

Transport of intensity equation: a tutorial

Chao Zuo^{a,b,*}, Jiayi Li^{a,b}, Jiasong Sun^{a,b}, Yao Fan^{a,b}, Jialin Zhang^{a,b}, Linpeng Lu^{a,b},
Runnan Zhang^{a,b}, Bowen Wang^{a,b}, Lei Huang^c, Qian Chen^{b,*}

^a Smart Computational Imaging (SCI) Laboratory, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

^b Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

^c Brookhaven National Laboratory, NSLS II 50 Rutherford Drive, Upton, NY 11973-5000, United States

ARTICLE INFO

Keywords:

Transport of Intensity Equation (TIE)
Quantitative Phase Imaging (QPI)
Phase Retrieval
Partial Coherence
Optical Diffraction Tomography (ODT)

ABSTRACT

When it comes to “phase measurement” or “quantitative phase imaging”, many people will automatically connect them with “laser” and “interferometry”. Indeed, conventional quantitative phase imaging and phase measurement techniques generally rely on the superposition of two beams with a high degree of coherence: complex interferometric configurations, stringent requirements on the environmental stabilities, and associated laser speckle noise severely limit their applications in optical imaging and microscopy. On a different note, as one of the most well-known phase retrieval approaches, the transport of intensity equation (TIE) provides a new non-interferometric way to access quantitative phase information through intensity only measurement. Despite the insufficiency for interferometry, TIE is applicable under partially coherent illuminations (like the Köhler’s illumination in a conventional microscope), permitting optimum spatial resolution, higher signal-to-noise ratio, and better image quality. In this tutorial, we give an overview of the basic principle, research fields, and representative applications of TIE, focus particularly on optical imaging, metrology, and microscopy. The purpose of this tutorial is twofold. It should serve as a self-contained introduction to TIE for readers with little or no knowledge of TIE. On the other hand, it attempts to give an overview of recent developments in this field. These results highlight a new era in which strict coherence and interferometry are no longer prerequisites for quantitative phase imaging and diffraction tomography, paving the way toward new generation label-free three-dimensional microscopy, with applications in all branches of biomedicine.

1. Introduction

In physics and mathematics, the “phase” of a periodic signal is a real-valued scalar that describes the relative location of each point on the waveform within the span of each full period. The phase is typically expressed as an angle (in degrees or radians), in such a scale that it varies by one full turn as the variable goes through each period. In the branch of wave optics, the concept of phase is only limited to the monochromatic coherent optical field, which can be described in terms of a *two-dimensional* (2D) complex exponent with its argument so-called the “phase” [1,2]. In comparison to phase, the amplitude component of an optical field is generally much easier to understand or comprehend. Because the square of the amplitude, also known as the “intensity” of light (is also referred to as “irradiance” in radiometry [3]), is the only visible component we can see with our eyes, representing the energy of the light [4]. Human eyes or the imaging sensors can only detect the intensity (or amplitude) component of light but the phase information cannot be accessed directly. One important reason is that the oscillation frequencies of light waves are around 10^{14} Hz [5], which is much higher

than the response speed of human eye (usually 30Hz) or photoelectric imaging device (the frame rate of the current high-speed camera can only reach 10^8 Hz level [6]).

The importance of phase can be illustrated by a simple and interesting example in Oppenheim’s article [7] (see Fig. 1): Given two different images, if we exchange their phase components in the frequency domain while keeping their amplitude components unchanged, then Fourier transforming their spectra back to spatial domain, the general appearance of the two images will almost interchange. This simple example shows that it is not the amplitude but the phase that bears most of the information in the image (note that this example may not be very appropriate because the phase here is the phase component of the image spectrum rather than that of the optical field). The importance of phase is particularly prominent in some specific fields, such as optical metrology, material physics, adaptive optics, X-ray diffraction optics, electron microscopy, and biomedical imaging. Most samples of interest belong to phase objects with very little absorption, but the spatial distribution of their *refractive index* (RI) or thickness is nonuniform. Therefore, they have small amplitude variations but strong phase components.

* Corresponding authors.

E-mail addresses: zuochao@njust.edu.cn (C. Zuo), chenqian@njust.edu.cn (Q. Chen).

<https://doi.org/10.1016/j.optlaseng.2020.106187>

Received 23 December 2019; Received in revised form 7 April 2020; Accepted 11 May 2020

Available online 19 June 2020

0143-8166/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

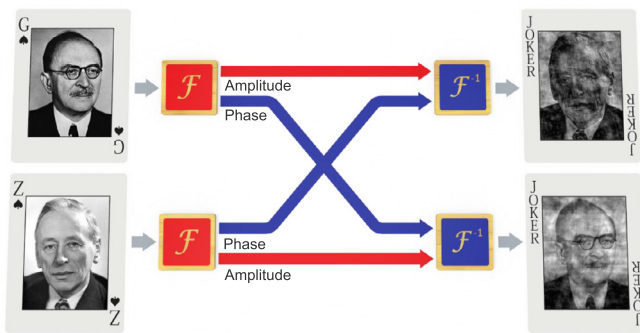


Fig. 1. A example illustrating the importance of phase of an image. A exchange of phase components in the frequency domain leads to the interchange of their general appearance.

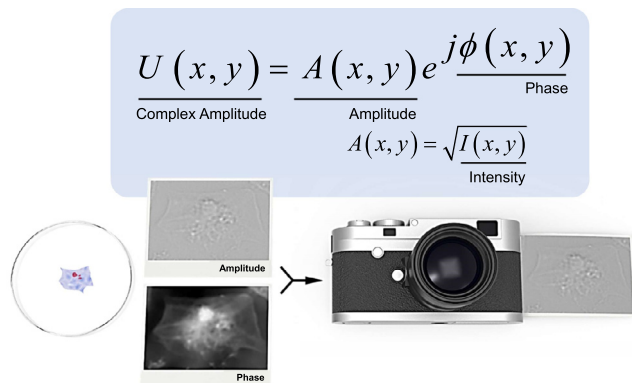


Fig. 2. The detector and the human eye are only sensitive to intensity/amplitude while phase information is lost.

Human eyes or photodetectors are only sensitive to the amplitude variation rather than the phase variation, so they cannot “see” phase object directly (Fig. 2). In other words, it is impossible to distinguish different components of the object with different thickness or RI. Thus, the acquisition of phase information is of particular importance in such cases.

Phase retrieval is a key research subject in the fields of optical metrology and imaging technologies. It plays an indispensable role in both industrial inspection and biomedical imaging applications. The phase imaging techniques, especially for biological samples and weak-absorptive semi-transparent objects, have a long history of development. The cytoplasm and most organelles of biological cells have very weak absorption, so they produce very little contrast under normal illumination in a traditional bright-field microscope. To overcome this difficulty, the most common approach is staining or labeling. Different components of the cell have various affinities to different dyes or fluorophores, producing enough intensity contrast or emitting different fluorescence spectra (Fig. 3). The most common modality utilizing exogenous contrast is fluorescence microscopy [8], in which a specimen is labeled with specific fluorescent molecules to provide targeted morphological information. The structure of the organelle and protein complex can also be visualized in *three dimension* (3D) with the help of the optical sectioning capability of laser confocal microscopy and multi-photon microscopy [9,10]. With the emergence of new fluorescent molecular probes and the improvement of optical imaging technologies, researchers have developed a variety of super-resolved fluorescence microscopic methods that exceed the resolution limit of conventional confocal microscopy, e.g., structured illumination microscopy (SIM) [11], stimulated emission depletion microscopy (STED) [12], photoactivated localization microscopy (PALM) [13], and stochastic optical

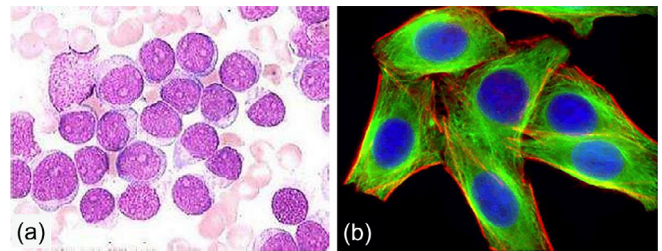


Fig. 3. Two typical cell labeling techniques. (a) chemical staining turns transparent phase objects into absorptive objects; (b) fluorescent labeling turns transparent phase objects into self-luminous objects.

reconstruction microscopy (STORM) [14]. These methods utilize the extremely high localization accuracy of single-molecule imaging or the fluorescence excitation and bleaching properties of photoactivatable fluorescent protein, significantly breaking the resolution limit of conventional optical microscopy and revealing nanoscale structures of fluorescently labeled live cells. The relative importance of fluorescence microscopy has recently been made evident by the 2014 Nobel Prize in chemistry for “the development of super-resolved fluorescence microscopy”. However, such methods still require fluorescent dyes and fluorescent proteins as biomarkers and are thus ill-suited for samples that are non-fluorescent or cannot be easily fluorescently tagged. Besides, the photobleaching and phototoxicity of the fluorescent agents prevent live cells imaging over extended periods of time [15].

Although most biological cells are transparent and do not change the amplitude of the light passing through them, they introduce phase delays due to different structural regions of nonuniform optical density (RI). In 1942, Zernike [16] invented the *Zernike phase contrast* (ZPC) microscopy [Fig. 4(a)] to visualize phase optically. The method uses a phase mask to shift only the unperturbed incident field component by a quarter of a wavelength, such that it interferes with scattered field with higher spatial frequencies, rendering the maximum contrast on the interference image. This provides a simple, efficient method for converting phase difference into intensity contrast, thus greatly improves the contrast of the transparent phase object under an optical microscope [Fig. 4(b)]. This led to a revolution in biological imaging, as the significant contrast gain enabled the observation of unstained biological cells and tissues that were nearly invisible before, and the invention earned Zernike the 1953 Nobel Prize in physics. Subsequently, Nomarski [17] invented the *Differential Interference Contrast* (DIC) microscopy based on the principle of polarization beam-splitting and shear interferometry in 1952. The phase contrast in DIC microscopy is proportional to the phase gradient of the sample along the shear direction, giving the appearance of a 3D physical relief corresponding to the variation of the optical density of a sample [Fig. 4(d)].

Although ZPC and DIC represent significant advancements in microscopy, these methods all suffer from one important drawback, that the measured intensity has a nonlinear, and thus non-invertible, relationship with the phase of the specimen. Without the phase information, morphologically relevant quantities of the sample, such as size, optical thickness, dry mass density, and RI cannot be quantified. On the other hand, “halo” [Fig. 4(b)] and “shadow off” [Fig. 4(d)] artifacts accompanied by these phase contrast methods also complicate the subsequent image analysis and processing (such as cell counting and segmentation). These limitations, when combined with the advent of digital image sensors and advances in information optics, have resulted in the emerging and burgeoning field of phase measurement and *quantitative phase imaging* (QPI) [18,19], which combines innovations in optics, imaging theory, and computational methods to image phase information of the sample quantitatively. Quantitative phase profiles of cells allow the determination of cellular structure and biophysical parameters with minimal sample manipulation. Especially in cases where conventional preparation techniques, such as fixation, staining, or fluorescent tagging, may

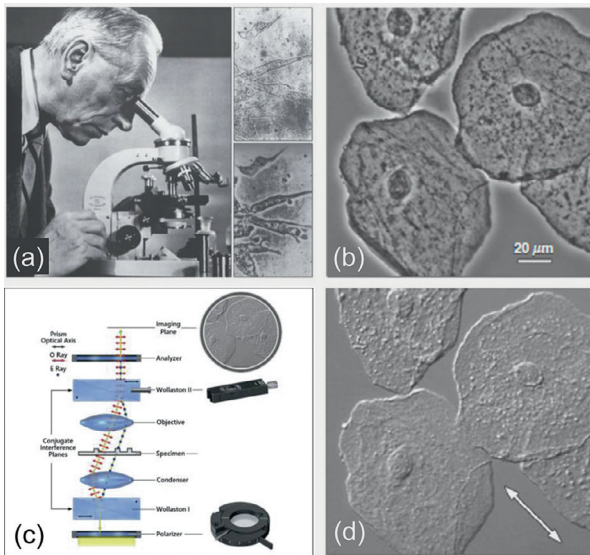


Fig. 4. Zernike phase contrast microscopy and differential interference contrast microscopy and their imaging results on unstained cheek cells. (a) Zernike phase contrast microscopy; (b) Zernike phase contrast microscopic image of unstained cheek cells; (c) DIC microscopy; (d) DIC microscopic image of unstained cheek cells.

affect cellular functions and limit biological insight, QPI offers an important alternative.

With the invention of the laser and the development of optical metrology for nearly half a century, the most well-established phase measurement method is through interferometry. Since the principle of optical coherent interference was first proven to be a potential measurement tool in the 1880s, interferometry has played a prominent role in optical metrology. Nowadays, a high-precision laser interferometer allows extremely sensitive phase measurements with the optical path measurement accuracy up to $\lambda/100$. The famous Laser Interferometer Gravitational-wave Observatory (LIGO) [20,21] that detected gravitational waves in 2016 is essentially a giant Michelson interferometer with an arm span of 4 km (Fig. 5), enabling it to detect a minute distance

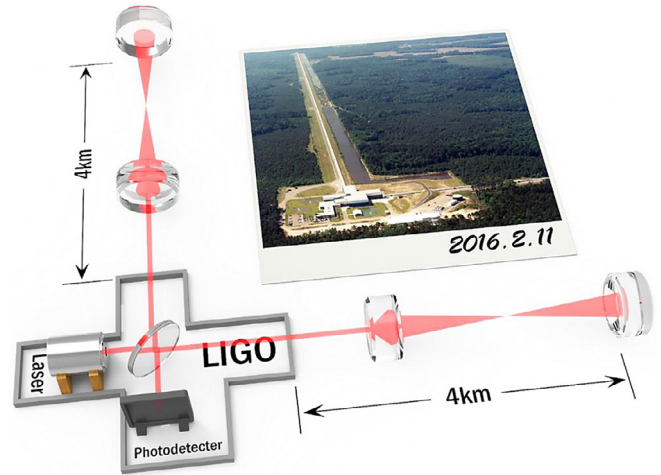


Fig. 5. Laser Interferometer Gravitational-wave Observatory (LIGO) in Louisiana and Washington, US.

change in length on the order of $1/10,000^{th}$ the width of a single proton, representing the incredibly small scale of the effects imparted by passing gravitational waves. Despite the advances of optical interferometry over the decades, its basic principle remains unchanged: by superimposing an additional coherent reference beam with the original object beam, the invisible phase information can be converted into a kind of intensity signal, *i.e.*, the interference pattern, which can be directly captured by traditional imaging devices. Through different types of fringe analysis algorithms, the phase can be demodulated from the interferogram. After decades of development, the interferometry has grown more matured and derived many branches, such as electronic speckle pattern interferometry (ESPI) [22,23], interferometric microscopy [24–26], and digital holography [27–31]. Their basic principles are very similar, and their developments are almost parallel. In particular, digital holography, due to its unique advantages and flexibilities of digital recording and numerical reconstruction, has made remarkable progress in the past two decades and has become a benchmark for quantitative phase measurement and microscopy [32–35] (Fig. 6). Nevertheless, the interferometric

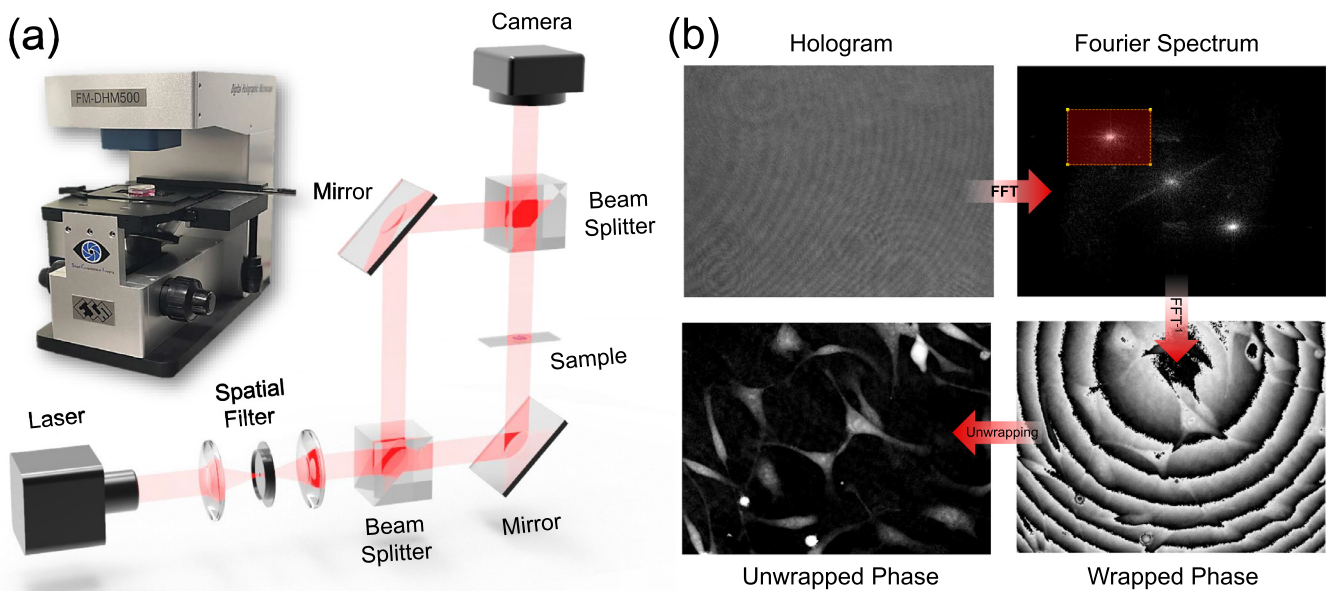


Fig. 6. Digital holography microscopy. (a) Schematic diagram of a typical digital holography microscope with a Mach-Zehnder configuration; (b) The basic principle of the digital recording and numerical reconstruction.

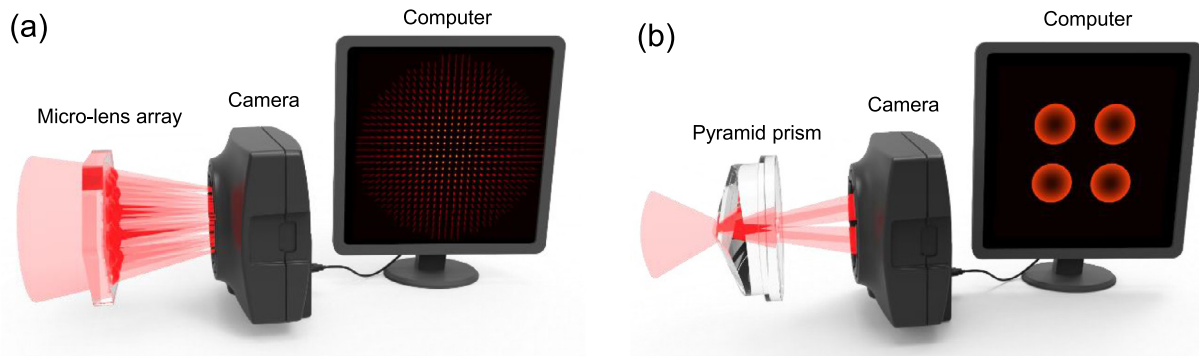


Fig. 7. Two typical wavefront sensing techniques. (a) Shack-Hartmann wavefront sensor; (b) pyramid wavefront sensor.

quantitative microscopic imaging represented by digital holography has not shaken the position of traditional microscopic imaging technique, failing to trigger an expected technological revolution in life science. The main reasons lie in the interferometric characteristics of “quantitative phase measurement”:

- 1) Digital holographic microscopy generally relies heavily on the light sources with a high degree of temporal coherence (e.g., laser) and spatial coherence (e.g., pinhole filtering) as well as complex interferometric optical configurations (i.e., object beam and reference beam);
- 2) Due to the laser illumination sources typically used, digital holographic microscopy suffers from speckle noise arising from stray interferences from imperfections in the optical system, which not only limits the imaging resolution but also strongly deteriorates the image quality;
- 3) Because of the high degree of spatial coherence of the light source, the imaging resolution is only limited to the coherent diffraction limit (half of the incoherent diffraction limit as in traditional incoherent microscope);
- 4) The additional reference beam path makes the measurement extremely sensitive to external disturbance (e.g., vibration, rapid airflow, etc.);
- 5) The phase demodulated from the interferogram is wrapped in the range of $(-\pi, \pi]$ (so-called “wrapped phase”), and additional phase unwrapping is needed to obtain the true absolute phase distribution.

In order to address or partially alleviate the above limitations of traditional interferometric phase imaging techniques, in recent years, many researchers have gradually turned their attention to low-coherence holography or white-light interferometric microscopy, e.g., spatial light interference microscopy (SLIM) [36], white-light diffraction phase microscopy (wDPM) [37], quadriwave lateral shearing interferometry (QWLSI) [38], and τ interferometry [39]. The combination of broadband illumination with the common-path geometries significantly alleviates the coherent noise problem and enhances the stability to mechanical vibrations and air fluctuations that typically affect any interferometric systems. However, most of them require relatively complicated optical systems which are not typically available to most bio/pathologists, prohibiting their widespread use in biological and medical science.

On a different note, the other broad category of phase measurement techniques do not rely on interferometry, which is known as non-interferometric phase measurement. A major branch of interferometric phase measurement techniques is called wavefront sensing, such as Shack-Hartmann wavefront sensor [40–42] [Fig. 7(a)], pyramid wavefront sensor [43–45] [Fig. 7(b)], and model-based wavefront sensor [46,47]. The Shack-Hartmann wavefront sensor is a geometric optics based phase measurement technique often used in adaptive optics systems with the main applications in telescope and astronomical observation. In such areas, “phase” is often referred to as the “wavefront aberration”, representing the difference between the reference wavefront of an ideal optical system and the distorted wavefront of an actual optical system.

The original prototype of the Shack-Hartmann wavefront sensor is an array of holes in a mask (so-called Hartmann screen) that was developed by the astronomer Hartmann [41] in 1900. Although Hartmann’s method has a simple design, its measurement accuracy and light utilization efficiency are quite low. Later, in 1971, Shack and Platt [40] modified the Hartmann screen by replacing the apertures in an opaque screen with an array of lenslet, which focuses incoming radiation to a spot array on the sensor [see Fig. 7(a)]. The integrated gradient of the wavefront across the lenslet is proportional to the displacement of the centroid relative to the reference regular spacing spot array (ideal aberration-free case), and then the wavefront (phase) distribution can be reconstructed through integration. Nowadays, the application of Shack-Hartmann wavefront sensor has expanded from astronomical imaging to beam quality diagnosis [48,49], optical testing [50,51], atmospheric laser communication [52,53], microscopic aberration correction [54,55], ophthalmology [56–58] etc.. However, due to the limitation of the physical size of the microlens, the Shack-Hartmann sensor does not make full use of all the pixel resolution of the image sensor, resulting in low spatial resolution of the recovered wavefront. Consequently, the Shack-Hartmann wavefront sensor is seldom directly used as an imaging tool, especially in the fields of QPI and optical microscopy.

Another very famous and widely used non-interferometric phase measurement technology is called phase retrieval. Since it is difficult to measure the phase of the light directly but very easy to measure the amplitude/intensity, we can consider the process of recovering (estimating) the phase from the intensity distribution as a mathematical “inverse problem”, and the solution to such kind of problem is so-called phase retrieval. Phase retrieval techniques can be divided into two categories: iterative methods and deterministic methods. The iterative phase retrieval has its origins in ideas in electron microscopy. In 1972, Gerchberg and Saxton [59,60] proposed the first iterative phase retrieval algorithm called *Gerchberg-Saxton* (GS) algorithm, which employs an iterative method in which the phase distribution within an object is found that is consistent with both an intensity image of it and with its far-field diffraction pattern. The proposed algorithm has been found to be able to converge on the correct solution, and subsequently, the underlying ideas have been applied to a wide variety of fields such as X-ray diffraction imaging [61–64] (also referred to as *coherent diffraction imaging*, CDI) [Fig. 8(a)], adaptive optics [65,66], and quantitative phase microscopy [67–71]. A practical issue in X-ray imaging is that the object scattering is so weak that the unscattered beam will swamp the signal of interest. This problem is mitigated by the introduction of a beam stop to prevent the undiffracted beam from destroying the detector at the expense of a substantial low spatial frequency information loss. Miao *et al.* [61] circumvented this problem by using an additional electron microscope image of the object to recreate the missing low spatial frequency data, and

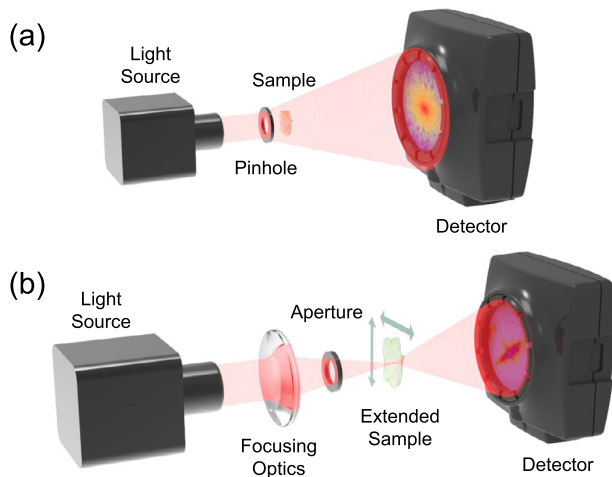


Fig. 8. Schematic layout of two main iterative phase retrieval methods. (a) Coherent diffraction imaging: a plane wave illuminates a sample, and the far-field diffraction pattern is measured by a detector; (b) Ptychography: a coherent probe is generated by an aperture of focusing optics. An extended sample is scanned through the probe on a 2D grid, and diffraction patterns are collected from a series of partially overlapping regions.

experimentally demonstrated the validity of this approach with X-ray data. The rapid growth of interest in CDI has largely been driven by the work of Miao *et al.* [61].

However, the GS algorithm also has some problems and limitations. Some of the problems may result from the iterative algorithm itself, for example, the convergence of the algorithm stagnates after early iterations or even get trapped into local optima [72–74]. Various modified algorithms have been proposed to improve the convergence of the original GS algorithm. In 1973, Misell [75] pointed out that the GS algorithm is not only applicable for the intensity distributions of the image plane and the far-field Fourier plane but also can be extended to two or more diffraction patterns of different propagation distances. He then used more intensity constraints to improve the accuracy and convergence of the iterative algorithm. Compared with the original GS algorithm, Misell’s modified algorithm is more flexible and practical, which opens up new ideas for improving the iterative phase retrieval algorithm. Up to now, new improved and optimized algorithms have still been emerging with a variety of techniques being used, such as multiple propagation distances [76,77], multiple illumination wavelengths [78], phase modulation [79,80], and speckle illumination modulation [81,82]. Other problems of the GS algorithm are inherent in solving inverse problems, such as the existence and uniqueness of the solution [83–86]. Clearly, if an object is so small that a near-field image cannot be acquired (only the far-field diffraction pattern is available), then reconstruction via the GS algorithm is not possible. In general, there is no unique mapping between a 2D signal and its Fourier magnitude (different objects could produce identical far-field diffraction patterns) and therefore, the problem is ill-posed. Bates [83] considered this problem and claimed (without proof) that an object can be uniquely determined by its autocorrelation function (*i.e.*, the far-field diffraction pattern) if its spatial extent is known (except for several unimportant cases, such as translation, phase conjugation, and inversion). In 1982, after carefully analyzing the optimization principle behind the GS algorithm, Fienup [87] pointed out that GS is equivalent to an error reduction algorithm, which is essentially the same as the steepest descent algorithm. Fienup’s work [87] has laid a solid theoretical foundation for the solvability of the phase retrieval problem from the perspective of convex optimization. It also provides a reasonable explanation for the stagnation problem: due to the non-convexity of the phase retrieval problem itself, there

is no guarantee that the error reduction algorithm can finally reach the global minimum. Fienup [87,88] demonstrated that a modification of the GS algorithm based on the theories of cybernetics, so-called *hybrid input-output* (HIO) algorithm, could help prevent the iterative algorithm from getting stuck in local minima where only support information is used at the object plane. This algorithm has led to a number of analyses and improvements, primarily in the manner in which the object estimate is updated. These include the hybrid projection reflection [89], difference map [90], relaxed averaged alternating reflections [91], and charge flipping [92]. Though all of these ideas seem simple to implement, their success in application to experimental data requires considerable skill and experience. A unified analysis of the methods has been published by Marchesini [93], suggesting that none of them have yielded an algorithm that has been found to converge in all cases. Another important innovation proposed by Marchesini *et al.* [63] is the so-called “shrink-wrap” algorithm, in which the support for the object is refined dynamically. This method is very useful for reconstructing the atomic distribution within nanocrystals.

Ptychography is another iterative phase retrieval approach that has its roots in electron microscopy [94]. In 2004, Rodenburg and Faulkner [95,96] introduced the idea of “ptychography” into the GS and HIO algorithms and presented a new iterative phase retrieval method so-called *Ptychographic Iterative Engine* (PIE). In the implementation of ptychography, the object is illuminated by a field that is limited in extent and the diffraction pattern is recorded, as illustrated in Fig. 8(b). This process is repeated until the area of interest has been completely scanned. The resultant spatially overlapped “sub-aperture” diffraction patterns are used as intensity constraints, and the complex amplitude distribution of the area covered by the beam can be reconstructed by iterative phase retrieval algorithm that is similar to GS and HIO. Compared with GS and HIO, the ptychographic phase retrieval is more stable and reliable due to the high data redundancy resulting from the large number of spatially overlapped intensity data collected. Consequently, the non-convexity of the phase retrieval problem is effectively alleviated, and the ambiguity between the true solution and its conjugate is eliminated [86]. In the ensuing decade, the ptychographic phase retrieval methods have been extensively studied and improved, and significant progress has been made in probe (illumination function) recovery [97–102], positional error correction [103–105], coherent mode decomposition [106–109], resolution improvement [110–113], and depth sectioning for multi-layer samples [114–117]. It has now become an important approach for high-resolution X-ray diffraction imaging [118–120], electron microscopy [121–123], and optical phase imaging and microscopy [124–126].

In 2013, Zheng *et al.* [127] extended ptychography from the spatial domain to frequency domain, presenting its dual counterpart named *Fourier ptychographic microscopy* (FPM). In FPM, “ptychography” exists in the Fourier domain: the sample is no longer scanned in the spatial domain by an aperture-limited illumination beam, but is successively illuminated by plane waves from different angles (usually with an LED array); the intensity images are not acquired in the far field (Fourier domain) but are low-resolution in-focus images captured directly in the spatial domain. The finite aperture of the imaging system becomes the “sub-aperture” in the frequency domain, while the changes of illumination angle enable the overlapping scanning of the sub-aperture in the frequency domain, as illustrated in Fig. 9. Because of the perfect duality, the key issues in FPM, such as aberration correction [128], positional error correction [129,130], coherent mode decomposition [131–133], depth sectioning for multi-layer samples [134,135], have been successively addressed in just a few years. At present, it has been developed almost in parallel with conventional (spatial) ptychography. The elegance of FPM lies in the fact that it not only retrieves the phase information of the sample but achieves synthetic aperture naturally along with the ptychographic updating process in the frequency domain, improving the imaging resolution effectively [130,136,137]. In general, a low-magnification objective with a low *numerical aperture* (NA) is used

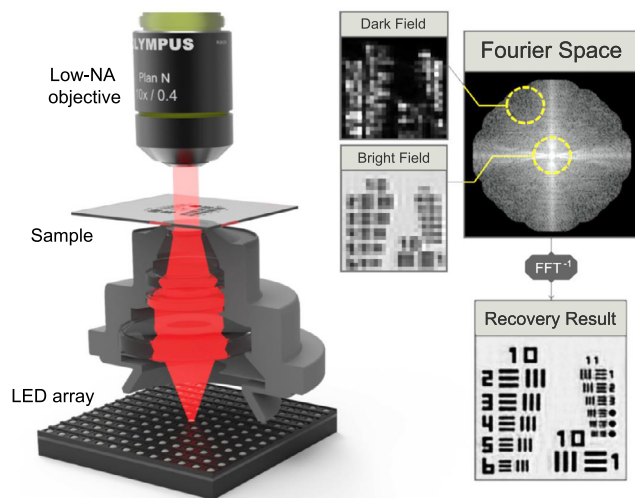


Fig. 9. Basic principle of Fourier ptychographic microscopy. The sample is successively illuminated from different directions with different LED elements in an LED array; the captured low-resolution images allow for simultaneous phase retrieval and synthetic aperture in the frequency domain.

in FPM to obtain an inherent large *field of view* (FOV), and then the high-angle illuminations are used to synthetic aperture. The final effective NA is the sum of the NA of the objective lens and the NA of the illumination at the largest angle, and thereby, FPM generally achieves a very large imaging *space-bandwidth product* (SBP). For example, in 2017, Sun *et al.* [138] reported a high-NA illumination based resolution-enhanced FPM (reFPM) platform, in which a LED-array-based digital oil-immersion condenser is used to create high-angle programmable plane-wave illuminations, endowing a $10\times$, 0.4 NA objective lens with the effective imaging performance of 1.6 NA (308 nm lateral resolution for coherent imaging) across a wide FOV of 2.34 mm^2 (Fig. 9).

As mentioned above, the iterative phase retrieval effectively overcomes the problem of “direct phase reconstruction from intensity measurements”, and promotes the evolution of phase measurement from interferometric to non-interferometric. In particular, the defocus-based iterative phase retrieval and ptychographic imaging have contributed significantly to the development and progress of adaptive optics, X-ray diffraction imaging, and electron microscopy. However, there are still two limitations: (1) a large amount of intensity data is often required to ensure stable convergence; (2) a large number of iterations are needed to obtain a reliable solution. These two shortcomings preclude their use for high-speed and real-time imaging applications. Besides, the phase retrieval algorithms relate the scattered field at the object to the scattered field at the detector via a Fourier transform, a relationship that is only valid for an optical field that is completely coherent. Thus, a reliable reconstruction was critically dependent on a very high degree of coherence [139], which is an assumption that does not hold for many partially coherent sources.

In contrast to iterative methods, another category of phase retrieval method uses propagation to recover phase directly, in a non-iterative manner. This idea was arguably first proposed by Teague [140,141]. In 1982, he first pointed out that under the paraxial approximation, one could employ the conservation of energy on propagation (Helmholtz equation) to derive a second-order elliptic partial differential equation that outlines the quantitative relationship between the variation of intensity along the optical axis to the phase of the optical field at the plane perpendicular to the optical axis [140]. Teague believed that the equation could be used as a phase retrieval approach. He termed this equation *transport of intensity equation* (TIE), and this will be the main theme of this tutorial. In his latter paper in 1983, Teague [141] proposed a tentative approach to the solution of the equation based on Green’s

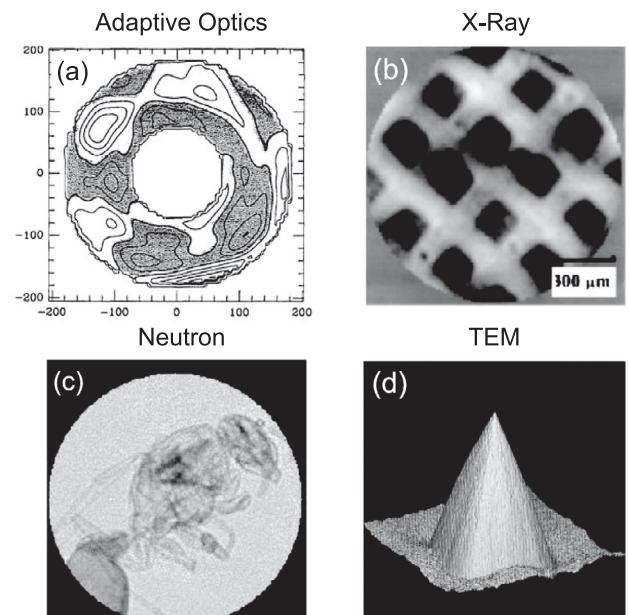


Fig. 10. The representative applications of TIE in the fields of (a) adaptive optics [148], (b) X-ray diffraction imaging [161], (c) neutron radiography [164], and (d) electron microscopy [166].

functions. Different from the iterative phase retrieval methods, TIE does not employ traditional scalar diffraction formula to recover the phase iteratively. When the intensity distribution at the in-focus plane (measured directly) and the axial intensity derivative (estimated by the finite difference between two defocused intensities) are known, the phase information can be retrieved directly (deterministically) by solving TIE without iterative operations. Compared with traditional interferometric methods, TIE has many unique advantages of being non-interferometric (without using reference beam), simple calculation (without iterations), capabilities to temporally/spatially partially coherent beams (e.g., LED illumination, halogen lamp, and built-in Köhler illumination of conventional microscopes), does not need phase unwrapping (directly obtain the absolute phase), complicated optical setups, and stable measurement environment. Shortly thereafter, in 1984, Streibl [142] found that TIE is applicable to optical microscopy as a tool for phase-contrast imaging of transparent phase objects. Since the numerical solution to TIE has not been proposed at that time, he only demonstrated the phase gradient enhancement effect of unstained cells by taking the axial intensity derivative. Nevertheless, this pioneering work laid a preliminary theoretical foundation for the subsequent extensive applications of TIE in the field of optical microscopy. In 1988, Ichikawa and Takeda [143] experimentally verified TIE for the first time. They solved TIE by Fourier transform method and obtained the quantitative phase distribution of the one-dimensional sample. However, rather than solving the partial differential equation, this method was only found to be a special case of grating shear interferometry [144]. In parallel, the research team of the University of Hawaii led by Roddier [144–148] was also exploring the possibilities of using TIE to correct the wavefront distortion introduced by atmospheric turbulence in adaptive optics [Fig. 10(a)]. Assuming that the amplitude of the wavefield is almost uniform, TIE can be simplified from a more complicated elliptic partial differential equation into a standard Poisson equation. This Poisson equation reveals that the axial intensity derivative of the optical field is proportional to its wavefront curvature (second derivative, i.e., Laplacian), so this technique is often called *curvature sensing* (CS). There is no doubt that Roddier’s research work has greatly promoted the TIE phase retrieval method. His contribution is not only applying TIE to adaptive optics but more importantly, he first proposed the numerical solutions to the simplified form of

TIE based on *successive over-relaxation* (SOR) method [145] and iterative Fourier transform method [149]. In 1988, Roddier estimated the axial intensity derivative by measuring the in-focus and slightly defocused intensity signals at the entrance pupil of a telescope and fed it directly back to the circuit of the deformable mirror to correct the atmospheric wavefront distortion in real-time [145,146]. Then, in 1991, Roddier [150] constructed the 13-unit low-order curvature sensing adaptive optics system and presented preliminary laboratory experimental results. The defocusing was realized through the rapid vibration of a curvature variable membrane mirror so that the overfocus and underfocus intensities can be measured simultaneously. The wavefront curvature signal was measured by an avalanche photodiode array, and the deformable mirror was driven based on the response matrix control method. Subsequently, the curvature sensing has been widely adopted in the US [the 3.6m-aperture CFHT (Canada-France-Hawaii Telescope) [151] and the 8m-diameter Gemini North telescope [152]], Europe [the 8.2m-aperture VLTI (Very Large Telescope Interferometer)] [153], Japan (the 8.2m-aperture Subaru Telescope) [154], and other countries' large-scale astronomical telescopes.

