 40.



(a) Phase space on the plane B (b) Phase space on plane C (c) Transverse intensity on plane C

Fig. 3-9. Phase space results of the two-array system for beam homogenizing given by Fig. 3-7 b. Phase space in (a) and (b) share the color bar in Fig. 3-8 a.

To close this section, we summarize that the Wigner function contains information about ray optics and wave optics. It facilitates the interpretation of intermediate optical effects, in particular for complex systems with partially coherent light and diffractive elements. We have discussed two types of discontinuous surfaces, those with a discontinuity in space and those with a discontinuity in slope. The phase step, grating and kinoform lens belong to the first type. For surfaces with a spatial discontinuity, the optical path difference is important to evaluate the destructive interference. For example, by varying this parameter we see the change of a kinoform lens from pure refraction to diffraction. This parameter also controls the energy

distribution among the multiple foci formed by the diffracted orders of a kinoform lens. The axicon, Fresnel lens with linear saw-tooth grooves and the lens array represent the second type of surface discontinuity, *i.e.* discontinuity in slope. This kind of discontinuity always causes pronounced interference effects.

3.1.2 Improved propagation algorithms for Wigner functions

Although the Wigner function is a helpful tool to analyze optical signals in phase space, the computation of the Wigner function has certain difficulties. We require a two-dimensional Wigner function to describe light fields with one transverse dimension. For light fields with two transverse dimensions, the Wigner function spans four dimensions. Consequently, the question of how to efficiently implement the Wigner function becomes an important issue in practice. The main goal of this section is to propagate light by using Wigner functions while saving computer memory and keeping computational accuracy. We concentrate on two aspects of the propagation algorithms.

The first aspect is the wavefront sampling. In optical systems, we often encounter strongly convergent or divergent beams. To represent such a beam in phase space requires a dense sampling grid. We introduce a method based on numerical collimation, to represent a convergent or divergent beam in phase space with as few sampling points as possible. Meanwhile, all the diffraction effects are preserved.

The second aspect is the free-space propagation. In paraxial regime, the propagation of the Wigner functions can be associated with the ABCD matrix formalism. A free-space propagation corresponds to a shearing of signals in the Wigner function. However, a straightforward implementation of shearing can severely burden the computer memory. We describe another algorithm based on Radon transform with an improved computational efficiency.

We assume that all the light is within a small numerical aperture (NA). The propagation operators are paraxial. We discuss the models with one transverse dimension. All the algorithms can be extended to fields with two transverse dimensions.

Removing a parabolic wavefront

We start with the sampling of the Wigner function. By definition of the Wigner function $W(x,u)$, the angle u and spatial distance Δx are Fourier conjugated to each other. The sampling on the spatial and angular axes follows the relation,

$$u_{max} = \frac{1}{\mathrm{d}(\Delta x)} \cdot \frac{\lambda}{2} = \frac{1}{2\mathrm{d}(x)} \cdot \frac{\lambda}{2} = \frac{(N_x - 2)}{2 \cdot 2x_{max}} \cdot \frac{\lambda}{2}, \tag{3.1}$$

where u_{max} denotes the maximum value on the angular axis, λ is the wavelength of the light field, the symbols $\mathrm{d}(\Delta x)$ and $\mathrm{d}(x)$ define the spatial distance between two adjacent sampling points on the Δx and x axes respectively, and N_x represents the total sampling points on the x axis. Here we use the term $(N_x - 2)$ because we define one less sampling point on the non-negative x values than on the negative x values. Based on Equation (3.1), the sampling on the angular axis u is automatically defined once the sampling on the spatial axis x is chosen.

If a beam is strongly convergent or divergent, a large angular range on the angular axis is required to fully describe the beam of a large angle cone. Thus, according to Equation (3.1), we are forced to have a dense sampling on the spatial axis x to fulfill the range requirement on the angular axis. Eventually, this leads to aliasing (Fig. 3-10). However, if the optical component does not need so many sampling points to describe its spatial structure, a dense sampling on the spatial axis x is a waste of computer memory. Therefore we convert the convergent or divergent beam into a quasi-collimated beam by removing a parabolic wavefront. For a quasi-collimated beam, the angular axis does not need a big range any more. Then the sampling density on the spatial axis x can be reduced.

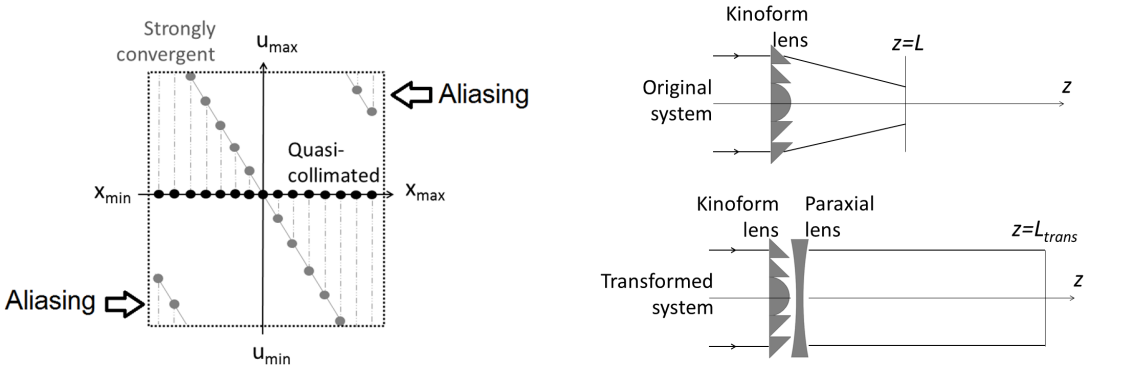
According to Siegman[52], a beam with a converging wavefront with a radius of curvature R propagating for a distance of L generates the same diffraction effects as this beam without the converging wavefront propagating for a transformed distance. The equation for the transformed distance is

$$L_{trans} = RL/(R + L). \tag{3.2}$$

After the propagation, we need to scale the collimated beam width by a factor of $R/(R + L)$, so that it is comparable to the original beam width.

Fig. 3-10 depicts a simple system as an example. In the original system, the kinoform lens with a focal length of 16 mm (*i.e.* $R = -16\,mm$) focuses a quasi-collimated input beam into a convergent beam. The kinoform lens has a diameter of 2 mm and a uniform groove height $\Delta z = 3.5\lambda$, where λ denotes the wavelength. We define the refractive index of the lens as $n_{rf} = 2.0$ with $\lambda = 0.6328\ \mu m$. The phase difference generated by the groove height is $2\pi(n - 1)\Delta z/\lambda = 7\pi$, leading to destructive interference. Thus the outgoing convergent beam carries additional diffraction effects. The beam propagates in free space with a distance L= 10 mm after

leaving the kinoform lens. In the transformed system, we place a paraxial negative lens right behind the kinoform lens to collimate the convergent beam. According to Equation (3.2), the collimated beam propagates in free space with a transformed distance $L_{trans} = 26.7\ mm$. After this propagation, we desire the same diffraction effects as in the original system.



(a) Under-sampling problem in phase space

(b) Collimating a convergent beam produced by a kinoform lens

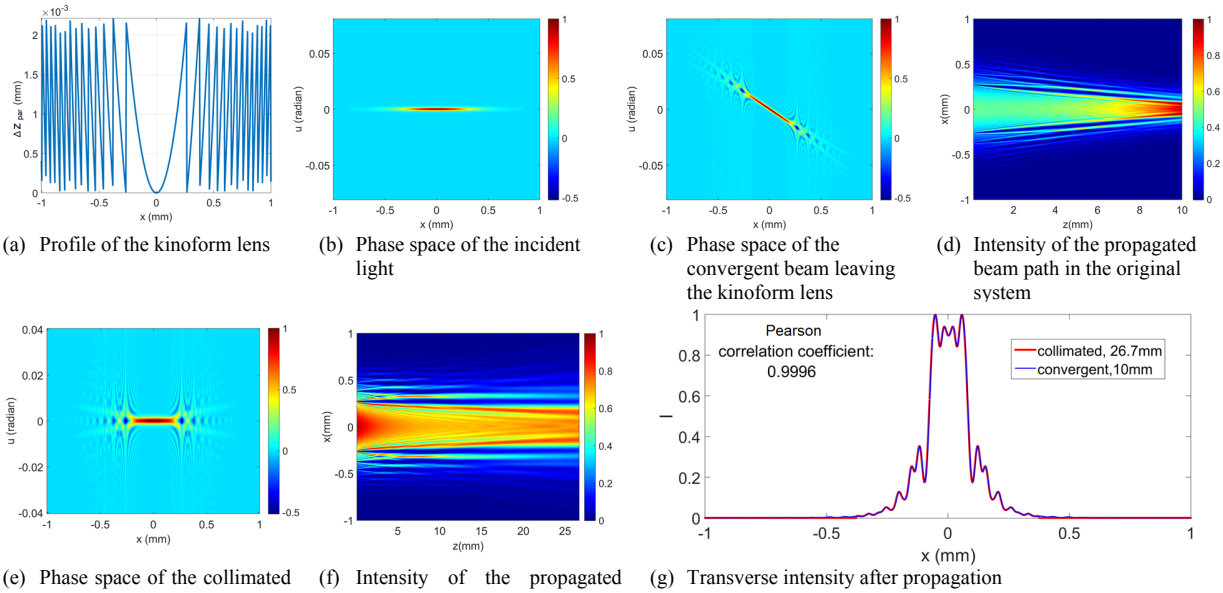
Fig. 3-10. Removing a parabolic wavefront from a convergent beam to overcome the under-sampling problem in phase space

To apply this transformation in the Wigner function, we insert a paraxial negative lens directly into Equation (2.27). This lens only removes a convergent curvature from the wavefront, without changing any other phase effects from the kinoform lens. It is expressed in an equation as follows,

$$W(x,u) = \int \Gamma_{source}\left(x+\frac{\Delta x}{2}, x-\frac{\Delta x}{2}\right)\cdot P_{kino}\left(x+\frac{\Delta x}{2}\right)\cdot P^{*}_{kino}\left(x-\frac{\Delta x}{2}\right)\cdot P_{par}\left(x+\frac{\Delta x}{2}\right)\cdot P^{*}_{par}\left(x-\frac{\Delta x}{2}\right)\exp\left(i\frac{2\pi}{\lambda}u\Delta x\right)d\Delta x, \tag{3.3}$$

$$P_j = \exp\left[-\frac{i2\pi}{\lambda}(n-1)\Delta z_j\right], j = kino, par, \tag{3.4}$$

where W denotes the Wigner function of light leaving the paraxial negative lens, Γ_{source} being the correlation function of the incident light onto the kinoform lens, P_j representing the phase modulation function of a surface (*i.e.* either the kinoform lens or the paraxial negative lens), Δz_j referring to the height of the corresponding surface. For the paraxial negative lens we define $\Delta z_{par} = \exp[-i\pi x^2/(\lambda f_{par})]$, where f_{par} is its focal length. The profile of the kinoform lens is shown in Fig. 3-11 a.



(a) Profile of the kinoform lens

(b) Phase space of the incident light

(c) Phase space of the convergent beam leaving the kinoform lens

(d) Intensity of the propagated beam path in the original system

(e) Phase space of the collimated beam leaving the paraxial negative lens

(f) Intensity of the propagated beam path in the transformed system

(g) Transverse intensity after propagation

Fig. 3-11. Phase space and propagated beam paths in the original and transformed systems. The lens profile in (a) does not have perfectly uniform groove heights due to finite sampling points.

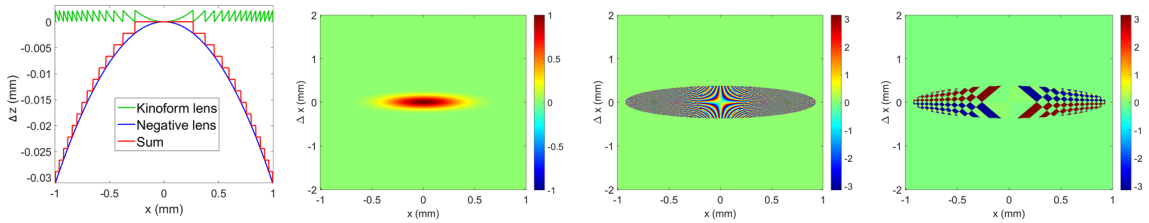
Fig. 3-11 compares the phase space of the original and the transformed systems given by Fig. 3-10b. In the original system, the focusing effect produced by the kinoform lens is expressed by a general tilt of all the signals in Fig. 3-11c. The additional diffraction effects are indicated by the oscillatory ripples in Fig. 3-11c. In the transformed system, the paraxial negative lens introduces an extra defocusing effect to the convergent beam. It brings the signals back to the horizontal orientation in Fig. 3-11e. Thus the convergent beam is collimated. We propagate this collimated beam with the distance $L_{trans} = 26.7\ mm$ and scale the transverse beam width after the propagation. Fig. 3-11g shows the nearly equal transverse intensities, with a Pearson correlation coefficient[53] of 0.9996, between the original and the transformed systems. It indicates that the diffraction effects caused by the kinoform lens are preserved in the transformed system.

Also, the angular axis in Fig. 3-11e has half the range of the angular axis in Fig. 3-11c. An angular range of $-0.04 \leq u \leq 0.04\ (rad)$ is sufficient for the collimated beam, whereas a range of $-0.08 \leq u \leq 0.08\ (rad)$ is required for the convergent beam. Thus the grid in the Wigner function is decreased from 1024 by 1024 pixels to 512 by 512 pixels by using the transformed system. This method saves the computer memory by a factor of 4.

Fig. 3-12 depicts more intermediate results to explain the phase effects introduced by the paraxial negative lens. Based on Equation (3.3) and (3.4)3.4, the product between P_{kino} and P_{par} includes a sum of the surface heights of the kinoform lens and the paraxial negative lens,

$$P_{kino}(x)P_{par}(x) = \exp\left[-i\frac{2\pi}{\lambda}(n-1)\Delta z_{kino}(x)\right]\exp\left[-i\frac{2\pi}{\lambda}(n-1)\Delta z_{par}(x)\right] = \exp\left\{-i\frac{2\pi}{\lambda}(n-1)\left[\Delta z_{kino}(x)+\Delta z_{par}(x)\right]\right\}. \quad (3.5)$$

The remaining height in Equation (3.5) (red curve in Fig. 3-12a) steps with a uniform height given by the kinoform lens. The paraxial negative lens eliminates the fast oscillatory phase (Fig. 3-12c) generated by the focusing effect of the kinoform lens. In the end, only the phase jumps produced by individual groove height of the kinoform lens are preserved in the correlation function, shown as quadrilaterals of constant phase in Fig. 3-12d. The edge of each quadrilateral indicates the groove position in the kinoform lens at the corresponding location given by the x and Δx axes. If the step height equals to a multiple of the wavelength, the quadrilaterals will vanish. The absolute values of the correlation function do not change between the surfaces because the surfaces only modify the phase of the incident beam.



(a) Surface height.

(b) Real values of the correlation function of the incident light, imaginary parts being zero.

(c) Phase of the correlation function of light leaving the Kinoform lens.

(d) Phase of the correlation function of light leaving the paraxial negative lens.

Fig. 3-12. Correlation functions of light inside the transformed system. The incident light is a Gaussian-Schell beam[17] with a beam width of 400 μm and a coherence length of 100 μm at the wavelength of 0.6328 μm.

Furthermore, our method offers an alternative to magnify the convergent beam near the focus region. As the collimated beam in the transformed system has a larger transverse width, we obtain a beam profile of more pixels filled with non-zero values. When we are interested in the exact focus of the convergent beam, we need to propagate the collimated beam to infinity (*i.e.* far field). A far-field propagation in phase space is performed by a 90° rotation, *i.e.* a Fourier transform.

It is worth noting that the Wigner function can describe non-paraxial light[24]. However, the numerical collimation method presented in this section has a limitation. Since we use the thin element approximation to process surfaces and apply ABCD matrices to propagate light in free space, our propagation operators are paraxial. Therefore, the algorithms are only useful for systems within a small NA. If the condition of a small NA is fulfilled, the convergent beam is allowed to contain a wavefront deviating from a parabola, *e.g.* a non-perfect focusing beam.

Shearing and Radon transform

The paraxial propagation of the Wigner function in free space is associated with an ABCD matrix $\begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix}$, where z defines the propagated distance. The signals in phase space before and after the propagation share a shearing relation[54], *i.e.*

$$W_o(x_o, u_o) = W(x + uz, u) \rightarrow \begin{cases} x_o = x + uz \\ u_o = u \end{cases}, \tag{3.6}$$

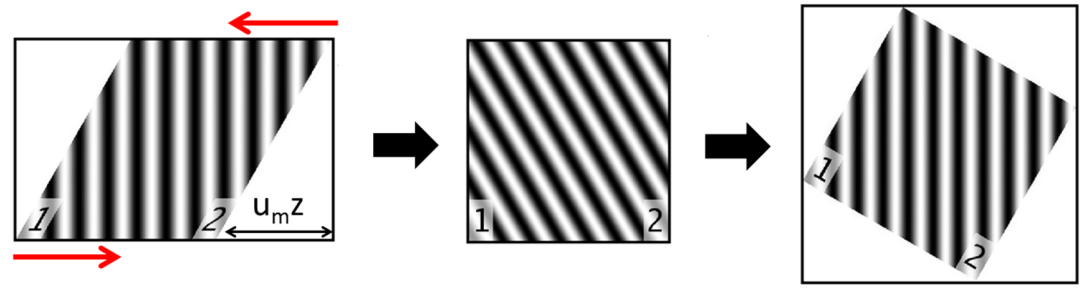
where W and W_o denote the Wigner functions before and after the free-space propagation. To perform shearing numerically, each row in $W(x, u)$ is shifted horizontally in phase space by a distance of $x' = uz$, where x' marks the shifted distance and u is the angular coordinate of a horizontal row. A shift in space corresponds to multiplying a phase factor in the Fourier domain. Thus the spatial shift of each row can be represented by two Fourier transforms,

$$W(x + x', u) = \boldsymbol{FT}_{k_x}^{-1}\{\exp[-i2\pi x' k_x] \cdot \boldsymbol{FT}_x\{W(x, u)\}\}, \tag{3.7}$$

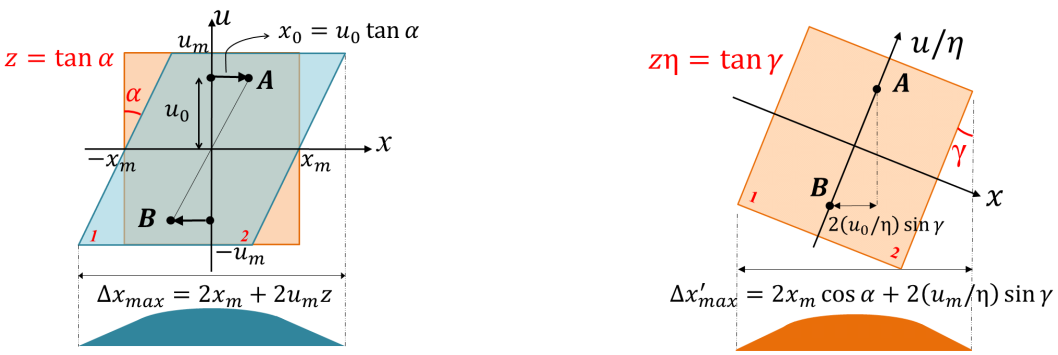
Where $\boldsymbol{FT}_x$ indicates a Fourier transform with respect to x, k_x is the spatial frequency along the x direction, and $\boldsymbol{FT}_{k_x}$ indicates a Fourier transform with respect to k_x.

The advantage of employing a Fourier transform in shearing is that it returns an accurate value at each shifted pixel. However, shearing with a Fourier transform also has a disadvantage. A large propagated distance z in free space leads to a large shifted distance of x_0 in phase space. Some signals in phase space are sheared outside the original region. The discrete Fourier transform mirrors these signals back into the original region and causes aliasing. To avoid this error, one must make sure that the computational region in phase space is wide enough to support shearing. Otherwise, zero-padding should be used to increase the computational region. As we are interested in light intensity along propagation, another method called Radon transform[13] reaches the same goal as shearing. A Radon transform projects an image along a radial line oriented at a specific angle to derive the intensity distribution. Fig. 3-13 a-c give a qualitative illustration. Fig. 3-13a is the phase space at the propagated distance z generated by

shearing the phase space at the propagated distance zero. The vertical lines in Fig. 3-13a indicate the integration direction of signals for the transverse intensity at the propagated distance z. Fig. 3-13b results from an inverse shearing of Fig. 3-13a in the directions of the red arrows. This inverse shearing corresponds to a back propagation from a distance z to zero. Thus Fig. 3-13b returns the phase space at the propagated distance of zero. The integration lines in Fig. 3-13a are sheared into a tilted orientation in Fig. 3-13b. They still imply the integration direction for the transverse intensity at the propagated distance z. Alternatively, we may perform a rotation of the phase space of Fig. 3-13b into Fig. 3-13c, and integrate the signals along the integration lines vertically in Fig. 3-13c to derive the transverse intensity at the propagated distance z. This is what we call as a Radon transform.



(a) Phase space at the distance of z (b) Reverse shearing of (a), phase space at the propagated distance zero (c) Rotation of (b)



(d) Details of the shearing geometry (e) Details of the rotation geometry

Fig. 3-13. Geometries of shearing and rotation.

Fig. 3-13d and e depict the schematic diagrams of the geometries for shearing and the Radon transform. Assume there are two points $A = (0, u_0)$ and $B = (0, -u_0)$ in phase space at the propagated distance zero. For the propagated distance z, the shearing angle is defined as $\alpha = \arctan(z)$. Points A and B in phase space are sheared to coordinates of $(u_0 z, u_0)$ and $(-u_0 z, -u_0)$ respectively. The distance between these two points in the transverse intensity at the propagated distance z is denoted by $\Delta x_{AB} = 2u_0 z$. The total width of the transverse

intensity is $\Delta x_{max} = 2x_m + u_m z$, where x_m and u_m define the maximum values on the spatial and angular axes in phase space at the propagated distance of zero.

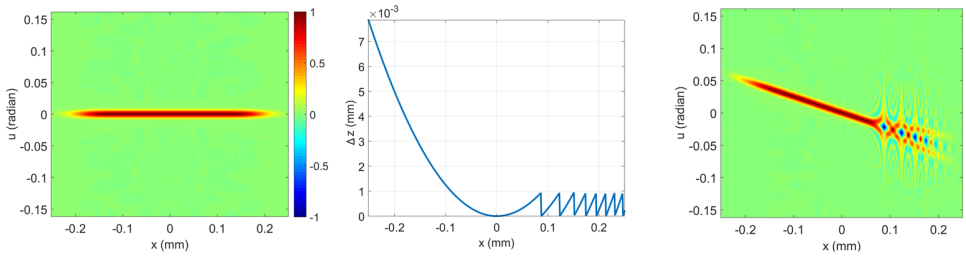
In the Radon transform (Fig. 3-13e), at first, we divide the angular axis by a factor of $\eta = u_m/x_m$, so that the two axes in phase space have the same unit and remain uniform during rotation. The projected intensity after a rotation of angle γ yields a maximum width of $\Delta x'_{max} = 2x_m \cos\gamma + 2(u_m/\eta)\sin\gamma$. Inside the projected intensity, the distance between A and B is $\Delta x_{AB}{}' = 2(u_0/\eta)\sin\gamma$. To implement the Radon transform, we need to find out the relation between the rotation angle γ and the propagated distance z. This is achieved by keeping the ratio between Δx_{max} and Δx_{AB}, and the ratio between $\Delta x_{max}{}'$ and $\Delta x_{AB}{}'$ equal,

$$\frac{2x_m+2u_m z}{2u_o z} = \frac{2x_m\cos\gamma+2(u_m/\eta)\sin\gamma}{2(u_o/\eta)\sin\gamma}. \quad (3.8)$$

From Equation (3.8) we derive the rotation angle $\gamma = \arctan(\eta z)$. The ratio factor for the transverse intensity width between shearing and the Radon transform is $\eta z/\sin\gamma$.

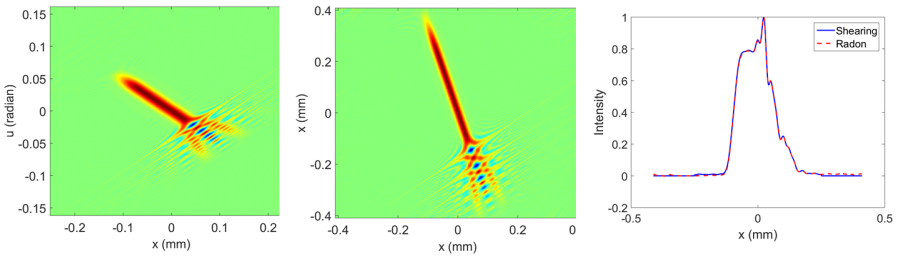
Fig. 3-14 and Fig. 3-15 show specific examples of using the two algorithms to propagate the Wigner function. We use a quasi-collimated rectangular Schell beam[20] as the incident light source. The beam travels through an asymmetric lens with half aspheric and half kinoform profiles (Fig. 3-15b) to generate an asymmetric phase space (Fig. 3-15c). In that case, the phase space orientation after a rotation is more distinguishable. Fig. 3-15 a-c compare the two algorithms for a short propagated distance ($z = 2mm$). In this case, both algorithms produce nearly identical results. However, with a larger propagated distance (Fig. 3-15 d-f, with $z = 8\ mm$), the shearing stretches the signals in phase space outside and causes aliasing (highlighted by the black parallelogram in Fig. 3-15 d). This kind of error can be avoided by using the rotation algorithm. A phase space with a width of $2\sqrt{2}x_m$ is large enough to support a rotation of any angle. For a propagated distance from zero to infinity, the rotation angle γ corresponds to 0° up to 90°. According to Larkin[55] and Lohmann[56], an arbitrary rotation matrix can be expressed as a product of three shearing matrices. When the fast Fourier transform is used to implement these shearing operations, this way produces accurate results for the rotated image.