In the mid-to-late 1990s, the theoretical aspect of TIE developed very rapidly and made fruitful achievements, wherein the X-ray diffraction imaging research group of the University of Melbourne was the primary driving force. In 1995, Gureyev *et al.* [155] first proven the well-posedness and uniqueness of the solution to TIE based on strict mathematics: if the intensity of the optical field is strictly greater than 0 (precluding the case of optical vortex) and under the appropriate boundary conditions, the solution of TIE exists and is unique (for the Neumann boundary condition, the solution is unique up to a trivial additive constant). This work has laid the theoretical foundation for the subsequent TIE phase retrieval and QPI applications. In 1996, Gureyev and Nugent [156,157] proposed to solve TIE based on the orthogonal expansion, such as Zernike polynomial expansion or Fourier expansion, and pointed out that when the intensity distribution is uniform, TIE can be solved by using *fast Fourier transform* (FFT) efficiently [158]. Shortly thereafter, Paganin and Nugent [159] extended the FFT-based solution so that it can be effectively accommodated to the case of the nonuniform intensity distribution. This simple and effective FFT-based method has now become the most widely used numerical solution to TIE. The establishment of the mathematical basis and the numerical solution of TIE have become a powerful vehicle that motivates its rapid development and widespread applications in different fields. In 1995, Snigirev *et al.* [160] recognize that propagation-based non-iterative phase retrieval techniques could be applied in the field of X-ray imaging. Soon after, in 1996, Nugent [161] *et al.* reported the quantitative phase image of a thin carbon foil using 16 keV hard X-rays based on TIE [Fig. 10(b)]. These papers triggered the rapid growth of interest of TIE in the applications of X-ray optics [162,163]. Subsequently, the applications of TIE have also been extended to the fields of neutron radiography [164,165] [Fig. 10(c)] and *transmission electron microscopy* (TEM) [166–174] [Fig. 10(d)]. The common feature of these fields is that it is difficult to obtain a highly coherent light source so that TIE became a simple and effective alternative to interferometry.

In the late 1990s, on the one hand, the theoretical framework of TIE was well established; on the other hand, researchers engaged in the field of QPI were increasingly aware of the importance of using low coherence illumination for improving the resolution and image quality. These two aspects then became the catalyst for the widespread application of TIE in the field of optical imaging and microscopy. In 1998, Paganin and Nugent [159] reinterpreted the meaning of “phase” of partially coherent fields and pointed out that it is a scalar potential whose gradient corresponds to the time-averaged Poynting vector. The importance of this work is that it gives a more generalized and meaningful definition of the “phase”, laying a theoretical basis for the subsequent TIE phase retrieval under partially coherent illuminations. In the same year, Barty *et al.* [175] briefly reported the QPI results of cheek cells and optical fibers based on TIE, which was the first appearance of TIE in quantita-

tive optical phase-contrast imaging. Then Barty *et al.* [176] extended this method to quantitative phase tomography, based on a similar approach as the “holotomography” earlier demonstrated by Cloetens *et al.* [163] in 1999. Owing to its successful applications in the field of QPI and optical microscopy, at the beginning of the 20th century, the theory of the TIE phase retrieval happened to coincide with the *optical transfer function* (OTF) theory of optical microscopy. In 2002, Barone-Nugent *et al.* [177] theoretically analyzed the image formation of weak phase objects under an optical microscope based on the 3D OTF theory developed by Striebl [178], further confirmed the applicability of TIE under partially spatially coherent illuminations. In the same year, Sheppard [179] analyzed the TIE phase retrieval under the first-order Born approximation and pointed out that the low-frequency component of the axial intensity derivative is proportional to the phase Laplacian (second-order derivative). The conclusion is consistent with the case of the simplified version of TIE under uniform intensity analyzed by Roddier [144–146]. The initial establishment of TIE in the field of optical microscopy and the experimental results of Barty *et al.* led the application of TIE into the fields of biological and biomedical imaging. Curl *et al.* [180,181] used TIE to quantitatively detect the morphology and growth rate of cells, indicating that quantitative phase information is very conducive to data processing and quantitative analysis, such as cell segmentation and counting. Ross *et al.* [182] utilized TIE to enhance the imaging contrast of unstained cells under microbeam irradiations. Dragomir *et al.* [183] successfully applied TIE to the quantitative measurement of the birefringence effect of unstained isolated cardiomyocytes. Not just limited to biological samples, TIE was also successfully applied to the quantitative measurement of phase and RI distribution of optical fibers and other technical phase objects [184–187].

From 2010 to date, the TIE phase retrieval reached another prime due to the brilliant achievements in visible light imaging and microscopy. During this period, the basic theory of TIE (especially for the theory of partially coherent imaging) and related applications (especially for the applications of quantitative phase microscopy and tomography) have made a great leap forward again, reaching new and unprecedented heights. Kou *et al.* [188] demonstrated that TIE could be directly implemented based on an off-the-shelf DIC microscope to realize QPI, and successfully obtained the quantitative phase images of human cheek cells. In the same year, Waller *et al.* [189] proposed a method named high-order TIE by using more intensity measurements at multiple defocused planes to enhance the accuracy and robustness of phase retrieval. The simple and straightforward descriptions of these work greatly boosted the popularity of TIE in the fields of optical imaging and microscopy. Meanwhile, Waller *et al.* also proposed single-shot TIE optical configurations based on volume holography [190] and chromatic aberration [191], enabling acquisition of intensity images at different defocus distances from a single exposure. In 2011, Kou *et al.* [192] put forward an OTF deconvolution method for QPI under partially coherent illuminations and quantitatively compared it with the conventional TIE. In 2012, Almoró *et al.* [193] used speckle field illumination to enhance the phase contrast of the defocused intensity images for smooth objects. In the same year, Gorthi and Schonbrun [194] first applied TIE to flow cytometry and realized the automatic collection of through-focus intensity images by tilting the microfluidic channel vertically. In the meantime, there are a large number of publications focusing on optimizing the finite-difference scheme phase retrieval to improve the accuracy and robustness of TIE under noisy conditions [195–200]. Since 2012, our group has carried out systematic research work on the TIE phase retrieval, and several key theoretical issues in TIE have been thoroughly and comprehensively studied:

- 1) Efficient numerical solution under inhomogeneous boundary conditions [201–203];
- 2) Phase discrepancy analysis and compensation [204];
- 3) Optimal axial intensity derivative estimation [205,206];
- 4) Phase space extension for partially coherent fields [207,208];

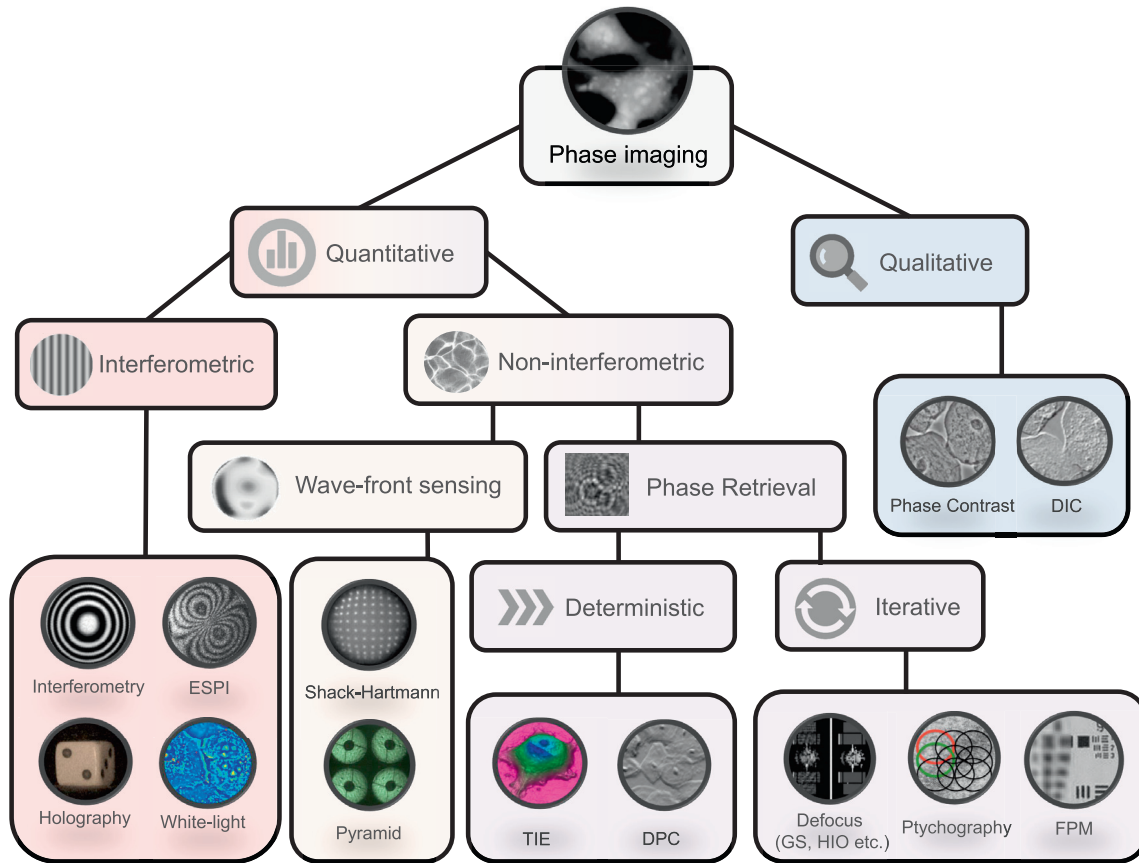


Fig. 11. Classification of the phase imaging techniques.

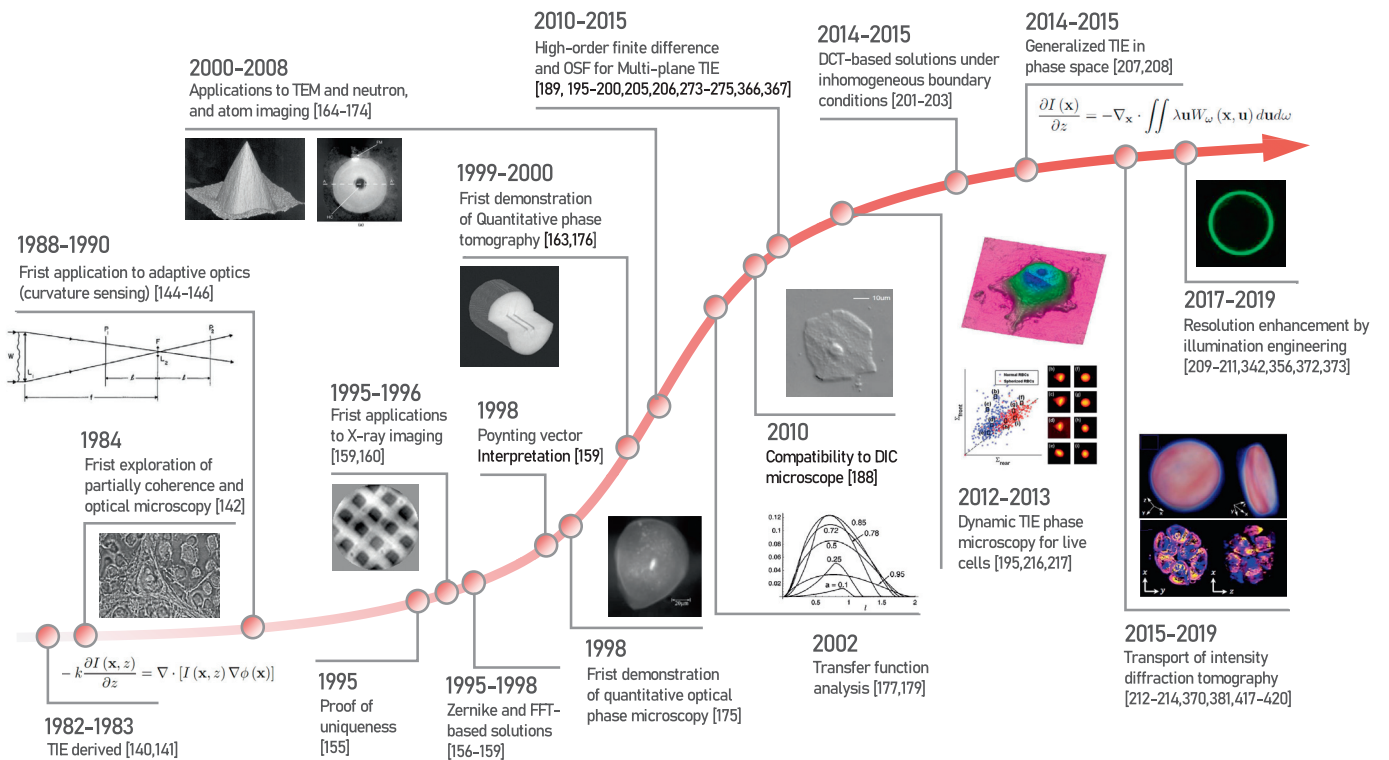


Fig. 12. Timeline of TIE research fields. Key developments in both theoretical and experimental aspects are shown.

- 5) Resolution enhancement via illumination engineering [209–211].
 6) Diffraction tomography under partially coherent illuminations [212–214];

In experimental aspects, instead of including a mechanical refocusing system (e.g., piezo-stage) to provide the through-focus scanning, we proposed to introduce an electrically tunable lens [215] or a *spatial light modulator* (SLM) [216] placed in the Fourier conjugated plane of the detector in a conventional wide-field transmission microscope. This provides fast high-speed acquisition of through-focus intensity stack, making the dynamic TIE phase imaging for live cells possible. We have successfully applied these systems for investigations of drug-induced morphology changes and phagocytosis of macrophages [216], imaging of cellular dynamics of cancer cells [215,217], and characterization of micro-optical elements [202]. To date, the TIE phase retrieval attracts more and more attention from both academia and industry, and related new theories and technologies are still constantly emerging.

As previously mentioned, we have briefly summarized the development of phase imaging and measurement technologies, suggesting that these technologies can be divided into two main categories: phase visualization method (phase-contrast imaging) and phase measurement method (quantitative phase imaging). The essential difference between these two categories is whether they can recover the “quantitative” phase information. For the phase measurement, it can also be divided into interferometry, wavefront sensing, and phase retrieval. Interferometry and phase retrieval also have their independent sub-branches. Fig. 11 illustrates the classification of phase imaging and measurement technologies. In the following parts, the emphasis will be focused on the non-interferometric phase measurement technology based on TIE. As discussed above, although the phase retrieval and quantitative phase microscopic imaging based on TIE have been extensively reported in recent years, there has been no new comprehensive review of TIE focusing on the basic principle, technical details, and practical applications of this emerging technology. Therefore, at present, there appears to be a moment in time for this tutorial, which describes the fundamental principles and latest progress of TIE in a rigorous and comprehensive way is of interest to the community of optical scientists and engineers, physicists, and biologist interested in optical microscopy and phase imaging. Before we start, a timeline of TIE research developments is outlined in Fig. 12, which provides an overview of the relevant research fields and summarizes the key developments in both theoretical and experimental aspects. We have also included Tables 1 and 2 at the beginning with a list of all the acronyms, notations, and symbols used in the tutorial.

The remaining parts of this paper are organized as follows: In Section 2, we start with the motivations behind the development of TIE, describe its fundamental principle, outline its derivations, and physical implications. From Sections 3 to Section 8, several key theoretical issues of TIE are discussed in detail, including its solutions, image formation model under coherent illumination, axial intensity derivative estimation, QPI under partially coherent illuminations, and 3D diffraction tomography based on the transport of intensity. In Section 9, the system configurations and typical applications of TIE are presented with an emphasis on optical imaging, metrology, and microscopy in the visible light band. Finally, the challenging problems, as well as future research directions, are outlined in Section 10.

2. Basic concept of TIE

TIE is a partial differential equation that describes the quantitative relationship between the phase and the axial intensity variation of the optical field. In this section, we will start from the simplest physical phenomenon without invoking complicated mathematics. Then we will use plain words to explain the transport of intensity phenomenon and its cause (Subsection 2.1). In the following Subsection 2.2, we will use three different approaches to derive TIE based on strict mathematical

Table 1
List of acronyms used in the tutorial

| Acronym | Full Name |
|---------|---|
| 1D | One-Dimensional |
| 2D | Two-Dimensional |
| 2.5D | Two-Point-Five-Dimensional |
| 3D | Three-Dimensional |
| 4D | Four-Dimensional |
| AOTF | Amplitude Optical Transfer Function |
| AF | Ambiguity Function |
| ATF | Amplitude Transfer Function |
| BC | Boundary Condition |
| CARS | Coherent Anti-Stokes Raman Scattering |
| CCF | Complex Coherence Factor |
| CDC | Complex Degree of Coherence |
| CDI | Coherent Diffraction Imaging |
| CT | Computerized Tomography |
| CTF | Contrast Transfer Function |
| CSD | Cross-Spectral Density |
| DC | Direct Current |
| DCT | Discrete Cosine Transform |
| DIC | Differential Interference Contrast |
| DOE | Diffraction Optical Element |
| DPC | Differential Phase-Contrast |
| FFT | Fast Fourier Transform |
| FOV | Field of View |
| FPM | Fourier Ptychographic Microscopy |
| FTS | Fourier Transform Spectrometer |
| GFST | Geometric-Flow Speckle Tracking |
| GTIE | Generalized Transport of Intensity Equation |
| GSM | Gaussian-Schell Model |
| LHS | Left Hand Side |
| LMI | Light Field Moment Imaging |
| MCF | Mutual Coherence Function |
| MI | Mutual Intensity |
| MSE | Mean Square Error |
| NA | Numerical Aperture |
| ODT | Optical Diffraction Tomography |
| OFS | Optimal Frequency Selection |
| OTF | Optical Transfer Function |
| PAM | Partitioned/Programmable Aperture Microscopy |
| POTF | Phase Gradient Transfer Function |
| PIE | Ptychographic Iterative Engine |
| POTF | Phase Optical Transfer Function |
| PTF | Phase Transfer Function |
| PSD | Power Spectral Density |
| PSF | Point Spread Function |
| PST | Phase Space Tomography |
| PWM | Pulse Width Modulation |
| QPI | Quantitative Phase Imaging |
| RHS | Right Hand Side |
| RI | Refractive Index |
| SBP | Space-Bandwidth Product |
| SCF | Self-Coherence Function |
| SGDF | Savitzky-Golay Differential Filter |
| SNR | Signal-to-Noise Ratio |
| TCC | Transmission Cross-Coefficient |
| TEM | Transmission Electron Microscopy |
| TIDT | Transport of Intensity Diffraction Tomography |
| TIE | Transport of Intensity Equation |
| TV | Total Variation |
| WDF | Wigner Distribution Function |
| WOTF | Weak Object Transfer Function |
| ZPC | Zernike Phase Contrast |
| GS | Gerchberg-Saxton |
| HIO | Hybrid Input-Output |
| PIE | Ptychographic Iterative Engine |
| CS | Curvature Sensing |
| SOR | Successive Over-Relaxation |
| IFFT | Inverse Fast Fourier Transform |
| NRR | Noise Reduction Ratio |
| GP | Gaussian Process |
| ROC | Radius of Curvature |
| LUT | Lookup Table |
| MFPI | Multi-Filter Phase Imaging |
| RMSE | Root Mean Square Error |

Table 2
List of notations and symbols used in the tutorial

| Notation or Symbol | Definition |
|---|--|
| $\mathbf{x} = (x, y)$ | Lateral 2D spatial coordinate (μm) |
| z | Longitudinal spatial coordinate (μm) |
| $\mathbf{r} = (x, z)$ | 3D spatial coordinate (μm) |
| $\mathbf{u} = (u, v)$ | Lateral 2D spatial frequency coordinate (μm^{-1}) |
| η | Longitudinal spatial frequency coordinate (μm^{-1}) |
| $\mathbf{f} = (\mathbf{u}, \eta)$ | 3D spatial frequency coordinate (μm^{-1}) |
| $r = \mathbf{x} = \sqrt{x^2 + y^2}, \theta = \tan^{-1}(y/x)$ | 2D polar coordinates |
| $\rho = \mathbf{u} = \sqrt{u^2 + v^2}, \theta = \tan^{-1}(v/u)$ | 2D spatial-frequency polar coordinates |
| $\mathbf{K} \equiv (k_x, k_y, k_z)$ | 3D spatial frequency coordinate in k space |
| $(u_x, u_y, u_z) \equiv (u, v, \eta)$ | 3D spatial frequency coordinate (μm^{-1}) |
| $(x, y, u, v) = (\mathbf{x}, \mathbf{u})$ | 4D coordinate in phase space |
| λ | Wavelength (μm) |
| k_0 | Wave number in vacuum (μm^{-1}) |
| k_m | Wave number in medium (μm^{-1}) |
| n_0 | Background refractive index |
| $n(\mathbf{r})$ | Refractive index distribution of the object |
| Δn | Refractive index difference between object and medium |
| j | Imaginary unit |
| δ | Delta function |
| ω | Angular frequency |
| (α, β, γ) | Direction angle |
| ∇ | Transverse gradient |
| Ω | Open bounded domain in 2D space |
| $\partial\Omega$ | Piecewise smooth boundary of Ω |
| $\bar{\Omega}$ | Closed bounded domain in 2D space |
| \mathbb{R}^2 | 2D real space |
| \mathbb{R}^N | N-dimensional real space |
| $\nabla_{\mathbf{x}}$ | Gradient in 3D space |
| \otimes | Convolution |
| $\otimes_{\mathbf{x}}$ | 2D spatial convolution in phase space |
| $\otimes_{\mathbf{u}}$ | 2D spatial-frequency convolution in phase space |
| $\otimes_{\mathbf{x}, \mathbf{u}}$ | 4D convolution in phase space |
| $\hat{\cdot}$ | Fourier transform of the corresponding viable |
| \sim | Estimated value of the corresponding viable |

and physical theories. At last, in [Subsection 2.3](#), the mathematical expression of TIE and its physical meaning are explained to echo the earlier [Subsection 2.1](#).

2.1. Transport of intensity effect

Since the traditional photodetectors (e.g., CMOS and CCD) are only sensitive to the intensity of the optical field and the phase information cannot be directly detected, most of the phase measurement methods need to convert the invisible phase information into the visible intensity information for recording and reconstruction. For example, the classical interferometry transforms the invisible phase information into visible fringe patterns by spatial superposition of two coherent light beams, and the fringe pattern can be then recorded by photodetectors. Finally, the quantitative phase distribution can be demodulated from the interferogram through fringe analysis algorithms.

The phase-to-intensity conversion does not only exist in the interference phenomena, and the propagation of the optical field is another kind of spontaneous phase-to-intensity conversion process. We can imagine that a swimming pool in the sunshine and the slight ripples appear on the surface of the pool because of the breeze or a dropped swimming ring. Looking towards the bottom of the pool, we can see the wave-like pattern created by the refraction of the light at the pool surface (see left of [Fig. 13](#)). The ripples on the water surface act as lenses in a variety of shapes, and the transmitted light is redistributed and may be concentrated in some regions of the pool bottom, instead of evenly distributed at the whole bottom. If the continuously changing curvature on the water surface is large enough, the water wave will focus the sunlight onto the bottom of the pool, forming some bright line network structure, as shown in the middle figure of [Fig. 13](#).

So, what is the relationship between the above phenomenon and phase retrieval? In fact, the water surface with ripples is essentially a phase object, and it only changes the phase of the incoming light. The light-and-dark network structure at the bottom of the swimming pool is exactly the “manifestation of phase via propagation”. We call this kind of phenomenon “transport of intensity effect”. In this context, there is neither laser nor interference, but the intensity pattern at the bottom of the swimming pool closely resembles the interference pattern in optical interferometry. This phenomenon suggests that the phase can be converted into intensity without resorting to interferometry, and this kind of intensity is often referred to as “phase contrast”. Similar to the situation that the light-and-dark intensity structure at the bottom of the swimming pool reflects the shape of the water surface, the aim of the non-interferometric TIE phase retrieval is to recover the quantitative phase distribution by measuring the phase-induced intensity variation (phase contrast signal) at the defocused plane. In this case, the morphology of the water surface is retrieved from the intensity pattern at the bottom of the swimming pool, as illustrated in the right image of [Fig. 13](#).

We are no strangers to the transport of intensity effect: the twinkling of stars in the night, the distorted scene outside the window in the rain, and the above-mentioned network of bright lines at the bottom of a swimming pool in the sunshine. These are all manifestations of phase, implying the inextricable ties between phase and intensity of the light wave. All these phenomena show that although phase objects cannot be observed directly, they incessantly, and implicitly manifest their existence. Nevertheless, the transport of intensity effect is not always observable. For instance, when the water surface is completely still and almost flat, the intensity patterns will disappear. The transport of intensity effect can be explained qualitatively based on geometric optics, as shown in [Fig. 14](#). If the observed optical field is a perfect plane

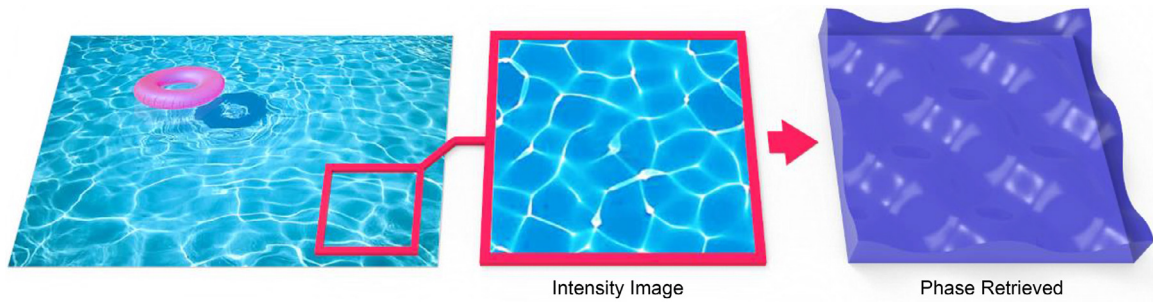


Fig. 13. The wave-like pattern at the bottom of a swimming pool in sunlight. The pool surface refracts the incident sunlight to produce the light-and-dark network structure at the bottom of the swimming pool.

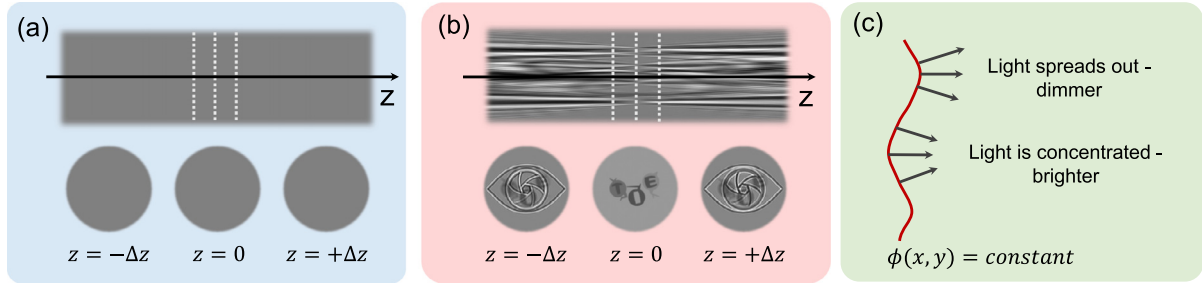


Fig. 14. The influence of phase on the beam propagation. (a) For on-axis plane waves, the intensity does not change during propagation; (b) non-planar phase change the intensity during propagation; (c) an intuitive geometric-optics interpretation of transport of intensity effect.

wave, the intensity will not change with the propagation distance Δz , as shown in Fig. 14(a). This is because the plane wave can be regarded as a cluster of parallel rays with the propagation direction parallel to the optical axis (perpendicular to the wavefront, *i.e.*, equiphase surface), as shown in Fig. 14(c). Since light rays spread along straight lines in free space, the intensity distribution maintains the same regardless of the location. However, for non-planar waves, the intensity will change with the propagation distance, as shown in Fig. 14(b). Fig. 14(c) gives the intuitive geometric interpretation of this phenomenon. Though wavefront distribution is not uniform, it can still be regarded as the combination of several piece-wise uniform functions. In each small area, the distribution of wavefront can be approximated as a plane wave, and the propagation direction of light rays in these areas is perpendicular to their wavefront. However, the phase distribution is not uniform for the whole region, and the distribution and propagation direction of light rays in different areas may be different from each other. The convergence or divergence of light rays in different areas will inevitably cause intensity variation during the propagation. On the other hand, there is a certain requirement of transport of intensity effect on the spatial coherence of the illumination. If it is cloudy, and the sunlight can no longer illuminate the pool surface directly due to the multiple scattering of the cloud, we can hardly observe the transport of intensity phenomenon (the relevant part will be discussed in detail in Section 6). Based on the transport of intensity effect mentioned above, the quantitative relationship between the axial intensity variation and the phase is established by TIE, and the quantitative phase distribution can be directly recovered by solving TIE. In the next subsection, we will introduce TIE and its detailed derivation.

2.2. Derivation of TIE

In 1982, Teague [140,141] first established the quantitative relationship between the longitudinal intensity variation and phase of a coherent beam with use of a second-order elliptic partial differential equation, so called TIE. Considering a paraxial monochromatic coherent beam propagating along z -axis, its complex amplitude can be expressed as $U(\mathbf{x}, z) = \sqrt{I(\mathbf{x}, z)} \exp[j\phi(\mathbf{x})]$, and TIE can be expressed as

$$-k \frac{\partial I(\mathbf{x}, z)}{\partial z} = \nabla \cdot [I(\mathbf{x}, z) \nabla \phi(\mathbf{x})] \quad (1)$$

where k is the wave number $2\pi/\lambda$; \mathbf{x} is the transverse coordinates $\mathbf{x} = (x, y)$; $I(\mathbf{x}, z)$ is intensity distribution at the plane located at the propagation distance z , and ∇ is gradient operator (Hamiltonian) over \mathbf{x} . In Teague's original TIE paper, he introduced the expression of complex amplitude into Helmholtz equation under the paraxial approximation (paraxial wave equation), and then separated the real part and the imaginary part to obtain TIE. In fact, TIE can be also derived from Poynting theorem or Fresnel diffraction formula Under the paraxial approximation and the limit of small propagation distance. Next, we will show three different ways to derive TIE.

2.2.1. Derivation of TIE from paraxial wave equation

TIE can be derived from the paraxial wave equation, which is also the approach adopted by Teague [140,141] when TIE was first derived. We know that the free space propagation of the monochromatic coherent beam obeys Helmholtz equation [5],

$$(\nabla^2 + k^2)U(\mathbf{r}) = 0 \quad (2)$$

where ∇ is the Hamiltonian in 3D space $\mathbf{r} = (x, y, z)$. The complex amplitude of coherent beams must satisfy Helmholtz equation, and plane waves and spherical waves are the basic solutions to Helmholtz equation. Considering a paraxial monochromatic coherent beam propagating along z -axis, it can be expressed as

$$U(\mathbf{r}) \approx U(\mathbf{x}, z)e^{jkz} \quad (3)$$

The scalar complex amplitude of the paraxial optical field is denoted as $U(\mathbf{x}, z)$, which is a slowly varying function with respect to z . Note in Eq. (3), the time-dependent part $e^{-j\omega t}$ (has no effect on intensity) has been omitted, where ω is the angular frequency of the light wave. The paraxial wave equation can be obtained by substituting Eq. (3) into Helmholtz equation

$$\nabla^2 U(\mathbf{x}, z) + 2jk \frac{\partial U(\mathbf{x}, z)}{\partial z} = 0 \quad (4)$$

It is the propagation law that the complex amplitude of a paraxial monochromatic coherent beam must obey. Note ∇ in Eq. (4) is now defined over \mathbf{x} . Next, we substituted the scalar complex amplitude expression $U(\mathbf{x}, z) = \sqrt{I(\mathbf{x})} \exp[j\phi(\mathbf{x})]$ into the paraxial wave equation, and

separate the real part (in the following derivation, the spatial coordinate is omitted for notation simplicity)

$$\nabla(U^*\nabla U - U\nabla U^*) + 2jk\left(U^*\frac{\partial U}{\partial z} + U\frac{\partial U^*}{\partial z}\right) = 0 \quad (5)$$

where

$$\nabla U = jU\nabla\phi \quad (6)$$

$$U^*\nabla U = U^*jU\nabla\phi = jI\nabla\phi \quad (7)$$

$$U^*\nabla U - U\nabla U^* = 2jI\nabla\phi \quad (8)$$

$$j\left(U^*\frac{\partial U}{\partial z} + U\frac{\partial U^*}{\partial z}\right) = j\left(\frac{\partial}{\partial z}UU^*\right) = j\frac{\partial I}{\partial z} \quad (9)$$

Substituting Eqs. (8) and (9) into Eq. (5), we can deduce TIE

$$-k\frac{\partial I}{\partial z} = \nabla \cdot (I\nabla\phi) \quad (10)$$

It is a second order elliptic partial differential equation about the phase function ϕ . Similarly, “transport of phase equation” can be obtained from the imaginary part of the paraxial wave equation

$$-2kI^2\frac{\partial\phi}{\partial z} = \frac{1}{2}I\nabla^2 I - \frac{1}{4}(\nabla I)^2 - I^2(\nabla\phi)^2 + kI^2 \quad (11)$$

Because the phase is not directly accessible, this equation is difficult to be applied directly. We just give the expression of this equation here considering the completeness of this tutorial.

2.2.2. Derivation of TIE from Fresnel diffraction formula

The diffraction law of coherent light field can be accurately described in the spatial domain or in the spatial frequency domain from the viewpoint of “system”. In this subsection, we use the Fresnel diffraction formula to derive TIE in the spatial and frequency domain, respectively.

1) Using Fresnel diffraction integral to derive TIE in the spatial domain

Based on the Huygens-Fresnel principle, Kirchhoff [218] derived the rigorous Kirchhoff diffraction formula based on the Helmholtz equation and Green’s theorem, under certain boundary conditions at the diffraction plane. The basic idea is based on the principle of linear superposition: since any complex light source can be regarded as a combination of individual point sources, the complex light waves can always be decomposed into linear combinations of simple spherical waves. And the linearity of the wave equation allows each spherical wave to apply the above principles respectively. Then their contributions on the diffraction plane are superposed to create the resultant diffraction field. Considering that the complex amplitude of the light wave in 3D space is $U_0(\mathbf{r}) = a(\mathbf{r})e^{j\phi(\mathbf{r})}$ (in order to simplify the representation, we denote $\mathbf{r} = (\mathbf{x}, z)$ as the 3D spatial coordinate vector), the complex amplitude of the optical field obtained after propagating a distance of Δz can be expressed as

$$U_{\Delta z}(\mathbf{r}) = \int U_0(\mathbf{r}')h(\mathbf{r}', \mathbf{r})d\mathbf{r}' \quad (12)$$

The above formula shows that wave propagation can be regarded as a linear system, and the impulse response function of the system $h(\mathbf{r}', \mathbf{r})$ is exactly the complex amplitude on the diffraction plane generated by the spherical wavelet propagated from \mathbf{r}'

$$h_{\Delta z}(\mathbf{r}', \mathbf{r}) = \frac{1}{j\lambda}K(\theta)\frac{\exp(jk|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} \quad (13)$$

where $\frac{\exp(jk|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|}$ is an ideal spherical wave emitted from \mathbf{r}' and propagated to \mathbf{r} . $K(\theta) = \frac{1}{2}(1 + \cos\theta)$ is the *obliquity factor*, where $\cos\theta = \frac{\Delta z}{|\mathbf{r} - \mathbf{r}'|}$ is the angle between z -axis and the line connecting \mathbf{r} and \mathbf{r}' .

Eq. (12) suggests that the complex amplitude distribution in the diffraction plane is the weighted linear (interferential) superposition of spherical wavelets from all points \mathbf{r}' in the original plane. However, it is known that Kirchhoff’s boundary conditions lead to physical solutions containing a grave mathematical inconsistency. Rayleigh and Sommerfeld [219,220] solved this problem by deriving two possible integration formulas with use of two different Green’s functions under “physically correct” boundary conditions. The Rayleigh-Sommerfeld integrals formulas only differ slightly with Kirchhoff’s formula, the resulting obliquity factors become $K(\theta) = \cos\theta$ for the first Rayleigh-Sommerfeld integral (RS-I type), and $K(\theta) = 1$ for the second Rayleigh-Sommerfeld integral (RS-II type). A comparison of these obliquity factors reveals that the Kirchhoff’s solution is the arithmetic average of the two Rayleigh-Sommerfeld solutions [2,5]. The existence of obliquity factor suggests that the amplitude contribution in the original plane of each spherical wavelet emitted from each point may not be isotropic. However, if the diffraction angle is so small that the paraxial approximation is satisfied, $K(\theta) \approx 1$ (regardless its form), the effect of the obliquity factor can be neglected. Then the diffraction becomes a “linear shift invariant system”

$$U_{\Delta z}(\mathbf{r}) = U_0(\mathbf{r}) * h(\mathbf{r}) \quad (14)$$

The system is uniquely determined by its spatially invariant impulse response function in the spatial domain

$$h(\mathbf{r}) = \frac{1}{j\lambda} \frac{\exp(jk|\mathbf{r}|)}{|\mathbf{r}|} \quad (15)$$

Here we assume that the original plane and the diffraction plane are located at $(\mathbf{x}, 0)$ and (\mathbf{x}, z) , respectively. The expression of the impulse response function can be further simplified and expressed as the 2D scalar form

$$h_{\Delta z}(\mathbf{x}) = \frac{1}{j\lambda} \frac{\exp\left(jk\sqrt{\Delta z^2 + |\mathbf{x}|^2}\right)}{\sqrt{\Delta z^2 + |\mathbf{x}|^2}} \quad (16)$$

For a paraxial beam propagating along z -axis, the expression of the impulse response function can be further simplified as follows

$$h_{\Delta z}(\mathbf{x}) = \frac{1}{j\lambda\Delta z} \exp(jk\Delta z) \exp\left\{\frac{j\pi}{\lambda\Delta z}|\mathbf{x}|^2\right\} \quad (17)$$

Substituting Eq. (17) into Eq. (14), we can get the Fresnel diffraction integral [2]

$$U_{\Delta z}(\mathbf{x}) = \frac{\exp(jk\Delta z)}{j\lambda\Delta z} \int U_0(\mathbf{x}_0) \exp\left\{\frac{j\pi}{\lambda\Delta z}|\mathbf{x} - \mathbf{x}_0|^2\right\} d\mathbf{x}_0 \quad (18)$$

If in addition to the Fresnel approximation the stronger *Fraunhofer approximation* is satisfied, then the quadratic phase factor inside the integral of Eq. (18) is approximately unity, and the diffraction field can be found (up to a multiplicative phase factor in \mathbf{x}) directly from a Fourier transform of the original field. Thus in the region of Fraunhofer diffraction (or equivalently, in the far field), aside from multiplicative phase factors preceding the integral, the diffraction field is simply the Fourier transform of the original field, evaluated at frequencies $\mathbf{f} = \mathbf{x}/\lambda\Delta z$.

Next, we derive TIE based on Fresnel diffraction integral. Considering a complex amplitude $U_0(\mathbf{x}) = a(\mathbf{x})e^{j\phi(\mathbf{x})}$ of a paraxial monochromatic coherent beam propagating along the optical axis, the distance between the diffraction plane and the object plane is Δz . According to Eq. (18), the intensity distribution in the diffraction plane can be expressed as

$$\begin{aligned} I_{\Delta z}(\mathbf{x}) &= U_{\Delta z}(\mathbf{x})U_{\Delta z}^*(\mathbf{x}) = \frac{1}{\lambda^2\Delta z^2} \int U_0(\mathbf{x}_1)U_0^*(\mathbf{x}_2) \exp\left[\frac{j\pi}{\lambda\Delta z}(|\mathbf{x} - \mathbf{x}_1|^2 - |\mathbf{x} - \mathbf{x}_2|^2)\right] d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \frac{1}{\lambda^2\Delta z^2} \int U_0(\mathbf{x}_1)U_0^*(\mathbf{x}_2) \exp\left[\frac{j\pi}{\lambda\Delta z}(|\mathbf{x}_1|^2 - |\mathbf{x}_2|^2)\right] \exp\left(2\pi j\mathbf{x} \cdot \frac{\mathbf{x}_2 - \mathbf{x}_1}{\lambda\Delta z}\right) d\mathbf{x}_1 d\mathbf{x}_2 \end{aligned} \quad (19)$$

Taking Fourier transform on Eq. (19), we get the intensity spectrum in the spatial frequency domain [221]

$$\begin{aligned} \hat{I}_{\Delta z}(\mathbf{u}) &= \mathcal{F}\{I_{\Delta z}(\mathbf{x})\} = \frac{1}{\lambda\Delta z} \int U_0(\mathbf{x}_1)U_0^*(\mathbf{x}_2) \\ &\exp\left[\frac{j\pi}{\lambda\Delta z}(|\mathbf{x}_1|^2 - |\mathbf{x}_2|^2)\right] \mathcal{F}\left\{\exp\left(2\pi j\mathbf{x} \cdot \frac{\mathbf{x}_2 - \mathbf{x}_1}{\lambda\Delta z}\right)\right\} d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \frac{1}{\lambda\Delta z} \int U_0(\mathbf{x}_1)U_0^*(\mathbf{x}_2) \exp\left[\frac{j\pi}{\lambda\Delta z}(|\mathbf{x}_1|^2 - |\mathbf{x}_2|^2)\right] \delta\left(\mathbf{u} - \frac{\mathbf{x}_2 - \mathbf{x}_1}{\lambda\Delta z}\right) d\mathbf{x}_1 d\mathbf{x}_2 \\ &\stackrel{\mathbf{x}_1 = \mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}, \mathbf{x}_2 = \mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}}{=} \int U_0\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right)U_0^*\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right) \exp\{-2j\pi\mathbf{x} \cdot \mathbf{u}\} d\mathbf{x} \end{aligned} \quad (20)$$

Note that the constant scale factor in above formula is omitted. When the propagation distance $\Delta z \rightarrow 0$, the first-order Taylor expansion of $U_0\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right)$ can be written as

$$U_0\left(\mathbf{x} \pm \frac{\lambda\Delta z\mathbf{u}}{2}\right) \approx U_0(\mathbf{x}) \pm \frac{\lambda\Delta z\mathbf{u}}{2} \cdot \nabla U_0(\mathbf{x}) \quad (21)$$

Substituting Eq. (21) into Eq. (20) and only retaining the linear term of Δz [210,216]

$$I_{\Delta z}(\mathbf{x}) = I_0(\mathbf{x}) - \frac{\Delta z}{k} \nabla \cdot [I_0(\mathbf{x})\nabla\phi(\mathbf{x})] \quad (22)$$

When $\Delta z \rightarrow 0$, $\frac{I_{\Delta z}(\mathbf{x}) - I_0(\mathbf{x})}{\Delta z} \approx \frac{\partial I(\mathbf{x})}{\partial z}$, and TIE is obtained.

2) Using Fresnel angular spectrum diffraction formula to derive TIE in the spatial frequency domain

On the other hand, the diffraction of the coherent beam can also be described from the viewpoint of “system” in the spatial frequency domain: the scalar coherent field $U(x, y, 0)$ (assuming $z = 0$) can be decomposed into the coherent superposition of different angular spectrum (plane wave) components $\hat{U}(u_x, u_y, 0)$ in the 2D spatial frequency domain. It can be presented by the following 2D Fourier transform [2]

$$U(x, y, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{U}(u_x, u_y, 0) e^{j2\pi(u_x x + u_y y)} du_x du_y \quad (23)$$

$$\hat{U}(u_x, u_y, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x, y, 0) e^{-j2\pi(u_x x + u_y y)} dx dy \quad (24)$$

where the exponential term $e^{j2\pi(u_x x + u_y y)}$ represents a monochromatic plane wave with unit amplitude, and the direction cosine vector of its propagation direction is $(\cos \alpha, \cos \beta)$, i.e.,

$$u_x = \frac{\cos \alpha}{\lambda}, u_y = \frac{\cos \beta}{\lambda} \quad (25)$$

The geometric propagation of a plane wave $e^{j2\pi(u_x x)}$ parallel to y -axis is illustrated in Fig. 15. The angle between the propagation direction and x -axis is denoted as α .

After propagation of a distance Δz , the optical path travelled in free space and the corresponding phase delay can be represented as

$$\Delta d = \Delta z \sin \alpha = \Delta z \sqrt{1 - (\cos \alpha)^2} = \Delta z \sqrt{1 - (\lambda u_x)^2} \quad (26)$$

$$\phi = k\Delta d = k\Delta z \sqrt{1 - (\lambda u_x)^2} \quad (27)$$

Therefore, the plane wave after the propagation of a distance Δz is $e^{j2\pi(u_x x)} e^{j\phi}$. It can be seen from the Fourier transform that the point in the Fourier spectrum corresponding to this plane wave is also multiplied by an exponential term, i.e.,

$$\begin{aligned} \hat{U}(u_x, 0, \Delta z) &= \hat{U}(u_x, 0, 0) e^{j\phi} \\ &= \hat{U}(u_x, 0, 0) e^{jk\Delta z \sqrt{1 - (\lambda u_x)^2}} \end{aligned} \quad (28)$$

Generally, for plane waves in arbitrary directions, we can get

$$\hat{U}(u_x, u_y, \Delta z) = \hat{U}(u_x, u_y, 0) e^{jk\Delta z \sqrt{1 - (\lambda u_x)^2 - (\lambda u_y)^2}} \quad (29)$$

To sum up, here we reveal the geometric significance of the angular spectrum propagation: the scalar coherent optical field $U(x, y, 0)$

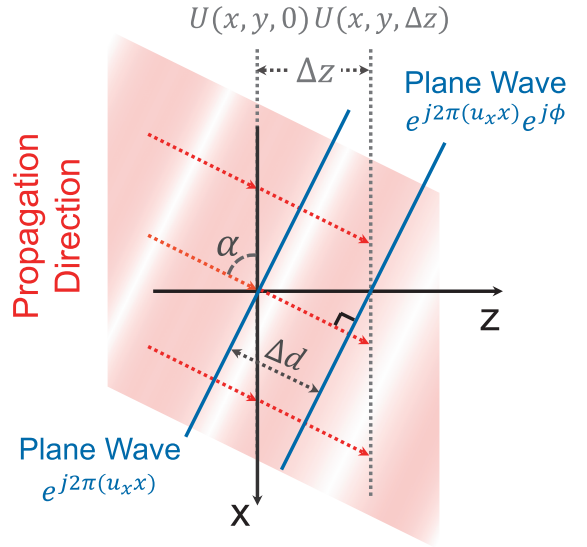


Fig. 15. The geometric illustration of the angular spectrum propagation.

located at the plane $z = 0$ can be regarded as the linear superposition of different plane wave components. After propagation of a distance Δz , the coherent field $U(x, y, z)$ is still composed of these plane waves propagating along their original directions. The amplitudes of these plane waves remain unchanged, but their optical paths traveled in each direction are different, i.e., their phase delays are different, which can be represented as $k\Delta z \sqrt{1 - (\lambda u_x)^2 - (\lambda u_y)^2}$. In the spatial frequency domain, each point in the Fourier spectrum is multiplied by an exponential term $e^{jk\Delta z \sqrt{1 - (\lambda u_x)^2 - (\lambda u_y)^2}}$, which is equivalent to the whole 2D spectrum multiplied by a 2D transfer function $H_{\Delta z}(u_x, u_y)$. Under the precondition of $1 - (\lambda u_x)^2 - (\lambda u_y)^2 > 0$, we get

$$H_{\Delta z}(u_x, u_y) = \frac{\hat{U}(u_x, u_y, \Delta z)}{\hat{U}(u_x, u_y, 0)} = e^{jk\Delta z \sqrt{1 - (\lambda u_x)^2 - (\lambda u_y)^2}} \quad (30)$$

From the perspective of linear shift invariant system, the propagation phenomenon may be regarded as a linear spatial filter with a finite bandwidth, i.e., $H_{\Delta z}(u_x, u_y)$, so so-called the *angular spectrum transfer function* [2,5]. The amplitude of $H_{\Delta z}(u_x, u_y)$ is zero outside a circular region of radius λ^{-1} in the frequency domain. Within that circular bandwidth, the modulus of the transfer function is unity but frequency-dependent phase shifts are introduced. It should be noted that as well as propagating plane waves, solutions of the wave equation are also given by evanescent waves (from the latin *evanescere*, to diminish). When $1 - (\lambda u_x)^2 - (\lambda u_y)^2 < 0$, the evanescent waves at high spatial frequencies diminishes exponentially with the increase of z . It decays to zero after the propagation of a distance about several wavelengths [2,5]. In most applications of optical imaging, evanescent waves have a very limited range and can therefore be neglected. It should be mentioned the remarkable fact that, Eq. (30) is proven to be the Fourier transform of the impulse response function [Eq. (13)] with the RS-I type obliquity factor, which suggests that the angular spectrum approach and the first Rayleigh-Sommerfeld integral are essentially identical in predictions of diffracted fields [2,5].

Under the paraxial approximation, the higher order terms in the Taylor expansion of the phase factor in Eq. (30) can be neglected

$$\begin{aligned} &\sqrt{1 - (\lambda u_x)^2 - (\lambda u_y)^2} \\ &= 1 - \frac{1}{2}\lambda^2(u_x^2 + u_y^2) + \frac{1}{8}\lambda^4(u_x^2 + u_y^2)^2 + \dots \\ &\approx 1 - \frac{1}{2}\lambda^2(u_x^2 + u_y^2) \end{aligned} \quad (31)$$

So we can get the angular spectrum transfer function under Fresnel (paraxial) approximation

$$H_{\Delta z}^F(u_x, u_y) = \exp(jk\Delta z) \exp\left[-j\pi\lambda\Delta z(u_x^2 + u_y^2)\right] \quad (32)$$

Similarly, it is not difficult to prove that Eq. (32) is the Fourier transform of the Fresnel impulse response function [Eq. (17)]. In Eq. (32), the first phase factor $\exp(jk\Delta z)$ is only related to the propagation distance Δz , so it can be neglected because it has no effect on the intensity and complex amplitude distribution. Furthermore, when the propagation distance $\Delta z \rightarrow 0$, the Fresnel angular spectrum transfer function can be further rewritten as

$$\begin{aligned} H_{\Delta z}^F(\mathbf{u}) &= e^{-j\pi\lambda\Delta z|\mathbf{u}|^2} \\ &= 1 - j\pi\lambda\Delta z|\mathbf{u}|^2 - \frac{(\pi\lambda\Delta z)^2|\mathbf{u}|^4}{2} - \dots \\ &\stackrel{\Delta z \rightarrow 0}{\approx} 1 - j\pi\lambda\Delta z|\mathbf{u}|^2 \end{aligned} \quad (33)$$

where \mathbf{u} is the spatial frequency vector (u_x, u_y). Considering the complex amplitude of an optical field to be recovered as $U(\mathbf{x}) = a(\mathbf{x})e^{j\phi(\mathbf{x})}$, its Fourier transform is $\hat{U}(\mathbf{u}) = \mathcal{F}[U(\mathbf{x})]$. So the intensity at a distance Δz can be expressed as

$$\begin{aligned} I(\mathbf{x}, \Delta z) &= \left| \mathcal{F}^{-1}\{\hat{U}(\mathbf{u})H_{\Delta z}^F(\mathbf{u})\} \right|^2 \\ &\stackrel{\Delta z \rightarrow 0}{\approx} \left| \mathcal{F}^{-1}\{\hat{U}(\mathbf{u})(1 - j\pi\lambda\Delta z|\mathbf{u}|^2)\} \right|^2 \\ &= \left| U(\mathbf{x}) - j\pi\lambda\Delta z\mathcal{F}^{-1}\{|\mathbf{u}|^2\hat{U}(\mathbf{u})\} \right|^2 \end{aligned} \quad (34)$$

Based on the differential properties of Fourier transform, the multiplicative frequency vector \mathbf{u} in the Fourier domain is equivalent to the gradient operator in the spatial domain, i.e., $j2\pi\mathbf{u} \rightarrow \nabla$. So the Laplacian operator can be represented as $\Delta = \nabla^2 \rightarrow -4\pi^2|\mathbf{u}|^2$. Substituting it into Eq. (34), we can get

$$\begin{aligned} I(\mathbf{x}, \Delta z) &= \left| U(\mathbf{x}) + \frac{j\pi\lambda\Delta z}{4\pi^2}\nabla^2 U(\mathbf{x}) \right|^2 \\ &= I(\mathbf{x}) - \frac{\Delta z}{k}\nabla \cdot [I(\mathbf{x})\nabla\phi(\mathbf{x})] \end{aligned} \quad (35)$$

When the propagation distance $\Delta z \rightarrow 0$, $\frac{I(\mathbf{x}, \Delta z) - I(\mathbf{x})}{\Delta z} \approx \frac{\partial I(\mathbf{x})}{\partial z}$, and TIE is obtained.

2.2.3. Derivation of TIE based on Poynting theorem

Poynting theorem [4] is in the nature a statement of conservation of energy for the electromagnetic field. In physics, the Poynting vector represents the directional energy flux of an electromagnetic field, and Poynting theorem states that the total dissipated energy flow through a closed surface s of a bounded volume V in the electromagnetic field is equal to the change over time of the complete electromagnetic energy inside V . Since energy may be neither created nor destroyed in free space, the time-averaged Poynting vector \mathbf{S} must obey a conservation equation expressed as

$$\oint_s \langle \mathbf{S} \rangle ds = 0 \quad (36)$$

Its corresponding differential form is

$$\nabla \cdot \langle \mathbf{S} \rangle = 0 \quad (37)$$

where ∇ is the Hamiltonian in 3D space, $\langle \mathbf{S} \rangle$ is the vectorial representation of time-averaged Poynting vector in 3D space. The time-averaged Poynting vector corresponds to the intensity (irradiance in definitions of radiometry) or flux density I of a surface element perpendicular to the Poynting vector. When the vector potentials of electromagnetism is neglected (in the absence of phase vortices), the time-averaged Poynting vector can be written in the form [159]

$$\langle \mathbf{S} \rangle = \frac{1}{k} I(\mathbf{x}, z) \nabla\phi(\mathbf{x}, z) \quad (38)$$

Transport-of-intensity equation (TIE) = Energy conservation law

$$\underbrace{-k \frac{\partial I(x, y)}{\partial z}}_{\text{Axial intensity derivative}} = \underbrace{\nabla \cdot [I(x, y) \nabla\phi(x, y)]}_{\text{Transverse energy}}$$

$$\approx \frac{I_{+\Delta z}(x, y) - I_{-\Delta z}(x, y)}{2\Delta z} \quad \text{Finite difference}$$

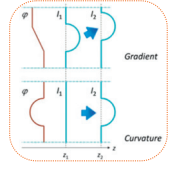
$$\nabla I(x, y) \nabla\phi(x, y) + I(x, y) \nabla^2\phi(x, y)$$


Fig. 16. The physical implication of TIE.

Under the paraxial approximation, the time-averaged Poynting vector can be explicitly decomposed into the transverse component and the longitudinal component

$$\langle \mathbf{S} \rangle = \frac{1}{k} I(\mathbf{x}, z) [\nabla\phi(\mathbf{x}) + k\mathbf{z}_0] \quad (39)$$

where \mathbf{z}_0 represents the unit vector along the z -direction. Substituting it into Eq. (37), we can get TIE

$$-k \frac{\partial I(\mathbf{x}, z)}{\partial z} = \nabla \cdot [I(\mathbf{x}, z) \nabla\phi(\mathbf{x})] \quad (40)$$

It should be noted that the time-averaged Poynting vector in Eq. (32) is well defined for both coherent fields and partially coherent fields. Paganin and Nugent [159] use this formulation as the definition of “phase”, which is a scalar potential of the time-averaged Poynting vector. If the wave is coherent, then this is equivalent to the conventional phase.

2.3. The physical implication of TIE

Before entering the next section about the solution of TIE, let us first examine this beautiful equation. Viewed as a whole, TIE is essentially an expression of the energy conservation law. Under the paraxial approximation, the longitudinal energy flux of the optical field can be approximated by the intensity. Transverse energy flux is determined by $I\nabla\phi$, and $I\nabla\phi$ represents the transverse component of the time-averaged Poynting vector, as shown in Eq. (39). Its divergence $\nabla \cdot (I\nabla\phi)$ represents for the total of energy variation in the transverse plane. Longitudinal and transverse energy dissipation should always be equal according to the energy conservation law (as shown in Fig. 16). Examining both sides of TIE, the left hand side (LHS) is the axial intensity derivative. In the simplest situation, it can be estimated by the finite difference between two intensity images collected at different propagation distances along the optical axis. More details about the axial intensity derivative estimation will be given in Section 5. Expanding the right hand side (RHS) of TIE, we obtain

$$-k \frac{\partial I}{\partial z} = \nabla \cdot (I\nabla\phi) = \nabla I \cdot \nabla\phi + I\nabla^2\phi \quad (41)$$

where the two terms contain the phase gradient (slope, the first derivative) term and phase curvature (the second derivative), respectively. Axial variation of intensity is determined by both phase slope and phase curvature. Phase slope induces the intensity translation, just like a prism, while phase curvature induces intensity convergence or divergence, just like a lens. So these two terms are also called prism term (slope and gradient) and lens term (curvature) [143] (as illustrated in Fig. 16).

At last, let us consider some special cases of TIE: if the phase is constant, the RHS of TIE is equal to 0, so there will be no transport of intensity effect. It is in accordance with the plane wave case discussed above. Another more meaningful situation is when the in-focus intensity is uniform $I \approx \text{constant}$, corresponding to the situation of a pure phase object, TIE can be simplified as

$$-k \frac{\partial I}{\partial z} = I\nabla^2\phi \quad (42)$$

This is a standard Poisson equation, revealing that a weak defocusing of a pure phase object produces intensity contrast that is proportional to

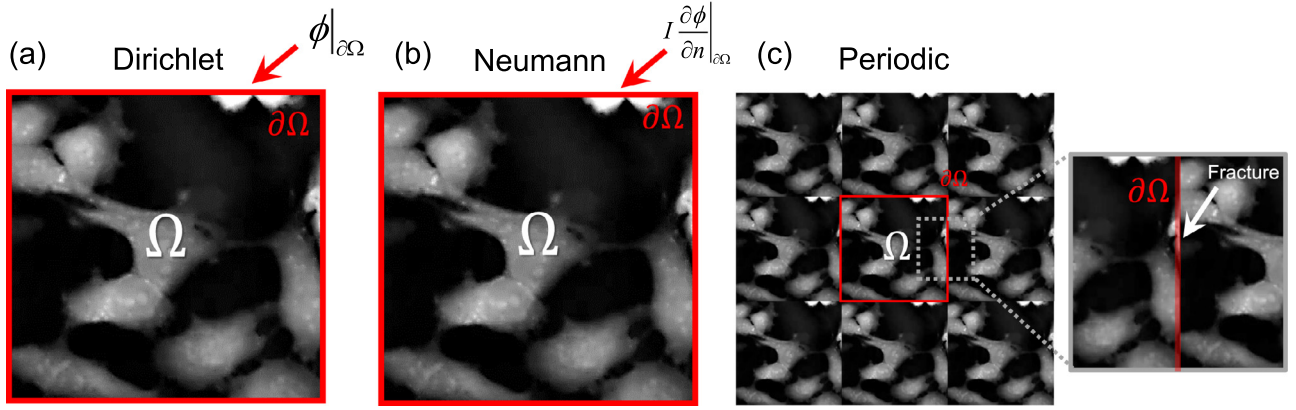


Fig. 17. Three typical boundary conditions used in TIE solvers.

the phase Laplacian (curvature). This is exactly the physical implication of the light-and-dark network structure at the bottom of the swimming pool in Fig. 13.

3. Solutions to TIE

In this section, we will discuss how to solve TIE, which is the fundamental premise for the success of TIE phase retrieval. In TIE, the intensity I on the RHS can be directly measured, and the LHS is the intensity derivative along the optical axis $\partial I/\partial z$ which cannot be directly measured. Fortunately, it can be approximated by finite differences with a minimum of two defocused intensity images taken at distinct planes separated by a small distance orthogonal to the optical axis (discussed in Section 5). Thus, the only one unknown quantity in TIE is the phase function ϕ . When the intensity I and intensity derivative $\partial I/\partial z$ are known, solving the phase function is a typical inverse problem.

3.1. Boundary conditions of TIE

There are two prerequisites for uniquely determining a solution to a partial differential equation: the partial differential equation and boundary conditions [222]. Thus, the TIE phase retrieval is essentially a boundary value problem: seeking a solution to TIE that also satisfies the given boundary conditions. We then adopt a more rigorous mathematical description to express this problem: we assume the region governed by TIE to be a general open and bounded domain $\Omega \subset \mathbb{R}^2$ with a piecewise smooth boundary $\partial\Omega$. The intensity distribution I is a continuous, nonnegative function defined on the enclosure $\bar{\Omega}$ (including the region inside Ω and the boundary $\partial\Omega$), and is continuous and strictly positive in Ω . The axial intensity derivative $\partial I/\partial z$ is assumed to be a continuous function in Ω . The phase ϕ is expected to be single-valued and smooth in $\bar{\Omega}$ (in the absence of phase vortices). In the following part, we shall consider three classes of possible boundary conditions that could be applied to solve TIE (see Fig. 17):

1) Dirichlet boundary conditions

The values of phase function ϕ are specified on the domain boundary

$$\phi|_{\partial\Omega} = g \quad (43)$$

Here g is a smooth function on the boundary $\partial\Omega$.

2) Neumann boundary conditions

The product of the normal derivative of ϕ is specified on the domain boundary

$$I \frac{\partial \phi}{\partial n} \Big|_{\partial\Omega} = g \quad (44)$$

Here g is a smooth function on the boundary $\partial\Omega$, and $\partial\phi/\partial n$ is the outward normal derivative.

3) Periodic boundary conditions

The phase at the boundary repeats cyclically, e.g., for a rectangular domain, the phase at the right boundary is the same as the phase at the left boundary.

3.2. Well-posedness and uniqueness of the solution

The “well-posedness” and the “uniqueness” of the solution are two essential problems of the TIE phase retrieval. The well-posedness means that the equation must exist at least one solution to ensure the solvability of the problem. The uniqueness of the solution means that the solution of the equation must be unique. Obviously, the phase of the object to be measured is well and uniquely defined. If the solution to TIE is not unique, it will be impossible to determine whether the obtained solution corresponds to the true phase of the object. In 1995, the well-posedness and uniqueness of the TIE solution were first proven by Gureyev et al. [223]: with the determined intensity I , axial intensity derivative $\partial I/\partial z$, and boundary conditions specified on $\partial\Omega$, the phase can be uniquely determined by solving the corresponding boundary value problem. For the case of Dirichlet boundary conditions, the solution to TIE always exists and is unique [222]. The case of Neumann boundary conditions demands special attention because a solution may or may not exist (depending on whether the compatibility condition [Eq. (45)] is satisfied [223,224]). The compatibility condition can be obtained by integrating both sides of TIE over the region Ω and then applying the divergence theorem

$$\oint_{\partial\Omega} I(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial n} ds = \iint_{\Omega} -k \frac{\partial I(\mathbf{x})}{\partial z} d\mathbf{x} \quad (45)$$

where s is a parameterization of $\partial\Omega$, and $\partial I/\partial z$ can be obtained by actual measurement. Thus, the RHS of Eq. (45) is known *a priori*, since the function $\partial I/\partial z$ can be experimentally obtained, and the LHS is determined solely by Neumann boundary conditions [Eq. (44)]. Mathematically, all these possible Neumann boundary conditions must satisfy Eq. (45) in order for there to be a nontrivial solution to TIE. In addition, notice that if ϕ is a solution to the Neumann boundary problem, then $\phi + C$ is also a solution for an arbitrary constant C . In other words, the solution of TIE subject to Neumann boundary conditions should be unique up to an arbitrary additive constant, assuming that the solution exists in the first place. This constant is not essential for the phase retrieval problem. The proof of the uniqueness theorem is straightforward: we can first assume that two qualified solutions have been exhibited and then prove they can only differ from one additive constant using the maximum principle for elliptic equations [225]. In general, the periodic boundary condition is widely adopted in the fields of lattice vibrations and electron

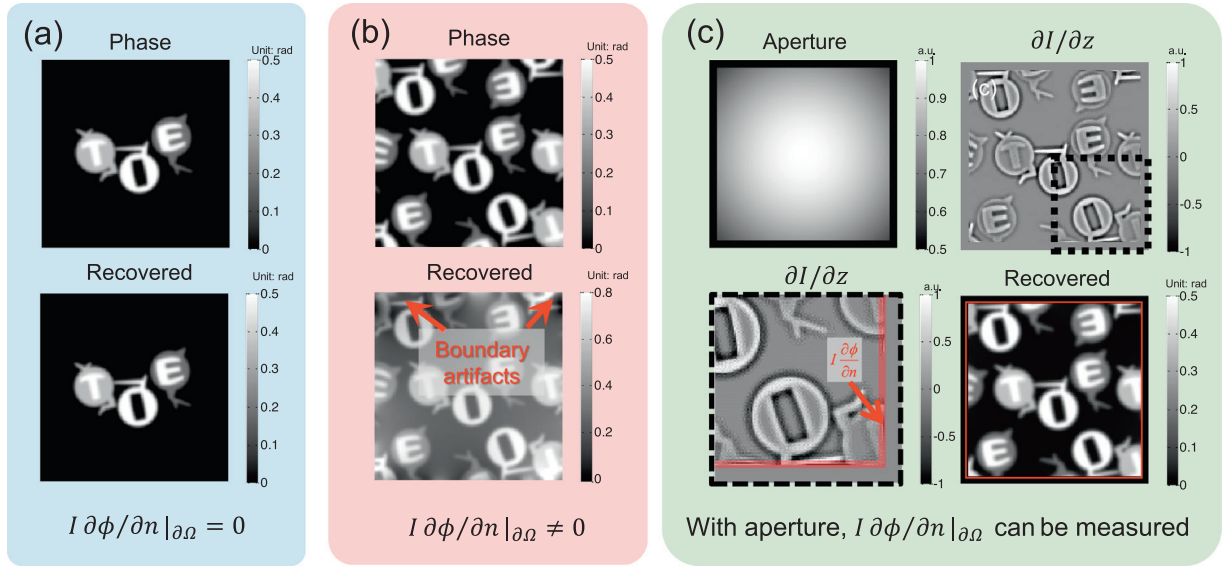


Fig. 18. Phase retrieval simulations for different types of objects. (a) An isolated object located in the central field of view (FOV) (the FFT-based solver gives accurate reconstruction); (b) a complex object extending outside the image boundary (the FFT-based solver produces large boundary artifacts); and (c) the DCT solver with a hard aperture (the inhomogeneous boundary conditions can be measured at the boundary, which produces accurate phase reconstruction even if the object is located at the aperture boundary).

theory of metals, wherein most objects of interest are characterized by periodicity. In most cases of QPI, the samples are not periodically distributed, so periodic boundary conditions are difficult to meet. The purpose of introducing periodic boundary conditions into the solution of TIE is to simplify the numerical implementation, which will be detailed in Subsection 3.3. For periodic boundary conditions, the solution to TIE is also unique up to an arbitrary additive constant.

In summary, regarding the well-posedness and uniqueness of the solution to TIE, the following conclusions can be drawn:

- 1) For Dirichlet boundary conditions, the solution to TIE always exists and is unique.
- 2) For Neumann boundary conditions, the solution to TIE may or may not exist, depending on whether the compatibility condition of [Eq. (45)] is satisfied. When the compatibility condition is satisfied, the solution to TIE exists and is unique up to an arbitrary additive constant.
- 3) For periodic boundary conditions, the solution to TIE always exists and is unique up to an arbitrary additive constant.

3.3. Compatibility condition and energy conservation law

It is very interesting to examine the physical picture described by the compatibility condition [Eq. (45)]: actually, it can be thought as an expression of local energy conservation law in bounded domains: the loss of energy (intensity) inside the domain arising from energy flow across the boundary. If the boundary condition does not satisfy Eq. (45), it is not physically grounded so that the correct solution cannot be obtained. If we extend the bounded domain Ω to unbounded free space, the contour integral vanishes and the compatibility condition becomes

$$\iint_{\mathbb{R}^2} \frac{\partial I(\mathbf{x})}{\partial z} d\mathbf{x} = 0 \quad (46)$$

It represents the law of energy conservation for unlimited free space. In essence, the energy conservation expressed by Eqs. (45) and (46) is a universal law of physics, so it imposes an implicit physical constraint on the solution of TIE.

3.4. Solutions to TIE

The essence of TIE retrieval is to solve the partial differential equation under appropriate boundary conditions, and how to solve TIE accurately and efficiently has become one of the most popular and challenging research directions since its inception. In the paper of Teague [141] in 1983, he tried to simplify TIE into standard Poisson equations by introducing an auxiliary function. This auxiliary function ψ is called “Teague’s auxiliary function”, which satisfies

$$I(\mathbf{x}) \nabla \phi(\mathbf{x}) = \nabla \psi(\mathbf{x}) \quad (47)$$

The physical meaning of this auxiliary function has been discussed in Subsection 2.2. $I \nabla \phi$ can be interpreted as the time-averaged transverse Poynting vector [159] and can be characterized by the scalar potential ψ . With this auxiliary function ψ , TIE can be simplified into two standard Poisson equations [76,141,201]

$$\frac{\partial I(\mathbf{x})}{\partial z} = -\frac{1}{k} \nabla^2 \psi(\mathbf{x}) \quad (48)$$

$$\nabla \cdot [I(\mathbf{x})^{-1} \nabla \psi(\mathbf{x})] = \nabla^2 \phi(\mathbf{x}) \quad (49)$$

The Teague’s auxiliary function ψ can be solved from Eq. (48), and then substituted into Eq. (49), the phase function can be uniquely determined by phase integration (up to an arbitrary additive constant). In order to solve these two Poisson equations, Teague [141] introduced the Green’s function method to derive the first analytical solution to TIE, although, to our knowledge, this has never been implemented. After that, various numerical TIE solvers have been proposed, such as the multi-grid method [76,226–228], the Zernike polynomial expansion method [156,229,230], the FFT-based method [76,149,158,159,231,232], and the discrete cosine transform (DCT) method [201–203]. Multi-grid method is a common method to solve partial differential equations, which is mainly used for the rectangular domain and relatively complicated to implement [76,226–228]. Since Zernike polynomials are orthogonal functions defined on the unit circle, the Zernike polynomial expansion method is used primarily in wavefront sensing and optical aberration characterization [156,229,230]. The FFT-based method is probably the most popular TIE solvers especially in the field of QPI for its simplicity, efficiency, and applicability for the case of non-uniform intensity distribution. The present widely used version of the FFT solver

is developed by Paganin and Nugent [159], which can be considered a generalization of the Gureyev and Nugent's approach [158] as it reduces to that form in the case of nonuniform intensity. Under general nonuniform intensity, it takes the following form [159]

$$\phi(\mathbf{x}) = -k\nabla^{-2}\nabla \cdot \left[I^{-1}(\mathbf{x})\nabla\nabla^{-2}\frac{\partial I(\mathbf{x})}{\partial z} \right] \quad (50)$$

where ∇^{-2} is the inverse Laplacian operator, denoting a simplified representation of the solution to the corresponding Poisson equation. Note that the inverse Laplacian operator is effectively implemented with FFT based on differential properties of the Fourier transform ($\nabla^{-2} \rightarrow -\frac{1}{4\pi^2|\mathbf{u}|^2}$) [76,159,232].

3.4.1. Solving TIE without boundary values

Despite the mathematical well-posedness, the rigorous implementation of TIE phase retrieval tends to be difficult because the associated boundary conditions are difficult to measure or to know *a priori*. For example, in Teague's Green function solution [141], one needs to know the phase value at the region boundary as the Dirichlet boundary condition, which is lack of feasibility because the phase function is exactly what we want to measure. Generally speaking, it is difficult to obtain any prior information about the measured object before taking any measurement (Teague recommends that the boundary value can be measured by using a Shack-Hartmann wavefront sensor, which is indeed very inconvenient for practical implementation). Parvizi *et al.* [233] pointed out that the Dirichlet boundary condition can be obtained by manually selecting the "smooth region" inside the FOV (the region without the object is considered 'flat' and has a constant phase value). However, this method also needs the prior information about the object and requires human intervention. The FFT-based method does not need to impose the boundary condition explicitly, but it assumes that the finite field is periodic and repetitive. This implies periodic boundary conditions due to the cyclic nature of the discrete Fourier transform. However, this situation is rather restrictive and does not reflect general experimental conditions. When the actual experimental situation violates those imposed assumptions, *e.g.*, objects are located at the image borders, severe boundary artifacts will appear, seriously affecting the accuracy of the phase reconstruction [201,202,234].