Fig. 3-16 sketches the intensity along the propagated beam based on the results in Fig. 3-15. With the aliasing error in Fig. 3-15d, the transverse beam width in Fig. 3-16a is always confined to a limited transverse extent when $z > 6\ mm$. This unphysical result is avoiding in Fig. 3-16b by using the Radon transform, as predicted by Fig. 3-15.

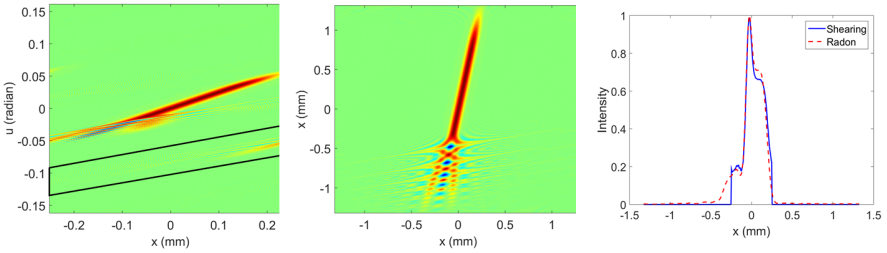


(a) Phase space of incident light (b) Profile of the lens (c) Phase space of outgoing lights

Fig. 3-14. Phase space results of a quasi-collimated Schell beam passing an asymmetric lens. (a) and (c) share the same colorbar.



(a) Shearing $\alpha = 63.43^{o}$ for $z = 2\ mm$ (b) Rotation $\gamma = 52.24^{o}$ for $z = 2\ mm$ (c) Transverse intensity at $z = 2\ mm$



(d) Shearing $\alpha = 82.87^{o}$ for $z = 8\ mm$ (e) Rotation $\gamma = 79.03^{o}$ for $z = 8\ mm$ (f) Transverse intensity at $z = 8\ mm$

Fig. 3-15. Comparison between shearing and the Radon transform in propagating. (a), (b), (d) and (e) have the same colorbar given by Fig. 3-14.

Both in shearing and the Radon transform, there are regions filled with zeros (shown as white areas in Fig. 3-13a and c). These areas do not influence the results of the calculation but require computer memory. The larger the zero regions are, the less efficient the calculation becomes. Fig. 3-17 compares the area of zero regions between the two algorithms based on the phase space given by Fig. 3-15. For a short propagated distance where zero-padding is unnecessary for shearing, shearing yields less zero area than the Radon transform. For example, Fig. 3-17 shows that for a propagated distance of $z < 3\ mm$, shearing has the advantage of requiring less memory than the Radon transform. For a larger propagated distance, the area of zeros in the Radon transform does not increase after exceeding a rotation angle γ of 45°. However, the area of zeros in shearing grows linearly with the propagated distance. It thus demands a linear rise of the computer memory, making this method impractical.

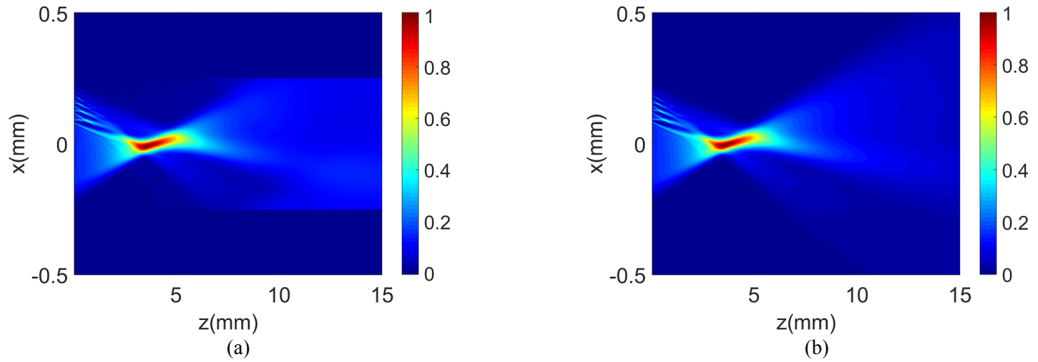


Fig. 3-16. Intensity of the propagated beam obtained by shearing (a) and the Radon transform (b).

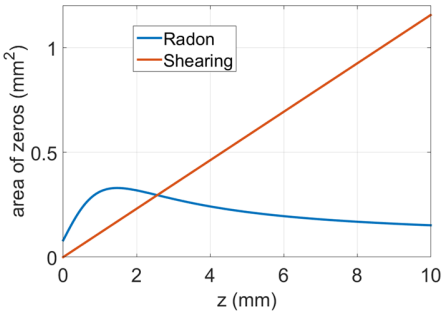


Fig. 3-17. Necessary zero-padding to propagate the phase space given by Fig. 5c without producing aliasing.

We summarize Section 3.1.2 as follows. We develop fast implementation methods to paraxially propagate beams in phase space, particularly for partially coherent light. We show how to apply a method known for working with coherent light to problems of partial coherence. A parabolic wavefront is subtracted from a convergent beam. The beam is converted into quasi-collimated form. Therefore, we reduce the required sampling density to represent this beam in phase space. The diffraction effects in propagation are preserved after the parabolic wavefront is removed. Furthermore, this approach offers an alternative to magnify the convergent beam near the focus region by observing the collimated beam at the physically equivalent distance. Besides, we compare the efficiency of two algorithms, shearing and Radon transform, for propagating the Wigner function in free space. For a large propagated distance, the shearing method requires zero-padding to avoid aliasing. Its demand on the computer memory grows linearly with the propagated distance. In contrast to this, the Radon transform keeps the computer memory within a finite limit.

The advantages of our proposed methods are even greater with the operation of a four-dimensional phase space, even though this is not specifically shown in this section.

3.1.3 Limitation of the Wigner function on propagating light inside a waveguide

So far, we employ the Wigner function with the thin element approximation to investigate optical elements. The paraxial implementation is efficient, given by the algorithms we propose. However, this method has certain limitations.

For example, we want to simulate the beam homogenizing effect of a waveguide. In this case, the propagation of partially coherent light cannot be easily represented by the Wigner function alone. A waveguide has a pipe structure and reflects light repeatedly inside its volume. We may understand this by a ray model illustrated in Fig. 3-18. We limit ourselves to having only one transverse dimension. The waveguide borders are considered as perfect mirrors. The green rays inside the waveguide keep being reflected and overlapped during the propagation. The rays reaching the output facet of the waveguide, denoted as facet 0, form a homogenized beam profile. In geometrical optics, we may consider the light path in another way. We mirror the original waveguide up and down to make several copies (denoted as ±3, ±2, ±1 in Fig. 3-18). In the next step, all the borders of the waveguides are considered completely transmissive. Rays emitted from the point source travel in straight lines (depicted as blue rays in Fig. 3-18) and reach the output planes marked by various integers. The absolute values of these numbers represent a number of reflection happening to the rays if they travel inside the original waveguide. In other words, the angles of the starting rays decide the amount of reflection they encounter. We then overlap the blue ray profiles on all output facets from -4 to +4. The overlapping result equals to the green-ray profile on facet 0.

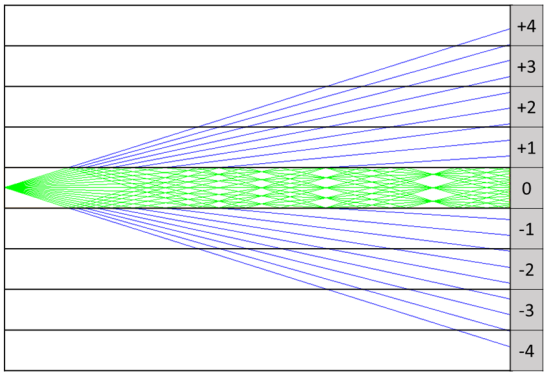


Fig. 3-18. Ray model of a light pipe.

Roelandt et al[57] propose a similar model to propagate partially coherent light inside a waveguide. They consider the beam as an incoherent sum of multiple Gaussian beams (*i.e.* shifted elementary modes). Each Gaussian beam propagates in free space towards the output facets without considering the waveguide borders, like the blue rays in Fig. 3-18. The propagation is computed by a Rayleigh-Sommerfeld diffraction integral. For each Gaussian beam, they sum up the amplitudes from individual output facets. Between the modes, light is overlapped incoherently. In the end, they obtain the homogenized beam profile, as if it travels inside the reflective waveguide.

This method combines the ideas of ray optics and partial coherence. It is an efficient model to investigate the beam homogenizing effect of a macroscopic light pipe. However, this model does not take the wave effects into account comprehensively. When the waveguide boundaries are assumed to be perfect mirrors, the Goos–Hänchen effect is not considered. Furthermore, this model is introduced with one transverse axis. The xz and yz planes are considered independent of each other. Consequently, this model excludes skew rays and the attenuation at each reflection.

We attempted to use the Wigner function to propagate light inside a waveguide. However, the thin element approximation cannot represent the unique structure of a waveguide. That complicates the propagation operator based on the Wigner function. Moreover, we want to accurately model the diffraction effects caused by the reflective borders of a waveguide. By only knowing the individual Wigner functions of two partially coherent beams, it is impossible to regenerate their interference effects[50]. That is because, additional information, *e.g.* a global phase, is missing in the phase space representation. Taking into account all these limitations, we decide to move from the phase space approach to modal expansion, to continue our investigation of partial coherence.

3.2 Partially coherent light in modal expansion

Compared to the discrete Wigner function describing light in each position and angle as delta functions, model expansion interprets light in separate modes. This is a more physical representation, since each mode propagates coherently as an individual. Therefore, this method is particularly advantageous for a beam with extra symmetry. In that case, we only need a few modes to fully represent the partially coherent beam, instead of computing numerous delta

functions in phase space. That greatly reduces the computational efforts, especially for light with two transverse dimensions.

In this section, we investigate beams with certain symmetry and use Schell beams as examples. We first compare two expansion methods, eigenmode and shifted elementary mode. Later on, we propose operators to propagate partially coherent beam in waveguides.

3.2.1 Modal properties of Schell beams

Before we discuss specific systems, we first present a comprehensive discussion of the basic properties of modal expansion. We take the Gaussian-Schell and flat-top beams as examples, to discuss their solutions and essential characters in eigenmode and shifted-elementary mode representations. The optical effects are visualized in the Wigner function.

Gaussian-Schell beam

i) Analytical computation of the eigenmodes

At first, we obtain the Wigner function of the Gaussian-Schell beam by Fourier transforming the correlation function,

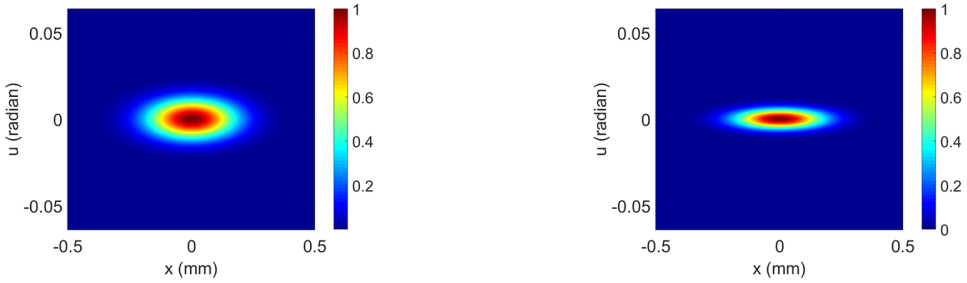
$$W(x,u) = \exp\{-x^2/{w_o}^2\}\exp\{-\pi^2u^2{l_c}^2/\lambda^2\} \tag{3.9}$$

where w_o denotes the beam waist, and l_c represents the correlation width. Equation (3.9) suggests that the Wigner function exhibits a Gaussian profile along the spatial and angular axes. If we integrate all the signals along the x or u axis, we obtain a Gaussian distribution. Note that, as a Gaussian function decays to zero at infinity, we need a sufficiently wide sampling region along the *x*-axis to avoid aliasing in numerical implementation.

Moreover, Equation (3.9) denotes that the angular extent Δu in the Wigner function is inversely proportional to the correlation length l_c. This coincides with the following physics understanding. When a beam has a larger correlation length, the fields emitted from two fixed source points are more correlated to each other. Then this beam has a higher degree of coherence. A more coherent beam yields a smaller angle of divergence. In the Wigner function, this divergence angle is illustrated as the signal extent along the u axis.

We show an example in Figure 3-20 of using the analytical eigenmodes to compose a Gaussian-Schell beam. Because the eigenmodes contain Hermite polynomials, we also denote the eigenmodes as Hermite modes, f_n. The parameters are as follows, wavelength $\lambda =$

0.5 μm, beam waist $w_0 = 0.25\ mm$, correlation width $l_c = 0.02\ mm$, mode index n from 0 to 49. Figure 3-20a is the transverse intensity profile of the full beam.



(a) Correlation width $l_c = 10$ μm. (b) Correlation width $l_c = 20$ μm.

Figure 3-19. Wigner functions of Gaussian-Schell beams with the wavelength $\lambda = 0.5$ μm and a beam waist $w_0 = 0.25\ mm$.

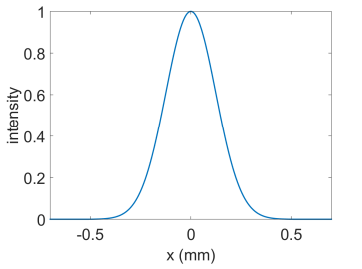
We illustrate the amplitudes and intensities of the Hermite modes (n from 0 to 4) in Figure 3-20b and c. We then define the power of a mode as an integration of the intensity along the spatial axis x. A higher-order mode gives a lower power contribution (Figure 3-20d).

With given parameters of the wavelength and beam waist, a smaller value of the correlation width yields a lower degree of coherence. Then the power ratio of the higher-order modes rises. In that case, we need a larger number of modes to represent the full beam.

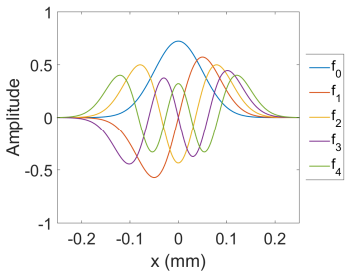
However, including more higher-order modes is not always necessary. Figure 3-20d implies that for any given beam, after reaching a certain mode index, in our example $n = 40$, the numerical accuracy does not improve significantly anymore. A further increase of the total mode number does not improve the accuracy further, but burdens the computer memory. Alternatively, this sampling criterion can be understood in Fourier optics as follows. High-order modes contain faster oscillations of the light fields, thus yielding a broader angular spectrum. In the Wigner function $W(x,u)$, those oscillations are expressed as the signal extent along the angular axis u. A denser sampling along the spatial axis Δx always results in a wider range of the angular axis u. With a finite angular spectrum of a given light source, only a finite range of the angular axis is required. Thus it is unnecessary to increase the spatial sampling on Δx infinitely.

Figure 3-21 illustrates the Wigner functions of four specific Hermite modes, given by the Gaussian-Schell beam in Figure 3-20. The amplitudes of each mode are overlaid onto these figures. Since the Hermite modes are orthogonal, a higher-order mode always contains amplitudes with a faster oscillation. In Fourier optics, a field with a more rapid oscillation in space produces a higher spatial frequency, *i.e.* a larger diffraction angle. Therefore, a higher-

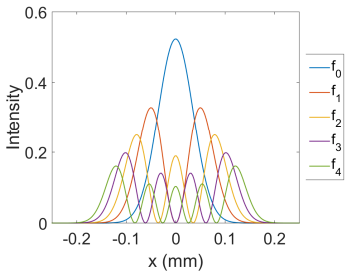
order mode in the Wigner functions has a larger angular extent. The frequency of oscillation is also indicated by the ring-shaped ripples. Moreover, the symmetry of the mode amplitude decides the sign of the Wigner function at the origin. An asymmetric field distribution, *i.e.* odd modes (*e.g.* f_5, f_{49}), produces a negative value with the largest absolute value at the origin. That implies a local destructive interference. While the even modes (*e.g.* f_6, f_{34}) show a maximum positivity at the origin. That suggests a local constructive interference. Furthermore, A sum of the Wigner functions of the 50 Hermite modes yields the Wigner function of the full beam (Figure 3-21e).



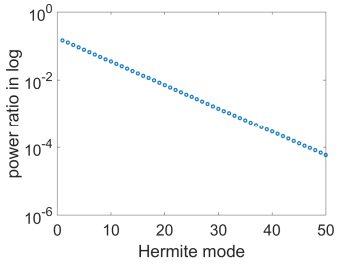
(a) The transverse intensity of the Gaussian-Schell beam, consisting of 50 Hermite modes.



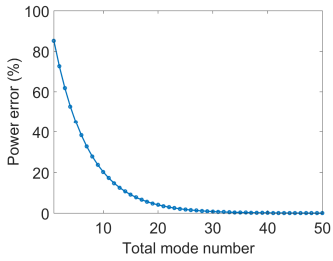
(b) Amplitudes of Hermite modes, n from 0 to 4.



(c) Intensities of Hermite modes, n from 0 to 4.

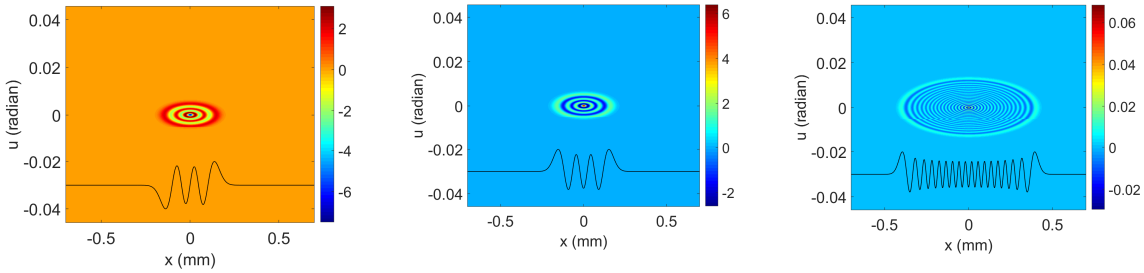


(d) The power ratio of each mode.



(e) Power error on the total number of eigenmodes.

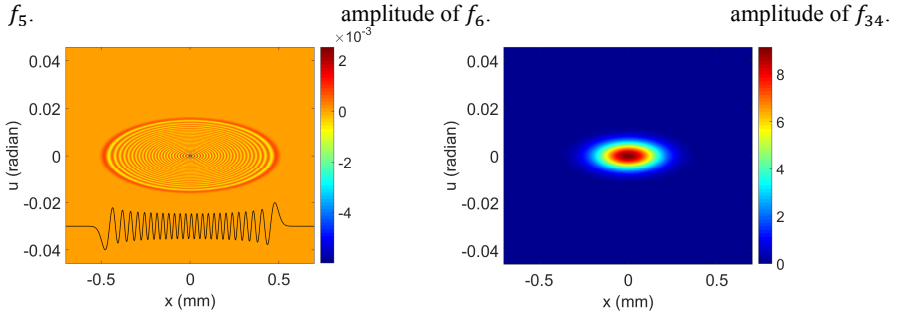
Figure 3-20. Eigenmode representation of a Gaussian-Schell beam, with parameters $\lambda = 0.50$ μ$m, l_c =$ 20 μ$m, w_0 = 0.25$ mm.



a) Wigner function of the Hermite mode n=5 overlaid with the amplitude of f_5.

(b) Wigner function of the Hermite mode n=6 overlaid with the amplitude of f_6.

(c) Wigner function of the Hermite mode n=34 overlaid with the amplitude of f_{34}.



(d) Wigner function of the Hermite mode n=49, overlaid with the amplitude of f_{49}.

(e) Wigner function of the Gaussian-Schell beam.

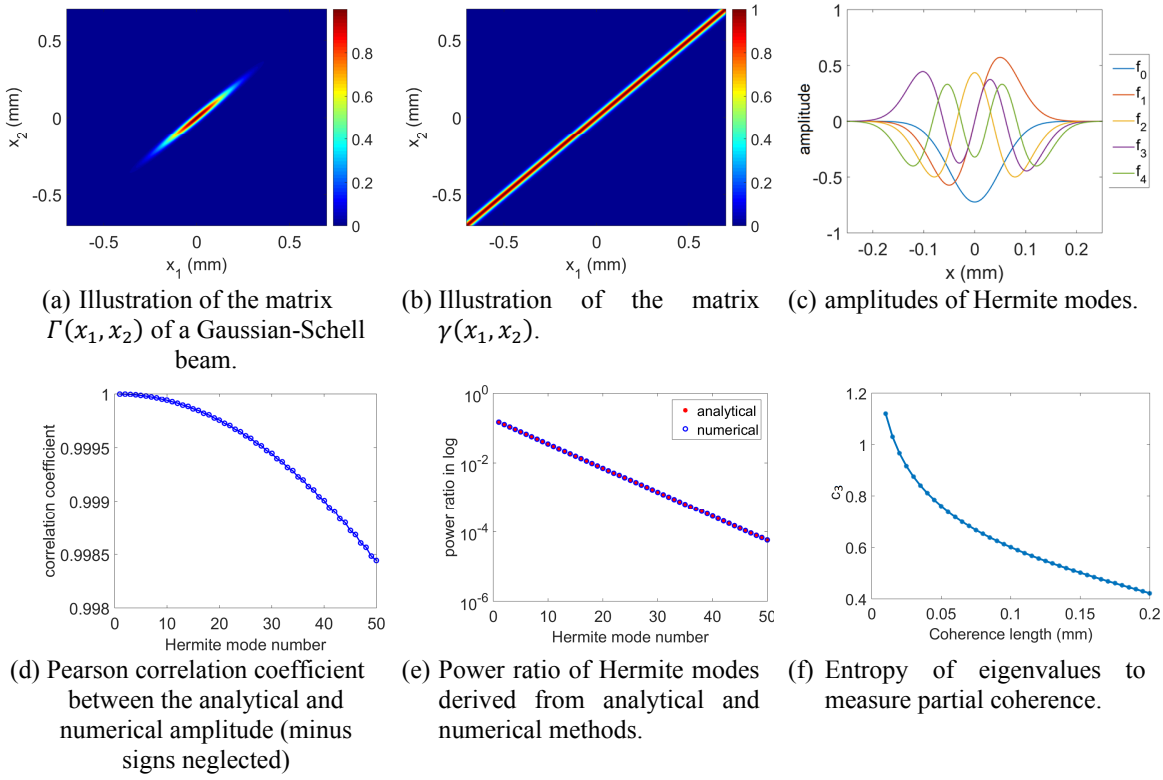
Figure 3-21. Wigner functions of individual Hermite modes. The magnitudes of the overlaid amplitudes are normalized only for visualization. This Gaussian-Schell beam is identical to the one in Figure 3-20.

ii) Numerical computation of eigenmodes

In the following, we use the linear algebraic method[32,33] to derive the eigenmodes of a Gaussian-Schell beam. Once we prove the validity of the numerical method, we can apply it to other beams, whose correlation functions have complicated forms, to obtain eigenmodes. That is particularly useful in some cases where the analytical eigenmodes are not known.

Figure 3-22a is the illustration of the discrete matrix $\Gamma(x_1, x_2)$ of a Gaussian-Schell beam at the waist plane. It contains only real values. The size of the matrix is 512 by 512. The diagonal elements of the Γ matrix form the transverse intensity of the full beam (Figure 3-20a). Moreover, we know that the degree of coherence of a Schell beam depends on the separation of two spatial coordinates, but not on their absolute positions. This is interpreted in the Γ matrix that the absolute value of each element depends on how far away it is from the diagonal line. Therefore, the elements in Γ matrix are symmetric about the diagonal, provided that all the elements are real. If there is a complex value inside the matrix, its symmetric counterpart about the diagonal is a complex conjugate. Furthermore, we illustrate the degree of coherence matrix $\boldsymbol{\gamma}$ in Figure

3-22b, given by $\gamma(x_1, x_2) = \exp\{-(x_1 - x_2)^2/(2{l_c}^2)\}$. Matrix $\boldsymbol{\gamma}$ is always a circulant matrix if the source belongs to a Schell type.



(a) Illustration of the matrix $\Gamma(x_1, x_2)$ of a Gaussian-Schell beam.

(b) Illustration of the matrix $\gamma(x_1, x_2)$.

(c) amplitudes of Hermite modes.

(d) Pearson correlation coefficient between the analytical and numerical amplitude (minus signs neglected)

(e) Power ratio of Hermite modes derived from analytical and numerical methods.

(f) Entropy of eigenvalues to measure partial coherence.

Figure 3-22. Hermite modes from the linear algebraic method. The parameters of the Gaussian-Schell beam are the same as Figure 3-20. The index count of the matrix elements starts from the bottom left.

In the next step, we derive the numerical eigenvalues and eigenvectors by the singular-value decomposition. Figure 3-22c displays the numerical results of individual Hermite-mode amplitudes. Compared to Figure 3-20b, there is a minus-sign difference for some modes. As the eigenvalues are always positive, the minus signs come from the eigenvectors that are derived numerically. However, these minus signs play no role physically and numerically. We are interested in the correlation of two fields, *i.e.* $\Gamma(x_1, x_2) = f(x_1)f^*(x_2)$. The magnitude of $\Gamma(x_1, x_2)$ is related to the absolute degree of coherence. The phase of $\Gamma(x_1, x_2)$ reveals the phase difference between the fields at x_1 and x_1. Although we add a minus sign to f, the product of two fields yields no change for $\Gamma(x_1, x_2)$. In the singular value decomposition of a correlation function with Hermitian properties, *i.e.* $\boldsymbol{\Gamma} = \Phi\boldsymbol{\Lambda}\Phi^{\mathrm{H}}$, we have $\boldsymbol{\Gamma}\Phi = \Phi\boldsymbol{\Lambda}$, where $\boldsymbol{\Gamma}$ denotes the correlation matrix, columns of Φ correspond to eigenvectors, and diagonals of $\boldsymbol{\Lambda}$ are eigenvalues. It also holds for $\boldsymbol{\Gamma}(-\Phi) = -\Phi\boldsymbol{\Lambda}$. That means if we have as an eigenvector defined

as ϕ, $-\phi$ is automatically an eigenvector too. Therefore, having a minus sign in the mode amplitudes between Figure 3-22b and Figure 3-20b does not influence our discussion about the numerical accuracy. We thus neglect these minus signs in the following comparison.