Because the boundary signals are not easy to acquire experimentally, some researchers sought new ways to solve TIE without taking any boundary measurements. Coincidentally, all these efforts aim to nullify the LHS of Eq. (45), making boundary conditions unnecessary

$$\oint_{\partial\Omega} I(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial n} ds = 0 \text{ or equivalently } \iint_{\Omega} \frac{\partial I(\mathbf{x})}{\partial z} d\mathbf{x} = 0 \quad (51)$$

This is actually equivalent to applying the energy conservation law for unbounded space to a bounded domain directly. Generally, Eq. (51) does not hold for a bounded region unless the contour integral [LHS of Eq. (45)] vanishes. Physically, it is equivalent to the special case that there is no overall energy transfer through the domain boundary. To solve TIE without boundary values, there are three common ways to satisfy Eq. (51):

1) "Soft-edged" illumination

The soft-edged illumination suggested by Gureyev and Nugent [157] represents a special case that the intensity is smooth and strictly greater than 0 inside the domain ($I > 0$) but strictly vanishes at the boundary $I_{\partial\Omega} = 0$. In this case, they proven that the solution to TIE exists and is unique to an arbitrary additive constant. Although Gureyev and Nugent [157] called this special case as a kind of "unconventional" boundary condition, by comparing it with Eq. (44), it can be found that it just corresponds to a special case of the homogeneous Neumann boundary conditions ($I\partial\phi/\partial n|_{\partial\Omega} = 0$). The disadvantage of this method is that defining the boundary of the domain tends to be very difficult in practice since "very-near zero" and "at zero" may seem almost the same but can be entirely different *vis-a-vis* the well-posed TIE.

2) "Flat" boundary

This is the simplest and the most common way to bypass the difficulty associated with the boundary condition problem. One can place the sample carefully so that it is centrally isolated in the image FOV, surrounded by an unperturbed plane wave, so that the boundary of the image is "empty" (sample-free). Then the phase at the boundary can be considered as "flat", and zero phase changes at the boundary (corresponding to homogeneous Neumann boundary conditions) can be assumed thus the Eq. (51) can be satisfied. In fact, in this case one can define not only homogeneous Neumann boundary conditions (constant phase $\phi|_{\partial\Omega} = C$ at the boundary), but any homogeneous Dirichlet boundary conditions ($I\partial\phi/\partial n|_{\partial\Omega} = 0$) or even periodic ones (the phase at the boundary repeats cyclically). In such a case, the phase can be recovered accurately regardless of the selected boundary conditions, for example, using the FFT-based method with periodic boundary conditions. However, this configuration does not reflect general experimental conditions, and is impractical when the object is larger than the camera FOV. This kind of boundary error is problematic when dealing with an extended object that must cover the image boundary, *e.g.*, microlens arrays [201,202,234].

3) Mirror padding

This method is proposed by Volkov *et al.* [231] to solve TIE without boundary value measurements. It has no additional requirements about the test object and experimental conditions, but only needs to symmetrically extend the intensity images into a four times larger size so that the homogeneous Neumann boundary conditions can be satisfied ($\partial\phi/\partial n|_{\partial\Omega} = 0$), and the LHS of Eq. (45) vanishes. Although Volkov *et al.* [231] demonstrated that this method performs better than the traditional FFT-based method in certain circumstances, it is only a pure mathematical trick to nullify the energy flow across the boundary through appropriate symmetrization of input images. It assumes there is no energy dissipation through the image boundary, which generally does not satisfy the compatibility condition [Eq. (45)], and thus, is not physically grounded. Note severe boundary errors can be observed when objects are located at the image boundary. Such kind of artifacts are not just limited to the boundary region but may propagate inside the domain and degrade the reconstruction accuracy prevalingly [201,202].

3.4.2. Solving TIE with boundary value measurements

The simplified solutions to TIE discussed in Subsection 3.4.1 can solve TIE without boundary value measurements, but it does not mean that we can solve TIE without imposing any boundary conditions. To be more specific, we have to restrict the test objects or experimental conditions to certain "implicit boundary conditions". On the other hand, for certain phase functions (such as tilt, defocus, astigmatism, and other phase aberrations with zero or constant curvature), the defocus-induced intensity derivative signals are all concentrated at the boundary region. If the boundary conditions are not considered, the phase can never be recovered [235–237]. The only solution to issues mentioned above is to solve TIE with correct boundary conditions, and the key problem is how to obtain the boundary signals experimentally. Roddier's research group [145,146] first explored this problem and found that the boundary values required for the simplified version of TIE [Eq. (42)] with uniform intensity can be obtained from the intensity measurements at the pupil boundary. Because the boundary signals can be acquired experimentally, Roddier *et al.* [145,146] used the SOR and the iterative Fourier transform [149] to solve the simplified TIE effectively. Later this method was adopted by Woods and Greenaway [238], and the Green's function method was used to solve the simplified TIE. Although Roddier's method addresses the problem of boundary value measurements, it is still necessary to distinguish the boundary signals from the interior intensity derivative, which has been known to cause serious difficulties [147,156,235,239]. Moreover, it relies on one major assumption: the intensity distribution should be uniform within the domain ($I \approx \text{constant}$).

This may be a reasonable assumption in the field of adaptive optics, but it is difficult to satisfy in the field of QPI, especially for objects with strong absorptions.

To solve the above-mentioned problems, Zuo et al. [201] found that even if the intensity distribution is non-uniform, the inhomogeneous Neumann boundary conditions can still be obtained when a hard-edged aperture (the aperture function $A_\Omega = 1$ when $\mathbf{r} \in \Omega$, $A_\Omega = 0$ when $\mathbf{r} \notin \Omega$) is introduced at the object plane ($z = 0$ or its conjugated plane). Substituting the aperture function into TIE, the intensity transport can be written in the following form

$$-k \frac{\partial I}{\partial z} = A_\Omega (I \nabla^2 \phi + \nabla I \cdot \nabla \phi) - I \frac{\partial \phi}{\partial n} \delta_{\partial\Omega} \quad (52)$$

where $\delta_{\partial\Omega}$ is the Dirac delta function around the aperture edge. Eq. (52) suggests that the axial intensity derivative signals consist of two non-overlapping components:

- 1) The intensity variation inside the domain due to the phase slope and curvature as if the aperture is not present.
- 2) A delta-function-like signal sharply peaked at the aperture boundary, which provides the exact Neumann boundary conditions for TIE [Eq. (44)].

Since the whole axial intensity derivative [the LHS of Eq. (52)] is experimentally measurable through finite-difference scheme (note the aperture should be smaller than the image FOV so that all the boundary signals can be captured), and the two RHS terms do not overlap in space, there is enough information to solve TIE uniquely without requiring prior knowledge of the boundary conditions. The inhomogeneous boundary value problem has been proven to be well-posed because it automatically satisfies the following compatibility condition [derived from integrating both sides of Eq. (52) over the whole space]

$$\iint_{\Omega} -k \frac{\partial I(\mathbf{x})}{\partial z} d\mathbf{x} = \iint_{\Omega} -k \frac{\partial I(\mathbf{x})}{\partial z} d\mathbf{x} - \oint_{\partial\Omega} I(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial n} ds = 0 \quad (53)$$

This is the same expression as the compatibility condition of the Neumann boundary problem [Eq. (45)], which leads to the inescapable conclusion that the solution to this Neumann boundary problem always exists and is unique up to an arbitrary additive constant. On this basis, Zuo et al. [201] developed the DCT-based TIE solver using experimentally measurable boundary conditions. The solution takes the following form

$$\phi(\mathbf{x}) = -k \nabla_{DCT}^{-2} \nabla_{DCT} \cdot \left[I^{-1}(\mathbf{x}) \nabla_{DCT} \nabla_{DCT}^{-2} \frac{\partial I(\mathbf{x})}{\partial z} \right]. \quad (54)$$

Eq. (54) is very similar to the FFT-based method proposed by Paganin and Nugent [Eq. (50)] [158,159], and the only difference is that the inverse Laplacian operator ∇_{DCT}^{-2} and the gradient operator ∇_{DCT} are implemented by the DCT rather than FFT (because the eigenfunctions of the Green's function associated with this Neumann boundary value problem is the Fourier cosine harmonics, rather than the general Fourier exponential harmonics). In addition, it should be noted that all the source data and related computations must be strictly defined on the closed rectangular region Ω , which includes both the aperture boundary and the region inside it (so that all the boundary signals can be enclosed in $\bar{\Omega}$). In this way, the measured intensity derivative can be treated as one entity without requiring special-purpose detection schemes to explicitly extract the boundary signals. Therefore, this approach solves a series of problems associated with the generation, acquisition, and separation of boundary signals in actual experimental conditions. The effectiveness of this method has been experimentally verified by the characterization of micro-optical elements [202]. Despite its success in solving TIE under inhomogeneous Neumann boundary conditions, the DCT-based solver still has two limitations:

- 1) The fast DCT algorithm is limited to the rectangular region, and thus, the DCT-based solver cannot be directly applied to non-rectangular (such as circle, ring, etc.) or irregular regions.

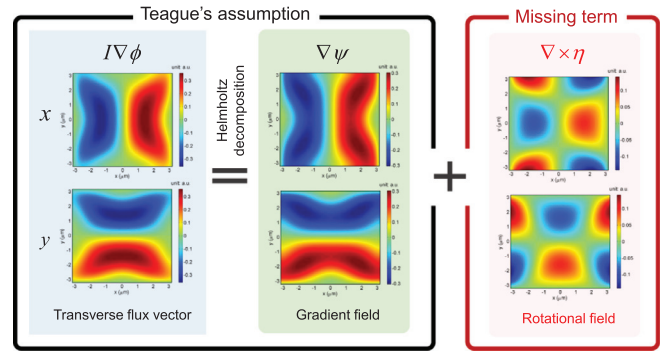


Fig. 19. Helmholtz decomposition of the transverse flux field. The x and y components of the vector fields are shown in the first row and the second row, respectively. The term $\nabla \times \eta$ is missing in Teague's assumption.

- 2) The effectiveness of the method relies on the hard-edged aperture. For the case of “soft-edged” aperture such as a Gaussian beam, although the energy conservation law can be satisfied, the algorithm is still not applicable in this case.

To address these problems, Huang et al. [203] further extended the DCT solver to an arbitrarily shaped aperture by an iterative compensation mechanism. This method takes the DCT-based solution as the initial value, and then it is refined by the iterative compensation algorithm to obtain an accurate solution. Thus, the DCT-based method can be extended to an arbitrarily shaped aperture. Ishizuka et al. [240,241] successfully applied the iterative DCT solver to recover the additional phase term corresponding to the curvature of field on the image plane in TEM. A similar iterative TIE solver based on FFT was also proposed by Mehrabkhani et al. [242].

3.5. Phase discrepancy analysis and compensation

Another important issue regarding the solution of TIE is “phase discrepancy”, which represents the difference between the ground truth phase and the phase retrieval by TIE. This is because most of TIE solvers, such as FFT and DCT, introduce the Teague's auxiliary function [141] [Eq. (47)] to simplify TIE into standard Poisson equations. The Teague's assumption suggests that the transverse flux is conservative so that a scalar potential ψ exists and satisfies Eq. (48). However, it is important to remark that the Teague's auxiliary function does not always exist in practical situations since the transverse energy flux may not be conservative [$I(\mathbf{x})\nabla\phi(\mathbf{x})$ is an ordinary 2D scalar field]. According to the Helmholtz's theorem, the transverse flux can be decomposed in terms of the gradient of a scalar potential ψ and the curl of a vector potential η

$$I(\mathbf{x})\nabla\phi(\mathbf{x}) = \nabla\psi(\mathbf{x}) + \nabla \times \eta(\mathbf{x}) \quad (55)$$

With certain (Dirichlet, Neumann, or periodic) boundary conditions, this decomposition is unique (or unique up to a vectorial constant that may float between the two components). Compared with Eq. (47), it is plain to see that the term $\nabla \times \eta(\mathbf{x})$ is ignored in Teague's assumption, making a silent hypothesis that the transverse flux $I(\mathbf{x})\nabla\phi(\mathbf{x})$ is irrotational (see Fig. 19). However, it is important to remark that the Teague's auxiliary function does not always exist in practical situations since the transverse energy flux may not be conservative, and consequently it would produce results that would not adequately match the exact solution.

In fact, Allen and Oxley [76] had noticed this phase discrepancy problem as early as 2001. Ten years later, Schmalz et al. [244] made a detailed theoretical analysis on the phase discrepancy problem based on Helmholtz decomposition theorem. They also provided a simulation example to demonstrate that the phase discrepancy induced by Teague's auxiliary function can be significant [mean square error (MSE) larger

than 9%]. In 2014, Zuo *et al.* [204] examined the effect of the missing rotational term on phase reconstruction and derived the necessary and sufficient condition for the validity of Teague's assumption

$$\nabla I(\mathbf{x})^{-1} \times \nabla^{-2} \{ \nabla \cdot [\nabla I(\mathbf{x}) \times \nabla \phi(\mathbf{x})] \} = 0 \quad (56)$$

Eq. (56) shows that if the in-focus intensity distribution is nearly uniform, the phase discrepancy resulting from the Teague's auxiliary function is quite small and negligible. This is why the issue of "phase differences" has not received enough attention before. However, when the measured sample exhibits strong absorption, the phase discrepancy may be relatively large and cannot be neglected, just as shown in the paper of Schmalz *et al.* [244]. To compensate the phase discrepancy owing to Teague's assumption, Zuo *et al.* [204] further developed a Picard-type iterative algorithm, in which the phase is gradually accumulated until a self-consistent solution is obtained. A similar iterative compensation method was also proposed by Shanker *et al.* [245]. Within two to four iterations, the phase discrepancy can be reduced to a negligible level, and the exact solution to TIE can be thus obtained.

At the end of this section, we can conclude that there are at least four issues need to be addressed for a desired TIE solver:

- 1) It should account for inhomogeneous boundary conditions with experimentally measured boundary signals.
- 2) It provides an accurate solution that is free from phase discrepancy.
- 3) For periodic boundary conditions, the solution to TIE always exists and is unique up to an arbitrary additive constant.
- 4) It should be computationally efficient and strictly convergent (if it is iterative).

A further attempt to this objective was recently reported by Zhang *et al.* [243], where a universal solution to TIE is proposed with the features of high accuracy, strict convergence guarantee, applicability to arbitrarily-shaped regions, and simplified implementation and computation. This method is formally similar to previous iterative solvers [203,204,242,245] except that the maximum intensity assumption is introduced to guarantee the rigorous convergence and simplify the implementation of the iterative algorithm. The comparisons of the different TIE solvers and the associated boundary conditions are summarized in Table 3.

To aid the reader, we have uploaded the **MATLAB source codes** for 4 typical TIE solvers described in this section: the FFT-based solver proposed by Paganin and Nugent [159], the DCT-based solver proposed by Zuo *et al.* [202], the iterative DCT-based solver proposed by Huang *et al.* [203], and the iterative universal solution proposed by Zhang *et al.* [243]. The code is composed of 4 simulation demo files, 3 experimental demo files, and supplemental experimental data and functions, as presented in the uploaded files.

In the simulations, the intensity measurements are designed at a wavelength of 0.633 μm , and the image is composed of 256×256 pixels with a pixel pitch of 2.2 μm . The defocused image is taken at a distance of 1 μm from the object plane. In the first simulation ("Main_Sim1_Isolated.m"), we create an isolated object located in the central FOV with different types of intensity distribution [e.g., uniform intensity (Case 1), nonuniform intensity without small intensity values (Case 2), nonuniform intensity with small intensity values (Case 3), nonuniform intensity with intensity zeros (Case 4)]. In the second simulation ("Main_Sim2_Extended.m"), the object is extended over the FOV boundary, and the intensity distributions are same as those in the first simulation. In the third simulation ("Main_Sim3_Wave front.m"), we simulate a smooth wavefront defined on a circular aperture with different types of intensity distribution [e.g., uniform intensity (Case 1), hard-edged Gaussian intensity without small intensity values (Case 2), and soft-edged Gaussian intensity with small intensity values (Case 3)]. In the fourth simulation ("Main_Sim4_Arbitrary.m"), we change the shape of the aperture arbitrarily, including elliptical aperture (Case 2), incomplete elliptical aperture (Case 3), annular aperture (Case 3), incomplete annular aperture (Case 4), telescope aperture (Case 5), and incomplete telescope aperture (Case 6).

Three experimental demo files are also provided to validate these TIE solvers. In the first experiment ("Main_Exp1_SingleCell.m"), the sample is an individual HeLa cell located in the central FOV. In the second experiment ("Main_Exp2_Microlens.m"), the sample is a microlens array limited by a (imperfect) rectangular aperture. In the third experiment ("Main_Exp3_MultipleCells.m"), several HeLa cells are randomly distributed within the FOV limited by an (imperfect) octagonal aperture. All experiment data are collected using an off-the-shelf wide-field microscope with the built-in Köhler illumination. The condenser aperture diaphragm of the microscope is closed down, and an interference filter (central wavelength 550nm) is inserted in the illumination path to produce quasi-monochromatic, spatially coherent illumination. For

Table 3
Comparison of the solutions and the boundary conditions of TIE

| Issues | Techniques | Pros | Cons |
|---------------------|--|---|---|
| TIE solvers | Green's function [141,238] Multi-Grid [76,227] Zernike polynomials [156,157] | Theoretical analysis Simple and fast Precisely represent the optical aberration | Lack of feasibility Low-frequency noise Circular domain only, difficult to follow details Imply periodic boundary conditions |
| | FFT [76,157–159] | Fast, easy to implement, incorporate regularization in reconstruction | |
| | DCT [201,202] | Fast, inhomogeneous Neumann boundary conditions | Rectangular domain only, aperture required to limit the FOV |
| | Iterative DCT [203] | Inhomogeneous Neumann boundary conditions, arbitrarily-shaped apertures | Need several iterations |
| | Iterative Universal Solution [243] | Inhomogeneous Neumann boundary conditions, arbitrarily-shaped apertures, strict convergence guarantee | Need several iterations |
| Boundary conditions | Homogeneous Dirichlet/Neumann [231,234] Periodic [76,157–159] Inhomogeneous Dirichlet [141] Inhomogeneous Neumann [201–203] | Easy to apply, can be implemented by different solvers Can be implemented by FFT-based solver - Can be measured by introducing a hard aperture | "Flat" (constant phase) at boundary Periodic phase at boundary Phase values at boundary required - |
| | Phase discrepancy | Picard-type iteration [204] | Can compensate the phase discrepancy Need 2-4 iterations |

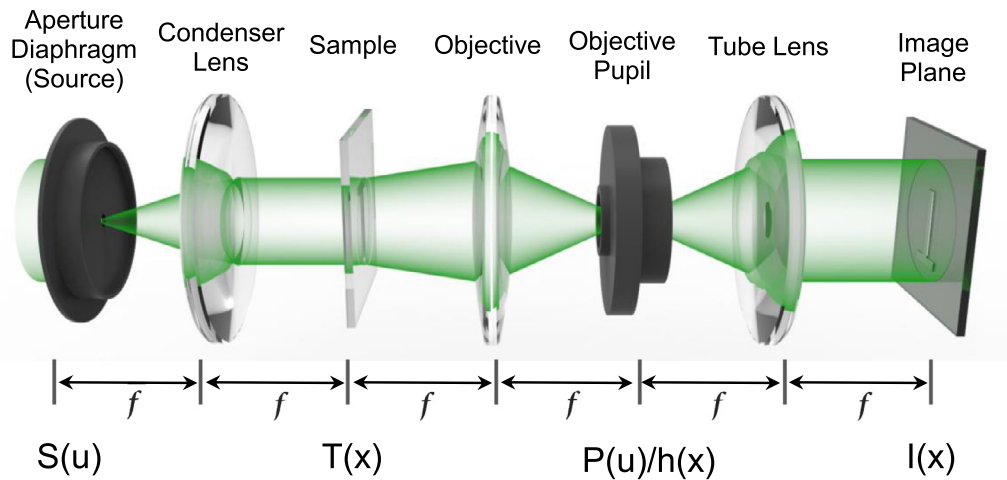


Fig. 20. Schematic of a wide-field microscope based on the Köhler illumination configuration - $6f$ imaging system.

more detailed explanations about the MATLAB source codes, readers can refer to the associated “Data in Brief” article entitled “MATLAB source codes and dataset for transport of intensity equation: a tutorial”.

4. Image formation of coherent imaging and microscopy

The basic idea of the computational phase microscopy based on TIE is to build an accurate forward image formation model and then implement the phase retrieval algorithm by solving the corresponding inverse problem. Although TIE describes the quantitative relationship between intensity and phase in the wave propagation, it depends on two restrictive assumptions: paraxial approximation and the weak defocusing. On the other hand, the derivation of TIE, as we discussed in Subsection 2.2, does not consider the finite aperture effect of the microscopic system. Therefore, in this section, we will consider a more practical microscopic imaging system and present the corresponding image formation models under coherent illumination. In addition, the relationship between these phase imaging models and the original TIE will be discussed.

4.1. Illumination coherence and apertures of a microscopic imaging system

A typical microscopic imaging system based on Köhler illumination can be described by the $6f$ system [142,178,246], as shown in Fig. 20. In Köhler illumination, an image of the light source is focused at the condenser aperture diaphragm to produce parallel light through the plane of the specimen or object. In this imaging system, the incoherent light source (the condenser aperture diaphragm) is collimated by the condenser lens to produce parallel light through the plane of the specimen or object. Then the object is imaged by the objective lens to form the frequency spectrum in the objective aperture plane. This spectrum is limited by the aperture diaphragm of the objective lens and then produces the image of the object on the image plane. In this imaging model, axial defocus is usually modeled as an angular spectrum transfer function applied to the objective aperture. In the following, the formal description of the imaging is limited to unit ($1.0\times$) magnification. An extension to arbitrary magnification is straightforward.

In a microscope based on the Köhler illumination configuration, there are two important apertures that should be considered. The first is the objective pupil defined by the NA of the objective lens, which is a dimensionless number that characterizes the range of light angles over which the imaging system can accept. The NA of the objective lens is a very important parameter because it determines the spatial resolution of the microscope imaging system. For a diffraction-limited optical imaging system, higher NA means it can gather light with larger angles, thereby having a more powerful ability to resolve fine specimen

detail. The condenser (illumination) aperture is another very important aperture of a microscope. Generally, people often only notice that the imaging resolution of a microscope is determined by the NA of the objective lens. In fact, the NA of the illumination has equal importance as well [247] (more detailed discussions can be found in Section 6). The size of the condenser aperture diaphragm can be used to control the NA of the light cone that illuminates the sample, or equivalently, the spatial coherence of the illumination.

Why is the concept of “coherence” so important in optical imaging? Because the coherence is the basic characteristic of light as a wave to form interference or diffraction phenomenon, it describes all properties of the correlation between physical quantities of a single wave, or between several waves or wave packets. Two beams of light are coherent when the phase difference between their waves is constant; they are incoherent if there is a random or changing phase relationship. Stable interference patterns are formed only by radiation emitted by coherent sources, ordinarily produced by splitting a single beam into two or more beams. A laser, unlike an incandescent source, can produce a beam in which all the components bear a fixed relationship to each other. When we discuss the propagation characteristics of light and the derivation of TIE in Section 2, the optical field is assumed to be perfectly coherent, *i.e.*, the object is illuminated with a plane wave obtained by collimating a monochromatic point source. However, in practice, the light source may not be strictly monochromatic and is of finite extent. These two aspects are related to the temporal coherence (spectrum) and spatial coherence (size) of light source, respectively [5,248]. More intuitively, temporal coherence is a measure of the correlation of light wave’s phase at different points along the direction of propagation (longitudinal direction). It tells us how monochromatic a source is. Generally, a laser with good monochromaticity can be considered as a temporally coherent source. In contrast, a broadband source is temporally incoherent, such as the light emitted from the incandescent lamp or the sun. The spatial coherence is a measure of the correlation of a light wave’s phase at different points transverse to the direction of propagation. It tells us how uniform the phase of a wavefront is and is related to the dimension of the light source. Thus, the light emitted from an ideal point source is spatially coherent, while the light from a highly extended source is spatially incoherent. Fig. 21 gives several typical examples of light sources with different degrees of temporal coherence and spatial coherence. In the general concept, the sunlight is considered completely incoherent, but the fact that it would not always be the case. The reason is that although sunlight is a broadband source (temporally incoherent), it can be approximated as a point source because the sun is far enough from us. When the sky is cloudless (the atmospheric scattering effect can be neglected), the sunlight reaching the earth’s surface can be considered

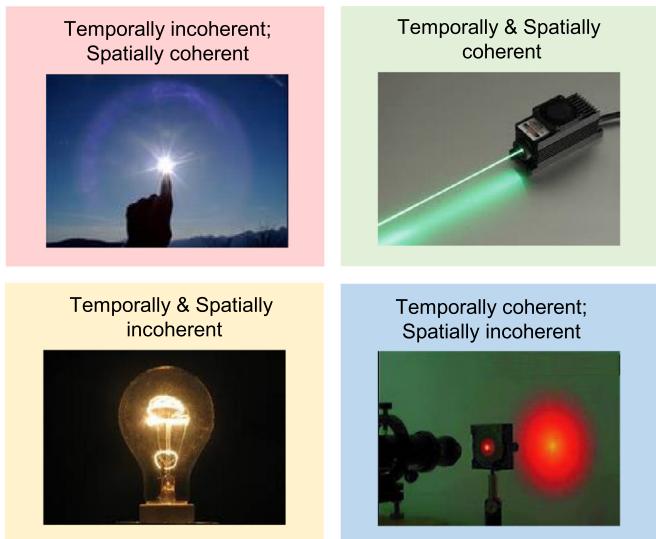


Fig. 21. Several typical examples of light sources with different degrees of temporal coherence and spatial coherence.

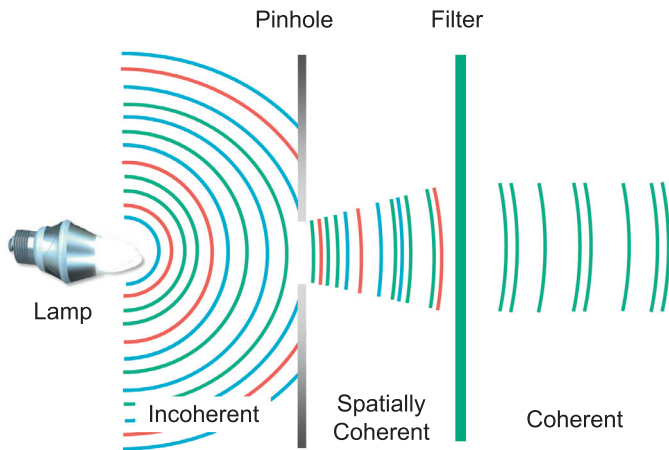


Fig. 22. Illustration of light wave coherence. An incoherent light source emits spherically shaped wave-fronts and propagates a certain distance, then passes through a pinhole so that the spatial coherence of light is improved. It then passes through a narrow wavelength filter to yield quasi-monochromatic light, achieving in this way temporal and spatial coherence.

spatially coherent. Similarly, the laser is not always completely coherent. It is only temporally coherent. In some special cases, e.g., when the laser light passes through strongly scattering medium (e.g., a rotating ground glass), the spatial coherence of the light will completely vanish.

Let us go back to the microscopic imaging system. In the simplified illumination path of Fig. 22, we assume the source is an extended source with a broad spectrum (temporally and spatially incoherent). The light emitted from the source is firstly limited by the condenser aperture diaphragm (a circular mask with diameter d), and then passes through an ideal monochromatic filter (wavelength λ) to form the illumination field in the object plane. At this time, if the aperture diaphragm is closed to the minimum (like an ideal pinhole), the illumination reaching the object plane can be regarded as spatially coherent (perfect spherical or plane waves). However, if the aperture diaphragm is gradually opened up, the spatial coherence of the illumination field will decrease accordingly. In a microscopic system, the spatial coherence of the illumination is often quantified by the coherence parameter, which is defined as the

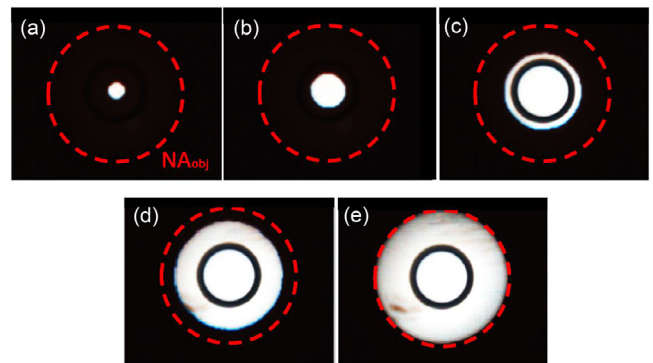


Fig. 23. Varying the illumination aperture to control the spatial coherence by opening up the condenser diaphragm of a phase contrast microscope. (a) $s = 0.15$; (b) $s = 0.3$; (c) $s = 0.55$; (d) $s = 0.8$; (e) $s = 1.0$.

ratio of the illumination NA to the objective NA [177,178,246,249]

$$s = \frac{NA_{ill}}{NA_{obj}} \tag{57}$$

The ratio of the illumination NA to the objective NA is equivalent to the radius ratio of the aperture diaphragm of the condenser to that of the objective lens. Therefore, by inserting a Bertrand lens or a phase telescope into the beam path to image the back focal plane of the objective lens, the coherence parameter can be estimated by measuring the ratio of the condenser aperture diaphragm to the objective aperture diaphragm (dashed line in Fig. 23). In fact, although s is called “coherence parameter”, it represents the degree of “incoherence” of the imaging system. Specifically, a larger s means weaker spatial coherence, so strict spatial coherence requires $s \rightarrow 0$, and $s \rightarrow \infty$ refers to complete incoherence. The coherence parameter of the microscopic imaging system is closely related to image formation. In the case of incoherent imaging, the sample’s phase component will no longer affect the intensity distribution of the image. For weak phase objects, the microscope behaves like a completely incoherent system when $s \geq 1$, which means the phase effect vanishes [178,179,249]. Therefore, $0 < s < 1$ is generally considered partially coherent, which is a beneficial setting for phase imaging. It can provide higher imaging *signal-to-noise ratio* (SNR) and improved imaging resolution. The relevant content will be discussed in detail in Section 6. In this section, we mainly discuss the image formation model and phase retrieval problem under an ideal coherent imaging system.

4.2. Ideal imaging model under coherent illumination

In the following section, we only consider the case of coherent illumination: the light source is a geometric point on the optical axis of the condenser aperture plane, and it has strict monochromaticity. The imaging process satisfies the ideal coherent imaging model, and can be described as a linear system for complex amplitude. Considering a thin object with complex transmittance $T(x) = a(x) \exp[j\phi(x)]$ illuminated by a monochromatic plane wave, the complex amplitude distribution in the object plane can be expressed as

$$U_0(x) = T(x) = a(x) \exp [j\phi(x)] \tag{58}$$

Without loss of generality, we assume that the plane wave illumination has unit intensity. The complex amplitude distribution in the image plane $U_i(x)$ can be represented as the convolution of the complex amplitude distribution of the ideal object field $U_0(x)$ and the *point spread function* (PSF) of the imaging system $h(x)$

$$U_i(x) = U_0(x) * h(x) \tag{59}$$

Thus, the captured image intensity can be expressed as

$$I_i(x) = U_i(x)U_i^*(x) = |U_0(x) * h(x)|^2 \tag{60}$$

Taking Fourier transform on both sides of Eq. (59), the following expression in the frequency domain can be obtained:

$$\hat{U}_i(\mathbf{u}) = \hat{U}_0(\mathbf{u})H(\mathbf{u}) \quad (61)$$

The Fourier transform of the intensity passing through the imaging system can be expressed as

$$\hat{I}_i(\mathbf{u}) = \hat{U}_i(\mathbf{u}) \otimes \hat{U}_i^*(\mathbf{u}) = \mathcal{F} \left\{ \left| \mathcal{F}^{-1} [\hat{U}_0(\mathbf{u})H(\mathbf{u})] \right|^2 \right\} \quad (62)$$

where $\hat{U}_0(\mathbf{u})$ and $\hat{U}_i(\mathbf{u})$ are the Fourier transforms of the complex amplitudes $U_0(\mathbf{u})$ and $U_i(\mathbf{u})$, respectively, and $H(\mathbf{u}) = \mathcal{F} \{h(\mathbf{x})\}$ is the transfer function under coherent illumination. For a diffraction-limited aberration-free imaging system, the coherent transfer function is just the pupil function of the objective $H(\mathbf{u}) = P(\mathbf{u})$, which is a standard circ function with unit amplitude across the passband (without attenuation) [2,5]. This indicates that we cannot get an ideal point when imaging of an ideal point object even with a perfect aberration-free lens, instead, we can only obtain a diffraction spot determined by the limited lens aperture. In addition, when the sample is located at a defocused plane with a defocus distance Δz , the coherent transfer function is determined by the product of the pupil function and the angular spectrum transfer function of the objective lens [Eq. (30)]

$$H(\mathbf{u}) = P(\mathbf{u})H_{\Delta z}(\mathbf{u}) = P(\mathbf{u})e^{jk\Delta z\sqrt{1-\lambda^2|\mathbf{u}|^2}},$$

$$P(\mathbf{u}) = \text{circ} \left(\frac{\mathbf{u}}{NA/\lambda} \right) = \begin{cases} 1 & |\mathbf{u}| \leq \frac{NA}{\lambda} \\ 0 & \text{else} \end{cases} \quad (63)$$

where NA denotes the numerical aperture of the objective lens. Compared with the pupil function, this expression includes one additional phase factor to account for the defocusing effect, thus this coherent transfer function is often referred to as the defocused pupil function. There are two points need to be further elaborated here: first, the coherent transfer function is sometimes written as $H(\mathbf{u}) = P(\mathbf{u})\exp \left[jk\Delta z \left(\sqrt{1-\lambda^2|\mathbf{u}|^2} - 1 \right) \right]$, in order to guarantee that it can be simplified into a form without the z -dependence [the term of $\exp(jk\Delta z)$]. Under the paraxial approximation without affecting the transverse intensity and phase distribution calculated from Eq. (32). In addition, microscopic systems are generally axisymmetric with respect to the optical axis, thus the coherent PSF and transfer function can be more conveniently expressed with the Fourier (inverse) transform in polar coordinates

$$h(r, \theta) = \int_0^{2\pi} \int_{\rho} H(\rho, \vartheta) e^{j2\pi r \rho \cos(\theta-\vartheta)} \rho d\rho d\vartheta \quad (64)$$

where (r, θ) are the polar coordinates corresponding to the spatial coordinates (x, y)

$$x = r \cos \theta$$

$$y = r \sin \theta \quad (65)$$

Similarly, (ρ, ϑ) are the polar coordinates corresponding to the spatial frequency coordinates (u, v) . When the imaging system is axisymmetric, the coherent transfer function is only related to the variable ρ , so Eq. (64) can be further simplified as

$$h(r) = \int_{\rho} H(\rho) J_0(2\pi r \rho) 2\pi \rho d\rho \quad (66)$$

where J_0 is a zero-order Bessel function of the first kind, which is defined as

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{jx \cos(\theta-\vartheta)} d\vartheta$$

$$\stackrel{\theta \text{ independent integration}}{=} \frac{1}{2\pi} \int_0^{2\pi} e^{jx \cos \theta} d\theta \quad (67)$$

Eq. (66) is so-called Hankel transform, which is the Fourier transform for the axially symmetric function, and is also a special case of

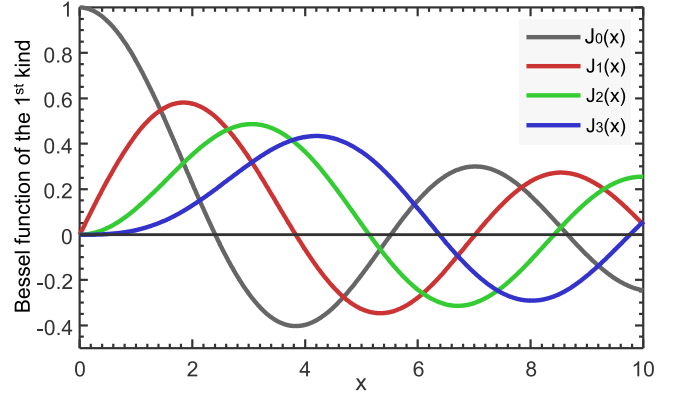


Fig. 24. The first four orders of Bessel function of the first kind.

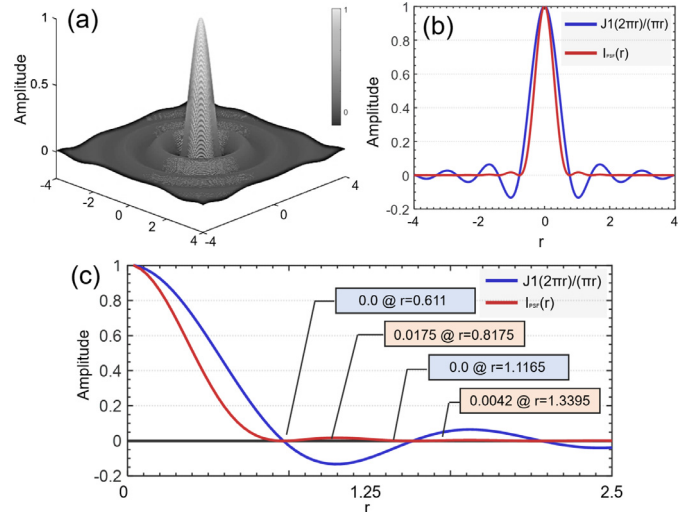


Fig. 25. Intensity distribution of Airy disk. (a) 2D intensity distribution of Airy disk; (b) cross-sections of Airy disk intensity and Airy function; (c) the zero-crossings of the intensity distribution of Airy disk.

the traditional Fourier transform. It is not difficult to prove that the Hankel forward transform has exactly same form as its inverse transform. For an ideal in-focus diffraction-limited optical system, the coherent transfer function is just the pupil function of the objective lens $H(\rho) = P(\rho) = \text{circ} \left(\frac{\rho}{NA/\lambda} \right)$, and the corresponding PSF can be expressed as

$$h(r) = \int_{\rho} \text{circ} \left(\frac{\rho}{NA/\lambda} \right) J_0(2\pi r \rho) 2\pi \rho d\rho = \frac{NA}{r\lambda} J_1 \left(2\pi r \frac{NA}{\lambda} \right)$$

$$= \pi \left(\frac{NA}{\lambda} \right)^2 \left[\frac{2J_1(\bar{r})}{\bar{r}} \right] \quad (68)$$

where J_1 is the first-order Bessel function of the first kind, as plotted in Fig. 24. \bar{r} is the normalized version of r , $\bar{r} = 2\pi \frac{NA}{\lambda} r$. Then the intensity distribution corresponding to Eq. (68) becomes

$$I_{PSF}(r) = |h(r)|^2 = \pi \left(\frac{NA}{\lambda} \right)^4 \left[\frac{2J_1(\bar{r})}{\bar{r}} \right]^2 \quad (69)$$

Eq. (69) is called intensity PSF, which is usually used to describe the imaging characteristics of an incoherent imaging system. Since $\frac{J_1(\bar{r})}{\bar{r}}$ is also an Airy function, this intensity distribution is also called ‘‘Airy disk’’, as shown in Fig. 25. It can be observed that 83.8% of the incident light energy is concentrated in the central bright spot, and the first intensity zero-crossing appears at the position of $\frac{0.611\lambda}{NA}$.

Similarly, when there is a defocus of distance Δz , the coherent transfer function [Eq. (63)] can be expressed as

$$H(\rho) = P(\rho)e^{jk\Delta z\sqrt{1-\lambda^2\rho^2}},$$

$$P(\rho) = \text{circ}\left(\frac{\rho}{NA/\lambda}\right) = \begin{cases} 1 & \rho \leq \frac{NA}{\lambda} \\ 0 & \text{else} \end{cases} \quad (70)$$

Under the paraxial approximation, the angular spectrum diffraction term in Eq. (70) can be approximated by the Fresnel angular spectrum transfer function [Eq. (32)]

$$H(\rho) = P(\rho)e^{-j\pi\lambda\Delta z\rho^2} \quad (71)$$

Here the z dependence factor has been omitted. It should be noted that in most literature concerning the OTF [250–252], Eq. (71) is often represented in a normalized coordinate system to simplify the relevant analysis and calculation, *i.e.*, the following normalized variables are introduced

$$\bar{r} = \frac{2\pi}{\lambda} NA \quad r = kNA \quad (72)$$

$$\bar{z} = 2k\left(1 - \sqrt{1 - NA^2}\right)z \approx^{\text{paraxial}} kNA^2z \quad (73)$$

$$\bar{\rho} = \frac{\rho}{NA/\lambda} \quad (74)$$

where k denotes the wave number, $P(\bar{\rho})$ is the pupil function of the objective lens, \bar{r} is the normalized radial transverse coordinate, \bar{z} is the normalized axial coordinate, and $\bar{\rho}$ is the normalized radial spatial frequency (more detailed discussions about the normalized coordinate system can be found in Section 8). In such a normalized coordinate system, the paraxial coherent transfer function of Eq. (67) can be rewritten as

$$H(\bar{\rho}) = P(\bar{\rho})e^{-\frac{1}{2}j\bar{z}\bar{\rho}^2} \quad (75)$$

Substituting Eq. (75) into Eq. (66), we can get the expression of the defocused PSF in the normalized coordinate

$$h_{\bar{z}}(\bar{r}) = \int_{\bar{\rho}} P(\bar{\rho})e^{-\frac{1}{2}j\bar{z}\bar{\rho}^2} J_0(\bar{r}\bar{\rho})2\pi\bar{\rho}d\bar{\rho} \quad (76)$$

In fact, the defocused PSF represented by Eq. (76) is also called 3D PSF because it describes the 3D complex amplitude distributions in the image space formed by an ideal point source. This expression will be discussed in detail in Section 8 when we talk about 3D phase imaging.

Under coherent illumination, the image formation process is linear about the complex amplitude. However, due to the effect of the complex conjugation operator [Eq. (60)], the captured intensity is linearly related neither to the amplitude nor to the phase of the sample, posing a major obstacle for phase retrieval. On the other hand, for an aberration-free imaging system with an infinite aperture, ($H(\mathbf{u}) \equiv 1$ and $h(\mathbf{x}) = \delta(\mathbf{x})$), $I_r(\mathbf{x}) = A^2(\mathbf{x})$, which means that in the in-focus plane the phase component of the object cannot be observed at all. Therefore, the basic idea of the TIE phase retrieval is to generate phase contrast by introducing defocusing into the pupil function, so that the phase information can be converted into the intensity information. Then the inverse problem of phase recovery can be solved by linearizing the quantitative relationship between the intensity and the phase distribution.

4.3. Image formation model, phase transfer function and linearization conditions

The major obstacle in the phase retrieval problem is the nonlinear relationship between the intensity (amplitude) and phase of the sample in the image formation process. Although this nonlinear inverse problem can be solved by the iterative phase retrieval algorithms [59,60,87], they rely on computationally intensive iterative operations, and there is no theoretical convergence guarantee. Therefore, deterministic phase retrieval methods are more attractive in this regard. The key to deterministic phase retrieval is how to linearize the quantitative relationship between the intensity and phase information.

4.3.1. Contrast transfer function model under the weak object approximation

In Subsection 2.2, we derived TIE in three different ways. However, all these derivations rely on two restrictive conditions: paraxial approximation and weak defocusing approximation. TIE is valid if and only if the two conditions are satisfied simultaneously, and thus, the linear relationship between the intensity and the quantitative phase can be established. However, these two conditions are sometimes difficult to meet in practical experimental conditions. For example, when imaging with a high NA objective, the paraxial approximation is untenable. On the other hand, when the defocus distance is too large to meet the weak defocusing approximation, the phase solved by TIE will also be deviated (detailed discussions about this issue can be found in Section 5).

It should be noted that there is more than one approach to achieve linearization. The phase contrast transfer function (CTF) method based on weak object approximation is another classical method for deterministic phase retrieval [163,221,253–256]. In the following section, we will derive the transfer functions of weak objects under coherent illuminations based on the scalar diffraction theory. When a weak object (the absorption and phase of the object is sufficiently small) is illuminated by a coherent plane wave, the complex amplitude of the field just behind the object can be simplified as

$$U_0(\mathbf{x}) = T(\mathbf{x}) = a(\mathbf{x}) \exp[j\phi(\mathbf{x})] \begin{cases} \phi(\mathbf{x}) \ll 1 \\ \approx [a_0 + \Delta a(\mathbf{x})][1 + j\phi(\mathbf{x})] \\ \Delta a(\mathbf{x}) \ll a_0 \\ \approx a_0 + \Delta a(\mathbf{x}) + ja_0\phi(\mathbf{x}) \end{cases} \quad (77)$$

where a_0 denotes the direct current (DC) component of the incident plane wave undisturbed by the object. The other two terms represent the diffracted light components. The second term can be further written as $\Delta a(\mathbf{x}) = a_0\eta(\mathbf{x})$, representing the contribution from the absorption variations $\eta(\mathbf{x})$; and $ja_0\phi(\mathbf{x})$ represents the contribution of the phase component. Taking Fourier transform of Eq. (77), the spectrum of the transmitted complex amplitude can be expressed as

$$\hat{U}_0(\mathbf{u}) = a_0[\delta(\mathbf{u}) + \hat{\eta}(\mathbf{u}) + j\hat{\phi}(\mathbf{u})] \quad (78)$$

Substituting Eq. (78) into Eq. (62) and after appropriate simplification, and the spectral distribution of the intensity image at the propagation distance z can be obtained as

$$\hat{I}_{\Delta z}(\mathbf{u}) = I_0\left\{\delta(\mathbf{u}) - 2\cos\left[k\Delta z\left(\sqrt{1-\lambda^2|\mathbf{u}|^2} - 1\right)\right]\hat{\eta}(\mathbf{u}) - 2\sin\left[k\Delta z\left(\sqrt{1-\lambda^2|\mathbf{u}|^2} - 1\right)\right]\hat{\phi}(\mathbf{u})\right\} \quad (79)$$

where $I_0 = a_0^2$. Noted that when deriving Eq. (79), the high-order diffraction term (the interaction between the intensity component and the phase component is neglected, leaving only their first-order terms) and the finite-aperture effect are omitted (assume the objective aperture is infinite, *i.e.*, $P(\mathbf{u}) \equiv 1$). By further invoking the paraxial approximation, the angular spectrum diffraction terms in Eq. (79) can be approximated by the Fresnel transfer function of Eq. (33), thus Eq. (79) can be further simplified as [163,221,253]

$$\hat{I}_{\Delta z}(\mathbf{u}) = I_0\left\{\delta(\mathbf{u}) - 2\cos(\pi\lambda\Delta z|\mathbf{u}|^2)\hat{\eta}(\mathbf{u}) + 2\sin(\pi\lambda\Delta z|\mathbf{u}|^2)\hat{\phi}(\mathbf{u})\right\} \quad (80)$$

It can be seen from the above equation that $\hat{I}_{\Delta z}$ consists of three components: the DC component $I_0\delta(\mathbf{u})$ (average intensity or background intensity), the intensity component associated with the absorption $2I_0\cos(\pi\lambda\Delta z|\mathbf{u}|^2)\hat{\eta}(\mathbf{u})$, and the intensity component associated with the phase $2I_0\sin(\pi\lambda\Delta z|\mathbf{u}|^2)\hat{\phi}(\mathbf{u})$. In this case, the intensity now has a linear relationship with the absorption function $a(\mathbf{x})$ and the phase function $\phi(\mathbf{x})$. The cosine and sine coefficients in front of the absorption function and the phase function are usually termed as *absorption transfer function* [ATF, $H_A(\mathbf{u})$] and *phase transfer function* [PTF, $H_P(\mathbf{u})$], respectively.

$$H_A(\mathbf{u}) = \cos(\pi\lambda\Delta z|\mathbf{u}|^2) \quad (81)$$

$$H_P(\mathbf{u}) = \sin(\pi\lambda\Delta z|\mathbf{u}|^2) \quad (82)$$

The ATF and PTF demonstrate the relative strengths of the information-bearing portion (amplitude or phase) of the image and the ever-present background. Although the linearization between the intensity and the phase has been achieved based on weak object approximation, the effects of absorption and phase are still coupled together in the captured intensity images. In order to recover the phase, these two components need to be decoupled. It can be found from Eqs. (81) and (82) that the absorption and phase transfer functions, *i.e.*, cosine and sine functions are the even and odd functions of the defocusing distance Δz , respectively. This indicates that the intensity variation induced by the equal but opposite defocusing is the same for the absorption components, but is opposite for the phase components. Therefore, it can be similar to TIE to collect an in-focus image and two defocused images with an equal amount and opposite direction as I_0 and $I_{\pm\Delta z}$ respectively. Thus, similar to the axial intensity derivative estimation in TIE, we can capture two intensity images with equal and opposite defocus. The subtraction of these two images gives a pure phase-contrast image with the effect of the absorption component canceled out

$$\frac{\hat{I}_{\Delta z}(\mathbf{u}) - \hat{I}_{-\Delta z}(\mathbf{u})}{4I_0} = H_p(\mathbf{u})\hat{\phi}(\mathbf{u}) \quad (83)$$

This leads to a linear relationship between the phase and the intensity, and the quantitative phase information can then be reconstructed by Fourier space deconvolution with the PTF $H_p(\mathbf{u})$. It should also be noted that since $H_p(\mathbf{u})$ tends to oscillate with the increase in the spatial frequency gradually, it contains zero-crossings, especially at higher spatial frequencies. In order to avoid the zero crossings in the pass-band, the defocus distance should be sufficient small $\Delta z \rightarrow 0$ [256,257], and thus the sinusoidal term in the PTF $H_p(\mathbf{u})$ can be approximated as $\sin(\pi\lambda\Delta z|\mathbf{u}|^2) \approx \pi\lambda\Delta z|\mathbf{u}|^2$, so that the oscillation can be effectively avoided. However, it may reduce the phase contrast for lower spatial frequencies, which compromises the SNR of the phase reconstruction [254,258]. In addition, based on this weak defocusing approximation, Eq. (83) can be further simplified to TIE with uniform intensity distribution [Eq. (42)]. We will study this case in detail in Subsection 5.1. Another possible solution to zero-crossings is to synthesize and optimize the PTF by using more than two images at multiple defocusing distances [200,256,259,260]. This method allows the response of PTF to be optimized over a wider range of spatial frequency at the expense of additional data acquisition and processing, which will be discussed in Subsection 5.2.

4.3.2. Mixed transfer function model under the slowly varying object approximation

As seen above, the TIE or CTF methods rely on different assumptions. TIE is valid for short propagation distances and the paraxial approximation. On the other hand, the CTF method effectively extends the linearization range from weak defocusing to an arbitrary defocusing distance (no longer requires $\Delta z \rightarrow 0$), but the derivation is based on the assumption that the object has weak absorption and phase distributions. When the object is thick or strongly absorptive, the approximation of Eq. (78) is no longer valid. In order to extend the validity of the approximations made in CTF, some researchers attempted to replace the weak object approximation with an alternative approximation so-called “slowly varying object” [221,256,259–262]. First, based on Eq. (20), the Fourier spectrum of the intensity distribution can be represented in the differential spatial domain [221]

$$\hat{I}_{\Delta z}(\mathbf{u}) = \int T\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right)T^*\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\exp\{-2j\pi\mathbf{x}\cdot\mathbf{u}\}d\mathbf{x} \quad (84)$$

where $T(\mathbf{x}) = a(\mathbf{x})\exp[j\phi(\mathbf{x})]$ is the complex transmittance of the object. In the slowly varying object approximation, the phase change is assumed to be much smaller than 1 within a small spatial neighborhood ($\pm\frac{\lambda\Delta z\mathbf{u}}{2}$)

$$\left|\phi\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right) - \phi\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\right| \ll 1 \quad (85)$$

In addition, we assume that the absorption is also slowly varying and satisfies the following first-order Taylor approximation

$$a\left(\mathbf{x} \pm \frac{\lambda\Delta z\mathbf{u}}{2}\right) \approx a(\mathbf{x}) \pm \frac{\lambda\Delta z\mathbf{u}}{2} \cdot \nabla a(\mathbf{x}) \quad (86)$$

Substituting Eq. (85) and Eq. (86) into Eq. (84) yields

$$\begin{aligned} T\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right)T^*\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right) \\ = a\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right)a\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\exp\left\{j\left[\phi\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right) - \phi\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\right]\right\} \\ \approx a\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right)a\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\left\{1 + j\left[\phi\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right) - \phi\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\right]\right\} \end{aligned} \quad (87)$$

Substituting Eq. (87) into Eq. (84) and then simplifying the equation using the weak absorption approximation

$$\begin{aligned} \hat{I}_{\Delta z}(\mathbf{u}) &= \int a\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right)a\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right) \\ &\quad \left\{1 + j\left[\phi\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right) - \phi\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\right]\right\}\exp\{-2j\pi\mathbf{x}\cdot\mathbf{u}\}d\mathbf{x} \\ &\approx \hat{I}_{\Delta z}^{\phi=0}(\mathbf{u}) + 2\sin(\pi\lambda\Delta z|\mathbf{u}|^2)\mathcal{F}\{I_0(\mathbf{x})\phi(\mathbf{x})\} \\ &\quad + \cos(\pi\lambda\Delta z|\mathbf{u}|^2)\frac{\lambda\Delta z}{2\pi}\mathcal{F}\{\nabla\cdot[\phi(\mathbf{x})\nabla I_0(\mathbf{x})]\} \end{aligned} \quad (88)$$

where

$$\hat{I}_{\Delta z}^{\phi=0}(\mathbf{u}) = \int a\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right)a^*\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\exp\{-2j\pi\mathbf{x}\cdot\mathbf{u}\}d\mathbf{x} \quad (89)$$

represents the Fourier transform of the intensity distribution at a distance Δz when the object phase component is neglected (the object is assumed to be purely absorptive). Eq. (88) is called the mixed transfer function proposed by Guigay et al. [259]. Compared with Eq. (80), these two equations share some similarities. For example, the sine and cosine terms of ATF and PTF [Eqs. (81) and (82)] are still clearly expressed in Eq. (88). In addition, the mixed model is more complicated, and the phase information is not completely linearized with the intensity. Therefore, the phase cannot be directly retrieved based on on-step deconvolution based on the PTF, and nonlinear iterative algorithms are needed to get the accurate solution. Unlike the weak object approximation, the derivation of the mixed model relies on the slowly varying object approximation, which imposes restrictions on the phase/absorption spatial gradients rather than their absolute values. Thus, the slowly varying object is a more relaxed and reasonable assumption than the weak object approximation [260]. In this case, if we expand the third term of Eq. (88) by the chain rule (according to $\nabla^2[\phi(\mathbf{x})I_0(\mathbf{x})] = \nabla\cdot[\phi(\mathbf{x})\nabla I_0(\mathbf{x})] + \nabla\cdot[\nabla\phi(\mathbf{x})I_0(\mathbf{x})]$) and combine similar terms, the following expression can be obtained [259]

$$\begin{aligned} \hat{I}_{\Delta z}(\mathbf{u}) &= \hat{I}_{\Delta z}^{\phi=0}(\mathbf{u}) + 2\sin(\pi\lambda\Delta z|\mathbf{u}|^2)\mathcal{F}\{I_0(\mathbf{x})\phi(\mathbf{x})\} \\ &\quad + \cos(\pi\lambda\Delta z|\mathbf{u}|^2)\frac{\lambda\Delta z}{2\pi}\mathcal{F}\{\nabla\cdot I_0(\mathbf{x})\nabla\phi(\mathbf{x})\} \end{aligned} \quad (90)$$

It is not difficult to prove that under weak defocusing approximation, Eq. (88) can be simplified to the standard TIE [Eq. (10)]. On the other hand, if the absorption is assumed to be weak $a(\mathbf{x}) = a_0 + \Delta a(\mathbf{x})$, $\Delta a(\mathbf{x}) \ll a_0$, Eq. (90) can be reduced to the CTF model [Eq. (80)]. This means that the weak phase approximation $\phi(\mathbf{u}) \ll 1$ used in the CTF model [Eq. (80)] can be relaxed to slowly varying phase approximation $\left|\phi\left(\mathbf{x} - \frac{\lambda\Delta z\mathbf{u}}{2}\right) - \phi\left(\mathbf{x} + \frac{\lambda\Delta z\mathbf{u}}{2}\right)\right| \ll 1$. In particular, the weak phase approximation requires that the phase value is much smaller than 1, which is difficult to satisfy for most biological samples. In contrast, the slowly varying object approximation only requires that the phase variation is much less than 1 within a small spatial neighborhood ($\pm\frac{\lambda\Delta z\mathbf{u}}{2}$), and thus is more universal than the weak object approximation.

4.3.3. Intensity difference model without approximations

In 2015, Sun et al. [206] further proposed an intensity difference model without approximations. This method is derived based on the angular spectrum diffraction formula, and invokes no approximations in

Table 4
Comparison of phase retrieval methods and required approximation conditions

| Methods | Approximation conditions | Phase reconstruction algorithms |
|---|--|--|
| TIE [141,159,202] | Paraxial approximation $\lambda^2 \mathbf{u} ^2 \ll 1$ Weak defocusing approximation $\Delta z \rightarrow 0$ | $-k \frac{\partial I(\mathbf{x})}{\partial z} = \nabla \cdot [I(\mathbf{x})\nabla\phi(\mathbf{x})]$ Fourier solution: $\phi(\mathbf{x}) = -k \mathcal{F}^{-1} \left\{ \frac{j2\pi\mathbf{u}}{4\pi^2 \mathbf{u} ^2 + \epsilon} \mathcal{F} \left[\frac{1}{I(\mathbf{x})} \mathcal{F}^{-1} \left\{ \frac{j2\pi\mathbf{u}}{4\pi^2 \mathbf{u} ^2 + \epsilon} \mathcal{F} \left[\frac{\partial I(\mathbf{x})}{\partial z} \right] \right\} \right] \right\}$ $\epsilon > 0$ is a small constant. |
| Weak absorption TIE [158,263] | Paraxial approximation $\lambda^2 \mathbf{u} ^2 \ll 1$ Weak defocusing approximation $\Delta z \rightarrow 0$ Weak absorption approximation $a(\mathbf{x}) = a_0 + \Delta a(\mathbf{x}), \Delta a(\mathbf{x}) \ll a_0$ | $-k \frac{\partial I(\mathbf{x})}{\partial z} = I_0 \nabla^2 \phi(\mathbf{x})$ Fourier solution: $\phi(\mathbf{x}) = \frac{k}{I_0} \mathcal{F}^{-1} \left\{ \frac{1}{4\pi^2 \mathbf{u} ^2 + \epsilon} \mathcal{F} \left[\frac{\partial I(\mathbf{x})}{\partial z} \right] \right\}$ $\epsilon > 0$ is a small constant. |
| CTF [163,221,253] | Paraxial approximation (optional) $\lambda^2 \mathbf{u} ^2 \ll 1$ Weak absorption approximation $a(\mathbf{x}) = a_0 + \Delta a(\mathbf{x}), \Delta a(\mathbf{x}) \ll a_0$ Weak phase approximation $ \phi(\mathbf{x}) \ll 1$ or Slowly varying phase approximation $\left \phi\left(\mathbf{x} - \frac{j\Delta z\mathbf{u}}{2}\right) - \phi\left(\mathbf{x} + \frac{j\Delta z\mathbf{u}}{2}\right) \right \ll 1$ | $\hat{I}_{\Delta z}(\mathbf{u}) = I_0[\delta(\mathbf{u}) - 2 \cos(\pi\lambda\Delta z \mathbf{u} ^2)\hat{\eta}(\mathbf{u}) + 2 \sin(\pi\lambda\Delta z \mathbf{u} ^2)\hat{\phi}(\mathbf{u})]$ Least-square solution: $\phi(\mathbf{x}) = \mathcal{F}^{-1} \left\{ \frac{D(\mathbf{u}) \sum_{j=1}^N \hat{I}_{\Delta z_j}(\mathbf{u}) \sin(\pi\lambda\Delta z_j \mathbf{u} ^2) - C(\mathbf{u}) \sum_{j=1}^N \hat{I}_{\Delta z_j}(\mathbf{u}) \cos(\pi\lambda\Delta z_j \mathbf{u} ^2)}{2I_0[C(\mathbf{u})D(\mathbf{u}) - E(\mathbf{u})^2] + \epsilon} \right\}$ where, $C(\mathbf{u}) = \sum_{j=1}^N \sin^2(\pi\lambda\Delta z_j \mathbf{u} ^2)$ $D(\mathbf{u}) = \sum_{j=1}^N \cos^2(\pi\lambda\Delta z_j \mathbf{u} ^2)$ $E(\mathbf{u}) = \sum_{j=1}^N \sin(\pi\lambda\Delta z_j \mathbf{u} ^2) \cos(\pi\lambda\Delta z_j \mathbf{u} ^2)$ $\epsilon > 0$ is a small constant. Solution of multi-plane symmetrical defocusing: $\phi(\mathbf{x}) = \mathcal{F}^{-1} \left\{ \sum_{j=1}^N \frac{ \sin(\pi\lambda\Delta z_j \mathbf{u} ^2) ^2 [\hat{I}_{\Delta z_j}(\mathbf{u}) - \hat{I}_{-\Delta z_j}(\mathbf{u})]}{4I_0 \sin(\pi\lambda\Delta z_j \mathbf{u} ^2) \sum_{j=1}^N [\sin(\pi\lambda\Delta z_j \mathbf{u} ^2) ^2 + \epsilon]} \right\}$ Solution of two plane symmetrical defocusing: $\phi(\mathbf{x}) = \mathcal{F}^{-1} \left[\frac{\sin(\pi\lambda\Delta z \mathbf{u} ^2) [\hat{I}_{\Delta z}(\mathbf{u}) - \hat{I}_{-\Delta z}(\mathbf{u})]}{4I_0 \sin^2(\pi\lambda\Delta z \mathbf{u} ^2) + \epsilon} \right]$ or $\phi(\mathbf{x}) = \mathcal{F}^{-1} \left[\frac{\hat{I}_{\Delta z}(\mathbf{u}) - \hat{I}_{-\Delta z}(\mathbf{u})}{4I_0 \sin(\pi\lambda\Delta z \mathbf{u} ^2)} \right]$ |
| Mixed CTF transfer function [259] | Paraxial approximation $\lambda^2 \mathbf{u} ^2 \ll 1$ Slowly varying absorption approximation $a\left(\mathbf{x} \pm \frac{j\Delta z\mathbf{u}}{2}\right) \approx a(\mathbf{x}) \pm \frac{j\Delta z\mathbf{u}}{2} \cdot \nabla a(\mathbf{x})$ | $\hat{I}_{\Delta z}(\mathbf{u}) = \hat{I}_{\Delta z}^{\phi=0}(\mathbf{u}) + 2 \sin(\pi\lambda\Delta z \mathbf{u} ^2) \mathcal{F} \{ I(\mathbf{x})\phi(\mathbf{x}) \}$ $+ \frac{j\Delta z}{2\pi} \cos(\pi\lambda\Delta z \mathbf{u} ^2) \mathcal{F} \{ \nabla \cdot [\phi(\mathbf{x})\nabla I(\mathbf{x})] \}$ Iterative solution: $\phi^{(n+1)}(\mathbf{x}) = \mathcal{F}^{-1} \left\{ \frac{\sum_{j=1}^N [2 \sin(\pi\lambda\Delta z_j \mathbf{u} ^2) \hat{I}_{\Delta z_j}(\mathbf{u}) - \hat{I}_{-\Delta z_j}^{\phi^{(n)}}(\mathbf{u}) - \Delta z_j^{(n)}(\mathbf{u})]}{\sum_{j=1}^N 4I_0 \sin^2(\pi\lambda\Delta z_j \mathbf{u} ^2) + \epsilon} \right\}$ where, $\Delta z_j^{(n)}(\mathbf{u}) = \frac{j\Delta z_j}{2\pi} \cos(\pi\lambda\Delta z_j \mathbf{u} ^2) \mathcal{F} \{ \nabla \cdot [\phi^{(n)}(\mathbf{x})\nabla I(\mathbf{x})] \}$ $\epsilon > 0$ is a small constant. $\phi^{(n)}(\mathbf{x})$ represents the phase of n^{th} iterations. |
| Intensity difference without approximations [206] | None | $\hat{I}_{\Delta z}(\mathbf{u}) - \hat{I}_{-\Delta z}(\mathbf{u}) = 4A_m^{(n)} \sin[\omega(\mathbf{u})] \hat{\phi}(\mathbf{u}) + \hat{R}(\mathbf{u})$ Iterative solution: $\phi^{(n+1)}(\mathbf{x}) = \arctan \left\{ \mathcal{F}^{-1} \frac{\sum_{j=1}^N \sin[\omega_j(\mathbf{u})] [\hat{I}_{\Delta z_j}(\mathbf{u}) - \hat{I}_{-\Delta z_j}(\mathbf{u}) - \hat{R}^{(n)}(\mathbf{u})]}{\sum_{j=1}^N [2A_m^{(n)} \sin[\omega_j(\mathbf{u})]^2 + \epsilon]} \right\}$ where, $\omega_j(\mathbf{u}) = k\Delta z_j(1 - \sqrt{1 - \lambda^2 \mathbf{u} ^2})$ $R^{(n)}(\mathbf{x}) = A_m^{(n)} \mathcal{F}^{-1} \{ \sin[\omega(\mathbf{u})] \mathcal{F} \{ A_e^{(n)}(\mathbf{x}) \tan[\phi^{(n)}(\mathbf{x})] \} \}$ $+ \mathcal{F}^{-1} \{ \cos[\omega(\mathbf{u})] \mathcal{F} \{ A_e^{(n)}(\mathbf{x}) \} \} \mathcal{F}^{-1} \{ \sin[\omega(\mathbf{u})] \mathcal{F} \{ A(\mathbf{x}) \sin[\phi^{(n)}(\mathbf{x})] \} \}$ $- \mathcal{F}^{-1} \{ \sin[\omega(\mathbf{u})] \mathcal{F} \{ A_e^{(n)}(\mathbf{x}) \} \} \mathcal{F}^{-1} \{ \cos[\omega(\mathbf{u})] \mathcal{F} \{ A(\mathbf{x}) \sin[\phi^{(n)}(\mathbf{x})] \} \}$ $A_m^{(n)} = \text{Mean} \{ A(\mathbf{x}) \cos[\phi^{(n)}(\mathbf{x})] \}$, Mean is the mean value. $A_e^{(n)}(\mathbf{x}) = A(\mathbf{x}) \cos[\phi^{(n)}(\mathbf{x})] - A_m^{(n)}$ with $A(\mathbf{x}) = \sqrt{I(\mathbf{x})}$ $\epsilon > 0$ is a small constant. $\phi^{(n)}(\mathbf{x})$ represents the phase of n^{th} iterations. |

the derivation process, further extend the validity and universality of the algorithm. From the above discussions, we can summarize the various linearization methods for phase retrieval. The phase retrieval problem is an ill-posed inverse problem. In general, we have four different models to solve the inverse problem, which rely on different assumptions: (1) TIE; (2) CTF; (3) mixed transfer function model; (4) intensity difference model without approximations. TIE is only established under the conditions of the paraxial approximation and weak defocusing. The advantage is that it imposes no restrictions on absorption and phase distributions. The establishment of the CTF method requires to satisfy the weak object approximation (or weak absorption and slowly varying phase approximation) and does not require the paraxial approximation and weak defocusing. The mixed model proposed by Guigay et al. [259] relaxes the limitations of weak defocusing and weak object approximation (but still needs the paraxial approximation). Furthermore, the method can be reduced to TIE when the weak defocus approximation is considered and is equivalent to the CTF method when incorporating the weak absorp-

tion approximation. Finally, the intensity difference model does not rely on any approximate conditions. However, the disadvantage is that they cannot recover the phase deterministically, and a nonlinear iterative solution is needed. In Table 4 and Fig. 26, we summarize the mathematical models, establishment conditions, the reconstruction algorithms, and the intrinsic relationship of these methods. Due to the limited space, this tutorial will not elaborate on their corresponding phase reconstruction algorithms. The basic idea behind these nonlinear iterative algorithms is straightforward: they are both based on the principle of ‘‘principal component linearization’’. More specifically, the principal components of the intensity model contributed by the DC term, and the phase term is initially retained, and more complicated higher-order nonlinear terms [such as the last cosine term in Eq. (88)] are neglected. In this way, the initial value can be solved in a linearized manner. Then, the residual intensity component is considered as a perturbation term and can be substituted back into the reconstruction equation to further refine the obtained phase distribution. In general, compared with the iterative

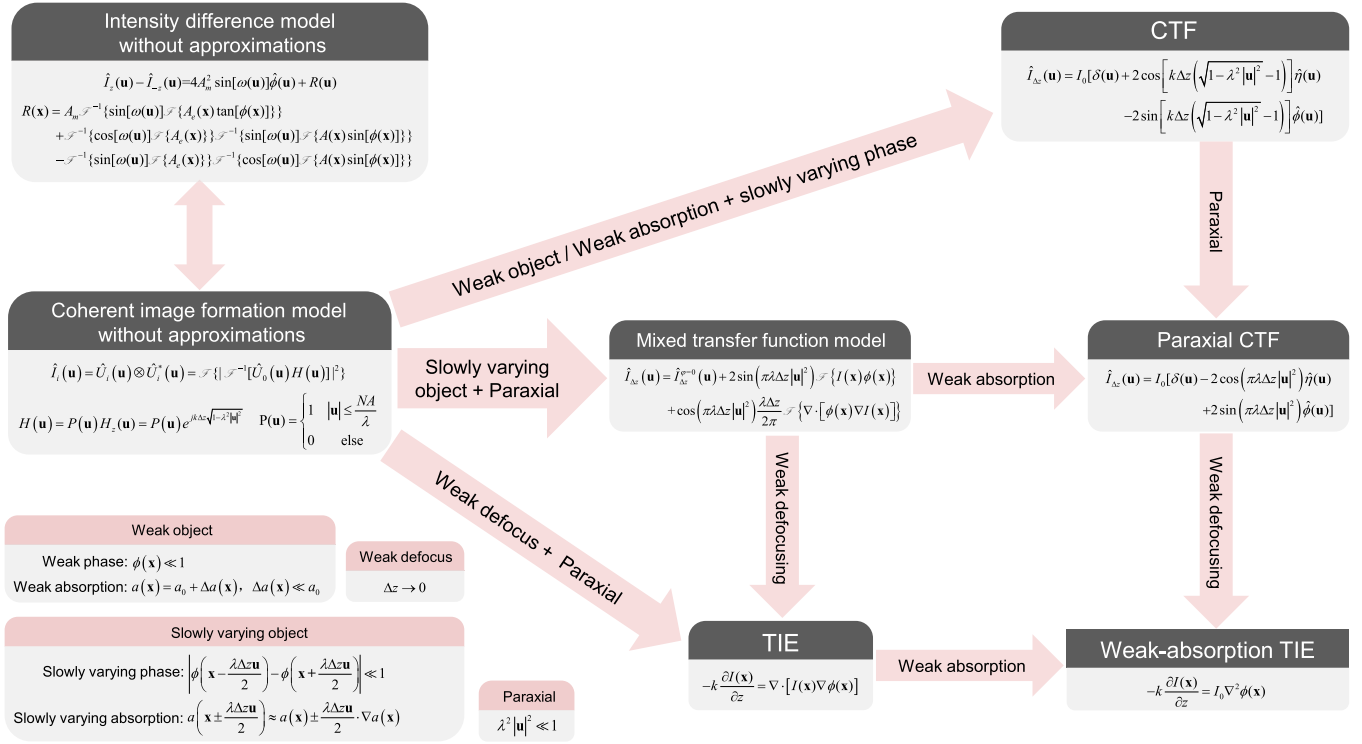


Fig. 26. Mathematical models, establishment conditions, and the intrinsic relationship of different phase retrieval algorithms

phase retrieval method, the convergence of such “principal component linearization” algorithms is much faster and more stable (usually converges within 3-5 iterations).

4.3.4. Finite aperture effect of imaging systems

When deriving the image formation models and the corresponding PTFs in previous subsections, the limited aperture effect of the objective lens is neglected [the objective aperture is assumed infinite, $i.e., P(\mathbf{u}) \equiv 1$ in Eq. (63)]. Obviously, for a practical imaging system, especially for a microscope, the limited aperture effect of the objective is one of the key factors that restrict the resolution of phase imaging, so it cannot be neglected. Fortunately, the effect of the objective NA is very simple for a coherent imaging system. It can be seen from Eq. (62) that the Fourier transform of the intensity at the image plane can be expressed as

$$\hat{I}_{\Delta z}(\mathbf{u}) = \hat{U}_{\Delta z}(\mathbf{u}) \otimes \hat{U}_{\Delta z}^*(\mathbf{u}) = \mathcal{F} \left\{ \left| \mathcal{F}^{-1} [\hat{U}_0(\mathbf{u})P(\mathbf{u})H_{\Delta z}(\mathbf{u})] \right|^2 \right\} \quad (91)$$

where $P(\mathbf{u})$ is the pupil function of the objective lens, and $H_{\Delta z}(\mathbf{u})$ is the angular spectrum transfer function [Eq. (63)]. It is not difficult to find that for an ideal coherent imaging system, the effect of the objective’s NA is equivalent to a low-pass filter in the frequency domain. The information with spatial frequencies lower than NA/λ can reach the image plane, and propagates a distance Δz to form a defocused intensity image [2,4,5]. Thus, when considering the NA of the imaging system, we can simply assume that the measured object is a low-pass filtered version of the original ideal object. Then all conclusions drawn in Subsection 4.3 can be directly applied by incorporating a pupil function $P(\mathbf{u})$ into the transfer functions. For example, when considering the limited NA of an imaging system, the ATF and PTF [Eqs. (81) and (82)] should be rewritten as

$$H_A(\mathbf{u}) = P(\mathbf{u}) \cos(\pi\lambda\Delta z|\mathbf{u}|^2) \quad (92)$$

$$H_P(\mathbf{u}) = P(\mathbf{u}) \sin(\pi\lambda\Delta z|\mathbf{u}|^2) \quad (93)$$

Their curves are plotted in Fig. 27. Limited by the lens aperture, the highest spatial frequency of the recovered phase under coherent illumination is limited to the coherent diffraction limit, $i.e., NA/\lambda$.

5. Axial intensity derivative estimation

In previous sections, it was shown that the in-focus intensity image and the axial intensity derivative are required for solving TIE. However, the axial intensity derivative $\partial I/\partial z$ cannot be measured directly, but it can be estimated from multiple intensity images captured at different defocus distances. The accuracy of the derivative estimation is critical for reliable phase retrieval based on TIE. In this section, we will discuss this issue in detail.