Figure 3-22d depicts the Pearson correlation between the analytical and numerical solutions of the Hermite modes. A value of 1 means the numerical values are completely identical to the analytical values. The coefficients for all the modes are very close to 1, meaning that the numerical amplitudes are very similar to the analytical amplitudes. Figure 3-22e shows the power contribution of each mode to the full beam. The numerical result coincides with the analytical result with less than 0.2% deviation.

Our example proves that the numerical approach is an efficient and accurate alternative to eigenmode expansion. This method is particularly useful for beams with a complicated form of the correlation function, especially when the analytical solutions of the eigenvalues and eigenvectors are not known.

As introduced in Chapter 2.3.1, the distribution of the eigenvalues (λ_n) reveals the degree of coherence. In the next step, we measure the entropy of eigenvalues given by Equation (2.41) with a changing correlation width l_c (Figure 3-22f). A higher degree of coherence gives a smaller value of entropy. If the beam is completely coherent, we have only one eigenvector (*i.e.* only one eigenmode). Therefore, only one of the eigenvalues is nonzero. Then we have the entropy $c_3 \to 0$, because $z \log z \to 0$ as $z \to 0$ [32]. This is demonstrated in Figure 3-22f. A larger coherence length imples a higher degree of coherence, thus the value of entropy c_3 decays.

So far we have presented the analytical and numerical solutions to the eigenmodes of a Gaussian-Schell beam. In the following, we evaluate the shifted-elementary mode approach for a Gaussian-Schell beam.

iii) Shifted-elementary modes

The elementary mode of a Schell beam is related to the degree of coherence, given by Equation (2.52). For a Gaussian-Schell beam, we have the degree of coherence

$$\gamma(\Delta x) = \exp\left[-\Delta x^2/\left(2{l_c}^2\right)\right] \tag{3.10}$$

Where $\Delta x = x_1 - x_2$ and u are Fourier conjugates. A Fourier transform of Equation (3.10) is given by

$$\tilde{\gamma}(u) = \boldsymbol{FT}_{\Delta \boldsymbol{x}}[\gamma(\Delta x)] = \exp\left(-u^2 {l_c}^2/2\right) \tag{3.11}$$

where a constant magnitude is neglected in Fourier transform, u and Δx are Fourier conjugates. Therefore, the elementary function, denoted as $f(x)$, takes the form

$$f(x) = \boldsymbol{FT_u}^{-1}\left[\exp\left(-u^2 {l_c}^2/4\right)\right] = \exp(-x^2/{l_c}^2) = \gamma(x)^2. \quad (3.12)$$

Equation (3.12) provides us with an important fact that, for a Gaussian-Schell beam, each elementary mode is a coherent Gaussian beam with a waist of l_c. We know that, for any coherent Gaussian beam, the phase space product follows

$$w_0 \cdot \Delta u = \lambda/\pi, \quad (3.13)$$

where w_0 is beam radius (a normalized intensity decaying to $1/e^2$), Δu is the diffraction angle of the Gaussian beam. Equations (3.13) implies that the elementary mode, $i.e.$ a coherent Gaussian beam, has a fixed area in phase space representation, *i.e.* a conserved etendue. When a Gaussian beam has a larger diffraction angle, given an unchanged wavelength, its beam waist must reduce. Now we let the beam waist equal to the correlation width, *i.e.* $w_0 = l_c$, we have

$$\Delta u = \lambda/\pi l_c \quad (3.14)$$

Equation (3.14) implies the connection between the angular extent of an elementary mode and the correlation width of the Gaussian-Schell beam. In Equation (3.9), the angular extent of a Gaussian-Schell beam, suggested by the exponential term $\exp\{-\pi^2 u^2 {l_c}^2/\lambda^2\}$ in Equation (3.9), also follows $\Delta u = \lambda/\pi l_c$. That means the Gaussian-Schell beam and every elementary mode must share a uniform diffraction angle. We may interpret this in the following two ways.

First, in the Wigner function, a Gaussian-Schell beam has a defined signal extension along the spatial and angular axes (Figure 3-23a). The spatial extension is given by w_0, while the angular extension is defined by Δu. Now we use the shifted elementary modes to represent the same Wigner function. Multiple Gaussian functions are placed one after another with a shifted distance along the x axis. Only when each mode has the same Δu as the Gaussian-Schell beam (Figure 3-23b), the signals of the Gaussian-Schell beam in the Wigner function can be restored by superposing the signals of individual modes. Imagine, if the Δu of an elementary mode does not reach the Δu of the full beam, no matter how dense we place these modes, they never fully cover the signal area of the Gaussian-Schell beam. As a result, given a partially coherent in the Wigner function, we need to first define the angular extent of the elementary modes. Then the beam waist of the elementary mode is automatically known by Equation (3.13). A larger diffraction angle indicates a low degree of coherence. It also yields a smaller Gaussian beam (*i.e.* elementary-mode) waist. We thus need a larger number of modes to represent the full beam.

In the Wigner function, the phase-space product of the full beam indicates the degree of coherence. The ratio of the phase-space product between the full beam and one elementary mode suggests the required number of modes.

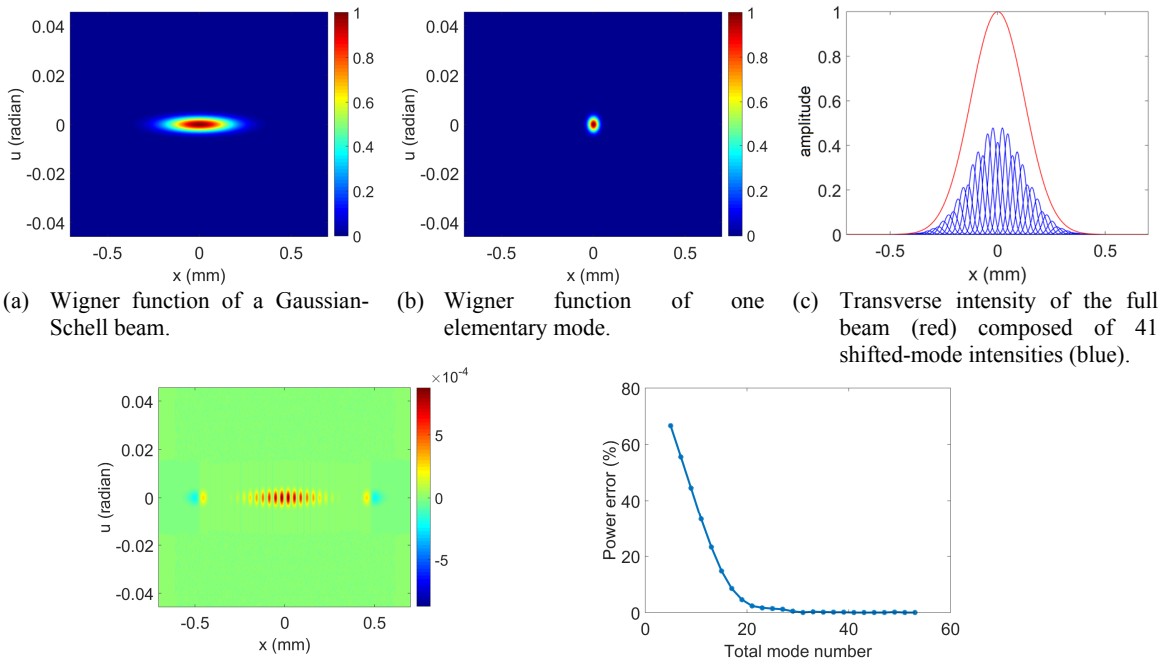
A second way to understand the requirement of Δu for the elementary modes is to use an intuitive physics picture. We consider a plane source in real space emitting a Gaussian-Schell beam. Each point on the source plane emits light within a certain angular cone. The transverse correlation width l_c defines the maximum distance of two correlated points. That means, as long as the separation of two points is smaller than l_c, their light fields interfere coherently. Now we replace this source by the multiple Gaussian beams (*i.e.* elementary modes). We assume no interference between every pair of two modes. In order to preserve the same coherence as the Gaussian-Schell beam, we must define the width of each Gaussian mode to be l_c. Only in that case, two source points that are within the support of one Gaussian mode, *i.e.* within the width of l_c, interfere coherently. When two points are beyond the distance of l_c, they cannot be covered by one Gaussian mode. Then these two points are not correlated to each other.

Figure 3-23c depicts the transverse intensity of a Gaussian-Schell beam ($\lambda = 0.5\ \mu m, l_c = 40\ \mu m, w_0 = 0.25\ mm$) consisting of 41 shifted modes. We optimize the magnitude of each mode so that their incoherent sum yields a Gaussian profile. One may also apply Equation (2.53) without optimizing the mode magnitudes. In that case the summed intensities exhibit a minor deviation from a Gaussian profile. The reason is that Equation (2.52) and (2.53) assume that the full-beam intensity changes slower than the elementary function. However, that is not always true, especially where the intensity function experiences a sharp slope change near the beam edge. We present the optimization method to determine the magnitudes of individual elementary modes in the next section.

We compare the original Wigner function of the Gaussian-Schell beam (Figure 3-23a), and the Wigner function composed by 41 shifted modes (central mode given by Figure 3-23b). Both Wigner functions are normalized to have the maximum magnitude of 1. The difference between these two Wigner functions (Figure 3-23d) is below 1%. This error results from two aspects. First, the elementary modes (*i.e.* Gaussian beams) are truncated, especially for those near the boders of the Schell beam. This problem can be overcome by increasing the sampling range of the spacial axis at the expense of the computer memory. Second, the elementary modes are placed with a finite distance to each other. This is, in fact, a numerical approximation assumed

in the shifted-elementary mode expansion, turning an integration into a sum (explained in Equations (2.49) and (2.54)). This error can be minimized by packing the shifted-elementary modes denser.

Similarly to the eigenmode method, the required mode number to compose a given Gaussian-Schell beam exhibits a limit (Figure 3-23e). After reaching a sufficient amount of modes (in our example, 35 modes), a further increase of mode number is not necessary. This coincides with a common sense that, the degree of coherence decides the total number of modes that we need to represent the partially coherent beam thoroughly.



(a) Wigner function of a Gaussian-Schell beam.

(b) Wigner function of one elementary mode.

(c) Transverse intensity of the full beam (red) composed of 41 shifted-mode intensities (blue).

(d) Difference between the Wigner function given by (a), and the Wigner functions summed by 41 shifted modes given by (b).

(e) Power error based on changing the number of total modes.

Figure 3-23. Shifted-elementary modes for a Gaussian-Schell beam, with parameters $\boldsymbol{\lambda = 0.5\ \mu m, l_c = 40\ \mu m, w_0 = 0.25\ mm}$.

We draw the following summary of the modal expansion for a Gaussian-Schell beam. The numerical solution of eigenmodes is a useful alternative when the analytical solution is unknown. Eigenmodes always form an orthogonal and convergent expansion. In general, an orthogonal expansion requires less modes than a non-orthogonal expansion. Although the shifted-elementary mode representation is a non-orthogonal expansion, it offers an intuitive way to represent a Gaussian-Schell beam. Compared to the eigenmodes, shifted-elementary modes

consider the source area by area, each emitting a coherent Gaussian beam. This is a vivid and straightforward picture for understanding a Schell beam.

Flat-top Schell beam

A flat-top Schell beam[20] in the near field is considered as a spatial overlapping of multiple Gaussian-Schell beam. In the far field, it exhibits a flat-top intensity profile. We present the mathematical and physical properties of such a beam. Afterwards, we discuss the modal expansion approaches. We introduce this beam as an example where the analytical solution of the eigenmodes is unknown.

We first discuss the properties of the near-field correlation function given by Equation (2.20). It consists of a linear sum of multiple Gaussian functions, *i.e.* $\gamma(\Delta x) = \binom{M}{m}\frac{(-1)^{m-1}}{\sqrt{m}}\exp\left[-\frac{\Delta x^2}{2m\delta^2}\right]$, where $\Delta x = x_1 - x_2$, δ denotes the correlation width, m and M being integers. The value of M controls the steepness of the flat-top beam profile. Each Gaussian function $\exp\left[-\frac{\Delta x^2}{2m\delta^2}\right]$ has a width varied by m. The magnitudes are governed by a sign-alternating binomial weighting term $\binom{M}{m}\frac{(-1)^{m-1}}{\sqrt{m}}$. Figure 3-24 illustrates these Gaussian functions with $M = 10$.

We sum up these M terms (Figure 3-25). The larger M is, the more their sum approaches a sinc function. The reason to compose a sinc profile for the correlation function can be easily understood in phase space. We explain this with an example in Figure 3-26.

We simulate a correlation function at the source plane, denoted as $\Gamma^{(0)}(x, \Delta x)$, with parameters $\lambda = 1\mu m, \sigma = 2\times10^{-3}, \delta = 10^{-3}, M = 40$. The Wigner function at the source plane defined as $W^{(0)}(x, u)$, is a Fourier transform of $\Gamma^{(0)}(x, \Delta x)$ with respect to Δx. Fourier transforming a sinc function yields a rectangular function. Hence, the more $\Gamma^{(0)}(x, \Delta x)$ resembles a sinc profile along the Δx axis (Figure 3-26a), a more flat-top distribution $W^{(0)}(x, u)$ exhibits along the angular axis u (Figure 3-26b). In our example, in order to obtain the correlation function in the far field, defined as $\Gamma^{\infty}(x_1^{\infty},\ x_2^{\infty})$, we propagate the source through a parabolic lens. The source represented by $\Gamma^{(0)}(x_1, x_2)$ is placed at the front focal plane of the lens. The correlation function at the back focal plane is a Fourier transform of $\Gamma^{(0)}(x_1, x_2)$, denoted as $\tilde{\Gamma}^{(0)}(k_1, k_2)$, with the k_1 axis flipped (Equation (2.15) and (2.16)). We then have

$$\Gamma^{\infty}(x_1^{\infty},\ x_2^{\infty}) \sim \tilde{\Gamma}^{(0)}(-k_1, k_2). \tag{3.15}$$

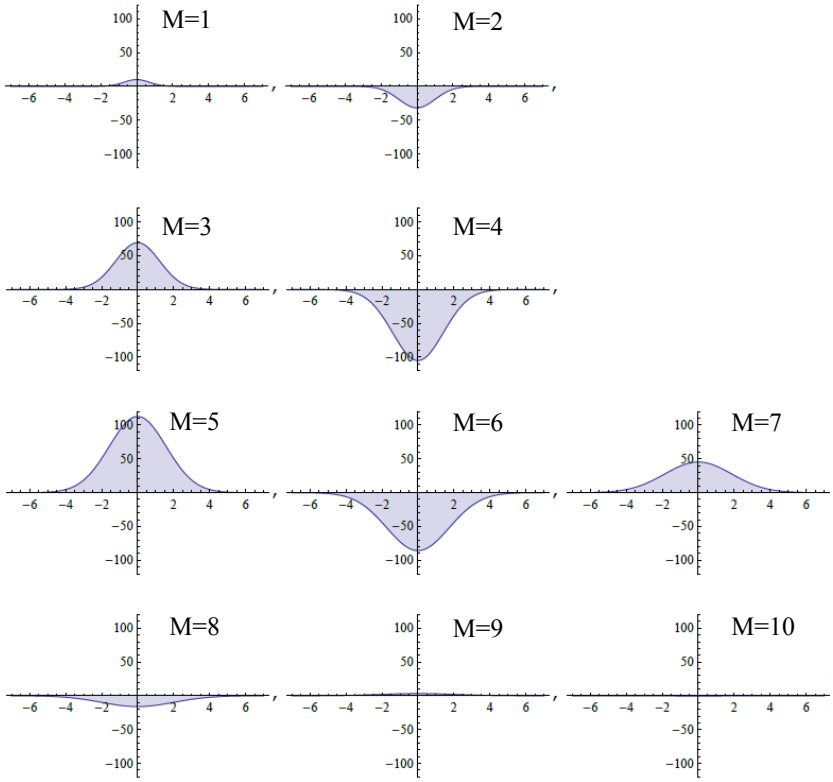


Figure 3-24. Illustration of individual $\gamma(\Delta x, m) = \binom{M}{m}\frac{(-1)^{m-1}}{\sqrt{m}}\exp\left[-\frac{\Delta x^2}{2m\delta^2}\right]$, with $M = 10$, horizontal axis being Δx, vertical axis being $\gamma(\Delta x, m)$.

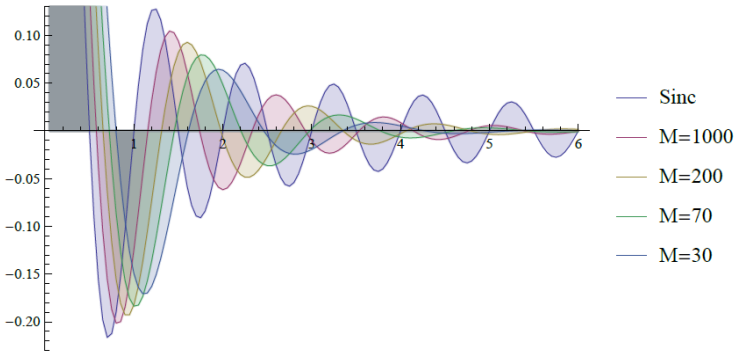


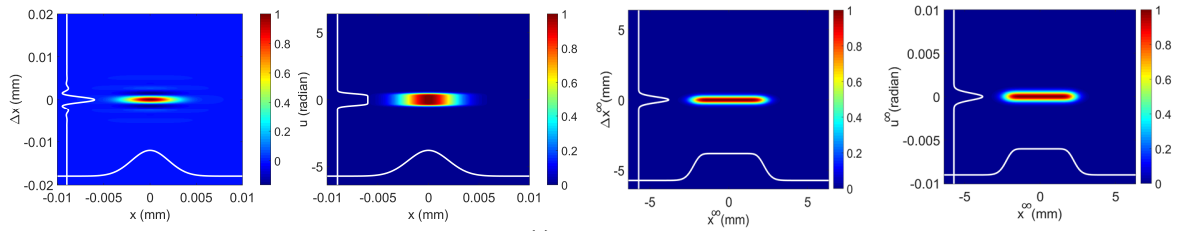
Figure 3-25. Illustration of $\gamma(\Delta x)$ given by varied integers of M, horizontal axis being Δx, vertical axis being $\gamma(\Delta x)$. The Δx axis is normalized so that the first zero of the sinc function falls at $\Delta x = 1$.

The coordinates at the back focal plane are scaled by

$$x^{\infty} = x \cdot \lambda \cdot f_{par} \cdot \left(\frac{n}{2} - 1\right) / [N_x \cdot \mathrm{d}x \cdot x_{max}], \tag{3.16}$$

where x and x^{∞} are the coordinates at the front and back focal planes, λ denotes the wavelength, f_{par} is the focal length of the paraxial lens, N_x represents the sampling points along the x axis, $\mathrm{d}x$ symbolizes the distance of two adjacent sampling points along the x axis, and x_{max} denotes the maximum coordinate at the x axis. $\Gamma^{\infty}(x^{\infty}, \Delta x^{\infty})$ and $\Gamma^{(0)}(x, \Delta x)$ are depicted in Figure 3-26a and c, where $x^{\infty} = (x_1^{\infty} + x_2^{\infty})/2\,, \Delta x^{\infty} = x_1^{\infty} - x_2^{\infty}\,, x = (x_1 + x_2)/2\,, \Delta x = x_1 - x_2$. As we have discussed in Chapter 2, the Wigner function in the far field (*i.e.* W^{∞}) is a 90° rotation of the Wigner function in the near field (*i.e.* $W^{(0)}$). This is apparent in Figure 3-26b and d, the flat-top profile along the angular axis in $W^{(0)}$ is now along the spatial axis in W^{∞}. We thus obtain a flat-top distribution of the irradiance intensity in the far field (horizontal white curve in Figure 3-26d). Furthermore, the phase space product of the Wigner function is conserved between the near field and the far field.

So far we have gained a general view of the correlation function of a flat-top Schell beam. In the next step, we investigate the modal expansion approaches for such a source.



(a) Correlation function $\Gamma^{(0)}(x, \Delta x)$ at the source plane. The overlaid white curves are $\Gamma^{(0)}(x = 0, \Delta x)$ vertically and $\Gamma(x, \Delta x = 0)$ horizontally.

(b) Wigner function $W^{(0)}$ at the source plane. The overlaid white curves are $\int W^{(0)}\,\mathrm{d}x$ vertically and $\int W^{(0)}\,\mathrm{d}u$ horizontally.

(c) Correlation function $\Gamma^{\infty}(x^{\infty}, \Delta x^{\infty})$ in the far field. The overlaid white curves are $\Gamma^{\infty}(x^{\infty} = 0, \Delta x^{\infty})$ vertically and $\Gamma^{\infty}(x^{\infty}, \Delta x^{\infty} = 0)$ horizontally.

(d) Wigner function W^{∞} in the far field. The overlaid white curves are $\int W^{\infty}\,\mathrm{d}x^{\infty}$ vertically and $\int W^{\infty}\,\mathrm{d}u^{\infty}$ horizontally.

Figure 3-26. Flat-top Schell beam at the source plane and in the far field. We obtain the light in the far field by focusing the source with a parabolic lens (with a focal length of 5 mm). The parameters are as follows, $\lambda = 1\mu m$, $\sigma = 2\times10^{-3}$, $\delta = 10^{-3}$, $M = 40$.

i) Numerical eigenmodes

As we are more interested in the far-field properties of the flat-top Schell beam, from now on we focus on the correlation function in the far field (Equation (2.21)).

By definition, a flat-top Schell beam is a superposition of multiple Gaussian-Schell beams. However, the eigenmodes of a flat-top Schell beam are not directly a linear sum of the eigenmodes of individual Gaussian-Shell beams. A qualitative proof is as follows. We take out any two components, e.g. m and $m+1$ from the sum in Equation (2.20), and denote them as $\Gamma^{(m)}(x_1, x_2)$ and $\Gamma^{(m+1)}(x_1, x_2)$. Each of them forms a Gaussian-Schell beam. Thus each follows the Fredholm integral equation (Equation 2.32). We sum them up,

$$\int_{-\infty}^{\infty} \Gamma^{(m)}(x_1, x_2)\phi_n^{(m)}(x_1)\mathrm{d}x_1 + \int_{-\infty}^{\infty} \Gamma^{(m+1)}(x_1, x_2)\phi_n^{(m+1)}(x_1)\mathrm{d}x_1 = \lambda_n^{(m)}\phi_n^{(m)}(x_2) + \lambda_n^{(m+1)}\phi_n^{(m+1)}(x_2). \tag{3.17}$$

Our goal is to find out a new set of eigenvectors Ψ and eigenvalues ε so that

$$\int_{-\infty}^{\infty} \left[\Gamma^{(m)}(x_1, x_2) + \Gamma^{(m+1)}(x_1, x_2)\right]\Psi_n(x_1)\mathrm{d}x_1 = \varepsilon_n \Psi_n(x_2) \tag{3.18}$$

However, according to Equation (2.33) and (2.34), $\phi_n^{(m)}$ and $\phi_n^{(m+1)}$ are two different Hermite-Gaussian functions modulated individually by the m^{th} and $(m+1)^{\text{th}}$ parameters. They do not share common terms to lead Equation (3.17) to Equation (3.18) in an analytical form. That means, the eigenmodes of a flat-top Schell beam is not a linear sum of the Hermite-Gaussian functions. Here is the end of the proof.

Note that, here we discuss the eigenmode expansion of the correlation function given by Equation (2.32). Regarding the beam propagation, we can still treat the full beam as an incoherent sum of the multiple Gaussian-Schell beams. Each Gaussian-Schell beam is expanded into its own set of eigenmodes. However, in that case, the total number of modes is significantly larger. That kind of expansion is inefficient.

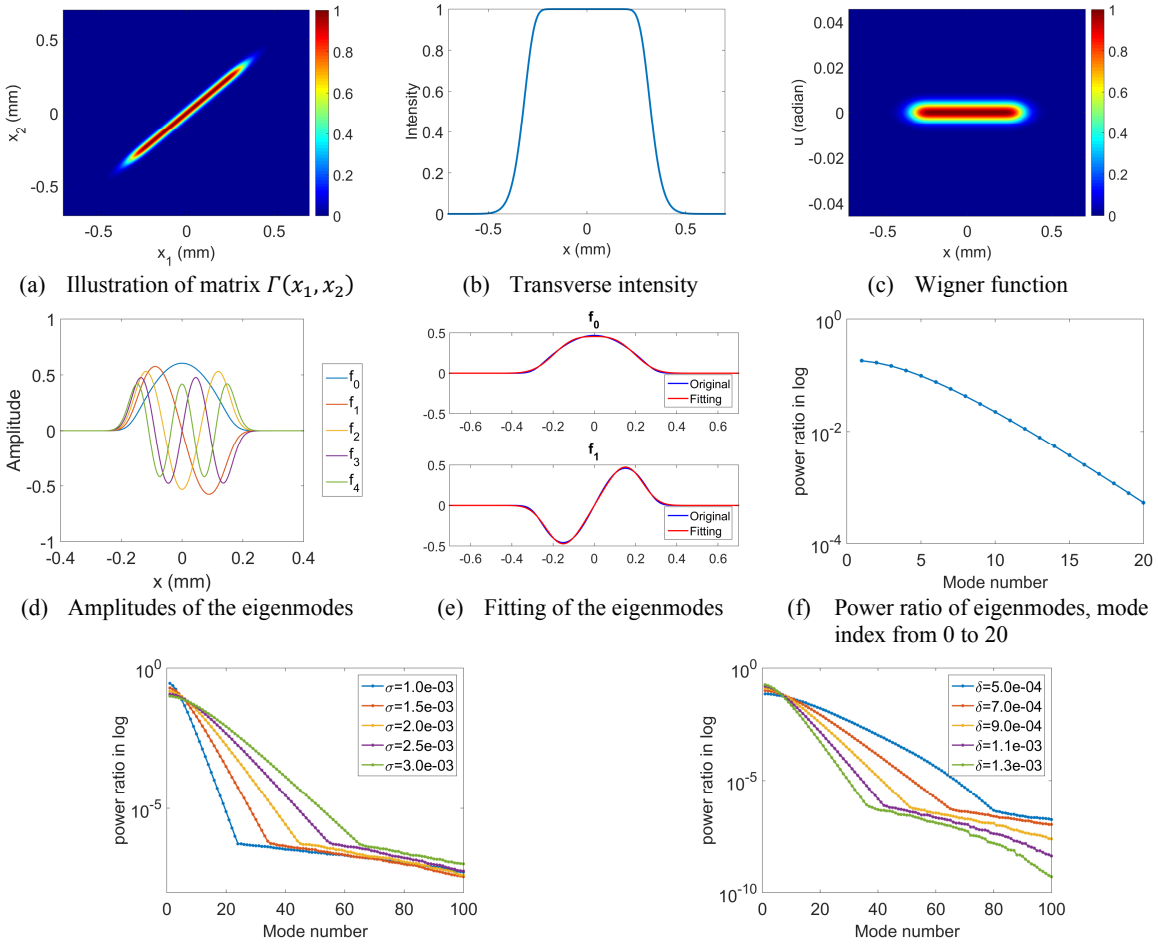
Now we raise a question: how do the eigenmodes of a flat-top Schell beam look like?

In the following, we use the linear algebraic method to derive the numerical eigenmodes of a flat-top Schell beam. Figure 3-27a is the visualization of the matrix $\Gamma(x_1, x_2)$. The diagonal of this matrix yields the transverse intensity (Figure 3-27b). As the parameter δ comes from the near-field correlation width, it determines the beam width in the far field. Another parameter σ, denoting the near-field beam width, governs the correlation width in the far field. We apply the singular-value-decomposition on $\Gamma(x_1, x_2)$ and derive the eigenmodes (Equation (2.36)). The first five eigenmodes, denoted as f_n, are illustrated in Figure 3-27d.

These eigenmodes resemble the eigenmodes of a Gaussian-Schell beam to some extent. However, they are not in the exact form of a Hermite-Gaussian function given by Equation (2.33). We use the following equation to fit the profiles of f_n given by Figure 3-27e,

$$f_n(x) = a_n H_n(b_n x)\exp[-c_n(bx)^{2d_n}] \tag{3.19}$$

where H_n denotes the n^{th} order Hermite polynomial, a, b, c and d are real values.



(a) Illustration of matrix $\Gamma(x_1, x_2)$ (b) Transverse intensity (c) Wigner function

(d) Amplitudes of the eigenmodes (e) Fitting of the eigenmodes (f) Power ratio of eigenmodes, mode index from 0 to 20

(g) Power ratio of the modes, σ changes, δ = 7e-4 mm. A larger σ indicates a smaller correlation width.

(h) Power ratio, δ changes, σ = 3e-3 mm. A larger δ indicates a smaller beam waist.

Figure 3-27. Eigenmodes of a flat-top Schell beam. (a)-(f) contain the following parameters: λ =5e-4 μm, σ = 3e-3 mm, δ =7e-4 mm.

We derive the following coefficients, $a_0 = 0.59, b_0 = -3.34, c_0 = 10.53, d_0 = 1.45$ for f_0, and $a_1 = 1.29, b_1 = -3.63, c_1 = 10.41, d_1 = 1.51$ for f_1. The accuracy of the fitting is estimated by the Pearson correlation coefficient. It reaches 0.9996 both for f_0 and f_1 (Figure 3-27f). The values of a, b, c and d vary according to the mode number. Most importantly, the value of d implies that the exponential term is not Gaussian, but with an additional power of

approximately 1.5. This term is by no means a linear sum of Gaussian functions. This proves indirectly that the eigenmodes of a flat-top Schell beam are not a linear sum of the eigenmodes of a Gaussian-Schell beam. The more flat-top the full beam is, the more its eigenmodes deviate from the Hermite-Gaussian functions. Figure 3-27f suggests that the mode power does not drop logarithmically as for the eigenmodes of a Gaussian-Schell beam. Instead, the lower-order modes play a bigger role in the power contribution.

Figure 3-27g and h visualize the impact of σ and δ on the power distribution of the eigenmodes. Note that we are interested in the far-field distribution. The parameter σ is defined as the beam width in the near-field. Due to the Fourier relation, it is inversely proportional to the correlation width in the far field. Therefore, a larger σ yields a lower degree of coherence in the far field. We thus need more power contribution from the higher-order modes to fully represent the far-field beam. Regarding the parameter δ, it denotes the correlation width in the near-field. Hence δ is inversely proportional to the far-field beam width. A larger δ indicate a smaller phase-space product in the far field. We thus need less modes to represent the full beam in the far field.

In the following, we investigate the flat-top Schell beam in shifted-elementary mode representation.

ii) Shifted-elementary modes

As the far-field correlation function given by Equation (2.21) has a complicated form. We obtain the profile of the elementary mode numerically.

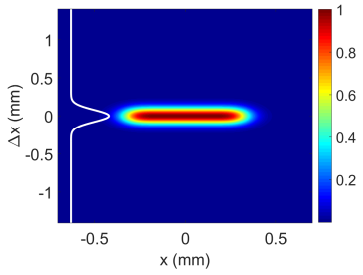
The elementary mode $f(x)$ is related the degree of coherence $\gamma(\Delta x)$ (Equation (3.12)). It can be directly extracted from the profile of $\Gamma(x, \Delta x)$,

$$f(\Delta x) = [\gamma(\Delta x)]^2 = [\Gamma(x = 0, \Delta x)]^2. \tag{3.20}$$

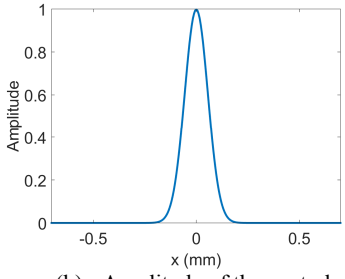
Figure 3-28 shows a simulation example of the shifted elementary modes of a flat-top Schell beam. The correlation function $\Gamma(x, \Delta x)$ is computed from Equation (2.21), visualized in Figure 3-28a. The overlaid white curve in Figure 3-28a denotes $\gamma(\Delta x)$. Since $\gamma(\Delta x)$ is a Gaussian function, the elementary mode is in a Gaussian form (Figure 3-28b). Figure 3-28c shows the spatial distribution of individual shifted modes. After adjusting the contribution of the individual modes, their incoherent sum produces the flat-top intensity profile of the full beam. Figure 3-28d is the Wigner function of the flat-top Schell beam. The angular extent of the full beam (Δu) is inversely proportional the correlation width (*i.e.* width of $\gamma(\Delta x)$). Figure 3-28e illustrates the Wigner function of the central elementary mode. Both functions in Figure 3-28b and c share the

same angular extent. That coincides with our expectation that, the elementary mode must have the same diffraction angle as the full beam. The phase-space product of one elementary mode, *i.e.* a coherent Gaussian beam, always follows λ/π. This value is given by the signal area in Figure 3-28e as 16 $\mu m{\cdot}radian$. We add up the Wigner functions of individual modes and scale it to have a maximum value of 1. Figure 3-28f shows the difference between this summed Wigner function and the original one in Figure 3-28a. The difference is in the range of 1e-4. It implies that the shifted modes represent the flat-top Schell beam well.

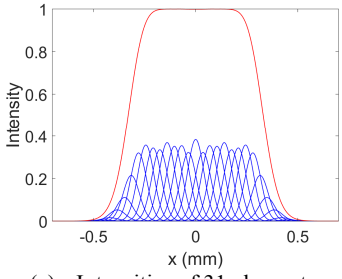
We draw the following summary of the model expansions of a flat-top Schell beam. The analytical eigenmodes of a flat-top Schell beam are not directly accessible. The numerical solutions can be derived by the singular value decomposition. They form a convergent and efficient expansion. These modes are not a linear sum of Hermite-Gaussian functions. In shifted-elementary mode expansion, the individual modes are coherent Gaussian beams. Although they are not an orthogonal expansion, shifted modes provides an intuitive information about the homogeneity and coherence of the source.



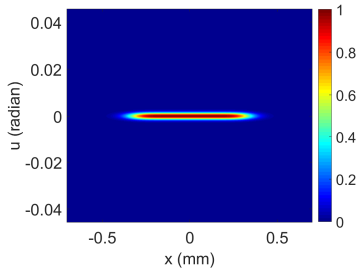
(a) Correlation function of the flat-top Schell beam, overlaid with $\gamma(\Delta x)$.



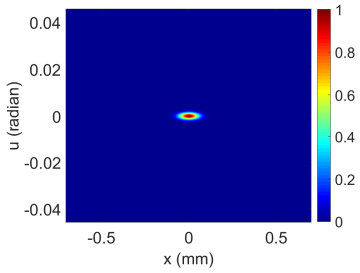
(b) Amplitude of the central elementary mode



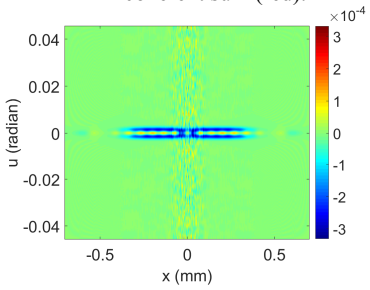
(c) Intensities of 31 elementary modes (blue) and their incoherent sum (red).



(d) Wigner function of the flat-top Schell beam



(e) Wigner function of the central elementary mode



(f) Difference between the Wigner function given by (b), and the Wigner function composed by shifted modes given by (e).

Figure 3-28. Shifted elementary modes of a flat-top Gaussian-Schell beam. Parameters are as follows, $\lambda = 0.5\ \mu m, \sigma = 1e-3\ mm, \delta = 7e-4\ mm, M = 45$.

In the following, we introduce an algorithm to determine the magnitudes of the shifted elementary modes, so that their incoherent sum approaches the intensity of the partially coherent source.

3.2.2 Using Wiener deconvolution to scale the shifted-elementary modes

We take the Gaussian-Schell beam as an example, In the shifted elementary mode expansion, the beam waist of each elementary mode is determined by the correlation width. Another parameter of the mode, *i.e.* the magnitude, should be adjusted, so that the incoherent sum of all the modes to be equal to the intensity of the full beam. In that case, a direct implementation of Equation (2.54) is not sufficient. We introduce an efficient approach to find out the accurate magnitudes of shifted elementary modes.

Equations (2.53) and (2.54) imply, a sum of shifted elementary modes comes from a numerical approximation of a convolution. That allows us to consider the problem as solving a deconvolution.

We consider the intensity profile of the Gaussian-Schell beam is a convolution between the one elementary mode intensity and a weighting function. This weighting function illustrates the local magnitudes of the given elementary mode at all spatial positions. Our task now is, by knowing an elementary mode and the intensity profile of the Gaussian-Schell beam, to regain the weighting function.

Here we briefly introduce the concept of a Wiener deconvolution. Let $I(x,y)$ denote the global distribution (*i.e.* intensity profile of the Gaussian-Schell beam), and $i(x,y)$ represent the intensity of one elementary mode. Their relation follows

$$I(x,y) = i(x,y) \otimes w_t(x,y) + n(x,y), \tag{3.21}$$

where $w_t(x,y)$ is the weighting function and $n(x,y)$ being the unknown noise. The reason we introduce noise in Equation (3.21) is because we use floating points in the numerical calculations. That introduces inaccuracy when we represent continuous numbers.

In the next step, we obtain the Fourier transforms of $I(x,y)$, $i(x,y)$ and $w_t(x,y)$, denoted as $\tilde{I}(k_x,k_y)$, $\tilde{\imath}(k_x,k_y)$ and $\widetilde{w}_t(k_x,k_y)$ respectively. Based on Equation (3.21) we may assume

$$\tilde{I}(k_x,k_y) \approx \tilde{\imath}(k_x,k_y) \cdot \widetilde{w}_t(k_x,k_y). \tag{3.22}$$

However, we cannot use an equal sign in Equation (3.22) because of the additional noise effect. Based on that, we then predict

$$\widetilde{w}_t(k_x,k_y) \approx \tilde{I}(k_x,k_y)/\tilde{\imath}(k_x,k_y). \tag{3.23}$$

Our goal now is to find out the exact equation for $\widetilde{w}_t(k_x, k_y)$. Therefore, we use the Wiener deconvolution and take into account the noise. Let

$$\widetilde{w}_t(k_x, k_y) = \tilde{I}(k_x, k_y) \cdot \tilde{g}(k_x, k_y), \tag{3.24}$$

where

$$\tilde{g}(k_x, k_y) = \frac{\tilde{\imath}^*(k_x, k_y)}{|\tilde{\imath}(k_x, k_y)|^2 + C_{ns}}. \tag{3.25}$$

The value C_{ns} represents the noise to signal ratio in the Fourier domain. By making C_{ns} a properly small constant, we take into account the noise effect. In fact, this constant C_{ns} is the exact subtlety of the Wiener deconvolution. Assume we let C_{ns} equal to zero, $\tilde{g}(k_x, k_y)$ simply ends up with being $1/\tilde{\imath}(k_x, k_y)$. Besides, we use the form of $\tilde{\imath}^*(k_x, k_y)/|\tilde{\imath}(k_x, k_y)|^2$ instead of $1/\tilde{\imath}(k_x, k_y)$ in Equation (3.25). That is because $\tilde{\imath}^*(k_x, k_y)/|\tilde{\imath}(k_x, k_y)|^2$ yields a real denominator, which is a more compatible way to numerically represent a complex value.

After we obtain $\widetilde{w}_t(k_x, k_y)$ from Equation (3.24), the weight function $w_t(x, y)$ can be obtained from an inverse Fourier transform of $\widetilde{w}_t(k_x, k_y)$.

During the implementation, there are two more steps we should pay special attention to. First, the sampling of $I(x, y)$ and $i(x, y)$ should avoid truncations or sharp edges, so that their Fourier transforms do not contain high-order signals. Our solution to this demand is to sample $I(x, y)$ and $i(x, y)$ with dense sampling points and additional zero-paddings outside the boundaries. Second, we add a low-pass filter Flt to Equation (3.24),

$$\widetilde{w}_t(k_x, k_y) = \tilde{I}(k_x, k_y) \cdot \tilde{g}(k_x, k_y) \cdot Flt(k_x, k_y), \tag{3.26}$$

where Flt is a circular-shape low-pass filter with smooth edges, to make sure the high-order artefacts resulting from the Fourier transforms are eliminated.

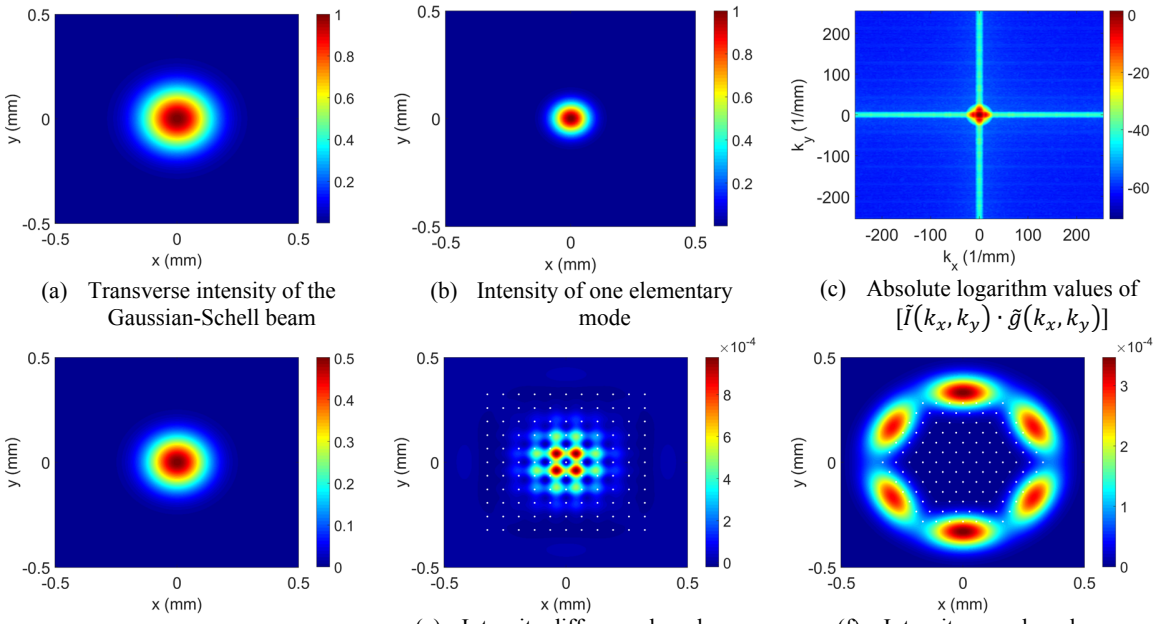
In the following, we show an example of using the Wiener deconvolution to compose a Gaussian-Schell beam with shifted elementary modes. Assume we have two transverse dimensions, *i.e.* x and y. A Gaussian-Schell beam has a beam waist of 0.2 mm, a transverse coherence length of 0.1 mm. The transverse intensity is shown in Figure 3-29a. We reconstruct the elementary modes on the same plane. Each elementary mode has a beam waist of 0.1 mm (Figure 3-29b)

We initially sample $I(x, y)$ and $i(x, y)$ from -0.5 to 0.5 mm in a 512 by 512 grid. To avoid aliasings in Fourier transforms, we add zeros to the borders of Figure 3-29a and b, to extend the

x and y axes from -1 to 1 mm. Thus the grid size becomes 1536 by 1536 pixels. Figure 3-29c depicts the absolute logarithmic value of $\left[\tilde{I}(k_x, k_y) \cdot \tilde{g}(k_x, k_y)\right]$. Non-zero values exit at the boundary of Figure 3-29c. We thus need a circular low-pass filter to remove all the high-order signals in the Fourier domain (Equation (3.26)). We Fourier transform $\tilde{w}_t(k_x, k_y)$ to obtain $w_t(x, y)$. Afterwards, we only take the signals in $w(x, y)$ from -0.5 to 0.5 mm in both x and y axes (Figure 3-29d).

In the next step, we center the elementary mode at individually shifted positions and scale its magnitude given by the local value of $w_t(x, y)$. We use two different grids to place the modes. One is a Cartesian grid, and another is a hexagonal grid. The intensity sum of all the elementary modes composes a Gaussian-Schell beam.

To test the accuracy of the Wiener deconvolution, we subtract the initial intensity of the Gaussian-Schell beam with the incoherent sum of all the elementary modes. The goal is to reach a difference of zero. Figure 3-29e and f show that the difference of a hexagonal sampling is a factor of two lower than the difference of a Cartesian sampling. The reason is that a hexagonal grid is isotropic. It provides the densest packing of circles and offers a better fit to a rotationally symmetric signal.



(a) Transverse intensity of the Gaussian-Schell beam

(b) Intensity of one elementary mode

(c) Absolute logarithm values of $[\tilde{I}(k_x, k_y) \cdot \tilde{g}(k_x, k_y)]$

(d) Weight function $w_t(x, y)$ with C_{ns} =1e-4.

(e) Intensity difference based on a Cartesian grid of elementary modes

(f) Intensity error based on a Hexagonal grid of elementary modes

Figure 3-29. Using Wiener deconvolution to compose a Gaussian-Schell beam with two transverse dimensions.

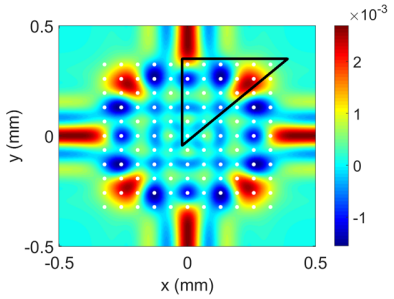


Figure 3-30. Intensity error by using a minimal search. The white points mark the location in individual modes. Only the magnitudes of the modes inside the black triangle are defined as variables.

Another outstanding advantage of using the Wiener deconvolution is the computational efficiency. The runtime to compute the results in Figure 3-29 is 0.6 second (Intel Core i7, 64 bit). We compare this method with a minimization method. A minimal search function in Matlab® is applied to optimize the mode magnitudes. Due to the rotational symmetry of a Gaussian-Schell beam, only the modes inside a π/4 azimuthal area are variables (black triangle in Figure 3-30). The constraint is to reach a zero difference between the incoherent sum of all modes and the initial Gaussian-Schell intensity. The runtime of the minimization approach takes 22 minutes (512 by 512 pixels, 21 variables). The accuracy is only half compared to the approach using Wiener deconvolution.

3.2.3 Partially coherent light in a step-index waveguide

In this section, we apply the modal theory to propagate a partially coherent beam inside a waveguide. By that, we can accurately model the diffraction effects generated by the waveguide structure.

We use a Gaussian-Schell beam as an example and propagate this beam inside a slab waveguide with a step index profile. The correlation function at the source plane, denoted as $z = 0$, is defined by a sum of multiple eigenmodes (i.e. Hermite modes),

$$\Gamma(x_1, x_2; z = 0) = \sum_n f_n^*(x_1; z = 0) f_n(x_2; z = 0), \tag{3.27}$$

We then decompose each eigenmode into a full set of the guided modes of the waveguide,

$$f_n(x; z = 0) = \sum_m c_{mn} u_m(x; z = 0), \tag{3.28}$$

where u_m is the m^{th}-order guided mode derived from the eigenvalue equation of a specific waveguide, and c is the waveguide mode coefficient. The propagation of an individual guided mode is described by a harmonic modulation along the optical axis and characterized by the propagation constant β_m [58],

$$u_m(x,z) = u_m(x;z=0)\exp(i\beta_m z), \tag{3.29}$$

The propagation of every coherent beam thus follows

$$f_n\ (x,z) = \sum_m c_{mn}u_m(x;z=0)\exp(i\beta_m z) = \sum_m c_{mn}^{(z)}u_m(x;z=0), \tag{3.30}$$

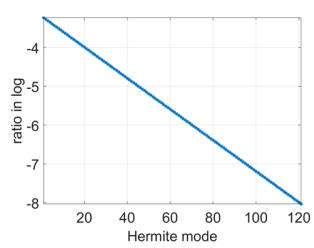
where $c_{mn}^{(z)} = c_{mn}\exp(i\beta_m z)$. Since the guided modes u_m remain unchanged in propagation, the propagation information of the Gaussian-Schell beam is carried out by $c_{mn}^{(z)}$. This method allows us to quickly access the light fields at any transversal plane inside the waveguide by simply varying the parameter z in Equation (3.30). We use this expansion tool to investigate the diffraction effects inside the waveguide.

We simulate a Gaussian-Schell beam with the wavelength of 0.5 μm, a beam waist of 50 μm and a correlation length of 1 μm. According to Equation (3.27), the full beam is decomposed into 121 eigenmodes. The power ratio of these modes is illustrated in Figure 3-31a. We propagate this beam into a step-index slab waveguide with an incidence angle of 4°. The waveguide has a diameter of 180 μm with $n_{core} = 1.5, n_{clad} = 1.45$, providing totally 277 guided modes.

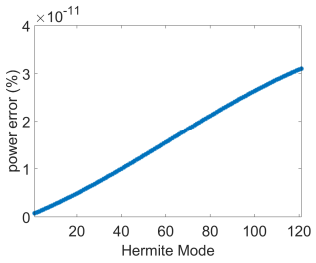
According to the Equation (3.30), the absolute values of $c_{mn}^{(z)}$ do not change in propagation, *i.et* $\left|c_{mn}^{(z)}\right| = |c_{mn}|$. The matrix of $|c_{mn}|$ is visualized in Figure 3-31d and e for different incidence angles of the Gaussian Schell beam. With 0^o incidence, only the odd-number guided modes contribute to the odd-number Hermite modes, also only the even-number guided modes contribute to the even-number Hermite modes. However, a titled incidence of the Gaussian Schell beam breaks this symmetry. The odd and even guided modes have mixed contribution to each Hermite mode. Furthermore, a tilted incidence, compared to the 0^o incidence, requires more contribution from the higher-order guided modes. We can imagine, for a beam with a sufficiently low degree of coherence, the given guided modes may fail to fully represent the higher-order Hermite modes. In that case, only a limited amount of Hermite modes propagate along the waveguide. Figure 3-31b shows the error of using guided modes to represent the Hermite modes of our defined Gaussian-Schell beam. The magnitude of the error implies that this full beam is well represented by the guided modes. That means the Gaussian–Schell beam is fully guided. The propagated intensity is depicted in Figure 3-31c.

Intensity fringes appear near the borders of the waveguide. That is because the step-index profile yields total internal reflection. The light propagating towards the border encounters the light

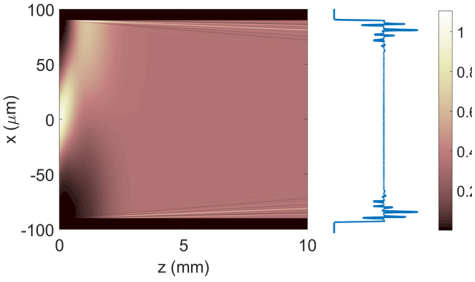
reflected by the border. They interfere with each other to a certain extent. The strength of interference is determined by the degree of coherence. If the beam is completely incoherent, all the light overlays on each other incoherently. In that case, there will be no visible intensity fringes.