5.1. Two-plane axial intensity derivative estimation

In order to estimate the axial intensity derivative $\partial I/\partial z$, Teague [141] proposed the classical two-plane finite-difference method in his original TIE paper. He suggested to capture two intensity images with slightly defocusing, and the defocus distances of the two images are equal and opposite relative to the central in-focus image. Then, the central finite-difference formula is used to estimate the axial intensity derivative

$$\frac{\partial I(\mathbf{x})}{\partial z} \approx \frac{I(\mathbf{x}, \Delta z) - I(\mathbf{x}, -\Delta z)}{2\Delta z} \quad (94)$$

In this section, we explicitly represent the intensity as a function of the defocus distance Δz . The two-plane-based finite-difference formula [Eq. (94)] is simple and easy to implement. But it also poses a problem: how to choose an appropriate defocus distance Δz [141,171,264]. As shown in the simulation results of Fig. 28, the accuracy of the derivative estimation increases with the decrease of the defocus distance Δz , and the spatial resolution of the corresponding reconstructed phase is also improved. However, when the defocus distance Δz becomes larger, the accuracy of the differential approximation will decrease, leading to “phase blurring” effect, $i.e.,$ the degradation of the spatial resolution, as

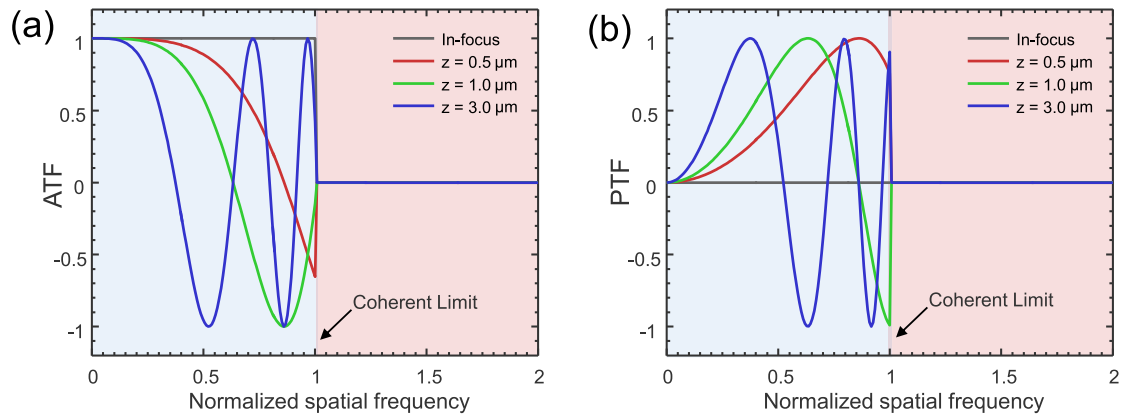


Fig. 27. Transfer functions of a limited aperture imaging system for different defocus distances under coherent illumination. (a) Amplitude transfer function; (b) Phase transfer function.

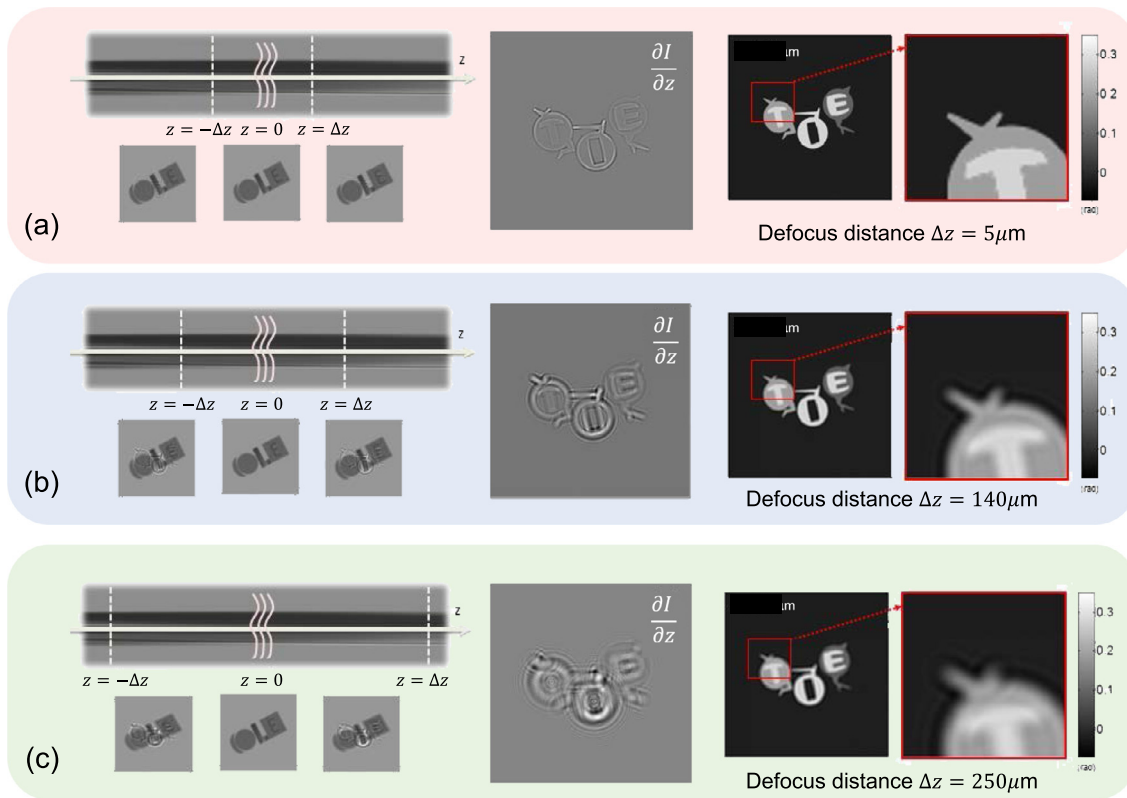


Fig. 28. The effect of defocus distances on the TIE phase reconstruction under noise-free condition. (a) Small defocus distance: original intensity image (left); axial intensity derivative (middle), and recovered phase distribution (right); (b) medium defocus distance; (c) large defocus distance; the square areas with red lines are magnified for clarity.

shown in Fig. 28(c). The “phase blurring” effect is also referred to as “non-linearity error”. This is because that the locally linear approximation is assumed in Eq. (94), and the increase in Δz will inevitably induce larger errors due to the nonlinear terms of the actual intensity signal. Therefore, mathematically, the smaller the defocus distance Δz is, the more accurate the two-plane finite difference can be approximated to the ideal derivative.

However, real measurements yield data with noise and discretization, which makes the problem more complicated. Under noisy conditions, the defocus distance Δz cannot be too small. Otherwise, the intensity derivative estimate will be overwhelmed by noise, as shown in Fig. 29(a). The cloud-like low-frequency noise will appear in the reconstructed phase, so the defocus distance Δz has to be large enough in order to secure an adequate level of SNR, as shown in Fig. 29(b). The

phase blurring effect becomes apparent when the defocus distance Δz becomes large. Thus a compromise is made where Δz is chosen to balance the high-order (or non-linearity) error and the noise effect. It is understandable that the optimal Δz should depend on both the maximum physically significant frequency of the object and the noise level.

5.1.1. Causes of low-frequency noise and high-frequency phase blurring

From simulation results shown in the previous section, it can be found that when the intensity images are noisy, the phase reconstructed by TIE will be contaminated by cloud-like low-frequency artifacts. This is a notorious problem in TIE, and the causes of the strong sensitivity to low-frequency artifacts will be analyzed in this subsection. For simplicity, let us consider the simplified TIE under the uniform

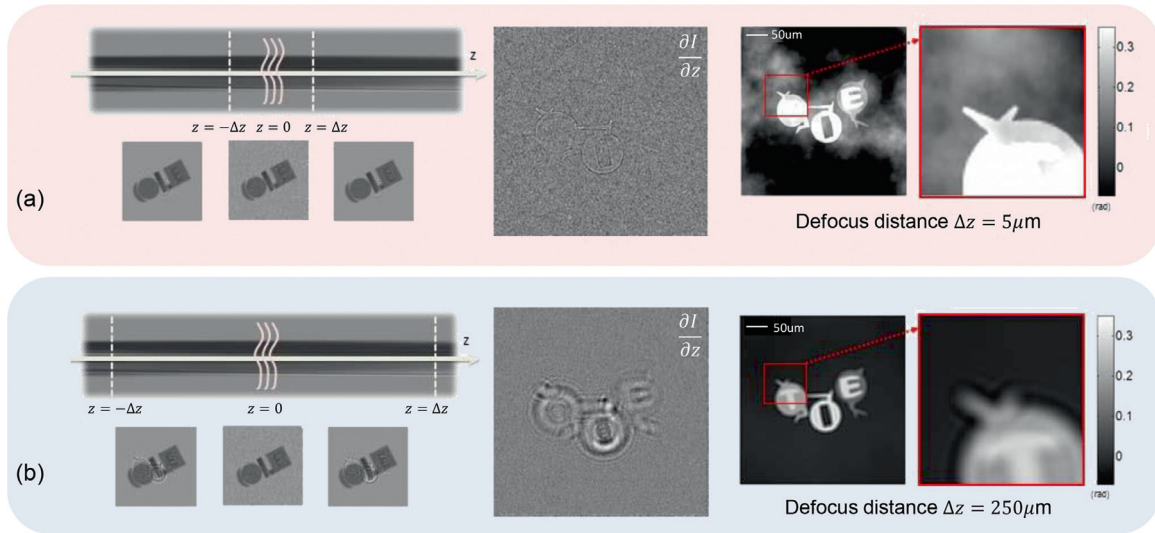


Fig. 29. The effect of defocus distance on the TIE phase reconstruction in the presence of noise. (a) Small defocus distance: original intensity image (left), axial intensity derivative (middle), and recovered phase distribution (right); (b) large defocus distance: original intensity image (left), axial intensity derivative (middle), and recovered phase distribution (right); the square areas with red lines are magnified for clarity.

intensity [Eq. (42)]

$$-k \frac{\partial I(\mathbf{x})}{\partial z} = I(\mathbf{x}) \nabla^2 \phi(\mathbf{x}) \quad (95)$$

It can be written as $\nabla^{-2} \leftrightarrow -\frac{1}{4\pi^2|\mathbf{u}|^2}$ in the Fourier domain

$$\hat{\phi}(\mathbf{u}) = \frac{1}{2\pi\lambda|\mathbf{u}|^2} \mathcal{F} \left\{ \frac{1}{I(\mathbf{x})} \frac{\partial I(\mathbf{x})}{\partial z} \right\} \approx_{\Delta z \rightarrow 0} \frac{1}{4\pi\lambda\Delta z|\mathbf{u}|^2} \frac{\hat{I}(\mathbf{u}, \Delta z) - \hat{I}(\mathbf{u}, -\Delta z)}{\hat{I}(\mathbf{u})} \quad (96)$$

where $\hat{\phi}(\mathbf{u})$ is the Fourier transform of the phase to be recovered, and the RHS is the Fourier transform of the normalized axial intensity derivative estimated by the two-plane finite difference. Comparing Eq. (96) with Eq. (83), $\pi\lambda\Delta z|\mathbf{u}|^2$ in the denominator of Eq. (96) can be regarded as the PTF of TIE

$$H_{TIE}(\mathbf{u}) = \pi\lambda\Delta z|\mathbf{u}|^2 \quad (97)$$

The PTF represents the weights of different spatial frequency components of the phase that can be transferred into the intensity via defocusing. The solution to TIE can be interpreted by an inverse filtering process in frequency domain based on the PTF (multiplying the inverse Laplacian $1/\pi\lambda\Delta z|\mathbf{u}|^2$). It should be noted that the response of the PTF $H_{TIE}(\mathbf{u})$ decrease to 0 with $|\mathbf{u}| \rightarrow 0$ quadratically, so the low-frequency phase components can hardly be transferred into the intensity via defocusing. For the limiting case of the zero-frequency component, the constant phase does not produce any phase contrast. In other words, since the low-frequency phase information in the captured intensity signal is weak, it requires a higher gain in the inverse filtering process in frequency domain (the inverse Laplacian has a singularity at the zero-frequency, and the gain near the zero-frequency tends to infinity accordingly). When there is noise, the low-frequency noise components (especially for the components close to the zero-frequency) will be amplified by the inverse Laplacian to create cloud-like artifacts superimposed on the reconstructed phase.

Next, let us turn to the high-frequency phase blurring effect shown in Fig. 30. For a weak phase object, it can be known from Subsection 4.3.2 that the PTF in the CTF model under weak object approximation can be represented as

$$H_{CTF}(\mathbf{u}) = \sin(\pi\lambda\Delta z|\mathbf{u}|^2) \quad (98)$$

Note that the establishment of the linearization does not impose any restrictions on the defocus distance Δz , so Eq. (83) is still valid for a large

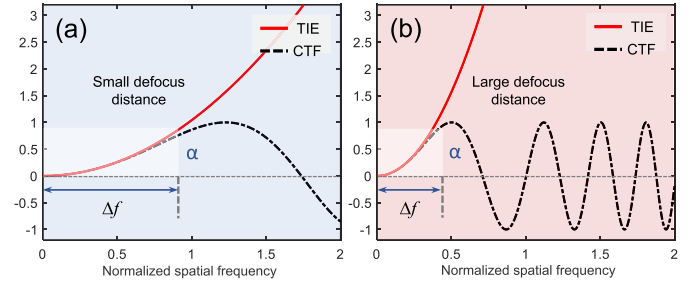


Fig. 30. Comparison of phase transfer functions of TIE and CTF under different defocus distances.

defocus distance. Comparing Eqs. (97) with (98), it can be found that since $H_{TIE}(\mathbf{u})$ is not equal to $H_{CTF}(\mathbf{u})$, the phases reconstructed from the CTF and TIE may not be identical. In Fig. 30, we compare the curves of $H_{TIE}(\mathbf{u})$ and $H_{CTF}(\mathbf{u})$ with different defocus distances. It can be seen that the two curves almost overlap at low frequencies as Δz approaches to 0. Nevertheless, with the increase in defocus distance Δz , the two functions begin to diverge. The underlying assumption of TIE's PTF $H_{TIE}(\mathbf{u})$ is the phase contrast increases linearly with defocus distance and quadratically with spatial frequencies. This assumption is obviously unreasonable since the phase contrast cannot arbitrarily increase with defocus distance or spatial frequencies (violate the energy conservation law). While $H_{CTF}(\mathbf{u})$ is valid for large defocusing, the curve tends to gradually oscillate with the increase in the spatial frequency and the defocus distance Δz . Therefore, the overestimation of the phase contrast at high spatial frequencies is the primary cause of the phase blurring effect if one misuses TIE beyond the small-defocus regime.

5.1.2. Unified framework based on Savitzky-Golay differential filter

In general, the multiple-plane axial intensity derivative estimation method can improve the high-order error and reduce the noise effect compared to the two-plane method. While these strategies have been shown to work well in many situations, their performance, however, depends heavily on the noise level and the characteristics of the experimental data [197–199]. With a given set of intensity measurements, making an appropriate choice of a suitable algorithm is difficult. For the high-order finite-difference with noise-reduction and the least-squares fitting methods, it is also necessary to determine the fitting order, which

is quite similar to the case that how to select the optimal defocus distance in the two-plane finite-difference method. Therefore, it is necessary to establish a theoretical framework that can more systematically understand, analyze, compare, and even improve the existing axial intensity derivative estimation methods.

To address this problem, Zuo et al. [205] proposed a unified framework of TIE axial intensity derivative estimation based on the Savitzky-Golay differential filter. The Savitzky-Golay filter was proposed by Savitzky and Golay [265] in 1964. They proven that fitting a polynomial to a set of input samples and then evaluating the resulting polynomial at a single point within the approximation interval is equivalent to discrete convolution with a fixed impulse response function (*i.e.*, the Savitzky-Golay filter). More specifically, the *Savitzky-Golay differential filter* (SGDF) is equivalent to the convolutional expression of the least-squares fitting method proposed by Waller et al. [189]. To evaluate the s th derivative at point t using a polynomial of m th degree on $2n + 1$ data points, the coefficients of the SGDF, *i.e.*, the weights of the convolution kernel can be calculated as [266]

$$a_i = \sum_{k=0}^m \frac{(2k+1)(2n)^{(k)}}{(2n+k+1)^{(k+1)}} P_k^n(i) P_k^{n,1}(0) \quad (99)$$

where $(a)^{(b)}$ is a generalized factorial function $(a)(a-1)\dots(a-b+1)$ with $(a)^{(0)} = 1$, and $P_k^n(t)$ is the Gram polynomial, and $P_k^{n,s}(t)$ is its s -order derivative defined as

$$P_k^{n,s}(t) = \left(\frac{d^s}{dx^s} P_k^n(x) \right)_{x=t} \quad (100)$$

Simplifying the Eq. (99) by setting the fitting order as $m = 2n, 1, m(m < 2n + 1)$ in Eq. (58), Zuo et al. [205] establish the correspondence between the different finite-difference methods and SGDFs, *i.e.*, least-squares fitting, in particular:

Conclusion 1: The high-order finite-difference method corresponds to the SGDF with degree $2n$.

Conclusion 2: The noise-reduction finite-difference method corresponds to the SGDF with degree 1.

Conclusion 3: The higher order finite-difference with noise-reduction method corresponds to the SGDF with degree m ($m < 2n + 1$).

Finally, it should be mentioned that the unequally spaced multi-plane methods [265–267] fall in the class of SGDF without exception, since the Savitzky-Golay filter has also been generalized for unequally or non-uniformly spaced data as its offspring. Even the two-plane method can be viewed as a special case of the SGDF when $n = 1, m = 1$.

The advantage of unifying all of the above derivative estimation strategies under the framework of SGDF is that it provides great insights into the behaviors, the shortcomings, and the performance of these existing intensity derivative estimation algorithms by only analyzing the properties of the corresponding SGDFs. SGDF has many good properties [268]:

- 1) The convolution operation is simple and is much faster and easier to implement than the standard least-squares fitting;
- 2) The convolution kernel (weight coefficients) can be easily obtained using a look-up table generated from the explicit solution, or pre-calculated using existing routines;
- 3) They are optimal differential filters that minimize the NRR subject to moments preservation constraints [267].

Among them, the property (3) is the most important one. From the perspective of probability and statistics: if the signal can be perfectly modeled by an m -order polynomial and the observed data are with independent Gaussian white noise, the m^{th} degree SGDF is an unbiased estimator of the derivative, and the estimation accuracy can reach the Cramer-Rao lower bound [268]. However, in practice, the signal order is unknown, and hence using a nonoptimal degree (*i.e.*, “underfitting” or “over-fitting”) will inevitably affect the accuracy and the effect of noise reduction. From the perspective of signal processing, the NRR $\sum_{i=-n}^n a_i^2$ determines the noise resistance ability of the SGDF. Through the frequency response decomposition, a SGDF (frequency response $H_{SG}(e^{j\omega})$)

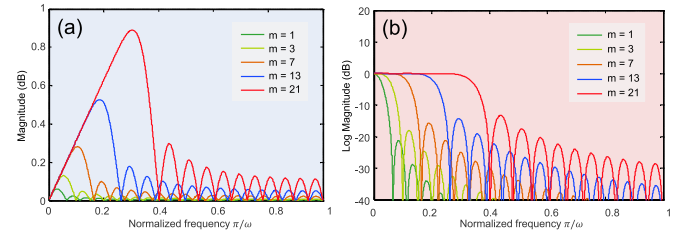


Fig. 31. Frequency domain properties of SGDFs with different orders ($n=15$). (a) Frequency domain response of the SGDF; (b) frequency domain responses of the low-pass filters implied in SGDFs.

can be regarded as the combination of an ideal first-order derivative filter (frequency domain response is $H_{ideal}(e^{j\omega}) = j\omega$) and a low-pass filter (frequency domain response is $H_{LP}(e^{j\omega})$):

$$H_{SG}(e^{j\omega}) = j\omega \frac{H_{SG}(e^{j\omega})}{j\omega} = H_{ideal}(e^{j\omega}) H_{LP}(e^{j\omega}) \quad (101)$$

Eq. (101) suggests that SGDF is equivalent to calculating the ideal first-order differentiation derivative on a smoothed (denoised) version of the data. Higher-order SGDFs can approach the ideal differential filter in the higher frequency range in the frequency domain response. However, the larger the noise suppression factor, the more sensitive it is to noise. On the contrary, although the response of the high-frequency part of the low-order filter is attenuated, its inherent low-pass filtering has a strong resistance to noise. From the frequency response (in log scale) of the low-pass filter plotted in Fig. 31, it can be seen the filters have very flat frequency responses in their passbands with modest attenuation in their stop band. The frequency response of SGDF is flatter and closer to the ideal derivative filter at low frequencies as the polynomial degree increases. The increase of the flatness is coincident with a relatively higher cut-off frequency or a wider low-pass range. Conversely, as the degree of fitting polynomial decreases, the effect of low-pass filtering becomes more evident.

In summary, when estimating the axial intensity derivative based on multi-plane intensity measurements, merely modifying of the finite-difference scheme in the spatial domain cannot fully utilize all the intensity information of multiple planes. Instead, it converts the contradiction in the selection of the defocus distance into the selection of the filter degree. A higher-degree SGDF gives a more intact phase (with a wider range of spatial frequencies can be accurately recovered) than a lower degree one. However, it results in larger NRR than the one assumes a smaller order. In other words, lower-order SGDFs always give estimates of the low-frequency phase with higher SNR, but they suffer from the problem of poor response for high-frequency phase variations. Therefore, the tradeoff between low-frequency artifacts and the high-order error can still not be fundamentally reconciled by only using a single degree of SGDF.

5.1.3. Convergence with the transfer function theory

TIE is sensitive to noise in the low-frequency range, due to the weak transfer of the low-frequency information in the beam propagation. On the other hand, another important aspect is the spatial resolution, which can be refined if the nonlinear contribution is taken into account. Instead of optimizing the finite-difference schemes in the spatial domain, many researchers tried to address the contradiction between low-frequency noise and high-order error by optimizing the PTF in the frequency domain. The optimal estimation of the axial intensity derivative is closely related to the transfer function theory discussed in Subsection 4.3. The PTF can be utilized to quantitatively analyze the relationship between phase contrast and defocus distance, thus providing theoretical guidance for optimizing the derivative estimation. There are two options for solving this problem:

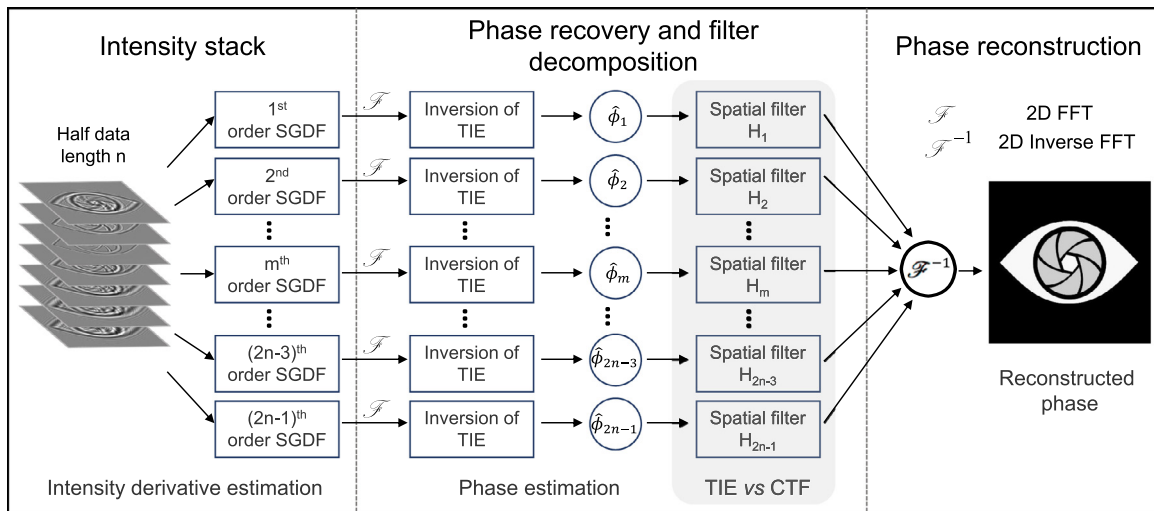


Fig. 32. The block diagram of the multi-plane TIE phase recovery method based on OFS.

1) Combining TIE with nonlinear phase retrieval algorithms

The nonlinear contributions in the image contrast formation are non-negligible when large propagation distances are used. The limitation due to the linearization can be refined by other methods which take into account the nonlinearity of the phase problem. One of such composite phase retrieval methods was proposed by Gureyev [269], in order to improve the phase solution. This method uses the phase map obtained with TIE as an initial approximation and then is refined by Gerchberg-Saxton and Fienup method. By using the linear TIE solution as a starting point, the amount of computations is reduced, and stagnation traps are avoided. Donnadiu et al. [270] pointed out that the PTF of TIE does not coincide with the CTF model under weak object approximation at large defocus distance, so the high-frequency phase information cannot be reconstructed correctly. He also found that iterative phase retrieval can effectively alleviate the sensitivity of TIE to low-frequency noise, and proposed a combined phase recovery scheme similar to Gureyev's method [269]. Subsequently, Gureyev et al. [256] unified the first-order Born approximation and TIE into the same theoretical framework, expanding the linearization range of TIE from the weak defocus to near Fresnel region. Guigay et al. [259] combined the CTF method with TIE to derive a mixed transfer function model (discussed in Subsection 4.3). The model reduces to the standard TIE in the small defocus regime and is equivalent to the CTF model under the weak object approximation. Langer et al. [260] compared TIE, CTF, and the mixed model by simulations and experiments. The mixed model seems to be more accurate and robust to noise than the other two methods, but TIE is most accurate without noise. The mixed model combines the advantages of TIE for strong absorptive samples and the CTF for large defocusing, effectively extending the valid range of TIE phase retrieval. However, the mathematical model becomes more complicated and cannot be fully linearized. Therefore, it is generally necessary to use TIE or the CTF method to get an initial solution, and then the initial solution is refined by an iterative procedure to get the final exact solution.

2) Optimum frequency selection based on multi-plane finite-difference schemes:

The simulation results in Subsection 5.1.1 show that in the traditional two-plane TIE, when the defocus distance is too small, high-frequency details can be well preserved, but the low-frequency noise is strong. In contrast, when the defocus distance is too large, the low-frequency noise can be suppressed, but high-frequency details vanish. Although it is difficult to make a good compromise between these two aspects, it inspires us with a straightforward idea: if the intensity images

at both small and large defocus distances are recorded, we can reconstruct two different phase images separately. Then, the high-frequency component of the reconstructed phase with small defocus and the high-frequency component of the reconstructed phase with large defocus can be combined to overcome low-frequency noise and high-order error simultaneously. Based on this idea, Paganin et al. [264] used a pair of complementary Gaussian low-pass/high-pass filters to decompose the phases recovered with both small and large defocusing. Based on the transfer function analysis, they derived the optimal cut-off frequency of the complementary filter. Although this method can effectively improve the noise robustness of TIE, it is only limited to the case of two defocus distances. As an extension of the traditional two-plane TIE, it does not make full use of all the captured intensity information. Regarding the multi-plane finite-difference schemes, Zuo et al. [205] analyzed the PTFs of TIE and CTF, and found that the phase recovered by TIE based on SGDF can be regarded as a low-pass filtered version of the ideal phase

$$\hat{\phi}(\mathbf{x}) = \phi(\mathbf{x}) * \mathcal{F}^{-1} \left\{ H_{LP}(e^{j\omega}) \Big|_{\omega=\Delta z\pi\lambda u^2} \right\} \quad (102)$$

The frequency response of the low-pass filter H_{LP} is exactly the same as Eq. (101), suggesting that the frequency characteristic of the retrieved phase using SGDF is determined by the frequency response of the low-pass filter that is implicit in the SGDF. Considering the phase information lies within a certain frequency bandwidth such that it can be correctly reconstructed by both a lower degree SGDF and a higher degree one (here "correct reconstruction" is quantitatively defined in [205] as to that the spatial frequencies of the phase information is below the 0.3dB cut-off frequency of the low-pass filter $H_{LP}(e^{j\omega})$ implied in the SGDF with a specific degree), the lower degree filter is preferred for its lower NRR, or in another word, higher SNR. Higher-order SGDFs are used if and only if none of the lower-order filters can correctly reconstruct the phase information for that given frequency band. Based on this idea, Zuo et al. [205] proposed the *optimal frequency selection* (OFS) method. Based on the unified framework based on SGDF, the optimal cut-off frequencies for SGDFs with different orders are derived, and the best frequency components of the phase images obtained from SGDFs with various degrees are extracted by a properly designed complementary filter bank. The extracted spatial frequency components of the phase are finally recombined into an optimal composite phase. Compared with the method of Paganin et al. [264], OFS is more flexible and can be easily extended to unequally-spaced and arbitrary number of measurement planes. In addition, SGDF makes full use of all the intensity measurements for the axial intensity derivative estimation so that better reconstruction results can be obtained. However, the disadvantage of this method is that it assumes

that the defocus distances are evenly distributed, so a large amount of intensity data is still needed. Besides, the selection of the defocus interval was not discussed. To address this problem, Martinez-Carranza et al. [271] used the noise-reduction finite-difference method of Soto and Acosta [272] as an example to discuss the optimal defocus distance selection for multi-plane finite-difference schemes. The selection criterion of the optimal defocus distance can be maximum derivative estimation accuracy [271] or minimum phase reconstruction error [271], and the latter one proven to provide better phase reconstruction accuracy. Based on the findings of Falaggis et al. [200], Zhong et al. [273] proposed a *Gaussian process* (GP) regression algorithm to accurately estimate the axial intensity derivative based on exponentially distributed defocus intervals. Thereby, the number of required images can be significantly reduced without compromising the reconstruction accuracy. Martinez-Carranza et al. [274] pointed out that the selection of the defocus distance should also be based on the minimum phase reconstruction error criterion rather than the maximum derivative estimation accuracy criterion, and further optimized the defocus distances (unequal intervals) based on the OFS principle. Sun et al. [206] further deduced the nonlinear intensity difference model under non-paraxial and non-weak object conditions. When the object to be measured has strong absorption, it firstly adopt the OFS scheme under unequal spacing to obtain the initial solution. Then the phase is iteratively optimized based on the nonlinear intensity difference model to obtain a more accurate reconstruction result.

5.2. Regularization methods for ill-posed inverse problems

No matter which axial intensity derivative estimation methods are used, the TIE phase retrieval is an inherently ill-posed inverse problem. Under uniform intensity approximation, TIE becomes a standard Poisson equation [Eq. (42)]. The phase retrieval by solving TIE is essentially a deconvolution based on inverse Laplacian in the frequency domain [Eq. (96)]. Similarly, in the CTF method discussed in Subsection 4.3, the defocus-induced phase contrast of a phase object is modeled as the product of the phase distribution and the PTF. The phase reconstruction also boils down to a frequency-domain deconvolution based on the corresponding PTF. Alternatively, the frequency-domain deconvolutions used in these phase retrieval algorithms can be discretized and represented as an energy minimization problem in the spatial domain

$$\hat{\Phi} = \arg \min_{\Phi} \frac{1}{2} \|\mathbf{H}\Phi - \mathbf{b}\|_2^2 \quad (103)$$

where $\mathbf{H} \in \mathbb{R}^{N \times N}$ represents the matrix representation of the discrete Laplacian or the corresponding PTF, $\Phi \in \mathbb{R}^N$ represents the phase vector to be solved, and $\mathbf{b} \in \mathbb{R}^N$ represents the measurable quantity (the normalized axial intensity derivative). In this subsection, we consider the image contains N pixels and is lexicographically row-stacked into a single column vector.

As mentioned earlier, due to the presence of noise and the zero-frequency singularity in the PTF of TIE $H_{TIE}(\mathbf{u})$, the deconvolution is ill-conditioned: the solution is highly sensitive to small perturbations (e.g., noise and system misalignment) in the measured data. The ill-posedness is more pronounced for the CTF method because the PTF $H_{CTF}(\mathbf{u})$ is strongly oscillatory at higher spatial frequencies, leading to numerous zero-crossings. The discussions in previous sections focus on how to minimize the ill-posedness by optimizing the defocus distance or using multiple intensity measurements. While in mathematics, statistics, and computer science, an alternative approach is so-called “regularization”. The principle of regularization is to convert the ill-posed problem into a set of well-posed problems by adding prior information to prevent unrealistic solutions

$$\hat{\Phi} = \arg \min_{\Phi} \frac{1}{2} \|\mathbf{H}\Phi - \mathbf{b}\|_2^2 + \tau R(\Phi) \quad (104)$$

where $R(\Phi) \in \mathbb{R}^N$ is the regularization function (also known as the penalty term), and τ is the corresponding regularization parameter. Fundamentally, the ill-posedness of the deconvolution optimization problem [Eq. (103)] is due to the fact that the solution space is too wide, thereby leading to instability. The regularization method, by introducing additional constraints, defines a compact set enclosing the true solution, then we can search a physically reasonable solution within the intersection between the original solution space and the defined compact set such that the obtained solution is stable and continuously dependent on the observed data. The well-known and commonly used regularization models, such as Tikhonov [159,275–277] and *total variation* (TV) [275,278,279] were successfully applied to TIE to suppress the low-frequency artifacts introduced by the ill-posedness of inverse Laplacian. The most popular Tikhonov regularization approximates the solution of the inverse problem through a minimization problem by adding a quadratic regularization term to Eq. (103). It is very simple and easy to implement. However, the denoising ability is limited and often accompanied by the excessive suppression of low-frequency components of the object. TV regularization is remarkably effective at simultaneously preserving edges whilst smoothing away noise in flat regions, even at low SNR. It is very suitable for the reconstruction of step-like phase samples. However, the algorithm is complex in calculation and sensitive to the selection of regularization parameters, and the reconstructed phase suffers from the so-called “staircase effect”. On the other hand, prior information about the sample, such as non-negativity constraints, support constraints, and the phase-attenuation duality commonly used in X-ray diffraction imaging [276,280] can also be effectively integrated into the optimization model. The solution space is further constrained based on the *a priori*, and thereby the phase reconstruction quality can be further improved.

6. Image formation under partially coherent illuminations

6. Image formation under partially coherent illuminations

No matter completely coherent fields, partially coherent fields, or completely incoherent fields, the intensity of the optical field is well-defined and directly accessible. Nevertheless, the definition of “phase” is only limited to fully coherent fields. Teague’s derivation of TIE was also based on the assumption of complete coherence [141], i.e., a monochromatic coherent beam with a well-defined phase. However, as discussed in Subsection 4.1, any physically achievable light sources are not strictly coherent, and in the field of optical microscopy, partially coherent illumination is beneficial to enhance the imaging resolution, improve the image quality, and suppress the coherent noise. Therefore, in this section, we will introduce the basic knowledge about statistics optics and coherence theory, and derive the image formation models under partially coherent illuminations. It provides a theoretical basis for the next section where the generalization of TIE for partially coherent fields is discussed.

6.1. Correlation function representations of partially coherent fields

The light emitted by most of the light sources in the real world is generated by different independent radiation oscillators. They produce fields that vary in time with highly complicated and irregular waveforms. Because of diffraction, these waveforms are greatly modified as the fields propagate. All photodetectors measure the time-averaged intensity over the waveform. This measurement depends on the integration time of the detector and the waveform of the light at the detector. Generally, this waveform is not precisely known. Coherence theory is a mathematical model that is very successful in describing the effects of this unknown waveform on the observed measurement of time-averaged intensity. It is based on the electromagnetic wave theory of light as formulated from Maxwell’s equations and uses statistical techniques to analyze the effects due to fluctuations in the waveform of the field in both time and space. As in scalar diffraction theory, it is much more convenient to treat monochromatic fields than it is to deal with fields that have complicated time dependencies. Therefore, each of these scalar components is usually represented at some typical points in space by a

superposition of monochromatic scalar components. Thus the field amplitude for a typical monochromatic field $U(\mathbf{x}, t)$ with angular frequency ω_0 is given by can be expressed as

$$U(\mathbf{x}, t) = U(\mathbf{x}) \exp(-j\omega_0 t) = A(\mathbf{x}) \exp[j(\phi - \omega_0 t)] \quad (105)$$

The amplitude of the optical field does not change over time, while the phase changes linearly with time. For a monochromatic deterministic optical field, the optical oscillation at each spatial point is the same in time and spatially infinitely extended, therefore, the time-dependent and fast-fluctuating part $\exp(-j\omega_0 t)$ of $U(\mathbf{x}, t)$ are usually ignored, it can be described deterministically by the 2D scalar complex amplitude $U(\mathbf{x})$ or a “phasor”. In addition, since the exponential term $\exp(-j\omega_0 t)$ is the eigenfunction of the linear time-invariant system, the Fourier transform of the time variable t for the monochromatic coherent field $U(\mathbf{x}, t)$ can be expressed as

$$\mathcal{F}\{U(\mathbf{x}, t)\} = A(\mathbf{x}) \exp[j\phi(\mathbf{x})] \delta(\omega - \omega_0) \quad (106)$$

It seems self-evident that there should be only one spectral peak for the monochromatic optical field. In the following part, we denote the 2D scalar complex amplitude of the monochromatic coherent field $U_\omega(\mathbf{x})$ with angular frequency ω as the subscript

$$U_\omega(\mathbf{x}) = A_\omega(\mathbf{x}) \exp[j\phi_\omega(\mathbf{x})] \quad (107)$$

However, for practical light sources, from the quantum nature of light emission, the amplitude and phase of the optical field are inevitably disturbed under the influence of external temperature, humidity, vibration, and other factors, together with the statistical fluctuations and inhomogeneous attenuation of the light-emitting atoms. Therefore, the 2D complex amplitude function $U(\mathbf{x})$ cannot fully describe the random disturbances at different moments or different spatial positions for partially coherent fields. For the simplest case, we know that for two beams of the same frequency $U_{\omega 1}(\mathbf{x})$ and $U_{\omega 2}(\mathbf{x})$ meet in space, and whether the interference fringes can be formed are determined by their interference terms $A_{\omega 1}(\mathbf{x})A_{\omega 2}(\mathbf{x}) \cos[\phi_{\omega 2}(\mathbf{x}) - \phi_{\omega 1}(\mathbf{x})]$. If the phase difference between the two beams cannot keep stable during the observation time due to the light-emitting mechanism or other factors, the interference term may be attenuated or even vanish by the random disturbances of the phase in the time average, thereby directly affecting to the formation of interference effects. In this case, the complex amplitude representation for deterministic fields is insufficient. Instead, we need to resort to statistical approaches to describe the coherence of the field. According to the theory of stochastic processes, the scalar function $U(\mathbf{x}, t)$ is regarded as a typical member from the “ensemble” that characterizes the statistical properties at point \mathbf{x} and time t of the optical field, or so-called a “realization”. Without loss of generality, the optical field is assumed to be a stationary and ergodic process, and its statistical properties do not change over time. Fig. 33 shows a typical “realization” in time (optical oscillations at a fixed point in space over time) in a partially coherent field and a typical “realization” in space (optical oscillations at different locations at a given time point).