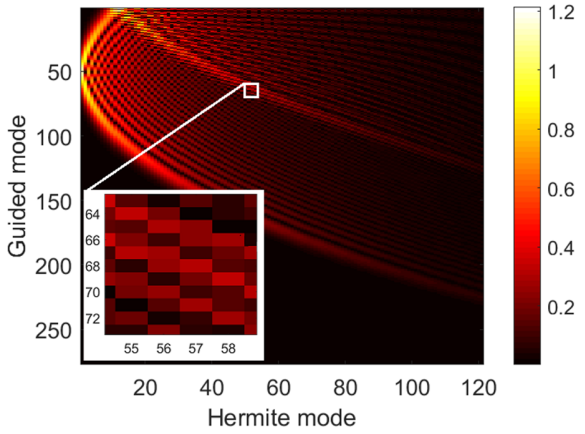
(a) Power ratio of the Hermite modes



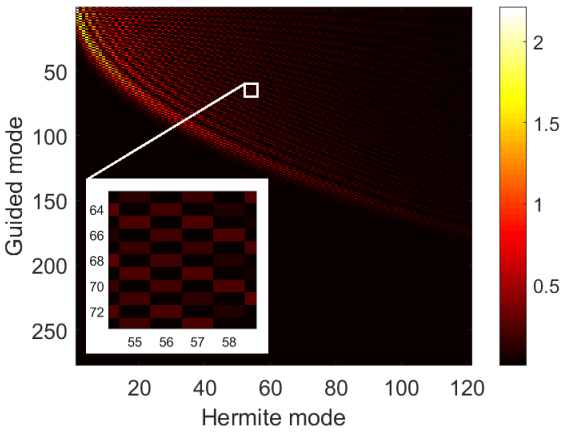
(b) Power error of using waveguide modes to represent Hermite modes.



(c) Propagated intensity inside the waveguide. The blue curve denotes the transverse intensity at $\boldsymbol{z = 10\ mm}$.



(d) Matrix of $|\boldsymbol{c_{mn}}|$ with $\mathbf{4^o}$ incidence of the Gaussian Schell beam.



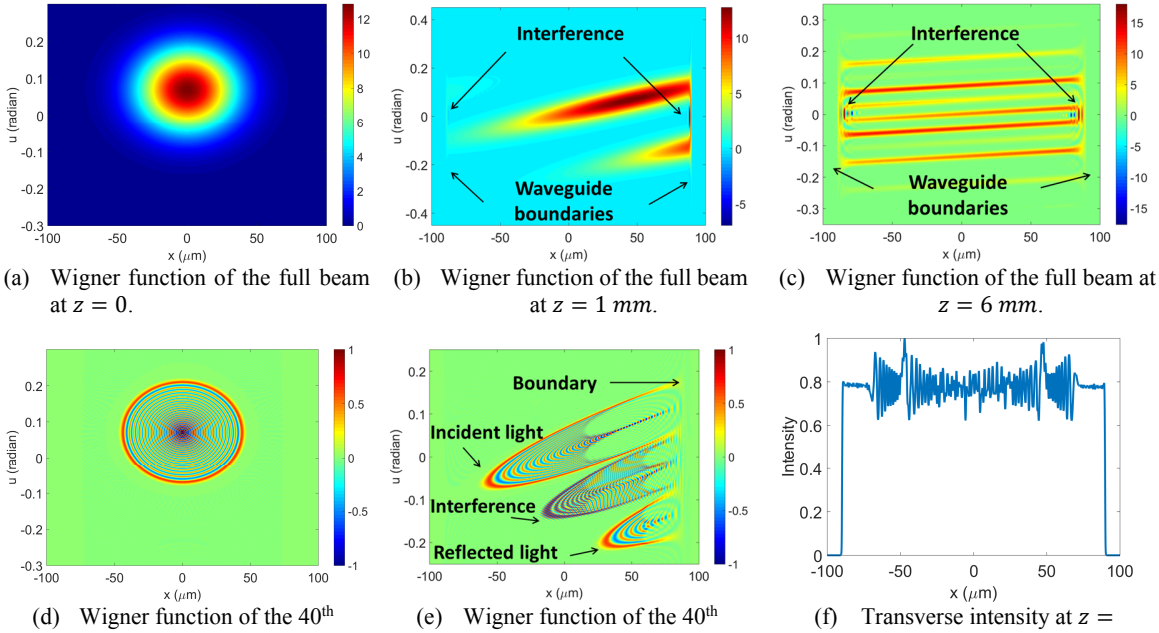
(e) Matrix of $|\boldsymbol{c_{mn}}|$ with $\mathbf{0^o}$ incidence of the Gaussian Schell beam.

Figure 3-31. Propagation of a Gaussian-Schell beam inside a step-index slab waveguide. The light source has the following parameters, $\boldsymbol{\lambda = 0.5\ \mu m, l_c = 1\ \mu m, w_0 = 50\ \mu m}$, 121 eigenmodes, (a)-(d) with an incidence angle of 4°, (e) with an incidence angle of 0°. The slab waveguide has a diameter of 180 μm, $\boldsymbol{n_{core} = 1.5, n_{clad} = 1.45}$ and 277 guided modes.

In our example (Figure 3-31c), the intensity fringes begin to appear once the beam reaches the borders at $z = 0.5\ mm$. As the beam travels further inside the waveguide, there is always light being reflected and interfered. Therefore, interference effects happen continuously. As a result, we see that the fringes also travel along the z direction (Figure 3-31c). The transverse intensity shows an asymmetric fringe profile (blue curve in Figure 3-31c) due to the titled incidence. We

may expect, as these fringes travel further, they move away from the border towards the center of the core (*i.e.* $x = 0$). As more and more fringes emerge at the border and propagate ahead, they overlap and smear out each other to a certain extent.

Figure 3-32 shows the Wigner functions of the light fields inside the waveguide. The titled incidence of the input beam is indicated by the signals in Figure 3-32a centered at $u = 0.07\ rad$. When the beam propagates further, its spatial extension increases until it reaches the waveguide borders. The reflected light at $z = 1\ mm$ (Figure 3-32b) is represented by the signals that are flipped back at the boundary position with an opposite angle. Between the incident and the reflected beams, diffraction ripples appear. They indicate the interference effects due to reflection. At $z = 6\ mm$ the light is reflected several times and fills the whole transverse region of the waveguide (Figure 3-31c). Those lines in the Wigner function (Figure 3-32b) reveals how many times the beam is reflected. Near the border position (*i.e.* $x = \pm 90\ \mu m$) and $u = 0$, the diffraction ripples become more visible (pointed by "Interference" in Figure 3-32c). They are the cross terms contributed by all the positive signals. In other words, the intensity fringes we see in Figure 3-31c result from the intensities between all the light beams that travel close to each other.



(a) Wigner function of the full beam at $z = 0$.

(b) Wigner function of the full beam at $z = 1\ mm$.

(c) Wigner function of the full beam at $z = 6\ mm$.

(d) Wigner function of the 40th eigenmode at $z = 0$.

(e) Wigner function of the 40th eigenmode at $z = 1\ mm$.

(f) Transverse intensity at $z = 2000\ mm$.

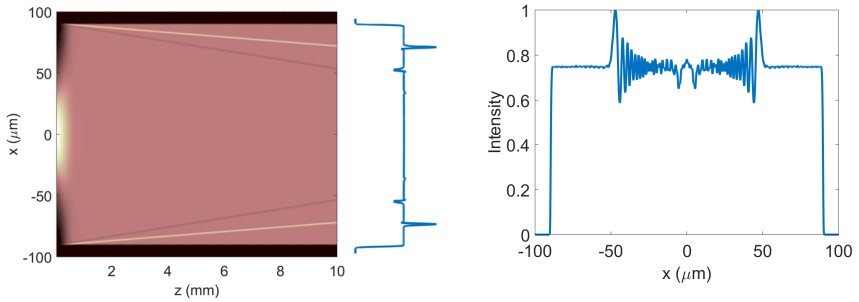
Figure 3-32. Wigner functions of light fields inside the waveguide. The parameters of the light source and the waveguide follow those in Figure 3-31.

Figure 3-32d and e show the Wigner functions of one eigenmode. That means, the signals represent completely coherent light. The densely sign-alternate rings in Figure 3-32d come from the oscillatory amplitudes of the eigenmodes (similar to Figure 3-21). At $z = 1\ mm$ (Figure 3-32e), the boundary reflects a part of the beam. The incident and reflected beams are clearly separate in angles. Due to complete coherence, the diffraction ripples are pronounced. It is obvious that the interference effects in Figure 3-32b are weaker than in Figure 3-32e. That is because, the distribution in Figure 3-32b is an overlapping of the Wigner functions of individual eigenmodes. This overlapping equals to an incoherent sum of multiple beams. Therefore, the interference fringes of a partially coherent beam are weaker than the interference fringes generated by a completely coherent beam.

Now we discuss the beam homogenizing effect produced by the waveguide. Figure 3-32f is the transverse intensity profile at $z = 2000\ mm$. Compared to the transverse intensity in Figure 3-31c, more fringes are produced with a larger distance of z due to multiple times of reflection. Theses intensity fringes are contributed by the propagated Hermite modes. We prove in Figure 3-33.

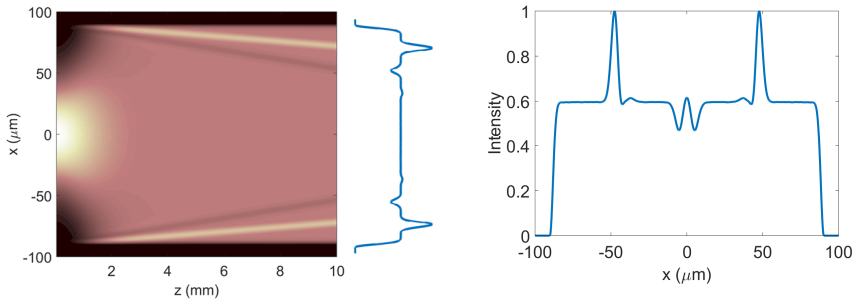
Figure 3-33 depicts two Gaussian-Schell beams with varied correlation lengths (l_c). A beam wth a smaller correlation width requires more Hermite modes for a full representation. The fringe width at $z = 10\ mm$ is proportional to the value of l_c (Figure 3-33a and c). It also implies that, with $l_c \to 0$ (*i.e.* with complete incoherence) the intensity fringes disappear. We compare these two beams (Figure 3-33b and d) at $z = 2000\ mm$. The beam with a lower degree of coherence gets more homogenized.

We compare Figure 3-32c and Figure 3-33b. They suggest, given the same distance of propagation, a tilted incidence results in more reflections. Therefore Figure 3-32c has denser fringes than Figure 3-33b. This coincides with the geometrical picture in Figure 3-18. A ray with a larger incidence angle encounters more reflection, given the same distance along the optical axis. However, the two profiles in Figure 3-32c and Figure 3-33b share a similar peak-to-valley ratio. It indicates the degree of coherence determines the level of homogenizing. A longer propagation distance does not improv the contrast of the intensity fringes.



(a) Correlation width $l_c = 1$ μm, 121 Hermite modes.

(b) Transverse intensity at $z = 2000\ mm$ with $l_c = 1$ μm.



(c) Correlation width $l_c = 4$ μm, 35 Hermite modes.

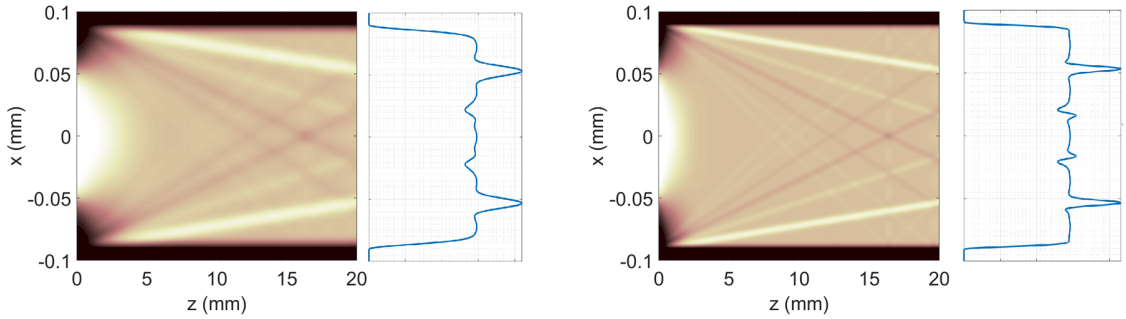
(d) Transverse intensity at $z = 2000\ mm$ with $l_c = 4$ μm.

Figure 3-33. A Gaussian-Schell beam with a different correlation width inside the waveguide. The blue curves are the transverse intensities at $z = 10\ mm$. The parameters are the same as those in Figure 3-31, except the value of l_c.

To achieve an entirely top-hat transverse intensity, we need an infinite amount of Hermite modes. This is similar to the Gibbs phenomenon. However, the property of a waveguide prohibits the possibility of having infinite eigenmodes. We know that an increase of the waveguide diameter leads to more guided modes. However, the maximum frequency of the guide mode is defined as $2\pi\sqrt{{n_{core}}^2 - {n_{clad}}^2}/\lambda$, independent of the waveguide diameter. That imposes a limit on the highest-order Hermite mode that is allowed to propagate inside the waveguide.

In reality we may expect that a larger propagation distance yields a more homogenized beam intensity. That corresponds to the concept of temporal coherence. As each guided mode has own propagation constant β, they travel at a slightly different speed along the optical axis. When the distance between the fastest and the slowest guided modes is larger than the coherence length (defined by temporal coherence, not the correlation length in spatial coherence), they do not interfere with each other anymore. However, regarding monochromatic spatial coherence, a larger propagation distance does not produce a more homogenized transverse intensity. In practice, a strictly monochromatic source does not exist. That means the temporal coherence length is always finite. In practice, we might still see a more homogenized effect when the

propagation distance is sufficiently increased. However, that is a contribution of temporal coherence, not spatial coherence.



(a) Propagated intensity of the flat-top Schell beam inside the waveguide, with $\lambda = 1\ \mu m$, $w_e = 8\ \mu m$, 41 shifted-elementary modes. The blue curve is the transverse intensity at $z = 20\ mm$.

(b) Propagated intensity of the flat-top Schell beam inside the waveguide, with $\lambda = 1\ \mu m$, $w_e = 4\ \mu m$, 121 shifted-elementary modes. The blue curve is the transverse intensity at $z = 20\ mm$.

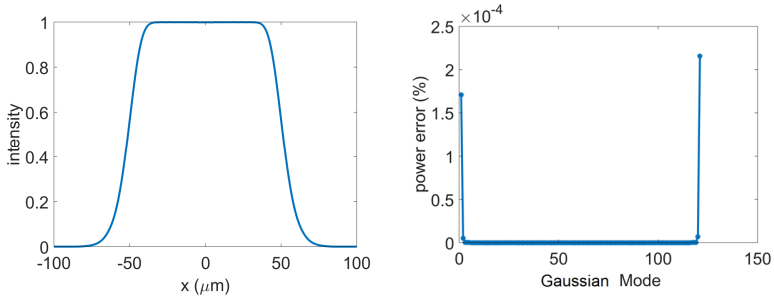
Figure 3-34. Flat-top Schell beam inside a step-index waveguide.

In the next step, we use a flat-top Schell beam as the input light source, to investigate the shifted-elementary modes inside a waveguide. The full beam is represented by multiple shifted-elementary modes, *i.e.* coherent Gaussian beams. Each coherent Gaussian beam is expanded into a linear combination of all the guided modes.

Figure 3-34 implies that the fringe size at the transverse intensity is proportional to the beam waist of the coherent Gaussian beam (denoted as w_e). To represent a full beam with a lower degree of coherence, we need elementary modes with a smaller beam waist. Therefore, a larger amount of elementary modes is required. Based on Figure 3-34, We may expect, when the beam is completely incoherent, *i.e.* $w_e \to 0$, intensity fringes disappear.

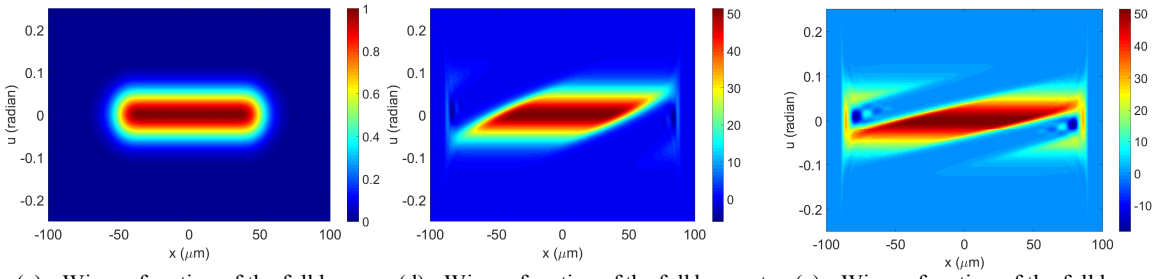
Figure 3-35 visualizes the Wigner functions the light given by Figure 3-34b. The full beam is represented locally by individual shifted modes. When the beam is truncated by the waveguide borders, only the modes near the borders are cut-off (Figure 3-35b). The propagation of the full beam is expressed as a shearing of the signals in the Wigner functions (Figure 3-35 c-e). Reflection at the borders happens at z=1 mm. The interference effects are distinctly indicated by the ripples near the waveguide borders at z=2 mm (Figure 3-35e). Due to the local representation of shifted modes, the inference effects are only generated by the modes that are near the borders (Figure 3-35f and g). The alternate negative-positive ripples in Figure 3-35g

indicate the interference effect generated by every Gaussian mode. Figure 3-35h shows the mode that has not been touched by the borders at $z = 1\ mm$. It propagates as if it is in free space.



(a) Transverse intensity of the full beam at z=0 mm

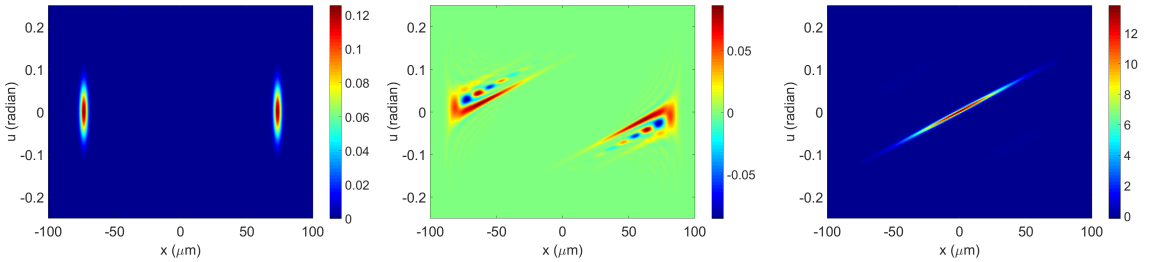
(b) Power error of using the guided modes to represent each elementary modes.



(c) Wigner function of the full beam at $z = 0\ mm$.

(d) Wigner function of the full beam at $z = 1\ mm$.

(e) Wigner function of the full beam at $z = 2\ mm$.



(f) Wigner function of the elementary modes f_4 and f_{116} at $z = 0\ mm$.

(g) Wigner function of the elementary modes f_4 and f_{116} at $z = 1\ mm$.

(h) Wigner function of the elementary mode f_{60} at $z = 1\ mm$.

Figure 3-35. Wigner functions of the beam given by Figure 3-34b.

3.2.4 Partially coherent light in a parabolic-index waveguide

In the following, we discuss the propagation of partially coherent light inside a waveguide with a parabolic index profile.

We define the parabolic index profile of the waveguide as

$$n_{rf}{}^2(\bar{x}) = n_{core}^2(1 - 2\Delta\bar{x}^2), \tag{3.31}$$

where $\bar{x} = x/\rho$, ρ is the radius of the waveguide, n_{core} denotes the refractive index of the core, Δ represents a unitless constant related to the index profile. There is another common way to express the same index profile:

$$n_{rf}{}^2(\bar{x}) = n_0 + n_2\rho^2(\bar{x})^2, \tag{3.32}$$

where $n_{core} = \sqrt{n_0}$ and $\Delta = \frac{n_2\rho^2}{-2n_0}$.

The guided modes of the waveguide take the form[58]

$$u_m(\bar{x}) = \exp(-V\bar{x}^2/2)\, H_m\left(V^{1/2}\bar{x}\right), \tag{3.33}$$

where m is the number of the guided mode starting from 0, H_m denotes the m^{th} order Hermite polynomial, $V = k\rho n_{core}\sqrt{2\Delta}$. The propagation constant for each guided mode follows

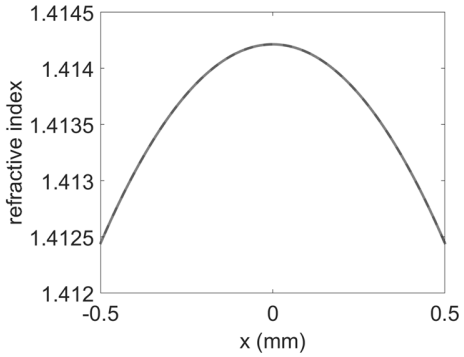
$$\beta_m(\bar{x}) = \sqrt{V^2/2\Delta - U^2}/\rho, \tag{3.34}$$

where $U = \sqrt{(2m+1)V}$.

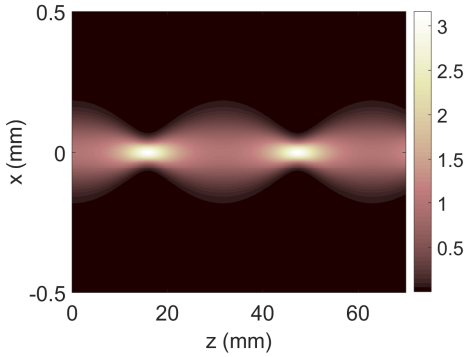
Note that Equation (3.33) describes a Hermite-Gaussian function. It is in a similar form as the eigenmodes of a Gaussian-Schell beam (Equation (2.33)). However, the dominant parameter V in Equation (3.33) depends on the waveguide structures, while the other dominant parameter c in Equation (2.33) is related to the coherence degree of the light source. As these two equations have different physical meanings, we should not confuse them with each other. For a clear differentiation, in the following example, we call the eigenmodes of a Gaussian-Schell beam as the source modes. The eigenmodes of the waveguide are still called guided modes.

We propagate a Gaussian-Schell model beam inside the waveguide. The full beam is represented by the orthogonal source modes (given by Equation (2.33) and (2.34)). Each source mode is expanded into a linear sum of the guided modes (given by Equation (3.33)). The propagation of each guided mode is described by Equation (3.30).

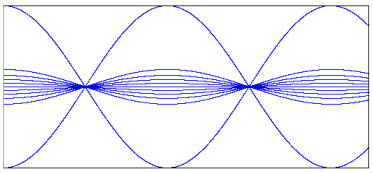
Figure 3-36 shows a simulated example. The waveguide has diameter of 1 mm, $n_{core} = 1.4142$, $\Delta = 0.0013$ for the wavelength $\lambda = 1$ μm. We include 70 guided modes. The parabolic index profile is shown in Figure 3-36a. The Gaussian-Schell beam in Figure 3-36b has a beam waist $w_0 = 150$ μm, a correlation width $l_c = 50$ μm, and 25 source modes. Due to the variation of the refractive index, light inside the waveguide travels in curved paths. Given a small NA, the evolution of light is periodic. It forms a fractional Fourier transform. In Figure 3-36b with the given z axis, light is focused at $z = 15.69$ mm and 31.38 mm. Every distance of 62.76 mm along the z axis includes one period of the evolution. The focus at $z = 15.69$ mm has a FWHM of 55.1 μm.



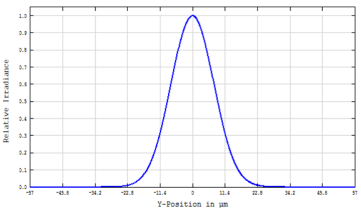
(a) Parabolic index profile at the transverse plane of a waveguide.



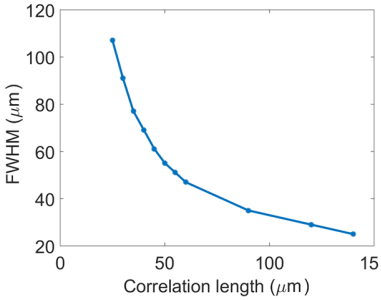
(b) Propagated intensity of the Gaussian-Schell beam with the following parameters, $w_0 = 150\ \mu m$, $l_c = 50\ \mu m$, 25 source modes.



(c) Ray paths in Zemax®, with a total length of 70 mm along the z axis and a Gaussian podization factor $\left(D_{wg}/2\right)^2/w_0^2 = 11.11$.



(d) The coherent PSF at $z = 15.69\ mm$ with a FWHM of 18 μm.



(e) FWHM of the focus ($z = 15.69\ mm$) with respect to a varied correlation length.

Figure 3-36. Propagation of a Gaussian-Schell beam inside a parabolic-index waveguide. Parameters are as follows, waveguide diameter $D_{wg} = 0.5\ mm$, , $n_c = 1.4142$, $\Delta = 0.0013$, 70 guided modes.

To investigate the impact of coherence on the focus size, we firstly look into the completely coherent case. We generate the ray paths in Zemax® with the parameters of wavelength and index profile (Figure 3-36c). The illumination in Figure 3-36c has an aperture diameter $D_{wg} = 1\ mm$, and a coherent Gaussian beam with a waist $w_0 = 150\ \mu m$. The first focus is located at $z = 15.69\ mm$, same as Figure 3-36b. The focus has a FWHM of 18 μm, which is a diffraction limited result.

We now vary the correlation width and evaluate beam size at the focus position (Figure 3-36d). An increase of the correlation width implies a higher degree of coherence. The FWHM of the focus thus decreases. The curve has the nature of an asymptote. Eventually, it reaches the FWHM of 18 μm for a completely coherent case.

It is worth noting that, in the paraxial regime, analytical solutions exist for the coherent light propagation inside a parabolic-index waveguide[59]. Our model serves as an alternative method.

We summarize the section of model expansion as follows. We have investigated the modal properties of the Gaussian-Schell beam and flat-top Schell beam. We expand the partially coherent beam into eigenmodes and shifted-elementary modes. Their optical effects are discussed in the Wigner function.