From the perspective of temporal frequency (spectrum), the random disturbance of the optical field $U(\mathbf{x}, t)$ can also be regarded as the inco-

herent superimposition of the deterministic monochromatic fields with different frequencies. So $U(\mathbf{x}, t)$ can be written as the following Fourier transform

$$U(\mathbf{x}, t) = \int U(\mathbf{x}, \omega) \exp(-j\omega t) d\omega \quad (108)$$

where $U(\mathbf{x}, \omega)$ is the (time-dependence-free) scalar complex amplitude of the deterministic monochromatic optical field of the frequency ω [to be distinguished from the monochromatic deterministic field $U_\omega(\mathbf{x})$, here $U(\mathbf{x}, \omega)$ explicitly includes the variable ω to represent polychromatic decomposition of the partially coherent field], and $\exp(-j\omega t)$ is the corresponding time-dependent part with high-frequency temporal oscillations. The integral represents the incoherent superposition of monochromatic optical components of different frequencies. For the description of the properties of partially coherent fields, $U(\mathbf{x}, \omega)$ and $U(\mathbf{x}, t)$ are of equal importance (because they are Fourier transform pairs), which represent a “realization” at the spatial point \mathbf{x} from the ensemble of an optical field with an arbitrary state of coherence in the temporal frequency domain, characterized by the angular frequency ω .

6.1.1. Mutual coherence function and cross-spectral density

As we mentioned in the beginning of this tutorial, the oscillation frequencies of light waves are very high, which is far beyond the temporal resolution of the current photoelectric imaging device. In fact, the quantity that can be measured is the time-averaged intensity of the optical field. For example, for a partially coherent field $U(\mathbf{x}, t)$, the intensity captured by photodetector can be represented as

$$I(\mathbf{x}) = \langle |U(\mathbf{x}, t)|^2 \rangle = \langle U(\mathbf{x}_1, t_1) U^*(\mathbf{x}_2, t_2) \rangle \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}, t_1=t_2} \\ \stackrel{t=t_1}{\tau=t_2-t_1} \langle U(\mathbf{x}_1, t) U^*(\mathbf{x}_2, t+\tau) \rangle \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}, \tau=0} \quad (109)$$

where the sharp bracket indicates the time average. Similarly, for coherent fields, we can create its two copies $U_1(\mathbf{x}_1, t_1)$ and $U_2(\mathbf{x}_2, t_2)$ by amplitude or wavefront splitting, and then superimpose them in space to produce interference. The interference pattern captured by the photodetector can be written as [281]

$$I(\mathbf{x}) = \langle |U(\mathbf{x}_1, t_1) + U(\mathbf{x}_2, t_2)|^2 \rangle \\ \stackrel{t=t_1}{\tau=t_2-t_1} \langle |U(\mathbf{x}_1, t) + U(\mathbf{x}_2, t+\tau)|^2 \rangle \\ = \langle |U(\mathbf{x}_1, t)|^2 \rangle + \langle |U(\mathbf{x}_2, t+\tau)|^2 \rangle + \\ 2\text{Re}\langle U(\mathbf{x}_1, t) U^*(\mathbf{x}_2, t+\tau) \rangle \\ = I(\mathbf{x}_1) + I(\mathbf{x}_2) + 2\text{Re}\langle U(\mathbf{x}_1, t) U^*(\mathbf{x}_2, t+\tau) \rangle \quad (110)$$

Comparing Eq. (109) with Eq. (110), we can find that there is a common correlation function in both of the intensity distributions. This function represents the cross-correlation of the light signals at two different points \mathbf{x}_1 and \mathbf{x}_2 , which is termed as the *mutual coherence function* (MCF) [282]. It plays a fundamental role in the coherence theory. For a stationary and ergodic optical field, MCF is more strictly defined as a

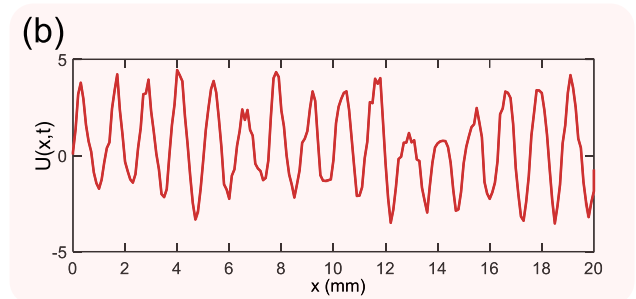
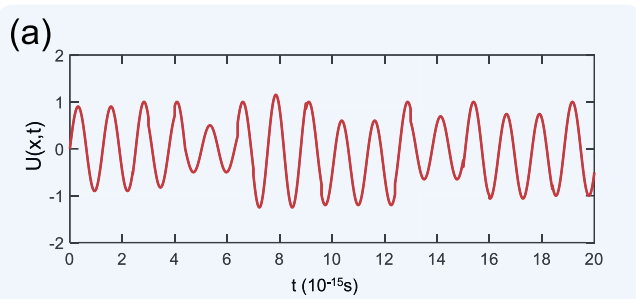


Fig. 33. The analytic signal: a typical realization (a) as a function of time for a fixed point in space; (b) as a function of space for one point in time.

correlation function of the optical oscillations at two points \mathbf{x}_1 and \mathbf{x}_2 with a relative time delay τ [282].

$$\Gamma_{12}(\tau) = \Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \langle U(\mathbf{x}_1, t_1)U^*(\mathbf{x}_2, t_2) \rangle = \langle U(\mathbf{x}_1, t)U^*(\mathbf{x}_2, t + \tau) \rangle \quad (111)$$

where $\tau = t_2 - t_1$, the sharp bracket indicates an ensemble average. For stationary and ergodic processes, the ensemble average is equivalent to time average, and thus, the correlation functions in Eqs. (109) and (110) are identical. Using the definition of MCF, the intensity can also be simply expressed as $I(\mathbf{x}) = \Gamma(\mathbf{x}, \mathbf{x}, 0)$, and the interference pattern generated by the superposition of two beams can be expressed as

$$I(\mathbf{x}) = \Gamma(\mathbf{x}_1, \mathbf{x}_1, 0) + \Gamma(\mathbf{x}_2, \mathbf{x}_2, 0) + 2\text{Re}\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) \quad (112)$$

It should be noted that the MCF is a measurable function that can be characterized through interferometry, as Eq. (112) shows. The coherence theory is to describe or characterize the physical characteristics of the optical field through its second-order or higher-order statistics that can be experimentally measured. The normalized version of MCF is also called *complex degree of coherence* (CDC), which can be defined as [283]

$$\gamma_{12}(\tau) = \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}} = \frac{\Gamma_{12}(\tau)}{\sqrt{I_1 I_2}} \quad (113)$$

MCF and CDC are two very important physical quantities that characterize the degree of correlation between optical oscillations at two different locations in the optical field. It is not difficult to understand from Eq. (110) that the physical meaning of CDC represents the contrast of the interferometric fringe pattern formed by the superposition of the two beams (for example, it can be explained by the Yang's double slit interference experiments of stationary optical fields [284]: \mathbf{x}_1 and \mathbf{x}_2 represent the position of two pinholes, and $\tau = t_2 - t_1$ represents the time delay between the two beams emitted from the two pinholes at a given location). The intensity at a given location is proportional to the beam intensity and the real part of the CDC. It can be proven by using the Schwarz's inequality that $0 \leq |\gamma_{12}(\tau)| \leq 1$. It is a properly normalized correlation coefficient, so that $\gamma_{11}(0) = \gamma_{22}(0) = 1$. This indicates that the field at a point in space must always be perfectly coherent with itself. All other values of $\gamma_{12}(\tau)$ are generally complex with an amplitude no more than one. This indicates that the fields at two different points, or at the same point after a time delay τ , are generally less than perfectly coherent with each other. The magnitude of the complete degree of spatial coherence (from zero to one) is a measure of the mutual coherence between the fields at the two test points and after a time delay τ . When the value of $|\gamma_{12}(\tau)|$ reaches the maximum, i.e., 1, the optical oscillations of the two points \mathbf{x}_1 and \mathbf{x}_2 with a time delay τ is completely coherent, and the fringe contrast reaches the maximum of 1. When the value of $|\gamma_{12}(\tau)|$ becomes 0, the optical oscillations of the two points \mathbf{x}_1 and \mathbf{x}_2 are completely incoherent, i.e., no interference fringes can be observed. In the case of $0 < |\gamma_{12}(\tau)| < 1$, the optical oscillations of the two points \mathbf{x}_1 and \mathbf{x}_2 are partially coherent.

We have already discussed in Subsection 4.1 that the coherence of the optical field is determined by the temporal characteristics (spectral distribution) and spatial characteristics (spatial extension) of the light source. Therefore, there exists two types of coherence quantities, namely, temporal coherence and spatial coherence, which are included in the concept of MCF. For example, when $\mathbf{x}_1 \neq \mathbf{x}_2$, the spatial coherence of the optical field is included in $\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)$, and when $\tau \neq 0$, the temporal coherence of the optical field is included in $\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)$. Although MCF completely describes the coherence of the field theoretically, the relationship between spatial coherence and temporal coherence are entangled together. In some cases, we need to analyze the effects of spatial coherence and temporal coherence separately. First, let us consider the temporal coherence. In the definition of MCF [Eq. (111)], when the two points \mathbf{x}_1 and \mathbf{x}_2 coincide with each other, MCF degenerates into *self coherence function* (SCF) at points \mathbf{x}_1 and \mathbf{x}_2 separately [282].

$$\Gamma_{11}(\tau) = \Gamma(\mathbf{x}_1, \mathbf{x}_1, \tau) = \langle U(\mathbf{x}_1, t + \tau)U^*(\mathbf{x}_1, t) \rangle \quad (114)$$

$$\Gamma_{22}(\tau) = \Gamma(\mathbf{x}_2, \mathbf{x}_2, \tau) = \langle U(\mathbf{x}_2, t + \tau)U^*(\mathbf{x}_2, t) \rangle \quad (115)$$

It can be seen that in the SCF, we only consider the influence of relative time delay on the correlation function. Therefore, the SCF is a direct reflection of the temporal coherence of the optical field. Obviously, when $\tau = 0$, $\Gamma_{11}(0)$ and $\Gamma_{22}(0)$ represent the intensities at the two points \mathbf{x}_1 and \mathbf{x}_2 and can be directly measured. It can be seen from Eqs. (114) and (115) that the SCF is an autocorrelation function that assumes optical oscillations as a temporal stochastic process, which can be experimentally measured with use of a Michelson interferometer. The *Wiener-Khinchin theorem* of the stochastic process tells us that the autocorrelation function of a stationary stochastic process is closely related to its *power spectral density* (PSD) [285]. For a process that is at least wide-sense stationary, the autocorrelation function $\Gamma_x(\tau)$ and PSD $S_x(\omega)$ form a Fourier transform pair

$$S_x(\omega) = \int \Gamma_x(\tau) \exp(j\omega\tau) d\tau \quad (116)$$

$$\Gamma_x(\tau) = \frac{1}{2\pi} \int S_x(\omega) \exp(-j\omega\tau) d\omega \quad (117)$$

where $\omega = 2\pi\nu = 2\pi c/\lambda$ is the angular frequency of the optical wave, where ν is the temporal frequency, c is the speed of light, and λ is the wavelength. PSD is commonly known as the *spectral distribution* of the light source, so the spectral distribution of the source directly determines the temporal coherence of the optical field. A light source of broad spectrum has a low degree of temporal coherence, whereas a light source with narrow linewidth has a high degree of temporal coherence, as illustrated in Fig. 34. The coherence time of a light source can be increased by using an optical filter to reduce its spectral width. The resultant gain of coherence comes at the expense of losing light energy.

In order to quantify the temporal coherence more straightforwardly, we can define the *coherence time* and the *coherence length* based on the half-power bandwidth (spectral width) of the source

$$\tau_c = \frac{1}{\Delta\nu} \quad (118)$$

$$l_c = c\tau_c = \frac{c}{\Delta\nu} = \frac{\lambda^2}{\Delta\lambda} \quad (119)$$

They are rough measures of the maximum allowable time delay and path difference between the beam and its replica for forming the interference. Though not as rigorous as the SCF, the coherence time and coherence length have been widely used due to their simple and intuitive physical implications (e.g., the coherence length of a 532 nm continuous laser (0.1 nm spectral width) is $532 \text{ nm} \times 532 \text{ nm} \div 0.1 \text{ nm} = 2.83 \text{ mm}$). Representative spectral bandwidths for different light sources, and their associated coherence times and coherence lengths are provided in Table 5. Conversely, we can also measure the PSD of the light source by measuring the SCF of the corresponding optical field (as illustrated in Fig. 35, the interference patterns generated by the Michelson interferometer with adjustable arm lengths are recorded, the interferometric signal is then converted into the spectrogram by Fourier transform), which is the basic principle of *Fourier transform spectrometer* (FTS).

Table 5
Spectral widths of typical light sources together with their coherence times and coherence lengths

| Source | $\Delta\nu_c(\text{Hz})$ | $\tau_c=1/\Delta\nu_c$ | $l_c=c\tau_c$ |
|--|--------------------------|------------------------|-------------------|
| Filtered sunlight ($\lambda_0=400 - 800\text{nm}$) | 3.74×10^{14} | 2.67fs | 800nm |
| Light-emitting diode ($\lambda_0=500\text{nm}, \Delta\lambda_0=50\text{nm}$) | 6×10^{13} | 16.7fs | 5 μm |
| Low-pressure sodium lamp | 5×10^{11} | 2ps | 600 μm |
| Multimode HeNe laser ($\lambda_0=633\text{nm}$) | 1.5×10^9 | 0.67ns | 20cm |
| Single-mode HeNe laser ($\lambda_0=633\text{nm}$) | 1×10^6 | 1 μs | 300m |

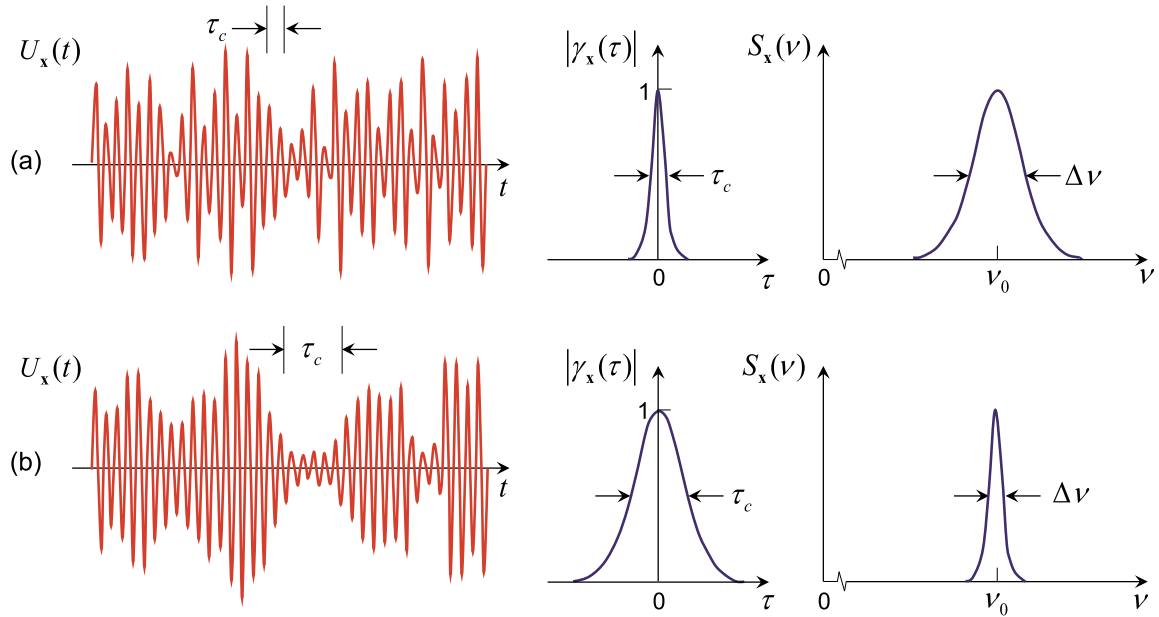


Fig. 34. Two random optical fields with different degrees of temporal coherence. (a) Low temporal coherence: optical oscillation, complex degree of temporal coherence, and power spectral density (from left to right); (b) high temporal coherence: optical oscillation, complex degree of temporal coherence, and power spectral density (from left to right).

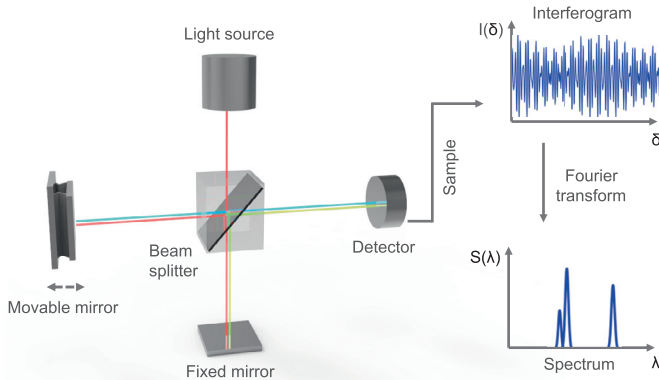


Fig. 35. The basic principle of Fourier transform spectrometer.

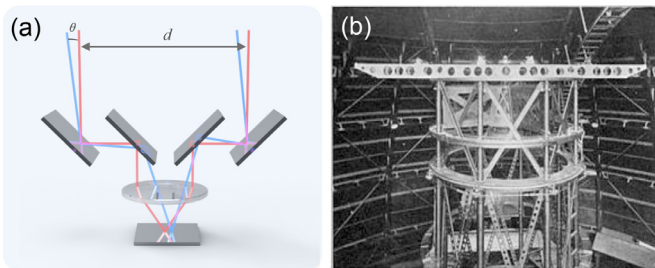


Fig. 36. Michelson stellar interferometer: measuring the spatial coherence of a quasi-monochromatic field by interferometry to infer the dimension of the light source. (a) Optical configuration; (b) photograph of a real system.

Next, we turn to the spatial coherence. Similar to the SCF, we can define the Fourier transform of the MCF as the *cross-spectral density* (CSD) [286]

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Gamma_{12}(\tau) \exp(j\omega\tau) d\tau \quad (120)$$

The above equation is also referred to as the generalized Wiener-Khinchin theorem. CSD is a central physical quantity in coherence theory, which is the ensemble-averaged correlation function between a typical monochromatic component of the field at two different locations \$\mathbf{x}_1\$ and \$\mathbf{x}_2\$. It can also be expressed as

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) \delta(\omega - \omega_0) = \langle U(\mathbf{x}_1, \omega_0) U^*(\mathbf{x}_2, \omega_0) \rangle \quad (121)$$

where \$U(\mathbf{x}, \omega)\$ is the Fourier transform of \$U(\mathbf{x}, t)\$, defined by Eq. (108), which represents a frequency-domain realization of the ensemble of the statistic properties of the optical field at the point \$\mathbf{x}\$, characterized by the angular frequency \$\omega\$. The sharp brackets indicate the ensemble average for different frequency components. Eq. (121) shows that the components of different frequencies in the partially coherent field are uncorrelated (unable to form stable interference), so CSD represents the correlation between the optical oscillations of the same frequency component at two different points in space. It quantitatively describes the spatial coherence of the optical field. A very important property of CSD is that it is a semi-definite Hermitian matrix. From the basic definition of CSD [Eq. (120)], the MCF can be expressed as an inverse Fourier transform of CSD

$$\Gamma_{12}(\tau) = \int_0^\infty W_\omega(\mathbf{x}_1, \mathbf{x}_2) \exp(-j\omega\tau) d\omega \quad (122)$$

Similar to Eq. (108), Eq. (122) indicates that the MCF of a partially coherent field results from the superposition of the CSD of different monochromatic components. Similar to MCF, we can also define the normalized CSD as *complex degree of spectral coherence* (at frequency \$\omega\$).

$$\begin{aligned} \mu_\omega(\mathbf{x}_1, \mathbf{x}_2) &= \frac{W_\omega(\mathbf{x}_1, \mathbf{x}_2)}{\sqrt{W_\omega(\mathbf{x}_1, \mathbf{x}_1) W_\omega(\mathbf{x}_2, \mathbf{x}_2)}} \\ &= \frac{W_\omega(\mathbf{x}_1, \mathbf{x}_2)}{\sqrt{S_\omega(\mathbf{x}_1) S_\omega(\mathbf{x}_2)}} \end{aligned} \quad (123)$$

where \$S_\omega(\mathbf{x})\$ is the PSD, which represents the energy of a particular monochromatic field component

$$S_\omega(\mathbf{x}) = W_\omega(\mathbf{x}, \mathbf{x}) \quad (124)$$

Because different monochromatic field components are uncorrelated and cannot interfere, the total intensity can be expressed as the sum of different monochromatic components

$$I(\mathbf{x}) = \int_{-\infty}^{\infty} S_{\omega}(\mathbf{x})d\omega \quad (125)$$

From the definition of complex degree of spectral coherence, it is plain to see that this function is a properly normalized correlation coefficient, which is always equal to unity if the two points are brought together, and is always less than or equal to unity as they are separated. If the magnitude of $|\mu_{\omega}(\mathbf{x}_1, \mathbf{x}_2)|$ is unity, it indicates that the monochromatic field component with angular frequency ω is perfectly coherent between the two points \mathbf{x}_1 and \mathbf{x}_2 . If the magnitude of this function is less than unity, it indicates less-than-perfect coherence. If the magnitude is zero, it indicates complete incoherence between the field amplitudes at the two points. For most partially coherent fields, CSD has significantly large values only for point separations which keep the two field points within the same coherence volume. This function depends only on the positions of the points and the single angular frequency that the field components at the two points share. Field components of different frequencies are always uncorrelated (and therefore incoherent), even at the same point.

6.1.2. Mutual intensity of quasi-monochromatic fields

In some earlier literature, in order to more conveniently analyze the spatial coherence, it is generally assumed that the optical field satisfies the so-called “quasi-monochromatic conditions”, i.e., the “narrow-band condition” (the spectral bandwidth $\Delta\omega$ of the light source is much smaller than its central frequency $\bar{\omega}$) plus “small path difference condition” (the optical path difference within the observation area is much smaller than the coherence length of the light source):

$$\Delta\omega \ll \bar{\omega} \quad (126)$$

$$|\mathbf{x}_1 - \mathbf{x}_2| \ll l_c \quad (127)$$

In such cases, Eq. (122) can be simplified as

$$\begin{aligned} \Gamma_{12}(\tau) &= \int_0^{\infty} W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) \exp(-j\omega\tau)d\omega \\ &= \exp(-j\bar{\omega}\tau) \int_0^{\infty} W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) \exp[-j(\omega - \bar{\omega})\tau]d\omega \\ &\stackrel{\exp(-j\Delta\omega\tau)=1}{\approx} \exp(-j\bar{\omega}\tau) \int_0^{\infty} W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) d\omega \\ &= \exp(-j\bar{\omega}\tau)\Gamma_{12}(0) \end{aligned} \quad (128)$$

The above equation shows that under quasi-monochromatic conditions, whether two points in the field can form interference fully depends on the spatial coherence. Therefore, the characteristics of the optical field at this time can be expressed by the zero-delay MCF $\Gamma_{12}(0)$, which is also called the *mutual intensity* (MI) J_{12} at points \mathbf{x}_1 and \mathbf{x}_2 .

$$\begin{aligned} J_{12} &= J(\mathbf{x}_1, \mathbf{x}_2) = \Gamma_{12}(\mathbf{x}_1, \mathbf{x}_2, 0) \\ &= \langle U(\mathbf{x}_1, t)U^*(\mathbf{x}_2, t) \rangle \end{aligned} \quad (129)$$

Therefore, a quasi-monochromatic optical field can be fully described by a single-frequency harmonic. It is not difficult to find that the MI J_{12} is also a Hermitian matrix. Similarly, the complex coherence of the quasi-monochromatic light is defined as the *complex coherence factor* (CCF), which is the CDC at zero time delay

$$\mu_{12} = \frac{\Gamma_{12}(0)}{[\Gamma_{11}(0)\Gamma_{22}(0)]^{1/2}} = \frac{J_{12}}{\sqrt{I_1 I_2}} = \gamma_{12}(0) \quad (130)$$

Taking Fourier transform of Eq. (128) with respect to the time variable yields

$$W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) = J_{12}\delta(\omega - \bar{\omega}) \quad (131)$$

The above equation shows that the CSD of the quasi-monochromatic field contains only one non-zero frequency component located at $\bar{\omega}$. This

means that the quasi-monochromatic field can be approximated to be of a single frequency component $\bar{\omega}$, or equivalently, its temporal coherence is close to that of the monochromatic (fully temporally coherent) field. Therefore, if we neglect the ω dependence of $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2)$, and interpret it as a quantity which represents the ensemble characteristics of a given monochromatic field component (ω is not a variable but a constant, i.e., $\bar{\omega}$), the MI representation J_{12} and CSD representation $W_{\bar{\omega}}(\mathbf{x}_1, \mathbf{x}_2)$ for a quasi-monochromatic field are equivalent in essence

$$W_{\bar{\omega}}(\mathbf{x}_1, \mathbf{x}_2) \equiv J(\mathbf{x}_1, \mathbf{x}_2) \equiv \Gamma_{12}(\mathbf{x}_1, \mathbf{x}_2, 0) \quad (132)$$

The temporal coherence is closely related to the spectral distribution of the light source. The spectral bandwidth of the light source is a direct reflection of the temporal coherence, e.g., the coherence length and the coherence time. Similarly, the spatial coherence is closely related to the size of the light source (for a quasi-monochromatic incoherent light source, the CCF of the corresponding far-field diffraction and the intensity of the light source constitute a Fourier transform pair; see discussions about the Van Cittert-Zernike theorem in Subsection 6.1.5 for more details). For an incoherent light source with evenly distributed intensity of arbitrarily-shaped area A_s , we can introduce a rough measure of spatial coherence, i.e., *coherent area* A_c at the distance z from the light source as

$$A_c = \frac{\bar{\lambda}^2 z^2}{A_s} = \frac{\bar{\lambda}^2}{\Omega_s} \quad (133)$$

Its physical implication is the maximum allowable area at a plane located at a distance of z away from the quasi-monochromatic incoherent source within which visible interference fringes can be formed. Conversely, we can also measure the spatial coherence of the quasi-monochromatic field through interferometry to infer the physical size of the light source (as shown in Fig. 35, the interference pattern of the light source is recorded by a Michelson interferometer with an adjustable arm. When the fringe contrast is reduced from the maximum to 0, the maximum arm distance is obtained, and thus the angular diameter of the light source can be calculated), which is the basic principle of the Michelson stellar interferometer.

6.1.3. Propagation of mutual coherence function

For a monochromatic coherent field, the optical field distribution can be fully described by the complex amplitude distribution, which is a function of spatial coordinate. When the complex amplitude is known at one plane, the field distribution at an arbitrary distance Δz can be determined based on the scalar diffraction theory discussed in Subsection 2.2.2. For a partially coherent field, the optical oscillations at one point in space change irregularly with time, and we need to focus on its statistical properties, i.e., investigate the correlation of optical oscillations at two different points in the time-space coordinate system. Thus, MCF is the fundamental physical quantity describing the properties of the statistical field. The MCF of the optical field may change during propagation. In this sense, the coherence of the optical field also propagates along with beam propagation. More specifically, for a partially coherent field propagating in free space, its temporal coherence does not change because it is only related to the spectral distribution of the source. However, the spatial coherence changes significantly with the propagation of the optical field. Therefore, this section focuses on the transmission characteristics of spatial coherence, i.e., the propagation of MI/CSD.

In Subsection 2.2.2, we know that the Huygens-Fresnel principle of the wave diffraction is based on the principle of linear superposition: since any complex optical wave field can be interpreted as a collection of point sources, it can always be decomposed into a linear combination of simple spherical wavelets. The linear property of the wave equation allows each spherical wave to propagate independently, and then their contributions on the diffraction plane are superimposed to obtain the entire diffraction field. Considering a monochromatic field with a complex amplitude distribution $U_0(\mathbf{x}, t)$, the complex amplitude of the

optical field after propagating a distance Δz can be expressed as

$$U_{\Delta z}(\mathbf{x}, t) = \int U_0(\mathbf{x}', t) h_{\Delta z}(\mathbf{x}', \mathbf{x}) d\mathbf{x}' = U_0(\mathbf{x}, t) \otimes h_{\Delta z}(\mathbf{x}) \quad (134)$$

where $h_{\Delta z}(\mathbf{x})$ is the impulse response function for free space propagation of coherent fields. Under the paraxial approximation, the oblique factor can be neglected, and the impulse response function becomes spherical wave represented by Eq. (16) (for quasi-monochromatic fields, λ should be replaced by $\bar{\lambda}$). Based on the scalar diffraction theory and the definition of MI, we can express the 4D MI function at the plane located at a distance Δz as

$$\begin{aligned} J_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2) &= \langle U_{\Delta z}(\mathbf{x}_1, t) U_{\Delta z}^*(\mathbf{x}_2, t) \rangle \\ &= \iint \langle U_0(\mathbf{x}_1, t) U_0^*(\mathbf{x}_2, t) \rangle h_{\Delta z}(\mathbf{x}'_1, \mathbf{x}_1) h_{\Delta z}^*(\mathbf{x}'_2, \mathbf{x}_2) d\mathbf{x}'_1 d\mathbf{x}'_2 \\ &= J_0(\mathbf{x}_1, \mathbf{x}_2) \otimes_{\mathbf{x}_1, \mathbf{x}_2} h_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2) \end{aligned} \quad (135)$$

where $h_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2)$ is the mutual point spread function for the free space propagation of coherent fields, which is defined as

$$h_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2) = h_{\Delta z}(\mathbf{x}_1) h_{\Delta z}^*(\mathbf{x}_2) \quad (136)$$

Therefore, when the 4D MI $J_0(\mathbf{x}_1, \mathbf{x}_2)$ at a given plane is known, the 4D MI at a different propagation distance Δz can also be obtained. Eq. (135) shows that the propagation of MI can also be regarded as a linear system in 4D space, and the response function of MI for each point pair in space $(\mathbf{x}_1, \mathbf{x}_2)$ is determined by 4D function $h_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2)$. Considering $J_0(\mathbf{x}_1, \mathbf{x}_2)$ as a weight factor, the MI $J_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2)$ at the plane Δz can be obtained by linear superimposition of all the response functions.

Similar to the angular spectrum diffraction theory discussed in Subsection 2.2.2, we can also analyze the propagation of partially coherent fields in the spatial frequency domain. By performing 4D Fourier transform of $(\mathbf{x}_1, \mathbf{x}_2)$ on both sides of Eq. (135), we can get

$$\hat{J}_{\Delta z}(\mathbf{u}_1, \mathbf{u}_2) = \hat{J}_0(\mathbf{u}_1, \mathbf{u}_2) H_{\Delta z}(\mathbf{u}_1, \mathbf{u}_2) \quad (137)$$

where $\hat{J}_{\Delta z}(\mathbf{u}_1, \mathbf{u}_2)$, $\hat{J}_0(\mathbf{u}_1, \mathbf{u}_2)$ and $H_{\Delta z}(\mathbf{u}_1, \mathbf{u}_2)$ are the corresponding 4D Fourier transforms of $J_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2)$, $J_0(\mathbf{x}_1, \mathbf{x}_2)$ and $h_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2)$, respectively. $(\mathbf{u}_1, \mathbf{u}_2)$ are the 4D spatial frequency coordinates corresponding to $(\mathbf{x}_1, \mathbf{x}_2)$ in the frequency domain. $H_{\Delta z}(\mathbf{u}_1, \mathbf{u}_2)$ is called the transfer function of MI for free space propagation, which is defined as

$$H_{\Delta z}(\mathbf{u}_1, \mathbf{u}_2) = H_{\Delta z}(\mathbf{u}_1) H_{\Delta z}^*(\mathbf{u}_2) \quad (138)$$

where $H_{\Delta z}(\mathbf{u})$ is the angular spectrum transfer function for free space propagation of coherent fields, defined by Eq. (30) [under the paraxial approximation, its form reduces to Eq. (32)].

6.1.4. Wave equations for the propagation of mutual coherence function

In the previous section, the propagation characteristics of spatial coherence was analyzed based on the scalar diffraction theory. However, it is of some general interest to examine the propagation problem at a more fundamental level. As we know that the light propagation essentially obeys the wave equation, we can derive a pair of scalar wave equations governing the propagation of the MCF

$$\nabla_1^2 \Gamma_{12}(\tau) = \frac{1}{c^2} \frac{\partial^2}{\partial \tau^2} \Gamma_{12}(\tau) \quad (139)$$

$$\nabla_2^2 \Gamma_{12}(\tau) = \frac{1}{c^2} \frac{\partial^2}{\partial \tau^2} \Gamma_{12}(\tau) \quad (140)$$

∇^2 is the Laplacian operator in 3D space (\mathbf{x}, z) . Based on the differential property of Fourier transform, a pair of Helmholtz equations for CSD propagation are obtained

$$\nabla_1^2 W_{12} + k^2 W_{12} = 0 \quad (141)$$

$$\nabla_2^2 W_{12} + k^2 W_{12} = 0 \quad (142)$$

It is easy to verify that MI propagates in accord with the same pair of Helmholtz equations

$$\nabla_1^2 J_{12} + \bar{k}^2 J_{12} = 0 \quad (143)$$

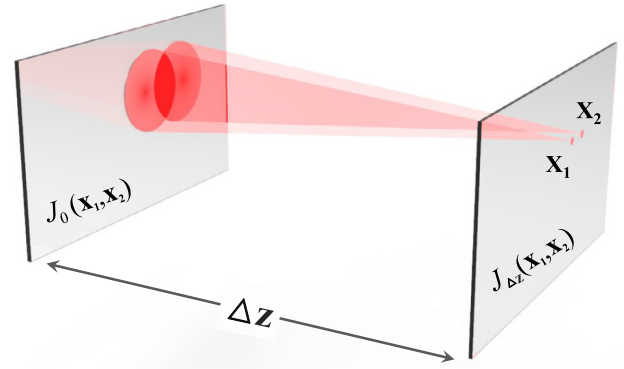


Fig. 37. Gain of coherence by propagation as a result of the spreading of light. Even if the light is completely incoherent in the source plane, the optical fluctuations at at points \mathbf{x}_1 and \mathbf{x}_2 shared a common region, and are therefore partially correlated.

$$\nabla^2 J_{12} + \bar{k}^2 J_{12} = 0 \quad (144)$$

where $\bar{k} = 2\pi/\bar{\lambda}$ is the average wave number. It can be seen that although the correlation functions of the partially coherent field are 6D function defined in two respective 3D spatial coordinates (\mathbf{x}_1, z_1) and (\mathbf{x}_2, z_2) , they also satisfy the Helmholtz equation [Eq. (2)] of the monochromatic coherent field in each 3D spatial coordinate. In addition, it should be noted that the Helmholtz equations for the propagation of CSD and MI share the same form. The only difference is that the general wave number k used in Eqs. (141) and (142) should be replaced by the average wave number \bar{k} in Eqs. (143) and (144). Because of the similarity between wave equations of CSD and MI, the properties of MI discussed in this section can be directly applied to CSD.

6.1.5. The Van Cittert-Zernike theorem

In general, light gains spatial coherence by the mere act of propagation. This is not surprising. Even if the source is completely incoherent, the radiation from each point spreads and overlaps with that from the neighboring points. The light reaching two points in the output plane comes from many points of the input plane, some of which are common (see Fig. 37). These common contributions create partial correlation between fluctuations at the output points. As one of the most important theorems of modern optics, the Van Cittert-Zernike theorem [283,287,288] describes the characteristics of the MI function produced by a quasi-monochromatic incoherent extended source. As the name implies, the theorem was first demonstrated in papers by Van Cittert [287] and Zernike [283,288]. In nearly all optical problems involving light that does not originate from a laser, the original optical source consists of an extended collection of independent radiators. Such a source can reasonably be modeled as incoherent in the sense

$$J_0(\mathbf{x}_1, \mathbf{x}_2) = I_0(\mathbf{x}_1) \delta(\mathbf{x}_1 - \mathbf{x}_2) \quad (145)$$

Note that if CSD is used to characterize the source distribution, Eq. (145) should be written as $W_0(\mathbf{x}_1, \mathbf{x}_2) = S(\mathbf{x}_1) \delta(\mathbf{x}_1 - \mathbf{x}_2)$. Starting from Eq. (135), and substituting Eq. (16) into Eq. (145), we can deduce

$$\begin{aligned} J_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2) &= \iint I_0(\mathbf{x}') h_{\Delta z}(\mathbf{x}', \mathbf{x}_1) h_{\Delta z}^*(\mathbf{x}', \mathbf{x}_2) d\mathbf{x}' \\ &= \frac{1}{\bar{\lambda}^2} \iint I_0(\mathbf{x}') \frac{\exp\left(jk\left(\sqrt{\Delta z^2 + |\mathbf{x}_1 - \mathbf{x}'|^2} - \sqrt{\Delta z^2 + |\mathbf{x}_2 - \mathbf{x}'|^2}\right)\right)}{\sqrt{\Delta z^2 + |\mathbf{x}_1 - \mathbf{x}'|^2} \sqrt{\Delta z^2 + |\mathbf{x}_2 - \mathbf{x}'|^2}} d\mathbf{x}' \end{aligned} \quad (146)$$

The phase factor in the integral can be interpreted as the interferometric superposition of two copies of the spherical wave at points \mathbf{x}_1 and \mathbf{x}_2 emitted from the light source at point \mathbf{x}' in the diffraction plane. The

contributions from the interference fields generated by all source points are weightedly added based on their respective intensities. Noted that the constant factor is omitted here for simplicity.