The are several advantages of using eigenmode expansion. First, the eigenmodes are orthogonal to each other. Thus the number of modes that we need to represent the beam fully is usually smaller than using the non-orthogonal mode expansion. Second, the power contribution of the eigenmodes decreases exponentially as the order increases. Therefore, by looking into the power ratio of high-order modes, we can estimate how many modes are sufficient to represent a full beam. Third, for any correlation function, we can derive the eigenmodes by using the linear algebraic method. This offers great convenience for analyzing complicated light sources.

However, there are disadvantages of using eigenmodes. The main concern falls on the sampling requirement because the higher-order modes exhibit faster oscillations. For a source with a low degree of freedom, we need not only more modes but also a denser sampling. Another complication of using eigenmode expansion is, the interference terms of individual modes can be complicated to interpret. We see this in Figure 3-32. Only when we sum up the Wigner functions of all the modes, the interference effects of the full beam are visualized.

If we use shifted-elementary modes, the above disadvantages of eigenmodes can be avoided. For a beam with a low degree of coherence, if one elementary mode is sampled properly, all the other modes follow the same sampling demand. The interference effect of a beam is locally described by the shifted modes. For example, inside a slab waveguide, only the modes near the borders are reflected and interfered. Therefore, the diffraction effects of a few modes at a specific spatial location already provides the local information of the partially coherent beam.

4 Coherence retrieval based on the inverse problems of modal expansion and the Wigner function

This chapter contains the application of the modeling methods based on the Wigner function and modal expansion in practice. They contribute to the coherence measurement in reality.

The first method corresponds to solving the inverse problem of the modal expansion (Section 4.1). It is solved by the iterative phase retrieval algorithms. We extend this scheme to a beam with an arbitrary wavefront. Cases with one transverse dimension (Section 4.1.1) and two transverse dimensions (Section 4.1.2) are presented. Afterwards, we discuss the potential and limit of this method.

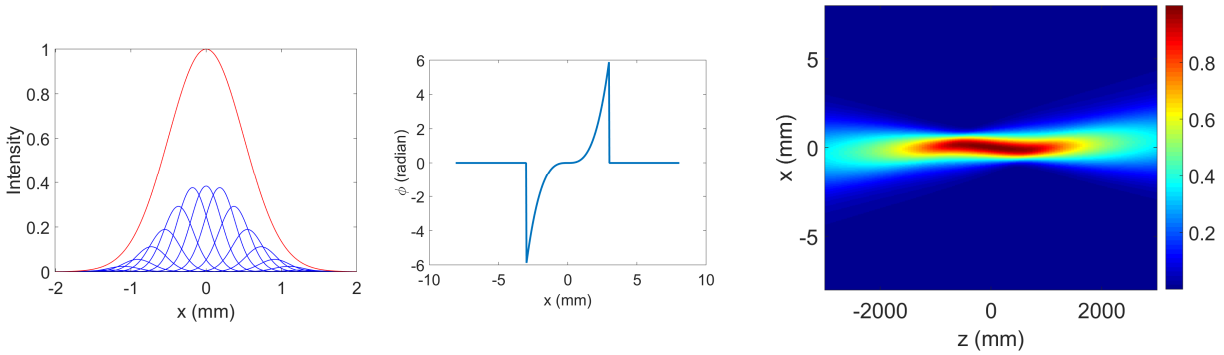
The second method for the coherence retrieval is the phase space tomography (Section 4.2). There are two main algorithmic schemes. One is the filtered-back projection. It recovers data directly in the Wigner function. Another is the Fourier slice theorem. It reconstructs the ambiguity function. We draw a short discussion of these two algorithms. Afterwards, we propose an efficient algorithm for the filtered back projection to recover the 4D Wigner function.

4.1 Mode recovery with an arbitrary wavefront

Rydberg[37] has applied the Gerchberg-Saxton algorithm to retrieval the correlation function of a Gaussian-Schell beam. In our work, we assign an arbitrary wavefront to the partially coherent beam and employ the similar algorithm to retrieve the correlation function. We first discuss the coherence reconstruction of a partially coherent beam with one transverse dimension and a 3rd-order phase term. Later on, we present a beam with two modes and recover their complex fields with two transverse dimensions.

4.1.1 Mode recovery with one transverse dimension

A Gaussian-Schell beam (with the wavelength $\lambda = 1\ \mu m$, a beam waist $w_0 = 1\ mm$ and a correlation width of 0.4 mm) is composed of 21 elementary modes (Figure 4-1a). Each elementary mode has a width of 0.4 mm. At $z = 0\ mm$, we insert a thin phase element containing a 3rd-order aberration to the Gaussian-Schell beam. The phase plate has a radius of $\rho = 3\ mm$ and a phase $\psi = 6(x/\rho)^3$ (Figure 4-1b), where x is the spatial coordinate at $z = 0\ mm$. Therefore, each mode of the Gaussian-Schell beam is added by this additional 3rd-order phase term, defined by $\exp(i2\pi\psi)$. We propagate the full beam in the entire space with z varying from -3000 to $3000\ mm$ (Figure 4-1c).



(a) The Gaussian-Schell beam composed of 21 elementary modes.

(b) Phase added to the Gaussian-Schell beam at $z = 0\ mm$.

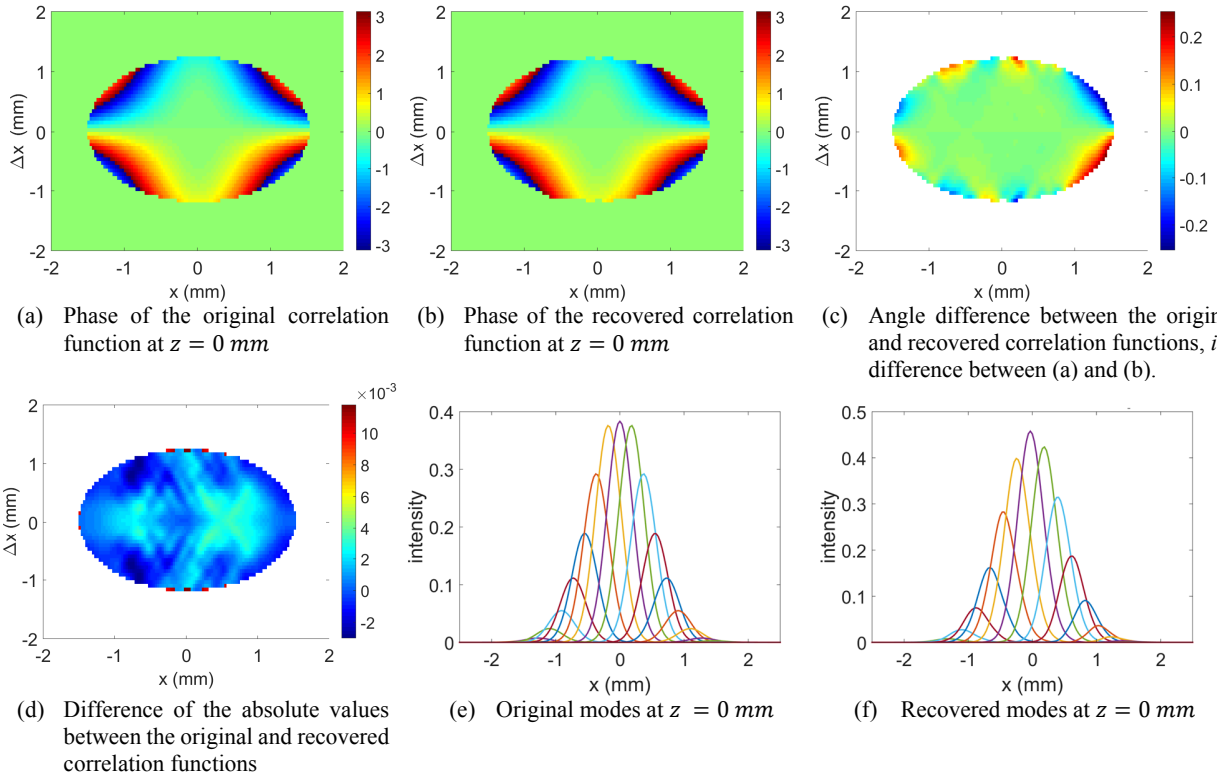
(c) Propagated intensities of the full beam in space.

Figure 4-1. Propagation of a Gaussian-Schell beam with a 3rd-order phase term.

We use the algorithm in Chapter 2.4.1 to retrieve the amplitude of individual elementary modes. In general, a larger amount of transverse planes produces a faster convergence in the Gerchberg-Saxton algorithm (explained in Equations (2.57-2.60)), because it provides more variation about the propagated phase. We take the transverse intensities at z = [-2800, -2000, -1400, -900, 0, 500, 1100, 1800, 2600].

mm as the reference intensities. Note that, taking more z planes over a large distance for the reference intensities are recommended. It provides more information of the phase evolution. That is the reason we take nine z planes over a distance of $5400\ mm$.

The iteration starts from $z = 0$, with 21 shifted Gaussian beams, each with a equal width of 0.2 mm, an equal magnitude and a constant phase. We consider the propagation through all the nine transverse planes as one iteration. Figure 4-2 shows the result after 30 iterations. The recovered correlation function at $z = 0$ (Figure 4-2b) converges to the original correlation function (Figure 4-2a). The 3rd-order phase aberration is illustrated inside. The angle difference (Figure 4-2c) between the original the recovered correlation shows a 10% peak-to-valley deviation. The error mainly concentrate on the borders of the beam, where the correlation function has absolute values below 1% of the maximum. Therefore, the angle error does not play a critical role in the numerical accuracy. Figure 4-2d depicts the absolute-value difference between the original and reconstructed correlation functions, both normalized to have a maximum of 1. The magnitude of the deviation is in a negligible range.



(a) Phase of the original correlation function at $z = 0\ mm$

(b) Phase of the recovered correlation function at $z = 0\ mm$

(c) Angle difference between the original and recovered correlation functions, *i.e.* difference between (a) and (b).

(d) Difference of the absolute values between the original and recovered correlation functions

(e) Original modes at $z = 0\ mm$

(f) Recovered modes at $z = 0\ mm$

Figure 4-2. Reconstructing the correlation function from the elementary mode retrieval, using the Gerchberg-Saxton algorithm with 30 iterations.

Figure 4-2e and f show the original and recovered modes at $z = 0$. It is important to point out that the recovered modes are not necessarily identical to the original modes. It depends on how we define the initial modes at the beginning of the iterations. Although the recovered modes are not unique, they yield a uniformly convergent correlation function. The reasons are as follows. For a beam with one transverse dimension, the intensity in the entire space uniquely determines the correlation function. However, a given correlation function does not uniquely determine the shifted mode expansion. That is because the shifted modes do not form an orthogonal expansion. The iterative algorithm is based on correcting the incoherent sum of the modes. Consequently, that modifies their individual phase, leading to a convergent solution to the phase difference of every two spatial coordinates. But the magnitudes of individual modes, the distances between every two adjacent modes, and their global phase are left free. We see that by comparing Figure 4-2e and f. The magnitudes of the recovered modes are not symmetric about $x = 0$. Each recovered mode has a slight change of magnitudes compared to those in Figure 4-2e. There are

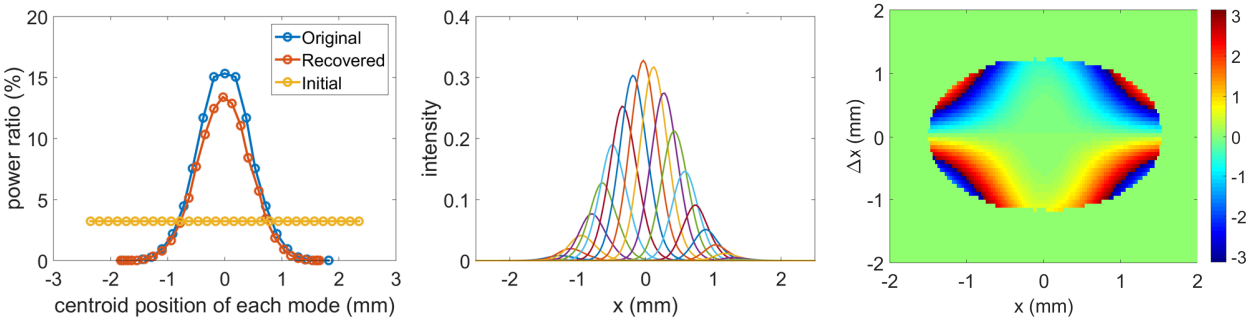
two features remaining unchanged. First, the width of the elementary modes must equal to the correlation length, because the coherence length defines the diffraction angle of the full beam. In our example, the recovered modes have a uniform width of 0.4 mm, same as the correlation width. Second, the total number of modes that give a main contribution to the full-beam power remain similar. That coincides with our previous conclusion that, with a given correlation function, there is a threshold for the demanded number of modes for a full presentation. After reaching a certain amount, a further increase of the mode number is unnecessary.

When the defined number of the recovered modes is different from the total number of the original modes, the algorithm is more likely to converge to a different solution. If the recovered modes are less than the original modes, we lose the physical accuracy of the recovered beam. One should avoid this situation.

We simulate another case where the recovered modes are more than the original modes (Figure 4-3). The original beam path is the same as Figure 4-1, with 21 original shifted-modes. The iteration starts with 31 initial modes with an equal magnitude and a constant phase. Therefore the power contribution of the initial modes is uniform (Figure 4-3a). After 30 iterations, the extra modes move towards the border of the full beam. The power contribution of the extra modes at the border is significantly low. The full-beam power mainly comes from the 13 modes near the origin of the x axis. The centroid positions of these 13 recovered modes along the x axis approach the centroid positions of the original modes. The deviation between the original and recovered modes in Figure 4-3a again proves that, the elementary-mode expansion is not a unique expansion. The intensities of the recovered modes at $z = 0$ are illustrated in Figure 4-3b. The width of each mode remains unchanged compared to Figure 4-2e and f. That is because the global degree of coherence remains the same, no matter how many recovered modes we define in the iterations. After 30 iterations, the recovered correlation function is shown in Figure 4-3c. It highly coincides with the original correlation function (Figure 4-2a).

Therefore, we draw the following conclusion of the mode recovery with one-transverse dimension. The Gerchberg-Saxton algorithm provides a convergent correlation function to a Gaussian-Schell beam with a phase aberration included. With one transverse dimension, the propagated intensities in space uniquely determine the correlation function. A correlation function only denotes the phase difference of fields at every two spatial coordinates without revealing the global phase. Moreover, the elementary mode decomposition is not an orthogonal

modal expansion for a partially-coherent beam. Depending on the initial parameters of the iterations, the phase retrieval algorithm can converge to a group of modes with a slightly different composition. Although the solution of the modes is not unique, they all give rise to a unique correlation function.



(a) Comparing the centroid and power of individual modes.

(b) Intensities of the 31 Recovered modes.

(c) Angle of the recovered correlation function, consisting of 31 modes.

Figure 4-3. Phase retrieval of 31 recovered modes for a partially coherent beam originally consisting of 21 modes.

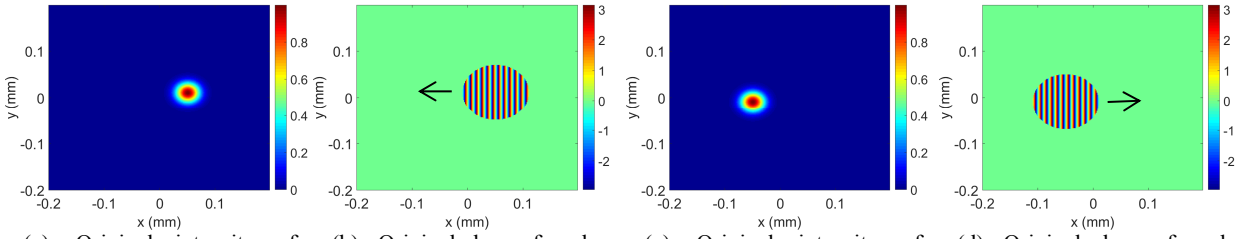
4.1.2 Mode recovery two transverse dimensions

When a beam contains two transverse dimensions, the information of partial coherence is described by a 4D correlation function, *i.e.* $\Gamma(x_1, y_1, x_2, y_2)$. The propagated intensities in the entire space provide a 3D distribution, *i.e.* $I(x, y, z)$. This is not adequate to uniquely determine the 4D correlation function. This ambiguity can be removed by inserting a rotationally asymmetric component into the beam path. It equals to introducing an additional diversification to the system. By rotating this asymmetric component, the propagated intensities change accordingly. It provides one more degree of freedom for the phase retrieval algorithm.

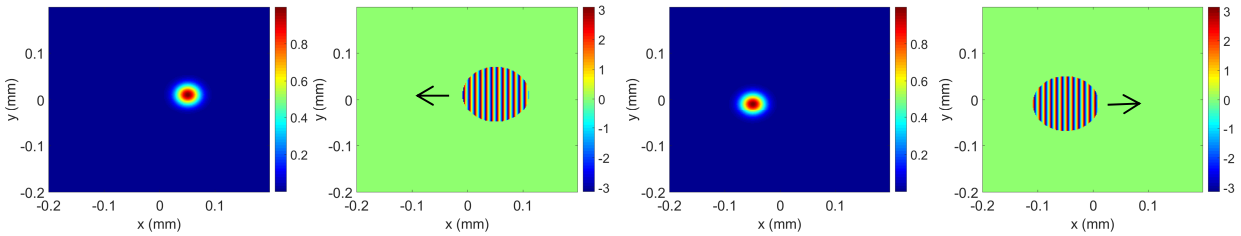
We simulate a beam with two shifted Gaussian modes. These two modes are spatially separate at $z = z_0$ (Figure 4-4a and c). We add a tilted wavefront with an opposite angle to both modes (Figure 4-4b and d). Therefore, two modes travel towards each other in propagation. We place a cylindrical lens at $z = z_0$. By rotating the cylindrical lens (with a focal length of 4 mm), the full beam consisting of two modes exhibit varied intensity distribution at each z distance (Figure 4-5). We take them as the reference intensities for the phase retrieval algorithm. The detail of the algorithm is shown in Figure 4-6. At each transverse plane z_i, three rotations of the asymmetric element yield three reference intensities. (Here the number of rotation is not restricted.) We thus propagate the fields between z_0 and z_i for three rounds. At each round of

propagation, when the fields return to the transverse plane of z_0, we switch the phase given by the rotated element. The steps inside the dash frame in Figure 4-6 form one iteration for one transverse plane.

In our simulated example, we take reference intensities at $z = 0.3, 0.6, 0.9\ mm$, each with $\theta = 0^o, 22^o, 45^o, 67^o$. After 200 iterations for each transverse plane, we obtain the recovered modes in Figure 4-4 e-h.

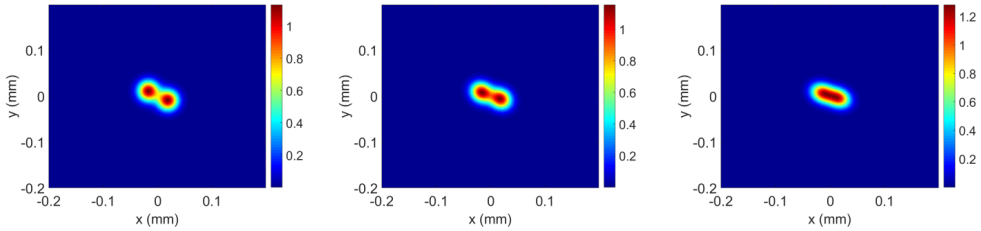


(a) Original intensity of mode 1 (b) Original phase of mode 1, arrow indicating the propagation direction. (c) Original intensity of mode 2 (d) Original phase of mode 2, arrow indicating the propagation direction.



(e) Recovered intensity of mode 1 (f) Recovered phase of mode 2 (g) Recovered intensity of mode 2 (h) Recovered phase of mode 2

Figure 4-4. Comparison between the original and recovered modes.



(a) Intensity at $z = 0.6\ mm$ with $\alpha = 0^o$ (b) Intensity at $z = 0.6\ mm$ with $\alpha = 22^o$ (c) Intensity at $z = 0.6\ mm$ with $\alpha = 45^o$

Figure 4-5. Reference intensities consisting of two modes.

The recovered phase, compared to the original phase, contains a global shift. This global phase does not affect the correlation function. We define an error measure for the 4D correlation function,

$$err = \frac{\iint |\Gamma(r_1, r_2) - \Gamma_o(r_1, r_2)|^2 \mathrm{d}r_1 \mathrm{d}r_2}{\iint |\Gamma(r_1, r_2)|^2 \mathrm{d}r_1 \mathrm{d}r_2 + \iint |\Gamma_o(r_1, r_2)|^2 \mathrm{d}r_1 \mathrm{d}r_2} \tag{4.1}$$

where $r = (x, y)$ at a given transverse plane, $\Gamma(r_1, r_2)$ and $\Gamma_o(r_1, r_2)$ denote the recovered and the original correlation function at a defined transverse plane. The error measure is between 0 and 1. In our case, we obtain err =1e-3 for the case in Figure 4-5. The mean phase error per pixel is 0.27 radian. That means the average peak-to-valley phase error is about 10%. Figure 4-7 visualizes the phase error of the recovered correlation function at centroid position of the two modes at z_0.

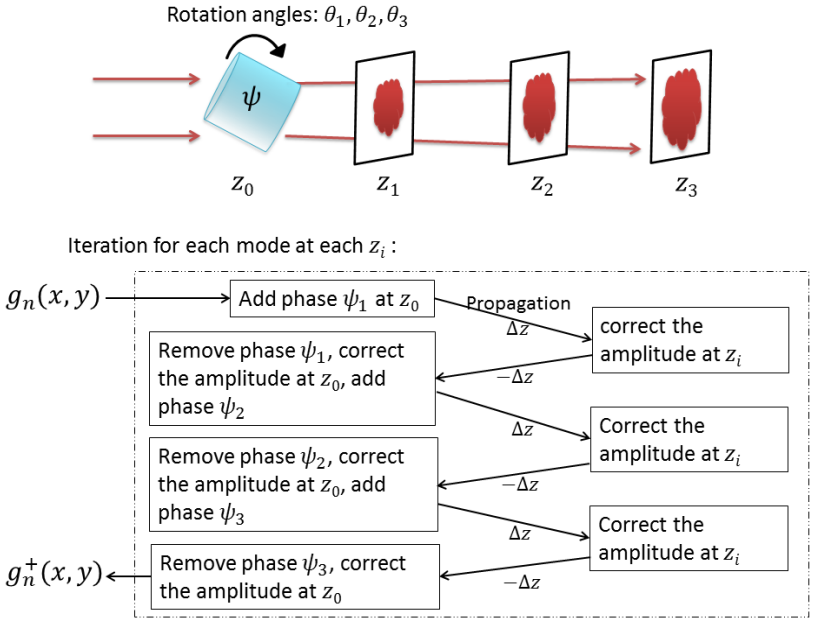
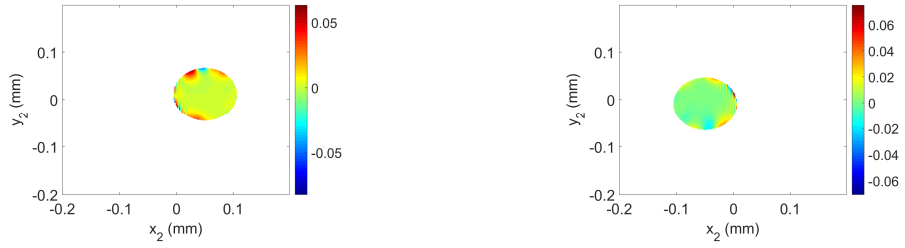


Figure 4-6. Iterative algorithm for partial coherent light with two transverse dimensions.



(a) Phase error of $\Gamma(r_1, r_2)$ @ r_1=(0.05,0.01) (b) Phase error of $\Gamma(r_1, r_2)$@ r_1=(-0.05,-0.01)

Figure 4-7. phase error of the recovered correlation function.

We draw the following conclusion on the coherence recovery using the iterative phase retrieval algorithms. First, it allows fewer intensity measurements, compared to the other method in the next section. Second, we can adapt the phase retrieval algorithm with one transverse dimension for a rotationally symmetric beam. In that case, we need to change the propagation operator based on Fourier transforms to another operator suitable for rotationally symmetric fields.

However, there are disadvantages of this approach. First, this iterative method yields a non-convex optimization[60]. In short, a convex optimization means that the iterations always converge to a unique solution, and this solution is the global minimum. To the contrary, the Gerchberg-Saxton algorithm may return a solution that is stuck at a local minimum. We may combine other algorithms, *e.g.* hybrid-input-output, to obtain a faster convergence. However, these algorithms are optimized for coherent light. For partially coherent light, it brings in the following issue. The constraint in each iteration is to correct the full-beam intensity. When multiple modes are overlaid spatially, it is necessary to differentiate their intensities by changing the parameters of the system (*e.g.* rotating asymmetric elements, varying z distance). We then raise the following questions. What kind of asymmetric element, in which rotation angle, brings the best azimuthal detection? Instead of inserting one asymmetric element, can we insert several elements to accelerate the convergence? These parameters are uncertain if we do not have any initial assumption of the light source. These questions lead us to consider the phase space tomography for coherence retrieval.

4.2 Investigation into phase space tomography

To perform phase space tomography, we need the inverse Radon transform. There are two algorithms for the inverse Radon transform, *i.e.* the filtered back projection and the ambiguity reconstruction. Although the computational effort of the filtered back projection is more demanding, it provides a higher numerical accuracy for the recovered Wigner function. We first present the sampling limit of the ambiguity function reconstruction, and later propose an improved algorithm for the filtered back projection in a 4D space.

Sampling limit in the ambiguity function reconstruction

To reconstruct the ambiguity function, it is inevitable to employ a regridding interpolation. That limits the accuracy of the recovered Wigner function. We use a 2D ambiguity function $Z(q_x, q_u)$ as an example to visualize this limit.

In the 2D case, every measured intensity fills one radial line of the ambiguity function $Z(q_x, q_u)$. That produces the blue points on a polar grid in Figure 4-8. Our goal is to obtain the Wigner function, *i.e.* a Fourier transform of the ambiguity function. We thus need to convert the ambiguity function from a polar grid to a Cartesian grid (*i.e.* the red points in Figure 4-8). The grid conversion requires that, every Cartesian point has at least one polar point in the

neighborhood. However, a polar grid has a dense sampling near the origin, and a sparse sampling near the border. Consequently, the largest distance between two polar points defines the equidistance of the Cartesian points. As shown in Figure 4-8, the grid conversion easily suffers from an unnecessary oversampling near the origin, and an under-sampling near the border. A direct interpolation in the ambiguity function modifies the Fourier space of the Wigner function. Hence, the interpolation scheme critically influences the numerical accuracy of the recovered Wigner function.

To circumvent this issue, in the following discussion, we propose an improved algorithm for the filtered back projection, to reach a high numerical accuracy. It mitigates the problem of interpolating the ambiguity function with re-gridding. Albeit at a greater computational cost, we perform the interpolation in the Wigner function space, where the interpolation artefacts have little influence on the outcome.

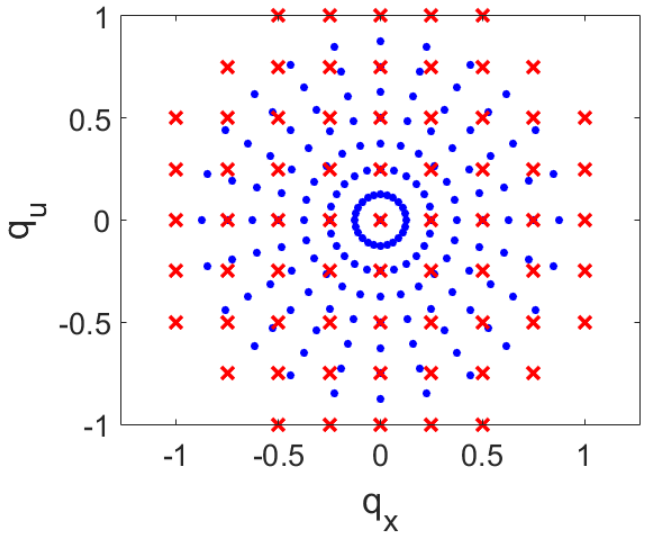


Figure 4-8 Ambiguity function from a polar sampling to a Cartesian sampling.

2D filtered back projection

To recover the 4D Wigner function, we require a 4D filtered back projection. A 4D filtered back projection is based on the 2D case. We first present our algorithm of a 2D filtered back projection. Later on, we propose the algorithm in a 4D space.

In our simulation of the 2D filtered back projection, we define an original Wigner distribution $W(x,u)$ in Figure 4-9a. We project $W(x,u)$ along an angle γ to obtain the projected distribution $S(\mu,\gamma)$. The projection is processed by a rotation of $W(x,u)$ followed by an integration along the μ axis. Each rotation is performed by three shearings[56],

$$\begin{bmatrix}\cos\gamma & -\sin\gamma\\ \sin\gamma & \cos\gamma\end{bmatrix}=\begin{bmatrix}1 & \alpha\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ \beta & 1\end{bmatrix}\begin{bmatrix}1 & \alpha\\ 0 & 1\end{bmatrix}, \tag{4.2}$$

where $\alpha=-\tan(\gamma/2)$ and $\beta=\sin\gamma$ denoting the maximum shearing distances. We implement each shearing by two Fourier transforms. For the horizontal shearing with a distance α, each row of $W(x,u)$ is shifted by a distance x',

$$W(x+x',u)=\boldsymbol{FT}^{-1}{}_{k_x}\{\exp[-i2\pi\alpha u k_x]\cdot\boldsymbol{FT}_x\{W(x,u)\}\}, \tag{4.3}$$

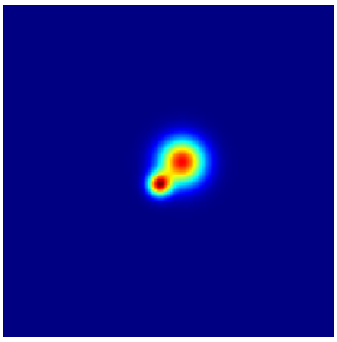
where k_x is the spatial frequency along x direction, $\boldsymbol{FT}_x$ denotes the Fourier transform with respect to x, $\boldsymbol{FT}^{-1}{}_{k_x}$ denotes the inverse Fourier transform with respect to k_x.

For the vertical shearing with a maximum distance β, each column of $W(x,u)$ is shifted by a distance u',

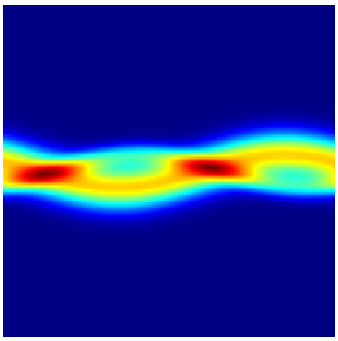
$$W(x,u+u')=\boldsymbol{FT}^{-1}{}_{k_u}\{\exp[-i2\pi\beta x k_u]\cdot\boldsymbol{FT}_u\{W(x,u)\}\}, \tag{4.4}$$

where k_u is the Fourier conjugate of u, related to the coherence length in x direction.

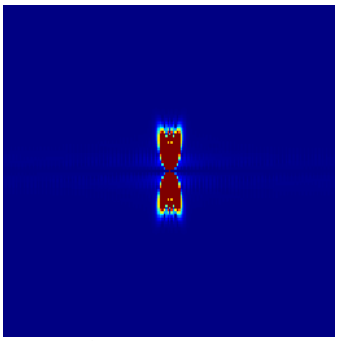
The sinogram $S(\mu,\gamma)$ is illustrated in Figure 4-9b. The γ value varies from 0 to 2π.



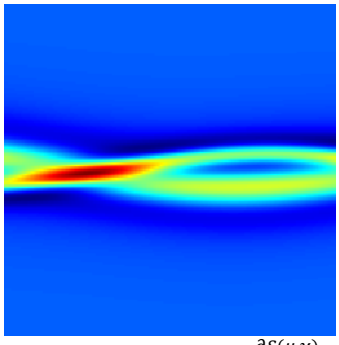
(a) Original distribution $W(x,u)$, horizontal axis being x, vertical axis being u.



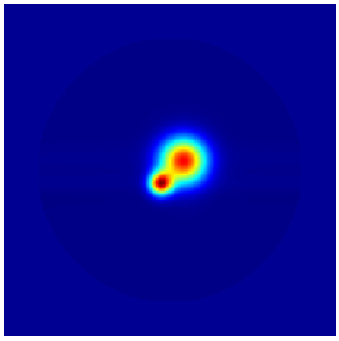
(b) Sinogram $S(\mu,\gamma)$, horizontal axis denoting γ from 0 to 2π, vertical axis being μ.



(c) $\tilde{S}(k_\mu,k_\gamma)$, *i.e.* the 2D Fourier transform of (b), horizontal axis being k_γ, vertical axis being k_μ.



(d) Distribution of $\frac{\partial S(\mu,\gamma)}{\partial\mu}$, horizontal axis being γ, vertical axis being μ.



(e) Recovered $W(x,u)$ by back projecting $\frac{\partial S(\mu,\gamma)}{\partial\mu}$ along each corresponding γ.

Figure 4-9. 2D inverse Radon transform.

To avoid unnecessary zero padding for each shearing, one can split the angle range of γ from 0 to π into four parts. When γ is between 0 and $\pi/4$, we directly apply Equations (4.3) and (4.4). With $\gamma \in (\pi/4, \pi/2)$, we first rotate $W(x,u)$ clockwise by 90 degree using matrix transpose and flipping. Then we employ the three shearings to obtain an additional rotation by a negative angle $\gamma_{new} = \pi/2 - \gamma$. With $\gamma \in (\pi/2, 3\pi/4)$, the procedure is similar. After a direct 90 degree rotation of $W(x,u)$, we perform another rotation with a positive angle $\gamma_{new} = \gamma - \pi/2$. With $\gamma \in (3\pi/4, \pi)$, we flip the matrix of $W(x,u)$ twice, up to down, left to right, to obtain a 180 degree rotation. Afterwards, an additional rotation with a positive angle of $\gamma_{new} = \pi - \gamma$ is processed by three shearings. The sinogram with $\gamma \in (\pi, 2\pi)$ is a repetition of the sinogram with $\gamma \in (0, \pi)$.

Figure 4-9c is the 2D Fourier transform of $S(\mu,\gamma)$, defined as $\tilde{S}(k_\mu, k_\gamma)$. The support of $\tilde{S}(k_\mu, k_\gamma)$ implies the sampling of $S(\mu,\gamma)$ along the γ and t axes. When the sampling density is sufficient, $\tilde{S}(k_\mu, k_\gamma)$ exhibits the shape of a bow tie[61], with an infinite extension along the vertical axis k_μ and zeros along the horizontal axis k_γ.

Before the back projection, we apply a filter to $S(\mu,\gamma)$. This filter is realized by taking a derivative of $S(\mu,\gamma)$ with respect to the t axis. The derivative is computed by

$$\partial S(\mu,\gamma_0)/\partial\mu = \boldsymbol{FT}_{k_\mu}^{-1}\{|k_\mu| \cdot \boldsymbol{FT}_\mu[S(\mu,\gamma_0)]\} \tag{4.5}$$

for each column of $\gamma = \gamma_0$. We then back project each 1D distribution of $\partial S(\mu,\gamma_0)/\partial\mu$ with an angle γ_0 to recover $W(x,u)$ (Figure 4-9e). The recovered $W(x,u)$ has a peak-to-valley error of less than 0.5%.

The above 2D back projection is applied as a sub-operator for the later 4D reconstruction. The benefit of this operator is that, the rotations and the derivatives are processed in the Fourier space. We thus circumvent any regridding interpolation schemes in the ambiguity function. That significantly improves the computational accuracy.

4D matrix rotation

We have introduced in Section 2.4.2 that a fractional Fourier transform corresponds to 4D phase space rotations. Therefore, a first-order system performing fractional Fourier transforms allows us to reconstruct a 4D Wigner function from the filtered back projections in a 4D space. Before discussing the detail of that, we first present the algorithm of a 4D matrix rotation. It helps to understand the structure of a 4D Wigner function.

To represent a fractional Fourier transform $\begin{bmatrix} \exp(i\gamma_x) & 0 \\ 0 & \exp(i\gamma_y) \end{bmatrix}$ in a 4D Wigner function $W(x, y, u, v)$, we perform a rotation at the xu plane by an angle γ_x, and at the yv plane by an angle γ_y. The overall rotation is performed by three shearings. The first shearing is along the x and y axes with $\alpha_x = -\tan(\gamma_x/2)$ and $\alpha_y = -\tan(\gamma_y/2)$ respectively,

$$W(x + x', y + y', u, v) = \boldsymbol{FT}^{-1}{}_{k_x,k_y}\{\exp[-i2\pi(\alpha_x u k_x + \alpha_y v k_y)] \cdot \boldsymbol{FT}_{x,y}[W(x, y, u, v)]\} \quad (4.6)$$

where k_x and k_y are spatial frequencies along the x and y directions respectively.

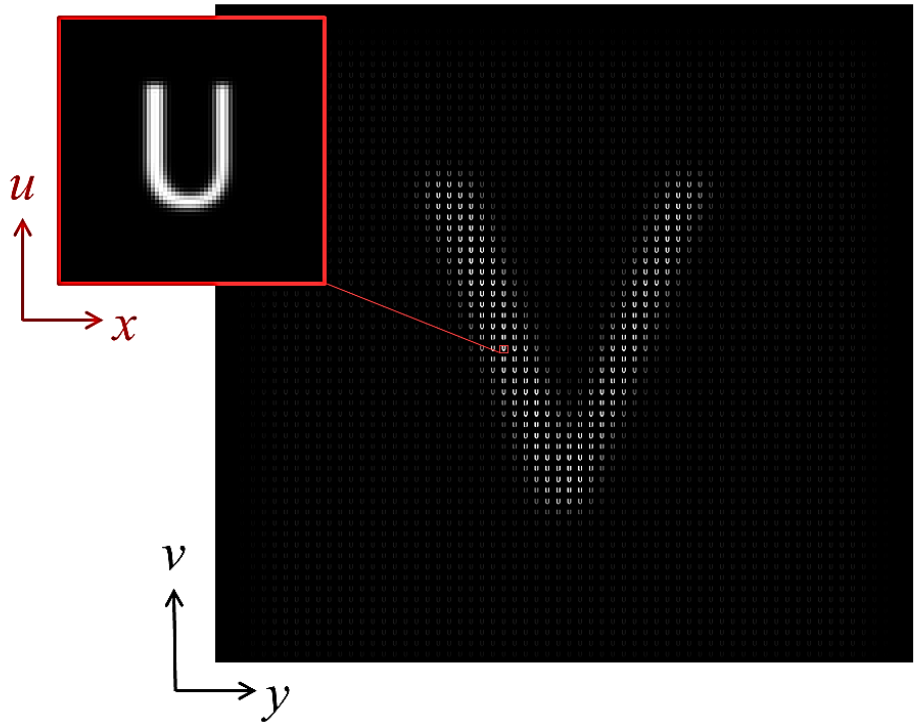
The second shearing is along the u and v axes with $\beta_x = \sin\gamma_x$ and $\beta_y = \sin\gamma_y$ respectively,

$$W(x, y, u + u', v + v') = \boldsymbol{FT}^{-1}{}_{k_u,k_v}\{\exp[-i2\pi(\beta_x x k_u + \beta_y y k_v)] \cdot \boldsymbol{FT}_{u,v}[W(x, y, u, v)]\} \quad (4.7)$$

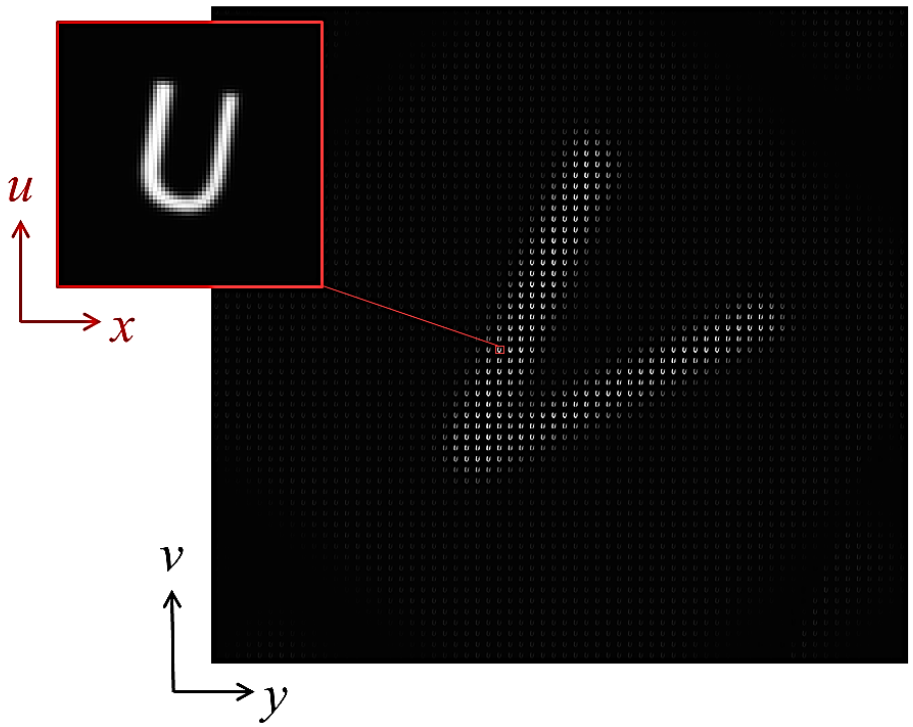
where k_u and k_v are the Fourier conjugates of u and v, meaning the correlation lengths along x and y directions respectively.

The third shearing is again along the x and y axes with $\alpha_x = -\tan(\gamma_x/2)$ and $\alpha_y = -\tan(\gamma_y/2)$, namely, a repeat of Equation (4.6).

A simulated example is shown in Figure 4-10. The 4D data set is displayed as a matrix on a 2D plane, with totally 64^4 pixels. We fill a symbol "U" at all the xu planes, and a symbol "V" at all the yv planes. Inside each point at the yv plane, there are sub-coordinates denoted by the u and x axes. We then perform a 4D rotation of this matrix. All the xu planes are rotated by 10 degree, and all the yv planes are rotated by 45 degree. The rotation is visualized by the letters of "U" and "V" in Figure 4-10b. Such a computation based on Equations (4.6) and (4.7) takes 9.8 seconds (Intel i7, 16GB RAM).



(a) 4D matrix before the rotation



(b) 4D matrix after the rotation

Figure 4-10. Visualization of the 4D matrix before and after the rotation. The xu planes are rotated by 10 degree, while the yv planes are rotated by 45 degree.

4D Wigner function reconstruction based on the filtered back projection

With the rotation algorithm from the previous section, the procedure of the 4D Wigner function reconstruction is as follows. The parameters of the fraction Fourier transform system define an ABCD matrix. This ABCD matrix provides the rotation angles γ_x and γ_y for the 4D Wigner function. Every time we change the system parameters, a new set of γ_x and γ_y are defined. The intensity captured by the camera at the output plane of the system is a 2D projection of the rotated 4D Wigner function. Therefore, all measured intensities contain four degrees of freedom, *i.e.* $(x_o, y_o, \gamma_x, \gamma_y)$, where *x* and *y* are the spatial axes on the camera plane. We then sort the measurements into a 4D data set.

We first fix an angle γ_x and a spatial coordinate *x*. From the measured 4D data, we then extract a 2D distribution with the axes of y_o and γ_y, denoted as $S_{yv}(\gamma_y, y_o)$. We consider $S_{yv}(\gamma_y, y_o)$ as a 2D sinogram with the angular coordinate γ_y and the spatial coordinate *y*. For every value of γ_y, the signals at the yv planes of the Wigner function are rotated with a specific angle. Note that, during the acquisition, the system parameters must be chosen to cover a π interval of γ_y with equidistant steps. Now we can apply the 2D filtered back projection. The signal reconstructed from the 2D sinogram $S_{yv}(\gamma_y, y_o)$ is denoted as $W_p(y, v)$, where the subscript "p" stands for "projection". It is an integration of the 4D Wigner function along the *u* axis (Figure 4-11).

For each parameter set of (γ_x, x_o), we repeat the above procedure to obtain $W_p(y, v)$. It yields a 4D matrix of $W_p(x_o, \gamma_x, y, v)$. Now we fix the coordinates of (y, v) in this 4D matrix, and define the remaining 2D distribution with the axes of x and γ_x as $S_{xu}(\gamma_x, x_o)$. We again look at $S_{xu}(\gamma_x, x_o)$ as a 2D sinogram with the angular coordinate γ_x and the spatial coordinate *x.* This sinogram is the projection of the Wigner function at the *xu* plane along the individual angles of γ_x. We perform 2D filtered back projection with each $S_{xu}(\gamma_x, x_o)$ in given coordinates of (y, v). The 4D Wigner function is eventually recovered. A summary of this algorithm is briefly illustrated in Figure 4-12.

Our proposed algorithm does not require any regridding interpolation in the space of $S_{yv}(\gamma_y, y_o)$. That improves the accuracy of the recovered Wigner function. However, this comes at the cost that during the back projection, a 4D data set needs to be filled with copies of a 2D data set. This is a more effort-demanding operation than interpolating a 2D space.

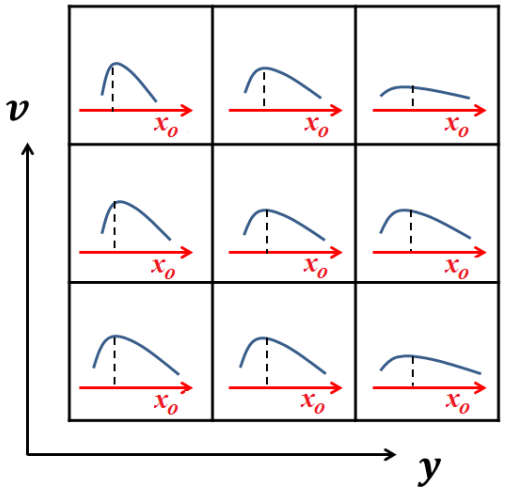


Figure 4-11. Visualization of a 3D projection of the 4D Wigner function, *i.e.* $W_p(x_o, \gamma_x, y, v)$ in Figure 4-12 with a fixed γ_x.

```
% The measured intensities based on two rotation angles form a 4D data set: S (x, y, γx, γy )

for x = 1 : nx  % count of the x pixel
  for γx = 1 :  nγx  % count of the variable γx
    Syv ( yo , γy ) = S (xo, yo, γx, γy ) ;
                % red symbols should be replaced by a colon in programing, here only for
                clarification.
                        % For fixed x and γx , we obtain the sinogram Syv with two axes of y and γy.
    Wyv ( y , v ) = filtered_back_projection_2d ( Syv ( : , :) , γy ) ;
                        % Use the algorithm of the 2D filtered back projection.
                        % Wyv is the distribution recovered from the sinogram Syv
    Wp (xo, y, γx, v) = Wyv (: , :) ;  % Fill the 4D matrix Wy with the recovered yv planes.
  end
end

for y = 1 : ny  % count of the y pixel
  for γy = 1 :  nγy  % count of the variable γy
    Sxu ( xo, γx ) = Wp (xo, y, γx, v) ;
                % red symbols should be replaced by a colon in programing, here only for
                clarification.
                        % For fixed y and v, we obtain the sinogram Sxu with two axes of x and γx.
    Wxu (x , u) = filtered_back_projection_2d ( Sxu ( : , : ) , γx ) ;
                        % Use the algorithm of the 2D filtered back projection.
                        % Wxu is the distribution recovered from the sinogram Sxu
    W(x, y, u, v) = Wxu (: , :) ;  % Fill the 4D matrix of Wigner function with the recovered uv planes.
  end
end
```

Figure 4-12. Algorithm to recover the 4D Wigner function by filtered back projections.

We draw the following conclusion of this chapter about the coherence retrieval. The mode recovery method with the iterative phase retrieval algorithm is suitable for a beam with a high degree of coherence. Compared to the phase space tomography in a 4D space, it requires less measurements and only 3D data processing (or 4D, if we index the modes with two numbers. But the mode numbers, in many cases, are much smaller than the typical sampling numbers of Cartesian grids). With one transverse dimension, this algorithm converges to a unique correlation function. This approach can be extended to a rotationally symmetric. With two transverse dimensions, this method is valid for a beam with spatially separable modes. When a large number of modes are spatially overlapped along propagation, we need a more sensitive detection of the azimuthal angles. This criterion is hard to fulfill by the mode recovery method. To overcome the limit of the mode recovery method, we investigate the phase space tomography. The main advantage of this approach is that it is suitable for a beam with any degree of coherence. Moreover, this is a non-iterative problem. The reconstruction returns a unique coherence. In addition, this method requires specific variations of the system parameters. That provides a complete azimuthal detection. Given prior knowledge about the support of the Wigner function, sampling requirements in the spatial (x, y) and angular (γ_x, γ_y) dimensions can be deduced. The reconstruction procedure can be performed in the ambiguity function with a direct point-to-point remapping. That involves regridding interpolation schemes. Our proposed algorithm performs the filtered back projection with indirect interpolations in the Fourier domain. Thus the numerical accuracy is improved.

5 Conclusions

We investigate the interaction of partially coherent light with optical systems, using the Wigner function and modal expansions.

Our first modeling tool for partial coherence is the Wigner function. We use the thin element approximation in the Wigner function to propagate partially coherent light through discontinuous surfaces. To gain an understanding of the Wigner function, we present and discuss numerous examples of optical systems. Our discussion classifies the surfaces into two types, those with a discontinuity in space, and those with a discontinuity in slope. For surfaces with a discontinuity in space, we present the phase step, grating, and the kinoform lens as examples. Interference at a phase step introduces negative values in the Wigner function. These negative values are always surrounded by oscillations. Using the Wigner function, we vividly illustrate how a periodic structure of phase steps leads to the well-known diffraction orders of gratings. We extend this argument to gratings with locally varying periods. That gives us an opportunity to discuss multiple foci of a kinoform lens generated by diffraction orders. The other type of surfaces contains discontinuous slopes. We discuss the axicon and lens array as examples. In the Wigner function, a change of the surface slope splits light into different angles and generates diffraction ripples. Based on the Wigner function of lenses arrays, we describe how the diffraction effects influence the light homogenization and show how a field lens eliminates the artifacts.

Moreover, we improve the propagation algorithms for the Wigner function. First, we employ a wavefront removal approach, which is known for coherent light, for the partially coherent light. That vastly reduces the sampling density. Meanwhile, the physical accuracy, especially the detail of the diffraction effects, is maintained. Second, we propose a fast implementation for the free-space propagation. In the paraxial regime, a free-space propagation is expressed as a shearing of the Wigner function. We implement shearing using Fourier transforms. This performs an artefact-free interpolation, giving numerically accurate results. We develop a similarly accurate method to implement rotation of the Wigner function. This operation is equivalent to a Fractional Fourier transform of the light fields, thus related to free-space propagation. In particular, a rotation can be expressed by three shearing operations. Benefiting

from the Fourier-based shearing, a rotation for any propagation distance can be performed within constant memory.

Our second modeling approach for partial coherence is the modal expansion. For a beam with a high degree of coherence or additional symmetry, it may be beneficial to represent the full beam by a few modes, instead of using a 4D Wigner function. We take Schell beams as examples and compare different expansion methods. For any form of a coherence function, a numerical solution of the eigenmodes can be derived by the singular value decomposition. Due to orthogonality, a higher-order eigenmode exhibits faster oscillation. The sampling thus requires higher density. Another expansion approach is the shifted elementary mode. It only applies to a beam that fulfills a Schell type in the far field. It can be beneficial for the Gaussian-Schell and flat-top Schell beams because the elementary modes are coherent Gaussian beams. The propagation operator is then simplified into independently solving multiple Gaussian beams. The Gaussian beam waist and the number of modes are determined by the degree of coherence. These parameters are directly indicated by the phase space of a Schell beam. Based on the modal properties, we develop an expansion tool to propagate partially coherent light inside waveguides. We expand a partially coherent beam into multiple modes. Each mode is decomposed into the guided modes of the waveguide. By that, we can efficiently propagate the light fields to any distance inside the waveguide. Also, the diffraction effects are accurately modeled. To demonstrate our expansion tool, we simulate waveguides with the step index and parabolic index profiles as examples.

Our modeling methods form the infrastructure for the experimental measurement of coherence. That equals to, (case A) solving an inverse problem of modal expansion, or (B) solving an inverse problem of the Wigner function (i.e. the phase space tomography). The common goal is to recover the coherence function from intensity measurements. In the mode recovery (case A), we retrieve the complex fields of individual modes by using the iterative phase retrieval algorithms. We extend this approach to a beam with an arbitrary wavefront. The solution converges fast in the one transverse dimensional case. That is because the propagated intensities in the entire space uniquely determine the coherence function. In the case with two transverse dimensions, an additional degree of freedom is necessary for the system. We employ a rotating cylindrical lens to realize the azimuthal detection. Our algorithm provides a convergent solution for two modes that are spatially separable in propagation. There are limitations of this algorithm.

First, the Gerchberg-Saxton algorithm is a non-convex problem. That means the optimization may get stuck easily at a local minimum instead of a further convergence. Second, when a beam contains a large amount modes that are overlapped spatially, we require a more careful detection on the azimuthal angles. However, this iterative approach does not provide a clear solution to determine the system parameters. This uncertainty of parameters is tackled in the phase space tomography (case B). To fully recover the 4D Wigner function, a certain variation of the system parameters needs to be fulfilled.

The systems parameters are related to the ABCD matrix of the first-order system. In order to fully reconstruct a 4D Wigner function, we need an astigmatic system to produce a varying ABCD matrix. With an Iwasawa decomposition, the ABCD matrix parameters can be chosen to cover a π interval of the phase space rotations. This approach offers us the insight, how to vary the system parameters in order to gain enough information for coherence reconstruction. The reconstruction is an inverse Radon transform. We compare the filtered back projection and the ambiguity function reconstruction. The latter requires a direct interpolation. To increase the numerical accuracy, we propose an improved algorithm for the filtered back projection. All the interpolations are performed indirectly in the Fourier space. Although this approach requires more computational memory compared to the ambiguity function reconstruction, it guarantees the numerical accuracy.

The Wigner function and the model expansion offer different views to identify a partially coherent beam. A proper choice among the modeling methods depends on the prior assumption of the tested beam.

This page is left blank intentionally.

References

1 Mehta, S. & Sheppard, C. Using the phase-space imager to analyze partially coherent imaging systems: bright-field, phase contrast, differential interference contrast, differential phase contrast, and spiral phase contrast. *Journal of Modern Optics* **57**, 718-739 (2010).

2 Bartels, R. *et al.* Generation of spatially coherent light at extreme ultraviolet wavelengths. *Science* **297**, 376-378 (2002).

3 Oh, C., Cao, Z., Tangdiongga, E. & Koonen, T. Free-space transmission with passive 2D beam steering for multi-gigabit-per-second per-beam indoor optical wireless networks. *Optics express* **24**, 19211-19227 (2016).

4 Mehta, S. & Sheppard, C. Partially coherent image formation in differential interference contrast (DIC) microscope. *Optics express* **16**, 19462-19479 (2008).

5 Peeters, M. *et al.* Spatial decoherence of pulsed broad-area vertical-cavity surface-emitting lasers. *Optics express* **13**, 9337-9345 (2005).

6 Wolf, E. New theory of partial coherence in the space-frequency domain. Part I: spectra and cross spectra of steady-state sources. *Journal of Optical Society of America, A* **72**, 343-351 (1982).

7 Wolf, E. New theory of partial coherence in the space-frequency domain. Part II: Steady-state fields and higher-order correlations. *Journal of Optical Society of America, A* **3**, 76-85 (1986).

8 Goodman, J. W. *Introduction to fourier optics*. (Roberts and Company Publishers, 2005).

9 Wigner, E. P. On the quantum correction for thermodynamic-equilibrium. *Physical Rev* **40**, 749 (1932).

10 Dolin, L. S. Beam description of weakly-inhomogeneous wave fields. *Izv. Vyssh. Uchebn. Zaved. Radiofiz* **7**, 559-563 (1964).

11 Walther, A. Radiometry and Coherence. *J Opt Soc Am* **58**, 1256 (1968).

12 Walther, A. Propagation of the Generalized Radiance through Lenses. *J Opt Soc Am* **68**, 1606-1610, doi:Doi 10.1364/Josa.68.001606 (1978).

13 Testorf, M., Hennelly, B. & Ojeda-Castaneda, J. *Phase Space Optics*. (Mc Graw-Hill, 2010).

14 Zhong, M. & Gross, H. Phase space of partially coherent light with discontinuous surfaces. *Journal of the European Optical Society-Rapid Publications* **12**, 1-8 (2016).

15 Wolf, E. *Introduction to the Theory of Coherence and Polarization of Light*. (Cambridge University Press, 2007).

16 Mandel, L. & Wolf, E. *Optical Coherence and Quantum Optics*. (Cambridge University Press, 1995).

17 Friberg, A. T. & Sudol, R. J. Propagation parameters of Gaussian Schell-model beams. *Optics Communications* **41**, 383-387 (1982).

18 Vahimaa, P. & Turunen, J. Finite-elementary-source model for partially coherent radiation. *Optics express* **14**, 1376- 1381 (2006).

19 Schell, A. A technique for the determination of the radiation pattern of a partially coherent aperture. *IEEE Transactions on Antennas and Propagation* **15**, 187-188 (1967).

20 Korotkova, O. Random sources for rectangular far fields. *Opt Lett* **39**, 64-67 (2014).

21 Bastiaans, M. J. The Wigner Distribution Function of Partially Coherent-Light. *Opt Acta* **28**, 1215-1224 (1981).

22 Scarpazza, D. A brief introduction to the wigner distribution. *Dipartimento di Elettronica e Informazione, Politecnico di Milano* (2003).

23 Eppich, B. Die Charakterisierung von Strahlungsfeldern mit der Wigner-Verteilung und deren Messung. *Technischen Universität Berlin* **Doctoral thesis** (1998).

24 K. Wolf, Alonso, M. & Forbes, G. Wigner functions for Helmholtz wave fields. *Journal of Optical Society of America, A* **16**, 2476- 2487 (1999).

25 Gross, H., Singer, W. & Totzeck, M. *Handbook of Optical Systems Vol. 2: Physical Image Formation*. Vol. 2 (Wiley-VCH, 2005).

26 Goldberg, R. R. *Fourier transforms*. Vol. 32 (Cambridge London, 1961).

27 Riesz, F. & Szokefalvi-Nagy, B. Functional Analysis Frederick Ungar Pub. *Co., New York* (1955).

28 Gori, F. Collett-wolf sources and multimode lasers. *Optical Communications* **34**, 301-305 (1980).

29 Wolf, A. S. a. E. Coherent mode representation of Gaussian Schell-model sources and of their radiation fields. *J. Opt. Soc. Am.* **72**, 923-928 (1982).

30 Gori, F., Guattari, G. & Padovani, C. Modal expansion for J0-correlated Schell-model sources. *Optics communications* **64**, 311-316 (1987).

31 Palma, C., Borghi, R. & Cincotti, G. Beams originated by J0-correlated Schell-model planar sources. *Optics communications* **125**, 113-121 (1996).

32 Ozaktas, H., Yüksel, S. & Kutay, M. Linear algebraic theory of partial coherence: discrete fields and measures of partial coherence. *JOSA A* **19**, 1563-1571 (2002).

33 Ozaktas, H., Gulcu, T. & Kutay, M. Linear algebraic theory of partial coherence: continuous fields and measures of partial coherence. *JOSA A* **33**, 2115-2124 (2016).

34 Tervonen, E., Turunen, J. & Friberg, A. T. Transverse laser-mode structure determination from spatial coherence measurements: experimental results. *Applied Physics B* **49**, 409-414 (1989).

35 Schäfer, B. & Mann, K. Determination of beam parameters and coherence properties of laser radiation by use of an extended Hartmann-Shack wave-front sensor. *Applied optics* **41**, 2809-2817 (2002).

36 Stoklasa, B., Motka, L., Rehacek, J., Hradil, Z. & Sánchez-Soto, L. Wavefront sensing reveals optical coherence. *Nature communications* **5** (2014).

37 Rydberg, C. & Bengtsson, J. Numerical algorithm for the retrieval of spatial coherence properties of partially coherent beams from transverse intensity measurements. *Optics express* **15**, 13613 (2007).

38 Rodrigo, J. A. & Alieva, T. Recovery of Schell-model partially coherent beams. *Opt Lett* **39**, 1030 (2014).

39 Raymer, M., Beck, M. & McAlister, D. Complex wave-field reconstruction using phase-space tomography. *Physical Review Letters* **72**, 1137-1140 (1994).

40 D. F. McAlister, M. B., L. Clarke, A. Mayer, and M. G. Raymer. Optical phase retrieval by phase space tomography and fractional order fourier transforms. *Opt Lett* **20**, 1811-1813 (1995).

41 A. Cámara, T. A., J. A. Rodrigo, M. L. Calvo. Phase space tomography reconstruction of the Wigner distribution for optical beams separable in Cartesian coordinates. *Journal of Optical Society of America, A* **26**, 1301-1306 (2009).

42 Luke, D. R., Burke, J. V. & Lyon, R. G. Optical wavefront reconstruction: Theory and numerical methods. *SIAM review* **44**, 169-224 (2002).

43 Gori, F., Guattari, G. & Santarsiero, M. Coherence and the spatial distribution of intensity. *JOSA A* **10**, 673-679 (1993).

44 Dragoman, D. Unambiguous coherence retrieval from intensity measurements. *JOSA A* **20**, 290-295 (2003).

45 Simon, R. & Mukunda, N. Twisted Gaussian Schell-model beams. *JOSA A* **10**, 95-109 (1993).

46 Schäfer, B. & Mann, K. Characterization of an ArF excimer laser beam from measurements of the Wigner distribution function. *New Journal of Physics* **13**, 043013 (2011).

47 Eppich, B., Mann, G. & Weber, H. in *Optical Systems Design 2005* (International Society for Optics and Photonics, 2005).

48 Wolf, K. B. *Geometric optics on phase space*. (Springer Science & Business Media, 2004).

49 Cámara, A., Rodrigo, J. & Alieva, T. Optical coherenscopy based on phase-space tomography. *Optics express* **21**, 13169-13183 (2013).

50 Hlawatsch, F. Interference terms in the Wigner distribution. *Digital signal processing* **84**, 363-367 (1984).

51 Sinzinger, S. & Testorf, M. Transition between diffractive and refractive micro-optical components. *Applied optics* **34**, 5970-5976 (1995).

52 Siegman, A. *Lasers*. 805-811 (University Science Books, Mill Valey, California, 1986).

53 Pearson, K. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* **58**, 240-242 (1895).

54 Rausch, D., Herkommer, A. Phase space approach to the use of integrator rods and optical arrays in illumination systems. *Advanced Optical Technologies* **1** (2012).

55 Larkin, K. G., Oldfield, M. A. & Klemm, H. Fast Fourier method for the accurate rotation of sampled images. *Optics communications* **139**, 99-106 (1997).

56 Lohmann, A. Image rotation, Wigner rotation, and the fractional Fourier transform. *Journal of Optical Society of America, A* **10**, 2181- 2186 (1993).

57 Roelandt, S., Tervo, J., Meuret, Y., Verschaffelt, G. & Thienpont, H. Propagation of partially coherent light through a light pipe. *Optics express* **21**, 17007-17019 (2013).

58 Snyder, A. & Love, J. *Optical Waveguide Theory*. (Chapman and Hall, 1983).

59 Saleh, B. E., Teich, M. C. & Saleh, B. E. *Fundamentals of photonics*. Vol. 22 (Wiley New York, 1991).

60 Bauschke, H. H., Combettes, P. L. & Luke, D. R. Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. *JOSA A* **19**, 1334-1345 (2002).

61 Rattey, P. & Lindgren, A. Sampling the 2-D Radon transform. *IEEE transactions on acoustics, speech, and signal processing* **29**, 994-1002 (1981).

This page is left blank intentionally.

List of frequently used symbols

Symbol	Unit	Property
$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$	-	2 by 2 submatrix of the 4 by 4 ray transform matrix
A_x, A_y	-	Diagonals of the 2 by 2 matrix $\boldsymbol{A}$ in a first-order system with two transverse dimensions
$A^{(0)}$	-	Angular correlation function
a_m, b_m, c_m, d_m $(m = 1, 2, 3, \dots)$	$m, m^{-1}, -, m^2$	Parameter in the correlation function of a flat-top Schell beam
$a_0,\ b_0,\ c_0$	-	Coefficients for the eigenmodes of a Gaussian-Schell beam
B_x, B_y	-	Diagonals of the 2 by 2 matrix $\boldsymbol{B}$ in a first-order system with two transverse dimensions
C_{ns}	-	Noise to signal ratio in the spatial frequency domain in Wiener deconvolution
C_o	-	Parameter in the correlation function of a flat-top Schell beam
c	$m \cdot s^{-1}$	Speed of light
c_1, c_2, c_3, c_4	-	Measure of coherence in linear algebra
$c_{mn}, c_{mn}^{(z)}$	-	Mode coefficient in waveguide propagation
c_R	m^{-1}	Curvature of a lens surface
cr	-	Correction term for the field magnitude in the iterative phase retrieval algorithm
D_{wg}	m	Diameter of a waveguide
d_g	m	Period of a grating
$\mathrm{d}(\Delta x), \mathrm{d}(x)$	m	Spatial distance between two adjacent sampling points on the Δx and x axes respectively
E	$V \cdot m^{-1}$	Electric field with time dependency
$\tilde{E}$	$V \cdot m^{-1}$	Electric field with temporal-frequency dependency
E'	$V \cdot m^{-1}$	Transmitted electric field through a thin element
err	-	Error of retrieved correlation function compared to the original correlation function
$\boldsymbol{FT}_{x,y}$	-	Fourier transform operator with respect to x, y
$\mathbf{f}$	$V \cdot m^{-1}$	Discretely sampled light field in a matrix form
f	$V \cdot m^{-1}$	Eigenmode of a source
f_{par}	m	Focal length of a paraxial lens
Flt	-	Circular-shape low-pass filter with smooth edges

g	$V \cdot m^{-1}$	Recovered mode from the iterative phase retrieval algorithm
g^+	$V \cdot m^{-1}$	Corrected mode in the iterative phase retrieval algorithm
$\tilde{g}(k_x, k_y)$	-	Wiener filter dependent on spatial frequency k_x, k_y
H_n	-	n[th] order Hermite polynomial
$h(x)$ with $x(t)$	-	Function h of a time (t)-dependent random process x
$\boldsymbol{I}$	-	Identity matrix
I	$W \cdot m^{-2}$	Intensity of light
I_o	$W \cdot m^{-2}$	Intensity at the output plane of a first-order system
I^∞	$W \cdot m^{-2}$	Far-field intensity
$\tilde{I}$	$W \cdot m^{-2}$	Intensity in spatial frequency domain
$i(x, y)$	$W \cdot m^{-2}$	Intensity of one elementary mode in space
$\tilde{\imath}(k_x, k_y)$	$W^{-1} \cdot m^2$	Fourier transform of the spatial intensity of one elementary mode
$\boldsymbol{J}$	-	Matrix satisfying $\boldsymbol{J} = i\begin{bmatrix} 0 & -\boldsymbol{I} \\ \boldsymbol{I} & 0 \end{bmatrix}$
J	$(V \cdot m^{-1})^2$	Mutual coherence function, dependent on time delay τ
$\boldsymbol{K}, \boldsymbol{G}, \boldsymbol{X}, \boldsymbol{Y}$	-	2 by 2 submatrices from the Iwasawa decomposition
k	m^{-1}	Wave number
k_x, k_y	m^{-1}	Spatial frequency along *x* or *y* direction
k_u, k_v	m	Fourier conjugates of the angular axis in the Wigner function, meaning the correlation length along x or y direction.
L	m	Propagated distance of a convergent beam
L_{trans}	m	Transformed distance with equivalent diffraction effects after removing a parabolic wavefront from a convergent beam
l_c	m	Transverse rms correlation width of a Gaussian-Schell beam
M, N, m, n	-	Integer indicating the mode number, or the sampling number
m_g	-	Diffraction order of a grating
N_x	-	Number of sampling points on the x axis
$n_{rf}, n_0, n_{core}, n_{clad}$	-	Refractive index (*e.g.* index in core and cladding of a waveguide)
$n(x, y)$	-	Noise function due to the float points in numerical calculation.
P, P_{kino}, P_{par}	-	Phase modulation function of a thin element (*e.g.* kinoform or paraxial lens)
$p(x)$	-	Probability density function of a random process x
$\vec{q}_x = (q_x, q_y)$	m^{-1}	Spatial frequency in a 4D ambiguity function
$\vec{q}_u = (q_u, q_v)$	m	Correlation width in a 4D ambiguity function
R	m	Wavefront radius of curvature of a beam

$r = (x, y, z)$	m	Spatial coordinate containing orthogonal components x, y, z
r'	m	Shifted spatial distance of a shifted-elementary mode
$\hat{\boldsymbol{r}}$	m	Unit vector in real space along the direction of $\boldsymbol{r} = (x, y, z)$
$\boldsymbol{S}$	-	4 by 4 first-order ray transformation matrix
$S(\mu, \gamma)$	-	Sonogram from the Radon transform, μ being the spatial axis, γ being the rotation angle.
$s(x)$	-	Arbitrary square-integrable function
T	s	Time period in a random process
$\boldsymbol{T}$	-	Symplectic matrix
$\boldsymbol{T_L}, \boldsymbol{T_M}, \boldsymbol{T_O}$	-	4 by 4 matrices from the Iwasawa decomposition, T_L being a lens transformation, T_M representing a magnification, T_O being an orthogonal simplectic matrix.
t, t_0	s	Time
U, V	-	Waveguide parameter
$u(x)$	$V \cdot m^{-1}$	Eigenmode of a waveguide
u_i, u_o	$radian$	Input and output angular coordinates of the 2D Wigner function through a first-order system
u_{max}	$radian$	Maximum value on the angular axis u in a Wigner function
$\vec{u} = (u, v)$	$radian$	Angular coordinates in a 4D Wigner function
u', v'	-	Shearing distance along u or v direction in a Wigner function
Δu	$radian$	Angular extent in the Wigner function
W	$W \cdot m^{-2} \cdot radian^{-2}$	Wigner function
W_o	$W \cdot m^{-2} \cdot radian^{-2}$	Wigner function of light at the output plane of a first order system
W'	$W \cdot m^{-2} \cdot radian^{-2}$	Wigner function of the transmitted light through a thin element
$w_t(x, y)$	-	Weight function of the elementary modes in space
$w(z)$	m	Transverse radius of a Gaussian Schell beam, as a function of the propagation distance z
w_0	m	Beam waist of a Gaussian-Schell beam
w_e	m	Beam waist of an elementary mode
$\widetilde{w}_t(k_x, k_y)$	-	Fourier transform of the weight function in the Wiener deconvolution
$x, y, z,$	m	Spatial coordinate

x_{max}	m	Maximum value on the x axis
x_i, x_o	m	Input and output spatial coordinates of the 2D Wigner function through a first-order system
x_1, x_2	m	Two arbitrary spatial positions on the x axis
$\vec{x} = (x, y)$	m	Spatial coordinates in a 4D Wigner function
x', y'	m	Shearing distance along x or y direction in phase space
Δx	m	Distance between two spatial points, $\Delta x = x_1 - x_2$
$\bar{x}$	m	Normalized spatial coordinate
x^{∞}	m	Spatial coordinate in the far field
Z	$W \cdot m^{-2} \cdot radian^{-2}$	Ambiguity function
z_0	m	Defined transverse plane in propagation
$\Delta z, \Delta z_{kino}, \Delta z_{par}$	m	Step height of a discontinuous surface (*e.g.* kinoform or paraxial lens)
α, β	m	Shearing distance decomposed from a rotation matrix
β_m	-	Propagation constant for the m^{th}-order guide mode
Γ	$(V \cdot m^{-1})^2$	Mutual intensity function, or correlation function for a monochromatic source
$\Gamma^{(0)}$	$(V \cdot m^{-1})^2$	Correlation function at the source plane
$\tilde{\Gamma}^{(0)}$	$(V \cdot m^{-1})^2$	Fourier transform of the correlation function at the source plane, *i.e.* angular correlation function.
Γ^{∞}	$(V \cdot m^{-1})^2$	Correlation function in the far field
$\boldsymbol{\gamma}$	-	A normalized correlation function in a matrix form with discrete spatial coordinates
$\gamma(r_1, r_2)$	-	Degree of coherence between fields at spatial positions r_1 and r_2
γ_A	-	Angular correlation function
γ_x, γ_y	-	Phase space rotation angle on *xu* and *yv* planes
$\tilde{\gamma}$	-	Fourier transform of the degree of coherence
Δ	-	Unitless constant in the parabolic-index waveguide model
δ	m	Near-field rms transverse correlation width of a flat-top Schell beam
η	$radian/m$	Scaling factor in the rotation of the Wigner function
ϑ, ϵ	-	Phase space rotation angle on *xy* and *uv* planes
κ	-	Conic constant of a lens curvature

$\boldsymbol{\Lambda}$	-	Diagonal matrix consisting of eigenvalues, obtained from the singular value decomposition of the correlation matrix.
λ	m	Wavelength
λ_n	-	Eigenvalue of a source mode
μ	m	Spatial axis of a sinogram in 2D Radon transform
ρ	m	Radius of the optical element
σ	m	Near-field rms source width of a flat-top Schell beam
$\tau = t_1 - t_2$	s	Time delay
Φ	$V \cdot m^{-1}$	Unitary matrix whose columns containing eigenvectors, obtained from the singular value decomposition of the correlation matrix.
ϕ_n	$V \cdot m^{-1}$	n^{th} order Guided mode of a source mode
$\Delta\varphi$	-	Normalized phase difference for discontinuous surfaces
ψ	-	Phase term of aberrations
ω	s^{-1}	Temporal frequency

This page is left blank intentionally.

Acknowledgements

Our research is partially funded by the German Federal Ministry of Education and Research within the separate projects KoSimO (FKZ031PT609X) and fo+ (03WKCK1D).

The first person I want to thank is my supervisor, Herbert Gross. He is always there to support me with advice and patience when I encounter all kinds of confusion and difficulties. His rich scientific knowledge and social skills have taught me how to be a better researcher and a better person. Staying in Herbert's group in the last four years, I have experienced the most fruitful and joyful time in my studies, scientifically and socially. Besides, I am grateful to my group members, especially Ralf Hambach, Sören Schmidt and Christoph Bösel for reading and correcting my thesis. I always learned a lot from their discussions. I also thank all the other group members, whose names are not mentioned here due to the page limit. They are always warm and supportive whenever I need help.

Also, I acknowledge Klaus Mann and Bernd Schäfer for generously sharing their results of phase space tomography. I greatly appreciate the discussions about partial coherence with Paul Urbach and his group members Chengliang Zhao and Sander Konijnenberg, especially. Sander is always efficient and bright when explaining to me the delicate details of phase retrieval. Furthermore, I give thanks to Konrad Schöbel for helping me to understand the Radon transform in a 4D space, to Franz Hlawatsch for supporting me with materials about cross terms in the Winger function, to Christer Rydberg who quickly sent me his Ph.D. thesis from Singapore although he has left this field for ten years.

In particular, I am grateful to my boyfriend, Martin Kielhorn. Without his mental support and technical discussions, I could not have finished my writing on time. I express eternal gratitude to my parents for their unconditional support all the time, no matter where I am and what I pursue.

This page is left blank intentionally.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbständig, ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise ~~entgeltlich~~/unentgeltlich geholfen:

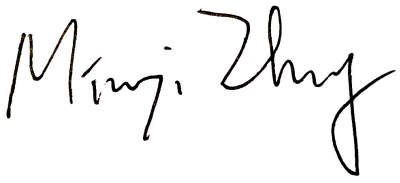
- Herbert Gross, Betreuer.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die geltende Promotionsordnung der Physikalisch-Astronomischen Fakultät ist mir bekannt.

Ich versichere ehrenwörtlich, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.



Jena, 21.12.2016

Ort, Datum Unterschrift d. Verfassers

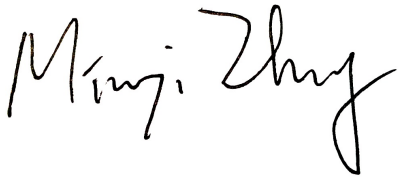
This page is left blank intentionally.

Curriculum Vitae

Zhong, Minyi

Date of Birth: 30.09.1987
Location of birth: Gaoyao, China

9.1994-7.2000	Taihe No. 1 Primary School, Guangzhou, China	
9.2000-7.2003	Guangzhou No. 115 Middle School, China	
9.2003-7.2006	Peiying Middle School, Guangzhou, China	
9.2006-7.2010	Bachelor of Science in Optical Information Science and Technology	Sun Yat-sen University, China
9.2010-7.2011	Master of Science in Applied Physics	Delft University of Technology, the Netherlands
10.2011-12.2012	Master of Science in Photonics	University of Jena, Germany
1.2013-12.2016	Ph.D. candidate	University of Jena, Germany
from 1.2017	Competence engineer	ASML, Veldhoven, the Netherlands



Jena, 21.12.2016
Location, Date

Signature

This page is left blank intentionally.