Under the paraxial approximation, the spherical wave of the impulse response function in the integral can be approximated by the Fresnel impulse response [Eq. (17)], and the following Van Cittert-Zernike theorem [283,287,288] can be obtained

$$J_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2) = J_{\Delta z}(\mathbf{x}_1 - \mathbf{x}_2) = \frac{1}{\lambda^2 \Delta z^2} \exp(j\psi) \iint I_0(\mathbf{x}') \exp\left(-\frac{2\pi}{\lambda \Delta z} \mathbf{x}' \cdot (\mathbf{x}_2 - \mathbf{x}_1)\right) d\mathbf{x}' \quad (147)$$

where the phase factor is

$$\exp(j\psi) = \exp\left[j \frac{\pi}{\lambda \Delta z} (|\mathbf{x}_1|^2 - |\mathbf{x}_2|^2)\right] \quad (148)$$

The phase factor can be interpreted as the interferometric superposition of two copies of the spherical wave at points \mathbf{x}_1 and \mathbf{x}_2 emitted from an on-axis point source. The Van Cittert-Zernike theorem, stated mathematically in Eq. (147), can be expressed in words as follows: aside from the factor $\exp(j\psi)$ and scaling constants, the MI in the diffraction plane is given by a 2D Fourier transform of the intensity distribution across the source. This relationship can be likened to the relationship between the field across a coherently illuminated aperture and the field observed in the Fraunhofer diffraction pattern of that aperture, although the physical quantities involved are entirely different.

We can also use the intensity in the diffraction plane $I_{\Delta z}(\mathbf{x}_1) = I_{\Delta z}(\mathbf{x}_2) = \frac{1}{\lambda^2 \Delta z^2}$ [i.e., $\mathbf{x}_1 = \mathbf{x}_2$ in Eq. (147)] to normalize $J_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2)$ and rewrite it in terms of CCF

$$\mu_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2) = \exp(j\psi) \frac{\iint I_0(\mathbf{x}') \exp\left(-\frac{2\pi}{\lambda \Delta z} \mathbf{x}' \cdot (\mathbf{x}_2 - \mathbf{x}_1)\right) d\mathbf{x}'}{\iint I_0(\mathbf{x}') d\mathbf{x}'} \quad (149)$$

Note that the phase factor $\exp(j\psi)$ in Eq. (149) does not affect the modulus of the CCF $|\mu_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2)|$, i.e., it does not affect the contrast of the interference fringes produced by two points \mathbf{x}_1 and \mathbf{x}_2 in Young's interference experiment, so $|\mu_{\Delta z}(\mathbf{x}_1, \mathbf{x}_2)|$ depends only on the difference of coordinates $\mathbf{x}_2 - \mathbf{x}_1$ in the diffraction plane. The mathematical statement of the Van Cittert-Zernike theorem is simply a precise statement of this relationship between the intensity distribution across the source and resulting fringe contrast for given locations of the pinholes. Just as a point source will create interference fringes of perfect visibility, each point on an incoherent source will create a separate fringe of high visibility. If the source size is too large, these elementary fringe patterns add with significantly different spatial phases, and the contrast of the overall fringe pattern is reduced. When the propagation distance Δz is large enough, the phase factor $\exp(j\psi)$ disappears, and the far field MI and the source intensity become a precise 2D Fourier transform pair.

When studying the properties of a partially coherent imaging system, it is often necessary to consider an illumination system with lenses (i.e., condenser), such as the most commonly used Köhler illumination configuration in microscopic imaging systems. In this case, the incoherent light source (condenser aperture) is placed in the front focal plane of the lens (condenser lens) and imaged at infinity. As a consequence, nonuniformities of the source brightness distribution are not imaged onto the object, and a highly uniform field of illumination is provided. It is not difficult to prove that in such an optical configuration, the relationship between the MIs in the front focal plane and the object plane (back focal plane) of the lens can be represented as

$$J_f(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\lambda^2 f^2} \iint J_0(\mathbf{x}'_1, \mathbf{x}'_2) \exp\left\{\frac{2\pi}{\lambda f} [\mathbf{x}'_2 \cdot (\mathbf{x}_2 - \mathbf{x}_1) - \mathbf{x}'_1 \cdot (\mathbf{x}_2 - \mathbf{x}_1)]\right\} d\mathbf{x}'_1 d\mathbf{x}'_2 \quad (150)$$

The above equation shows the MIs in the front and back focal planes of the thin positive lens form a 4D Fourier transform pair, i.e., $J_f(\mathbf{x}_1, \mathbf{x}_2) = \mathcal{J}_0(\mathbf{u}_1, \mathbf{u}_2) \Big|_{\mathbf{u}_{1,2} = \pm \frac{\mathbf{x}_{1,2}}{\lambda f}} = \mathcal{F}\{J_0(\mathbf{x}_1, \mathbf{x}_2)\} \Big|_{\mathbf{u}_{1,2} = \pm \frac{\mathbf{x}_{1,2}}{\lambda f}}$. This relationship is similar to the relationship between the coherent fields in

the front and back focal planes of a thin lens, although the physical quantities involved are entirely different. We further consider the source is quasi-monochromatic and incoherent by substituting Eq. (145) into Eq. (150)

$$J_f(\mathbf{x}_1, \mathbf{x}_2) = J_f(\mathbf{x}_1 - \mathbf{x}_2) = \frac{1}{\lambda^2 f^2} \iint I_0(\mathbf{x}') \exp\left(-\frac{2\pi}{\lambda f} \mathbf{x}' \cdot (\mathbf{x}_2 - \mathbf{x}_1)\right) d\mathbf{x}' \quad (151)$$

Aside from the scaling constants, the MI in the object plane is precisely the Fourier transform of the source intensity. Note that comparing with Eq. (147), the phase factor $\exp(j\psi)$ disappears. The MI in the object plane is a function only of the differences of coordinates in the object plane and can easily be found by 2D Fourier transforming the source intensity distribution.

6.1.6. Coherent mode decomposition

Although MCF and CSD can characterize the basic properties of partially coherent fields, including propagation and diffraction. However, they are both 4D functions, so the relevant analysis and calculation are quite complicated. *Coherent mode decomposition* [286,289–291] provides an effective tool to simplify this problem. It has shown that any partially coherent field can be represented as the sum over component fields that are each perfectly self-coherent, but mutually incoherent with each other. Thus, the CSD for any field can be represented in the form

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = \sum_n \lambda_n(\omega) \psi_n(\mathbf{x}_1, \omega) \psi_n^*(\mathbf{x}_2, \omega) \quad (152)$$

where $\psi_n(\mathbf{x}, \omega)$ is a complex amplitude called “coherent mode”, and is mutually incoherent. λ_n is a positive number representing the weight of the corresponding mode. A coherent mode $\psi_n(\mathbf{x}, \omega)$ can be interpreted as a monochromatic deterministic field of frequency ω with a well-defined amplitude and phase, and its propagation obeys the Helmholtz equation described by Eq. (2). When $\mathbf{x}_1 = \mathbf{x}_2$, Eq. (152) just represents that the PSD of the partially coherent field is the incoherent superposition of PSDs of different coherent modes

$$S_\omega(\mathbf{x}) = \sum_n \lambda_n(\omega) \psi_n(\mathbf{x}, \omega) \psi_n^*(\mathbf{x}, \omega) = \sum_n \lambda_n(\omega) S_{\text{om}}(\mathbf{x}) \quad (153)$$

so λ_n can be interpreted as the proportion of the PSD occupied by the corresponding coherent mode.

Coherent mode decomposition provides a simple and intuitive way to understand the underlying physical mechanism of partially coherent fields. Nevertheless, the coherent mode is derived from the CSD. When the CSD is unknown, it is very difficult to obtain the coherent mode of a partially coherent field. In order to apply the idea of coherent mode decomposition more conveniently, researchers are not only limited to standard the coherent modes extracted from the CSD, but directly use some pre-defined non-orthogonal complete basis functions as “suboptimum” coherent modes to represent the partially coherent field. For example, when studying the multimode output beam of a laser, Hermite-Gaussian modes [292] or Laguerre-Gaussian modes [293] are often adopted. Hermite-Gaussian modes form an orthogonal basis for the solutions to the paraxial wave equation in the Cartesian coordinate system, and the Laguerre-Gaussian modes form an orthogonal basis for the solutions to the paraxial wave equation in the cylindrical coordinate system. The proration of these two types of modes can be represented based on the Huygens-Fresnel scalar diffraction theory. Thus, coherent mode decomposition provides a more convenient way to characterize the propagation properties of some special partially coherent fields.

6.1.7. Various models for partially coherent fields

From Subsection 6.1.5, we know that under the paraxial approximation, the CSD in the diffraction plane generated from an incoherent extended light source follows the Van Cittert-Zernike theorem [Eq. (147)]. Specifically, the CSD in the diffraction plane can be generally represented as

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = \psi_\omega(\mathbf{x}_1) \psi_\omega^*(\mathbf{x}_2) g(\mathbf{x}_1 - \mathbf{x}_2) \quad (154)$$

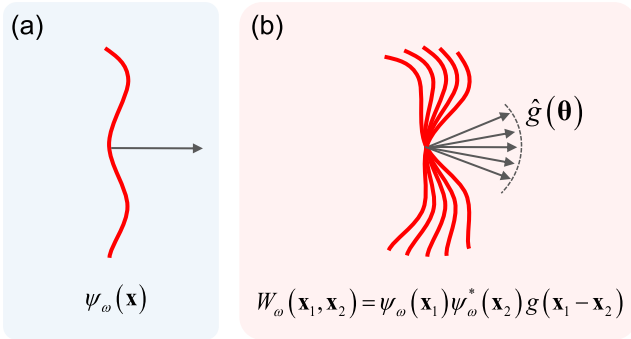


Fig. 38. Physical picture for understanding the physical meaning of the various models for partially coherent fields [296]. (a) A coherent field; (b) a partially coherent field described by the generalized Schell model, where $\hat{g}(\boldsymbol{\theta})$ is the Fourier transform of the source distribution (direction factor).

Eq. (154) shares certain similarities with the coherent mode decomposition expressed by Eq. (152). We call $\psi_\omega(\mathbf{x})$ in Eq. (154) *coherent component wave*. Compared with Eq. (147), its form can be represented as

$$\psi_\omega(\mathbf{x}) = -\frac{j}{\lambda\Delta z} \exp\left(j\frac{\pi}{\lambda\Delta z}|\mathbf{x}|^2\right) \quad (155)$$

It can be regarded as the spherical wave at the diffraction plane formed by an on-axis point source. Different from coherent mode decomposition that the mode weight can only be a positive constant, the “weight” $g(\mathbf{x}_1 - \mathbf{x}_2)$ in Eq. (154) is a function of coordinate difference $\mathbf{x}_1 - \mathbf{x}_2$, we call it *direction factor*. Compared with Eq. (147), the form of direction factor $g(\mathbf{x})$ is given by

$$g(\mathbf{x}) = \iint I_0(\mathbf{x}') \exp\left(-\frac{2\pi}{\lambda\Delta z}\mathbf{x}' \cdot \mathbf{x}\right) d\mathbf{x}' \quad (156)$$

In fact, it is just the Fourier transform of the intensity distribution of the light source, so the direction factor and the CCF are closely related. Generally speaking, the maximum value of $g(\mathbf{x})$ is located at $\mathbf{x} = 0$, and the decay rate with the two-point distance determines the spatial coherence of the field.

In Eq. (156), point \mathbf{x}' at the source plane represents the inclined plane wave component in the corresponding angular spectrum. Although the coherent component wave in Eq. (154) is fixed, when multiplied with $\psi_\omega(\mathbf{x})$, its direction is modulated thus it becomes another component wave of a different direction. According to this idea, we can reinterpret the Van Cittert-Zernike theorem of Eq. (147): different point radiators of the incoherent source produce spherical component waves with different tilted angles. The spherical component waves are weighted by the direction factor determined by the source intensity distribution $I_0(\mathbf{x}')$, and then incoherently superposed to create the CSD of the entire field, as illustrated in Fig. 38. When the CSD can be written in the form of Eq. (154), we consider it satisfying the *generalized Schell model* [223,294,295].

If the distance between the source and the diffraction plane is far enough to satisfy the Fraunhofer approximation, $\psi_\omega(\mathbf{x})$ can be simplified as a plane wave. In this case, the generalized Schell model reduces to the *Schell model* [294].

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{S_\omega(\mathbf{x}_1)S_\omega(\mathbf{x}_2)}g(\mathbf{x}_1 - \mathbf{x}_2) \quad (157)$$

Combined with the definition of Eq. (123), it can be found that in Schell model, the direction factor $g(\mathbf{x}_1 - \mathbf{x}_2)$ becomes the CCF $\mu_\omega(\mathbf{x}_1, \mathbf{x}_2)$, which only depends on the two-point coordinate difference. It is easy to verify $g(\mathbf{x}) = g^*(-\mathbf{x})$ and $g(\mathbf{0}) = 1$.

In the field of laser technology, the *Gaussian Schell-model* [286,291] is more commonly used to describe the partially coherent fields emitted from multimode lasers. GSM is a special case of the Schell model [Eq. (157)], in which $S_\omega(\mathbf{x})$ and $g(\mathbf{x})$ both take the forms of Gauss

function

$$S_\omega(\mathbf{x}) = S_{\omega 0} \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma_s^2}\right) \quad (158)$$

$$g_\omega(\mathbf{x}) = \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma_g^2}\right) \quad (159)$$

where $S_{\omega 0}$ is a constant representing the central (maximum) PSD value; σ_s and σ_g denote the spatial correlation length and the beam waist width of the GSM beam, respectively. When $\sigma_s \gg \sigma_g$, the source is spatially coherent, and when $\sigma_s \ll \sigma_g$, the source is almost spatially incoherent. The ratio between the two factors $q = \sigma_g/\sigma_s$ is related to the well-known M^2 (beam quality factor) [297]

$$M^2 = \sqrt{1 + \frac{2}{q^2}} \quad (160)$$

When the intensity of the partially coherent field is relatively uniform, and the spatial coherence of the source is relatively low [$|g(\mathbf{x})|$ decays rapidly with the increase of the two-point distance], then $S_\omega(\mathbf{x})$ is a slowly varying function compared with $g(\mathbf{x})$ so that the following approximation can be satisfied

$$\sqrt{S_\omega(\mathbf{x}_1)S_\omega(\mathbf{x}_2)} \approx S_\omega\left(\frac{\mathbf{x}_1 + \mathbf{x}_2}{2}\right) \quad (161)$$

Combining Eq. (161) with the Schell’s model [Eq. (157)], the *quasi-homogeneous model* is obtained. In the early literature, the sources satisfying the quasi-homogeneous model are more intuitively called “slowly varying homogeneous source”.

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = S_\omega\left(\frac{\mathbf{x}_1 + \mathbf{x}_2}{2}\right)g(\mathbf{x}_1 - \mathbf{x}_2) \quad (162)$$

Furthermore, at the end of Subsection 6.1.5, we also consider the case when the incoherent light source (condenser aperture diaphragm) is placed at the front focal plane of the lens and imaged at infinity. If the finite aperture effect of the lens is neglected, the intensity of the illumination field is perfectly homogeneous. Compared with Eq. (147), it can be found that $\psi_\omega(\mathbf{x})$ can be reduced to an plane wave with uniform intensity distribution. In this case, we consider the optical field satisfying the *spatially stationary model* (also called *statistically stationary model*)

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = S_{\omega 0}g(\mathbf{x}_1 - \mathbf{x}_2) \quad (163)$$

For a spatially stationary field, it can be interpreted as the incoherent superposition of plane component waves of different directions with weighting coefficients determined by the source intensity distribution. The highly homogeneous illumination field produced by the built-in Köhler illuminations of most current microscopes can be well described by the spatially stationary model. The coherence models discussed in this subsection is summarized in Table 6.

6.2. Phase space representations of partially coherent fields

In addition to correlation functions, another powerful new tool for understanding and characterizing partially coherent fields is called “*phase-space optics*”. As we know, Fourier transform is widely used in the research of deterministic signals. However, for non-stationary signals (e.g., partially coherent optical fields), the joint space and spatial-frequency representation, i.e., the phase-space representation should be used. The phase-space optics refers to a representation of optical signals in an artificial configuration space simultaneously providing information about spatial properties of the signal and its angular spectrum. In 1932, Wigner [298] introduced the *Wigner distribution function* (WDF) in mechanics that permitted a description of mechanical phenomena in phase space. Such a Wigner distribution was introduced in optics by Dolin [299] and Walther [300,301] in the 1960s, to relate partial coherence to radiometry. A few years later, the Wigner distribution was introduced in optics again by Bastiaans [302–304]. He illustrated how

Table 6
Coherence models of different partially coherent fields

| Type of models | Cross spectral density | Coherent Component |
|---|---|--|
| Generalized Schell model (Van Cittert-Zernike theorem) | $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) = \psi_{\omega}(\mathbf{x}_1)\psi_{\omega}^*(\mathbf{x}_2)g(\mathbf{x}_1 - \mathbf{x}_2)$ | Spherical wave $\psi_{\omega}(\mathbf{x}) = -\frac{j}{\lambda z} \exp(j\frac{\pi}{\lambda z} \mathbf{x} ^2)$ |
| Schell model (far-field Van Cittert-Zernike theorem) | $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{S_{\omega}(\mathbf{x}_1)S_{\omega}(\mathbf{x}_2)}g(\mathbf{x}_1 - \mathbf{x}_2)$ | Plane wave (nonuniform intensity) $\psi_{\omega} = \sqrt{S_{\omega}(\mathbf{x})}$ |
| Gaussian Schell-model | $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{S_{\omega}(\mathbf{x}_1)S_{\omega}(\mathbf{x}_2)}g(\mathbf{x}_1 - \mathbf{x}_2)$ | Gaussian Beam $S_{\omega}(\mathbf{x}) = S_{\omega 0} \exp(-\frac{ \mathbf{x} ^2}{2\sigma_x^2})$ |
| Quasi-homogeneous model | $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) = S_{\omega}(\frac{\mathbf{x}_1 + \mathbf{x}_2}{2})g(\mathbf{x}_1 - \mathbf{x}_2)$ | Plane wave (smooth intensity) $\psi_{\omega} = \sqrt{S_{\omega}(\mathbf{x})}$ |
| Spatially stationary model (Statistically stationary model) | $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2) = S_{\omega 0}g(\mathbf{x}_1 - \mathbf{x}_2)$ | Plane wave (uniform intensity) $\psi_{\omega} = \sqrt{S_{\omega 0}}$ |

space-frequency representations are well-suited for gaining physical insight and developing novel engineering applications for partially coherent imaging. On the other hand, the concept of *ambiguity function* (AF) was introduced by Woodward [305] in the signal processing of radar and sonar measurements in 1953. Twenty years later, Papoulis [306] reintroduced this concept into the research of optical imaging system. In fact, AF and WDF form a Fourier transform pair. Phase space, and in particular WDF and AF, can be recognized as one common platform for understanding and applying the physics of more traditional models for describing electromagnetic signals as they evolve and propagate through an optical system [307].

6.2.1. Wigner distribution function and ambiguity function

In the classical coherence theory, MCF in the space-time domain and the CSD in the space-frequency domain are the most commonly used tools to describe partially coherent fields. However, their inherent bilinear, stochastic, and wave-optical nature often lead to complicated mathematics and difficulties in comprehension. The WDF representation can effectively overcome these shortcomings and provide a particularly convenient tool for modeling and analyzing of partially coherent fields due to its simplicity and intuitiveness. For a partially coherent field $U(\mathbf{x}, t)$, its WDF is defined as the Fourier transform of the corresponding CSD in a differential space coordinate system

$$W_{\omega}(\mathbf{x}, \mathbf{u}) = \int W_{\omega}\left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2}\right) \exp(-j2\pi\mathbf{u}\mathbf{x}') d\mathbf{x}' \quad (164)$$

where \mathbf{u} is the coordinate in frequency domain corresponding to \mathbf{x} . By convention, we denote WDF with the same character W_{ω} , but it should be distinguished from CSD by its variables (\mathbf{x}, \mathbf{u}) . There are two points need to be noted. On the one hand, similar to the CSD $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2)$, the 4D WDF $W_{\omega}(\mathbf{x}, \mathbf{u})$ only describes the spatial coherence of one single frequency component ω in a polychromatic partially coherent field. On the other hand, the WDF $W_{\omega}(\mathbf{x}, \mathbf{u})$ is defined in phase space, i.e., it has joint coordinates of both the spatial domain and the spatial frequency domain (\mathbf{x}, \mathbf{u}) , which should be distinguished from the CSD $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2)$ where two 2D spatial coordinates are involved. The relationship between the differential space coordinate system $(\mathbf{x}, \mathbf{x}')$ used in WDF and the original space coordinate system $(\mathbf{x}_1, \mathbf{x}_2)$ is defined as

$$\begin{cases} \mathbf{x} = \frac{\mathbf{x}_1 + \mathbf{x}_2}{2} \\ \mathbf{x}' = \mathbf{x}_1 - \mathbf{x}_2 \end{cases} \quad \text{or} \quad \begin{cases} \mathbf{x}_1 = \mathbf{x} + \frac{\mathbf{x}'}{2} \\ \mathbf{x}_2 = \mathbf{x} - \frac{\mathbf{x}'}{2} \end{cases} \quad (165)$$

Note that this is a unitary transformation and the corresponding Jacobian is equal to unity. It can be seen that the WDF arises “midway” between the CSD $W_{\omega}(\mathbf{x}_1, \mathbf{x}_2)$. The physical meaning of WDF can be strictly understood as the probability density distribution of photon position and momentum. In Subsection 2.2.2, we know that the spatial frequencies of the scalar coherent field correspond to the plane wave components (angular spectrum) propagating in different directions, so we can (approximately) interpret \mathbf{u} of the WDF $W_{\omega}(\mathbf{x}, \mathbf{u})$ as the direction of light rays at point \mathbf{x} . Considering the 1D simplified case shown in Fig. 39, the spatial frequency \mathbf{u} of WDF and the ray direction (the angle θ between the ray and optical axis) has the following relationship

$$u = \frac{\cos \alpha}{\lambda} = \frac{\sin \theta}{\lambda} \stackrel{\text{paraxial}}{\approx} \frac{\theta}{\lambda} \quad (166)$$

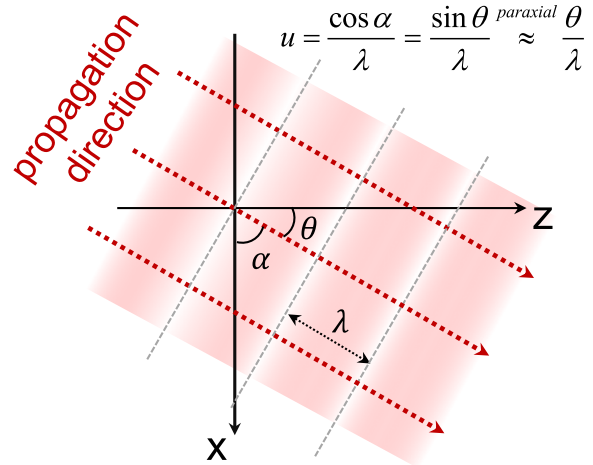


Fig. 39. The relation between the spatial frequency and the direction of propagation under the paraxial approximation.

In geometric optics, an optical ray through one point can be uniquely determined by its position and direction. Therefore, the WDF can be understood as a more rigorous ray model, defined as “generalized radiance” [300,301]. However, different from traditional radiance, generalized radiance can be negative. It not only considers that light ray travels in straight lines but also accurately describes the wave optical effects of light waves, such as interference and diffraction. More discussion about the physical meaning of WDF can be found in Subsection 6.2.6.

When ω is fixed, the WDF $W_{\omega}(\mathbf{x}, \mathbf{u})$ can only describe the spatial coherence of one single frequency component. For the sake of brevity, we will omit the subscript ω and only consider the case of (quasi-) monochromatic field (ignore the temporal coherence). Note that for a polychromatic field, the temporal coherence can be simply incorporated by the integral over all optical frequencies.

Next, we consider a perfectly coherent (both spatially and temporally) field, its CSD can be expressed as $W(\mathbf{x}_1, \mathbf{x}_2) = U(\mathbf{x}_1)U^*(\mathbf{x}_2)$, the corresponding WDF can be expressed as

$$W(\mathbf{x}, \mathbf{u}) = \int U\left(\mathbf{x} + \frac{\mathbf{x}'}{2}\right)U^*\left(\mathbf{x} - \frac{\mathbf{x}'}{2}\right) \exp(-j2\pi\mathbf{u}\mathbf{x}') d\mathbf{x}' \quad (167)$$

Similarly, we can define the AF

$$A(\mathbf{u}', \mathbf{x}') = \int W\left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2}\right) \exp(-j2\pi\mathbf{u}'\mathbf{x}) d\mathbf{x} \quad (168)$$

It is not difficult to prove that AF and WDF form a Fourier transform pair:

$$\begin{aligned} A(\mathbf{u}', \mathbf{x}') &= \iint W(\mathbf{x}, \mathbf{u}) \exp[-j2\pi(\mathbf{u}'\mathbf{x} - \mathbf{u}\mathbf{x}')] d\mathbf{x}d\mathbf{u} \\ &= \mathcal{F}\{W(\mathbf{x}, \mathbf{u})\}(\mathbf{u}', \mathbf{x}') \end{aligned} \quad (169)$$

Therefore, as two central physical quantities in phase space optics, the WDF and the AF representations are equivalent in essence. In the next subsection, we will focus on the properties of WDF.

6.2.2. Properties of Wigner distribution function

The popularity of WDF for characterizing partially coherent fields roots from its unique properties. In this subsection, we give a brief introduction to the most commonly used properties of WDF.

1) Realness

$W(\mathbf{x}, \mathbf{u})$ is always a real function

$$W(\mathbf{x}, \mathbf{u}) \in \mathbb{R}, \quad \forall \mathbf{x}, \forall \mathbf{u} \quad (170)$$

This property can be derived from the positive semidefiniteness and Hermitian symmetry of CSD. Note that, $W(\mathbf{x}, \mathbf{u})$ is not necessarily non-negative; this prohibits a direct interpretation of the WDF as an energy density function (or radiance function).

2) Spatial marginal property

$$\int W_{\omega}(\mathbf{x}, \mathbf{u}) d\mathbf{u} = W_{\omega}(\mathbf{x}, \mathbf{x}) = S_{\omega}(\mathbf{x}) \quad (171)$$

The equation shows that the integral of WDF over the frequency variable represents the PSD at optical frequency ω and position \mathbf{x} . When the signal is quasi-monochromatic, the spatial marginal of WDF is just the intensity. More strictly speaking, for polychromatic fields, the intensity is the total energy obtained by integrating the PSD over all wavelengths

$$\iint W_{\omega}(\mathbf{x}, \mathbf{u}) d\mathbf{u} d\omega = \int S_{\omega}(\mathbf{x}) d\omega = I(\mathbf{x}) \quad (172)$$

3) Spatial frequency marginal property

$$\int W_{\omega}(\mathbf{x}, \mathbf{u}) d\mathbf{x} = G_{\omega}(\mathbf{u}) \quad (173)$$

$G_{\omega}(\mathbf{u})$ is the Fourier transform of the spectral density, called directional power spectrum. For a monochromatic coherent field $U(\mathbf{x})$, $G(\mathbf{u})$ is the power spectrum in the spatial frequency domain, i.e., $|U(\mathbf{u})|^2$.

4) Convolution property

When $U(\mathbf{x}) = U_1(\mathbf{x})U_2(\mathbf{x})$, then

$$W(\mathbf{x}, \mathbf{u}) = W_1(\mathbf{x}, \mathbf{u}) \otimes_{\mathbf{u}} W_2(\mathbf{x}, \mathbf{u}) \quad (174)$$

where $\otimes_{\mathbf{u}}$ indicates convolution over \mathbf{u} , $W_1(\mathbf{x}, \mathbf{u})$ and $W_2(\mathbf{x}, \mathbf{u})$ are the WDFs of $U_1(\mathbf{x})$ and $U_2(\mathbf{x})$, respectively.

When $U(\mathbf{x}) = U_1(\mathbf{x}) \otimes_{\mathbf{x}} U_2(\mathbf{x})$, then

$$W(\mathbf{x}, \mathbf{u}) = W_1(\mathbf{x}, \mathbf{u}) \otimes_{\mathbf{x}} W_2(\mathbf{x}, \mathbf{u}) \quad (175)$$

$\otimes_{\mathbf{x}}$ indicates the convolution over \mathbf{x} .

5) Instantaneous frequency

For monochromatic coherent fields $U(\mathbf{x}) = a(\mathbf{x})e^{j\phi(\mathbf{x})}$, the instantaneous frequency of $U(\mathbf{x})$ is associated with the phase gradient $\nabla\phi(\mathbf{x})$ (the concept of ‘‘instantaneous frequency’’ for 1D signal [308] is adopted here to represent the phase gradient, see Subsection 6.2.2 for detailed discussions), which has the following relationship with the WDF

$$\frac{\int \mathbf{u} W(\mathbf{x}, \mathbf{u}) d\mathbf{u}}{\int W(\mathbf{x}, \mathbf{u}) d\mathbf{u}} = \frac{1}{2\pi} \nabla\phi(\mathbf{x}). \quad (176)$$

For the proof of Eq. (176), readers can refer to [307].

6.2.3. Optical transformation of Wigner distribution function

The concept of WDF is very close to the optical ray in geometric optics. Although it cannot be fully equivalent to energy density function in a strict sense, its propagation and optical transformation properties strictly follow the ray model, making it a bridge between geometric optics (radiometry) and wave optics. In this subsection, we outline the optical transformation properties of WDF.

1) Fresnel propagation

Under the paraxial approximation, the Fresnel propagation property of WDF can be derived from the Fresnel propagation formula of coherent fields [Eq. (18)] and partially coherent fields [Eq. (135)]

$$W_{\Delta z}(\mathbf{x}, \mathbf{u}) = W_0(\mathbf{x} - \lambda\Delta z\mathbf{u}, \mathbf{u}) \quad (177)$$

which is an \mathbf{x} -shear of the input WDF, and the amount of shear is linearly proportional to the propagation distance.

2) Chirp modulation (lens)

When the optical field passes through a lens or a quadratic phase mask, a quadratic phase factor $\exp\left(j\frac{\pi}{\lambda f}|\mathbf{x}|^2\right)$ is induced, where f is the focal length representing the curvature of the phase factor. The WDF of the output field can be expressed as

$$W_f(\mathbf{x}, \mathbf{u}) = W_0\left(\mathbf{x}, \mathbf{u} + \frac{\mathbf{x}}{\lambda f}\right) \quad (178)$$

which is a \mathbf{u} -shear of the input WDF, and the amount of shear is inversely proportional to f .

3) Fourier transform (Fraunhofer diffraction)

The WDF corresponding to the Fourier transform $\hat{U}(\mathbf{u})$ of the original signal $U(\mathbf{x})$ can be realized by interchanging the space and spatial frequency coordinates of the original WDF.

$$W_{\hat{U}}(\mathbf{x}, \mathbf{u}) = W_U(-\mathbf{u}, \mathbf{x}) \quad (179)$$

4) Fractional Fourier transform

Fractional Fourier transform is an extension of traditional Fourier transform to fractional order [309–312], which is generally defined as a function of rotation angle θ

$$f_{\theta}(\mathbf{x}) = \mathcal{F}_{\theta}\{f(\mathbf{x}_m)\} = \left[\frac{\exp\left(j\frac{1}{2}\theta\right)}{\sqrt{j\sin\theta}} \right] \int f(\mathbf{x}_m) \exp\left[j\pi \frac{(|\mathbf{x}_m|^2 + |\mathbf{x}|^2)\cos\theta - 2\mathbf{x}_m \cdot \mathbf{x}}{\sin\theta} \right] d\mathbf{x}_m \quad (180)$$

where θ is the rotation angle and $\theta \neq n\pi$. The fractional order of the transform is defined as $\theta/(\pi/2)$. In particular, when $\theta = 0$, $f(\mathbf{x}) = \mathcal{F}_0\{f(\mathbf{x})\}$ is the original signal; when $\theta = \pi$, $f(-\mathbf{x}) = \mathcal{F}_{\pi}\{f(\mathbf{x})\}$ is the mirror image of the original signal; when $\theta = \pi/2$, $\hat{f}(\mathbf{u}) = \mathcal{F}_{\pi/2}\{f(\mathbf{x})\}$ is the traditional (1st order) Fourier transform of the original signal; when $\theta = -\pi/2$, $f(\mathbf{x}) = \mathcal{F}_{-\pi/2}\{\hat{f}(\mathbf{u})\}$ is the traditional inverse (-1st order) Fourier transform of the original signal. In optics, the fractional Fourier transform can be realized by single lens or double lens setups [respectively referred to as type I (RQR setup) and type II (QRQ setup)] [310,311], as shown in Fig. 40. By changing the focal length or spacing of the lens [for type I setup $R = \tan(\theta/2)$; $Q = \sin\theta$, for type II setup

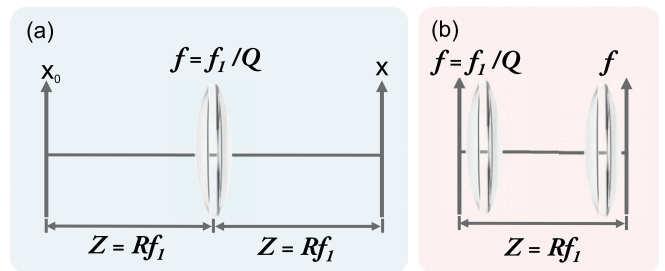


Fig. 40. Fractional Fourier transform. (a) Setup (Type I) for performing a fractional Fourier transform. Parameters R and Q determine the degree P and the angle $\phi = P\pi/2$. The signals are 2D; the lens is spherical; (b) setup (Type II) for performing a fractional Fourier transform.

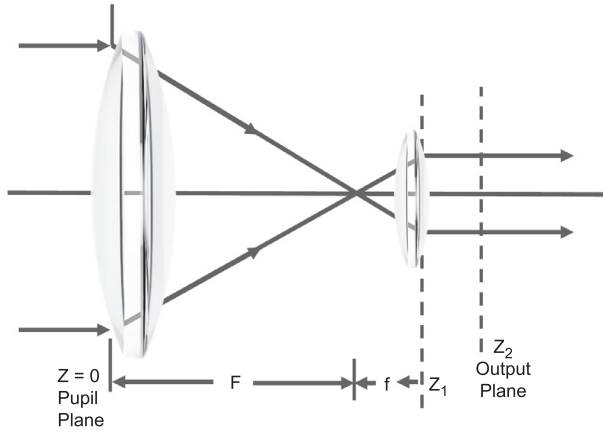


Fig. 41. Beam amplification (compression). Output plane is located at $z_2 = z_1 + (1 + f/F)f$. The output-plane wave amplitude is $U_{z_2}(\mathbf{x}) = (-1/M)U_0(-\mathbf{x}/M)$, where $M = f/F$.

$Q = \tan(\theta/2)$; $R = \sin \theta$], fractional Fourier transform pairs of different orders can be obtained at the input plane and the output plane.

It can be proven that the WDF corresponding to the fractional Fourier transform $\hat{U}_\theta(\mathbf{u})$ of the original signal $U(\mathbf{x})$ can be expressed as phase-space rotation of the original WD

$$W_{\hat{U}_\theta}(\mathbf{x}, \mathbf{u}) = W_U(\mathbf{x} \cos \theta - \mathbf{u} \sin \theta, \mathbf{u} \cos \theta + \mathbf{x} \sin \theta) \quad (181)$$

Similarly, we can take a projection of $W_{\hat{U}_\theta}(\mathbf{x}, \mathbf{u})$ over \mathbf{u} , the obtained intensity $\int W_{\hat{U}_\theta}(\mathbf{x}, \mathbf{u}) d\mathbf{u}$ is actually the Radon transform of the original WDF W_U at angle θ (see Subsection 6.2.5 for details) [313]. $\int W_{\hat{U}_\theta}(\mathbf{x}, \mathbf{u}) d\mathbf{u}$ is also known as the generalized marginal or generalized projection of WDF, which is proven to be nonnegative. If we change the fractional order (rotation angle θ) of the Fourier transform and collect the corresponding intensity distributions, we can reconstruct WDF based on the principle similar to *computerized tomography* (CT). Such a coherence measurement technique, which reconstructs WDF by phase-space rotation and projection, is called phase-space tomography [314–317], (relevant contents will be discussed in Subsection 6.2.6).

5) Beam amplifier (compressor)

If the optical signal passes through a $4f$ optical system with two lens of different focal lengths (Fig. 41), the size of the beam will be magnified (or compressed) M times (M is the focal length ratio between two lenses), then the WDF of the output field can be represented by the scaling of the original WDF

$$W_M(\mathbf{x}, \mathbf{u}) = W_0\left(M\mathbf{x}, \frac{\mathbf{u}}{M}\right) \quad (182)$$

6) First order optical system

The first-order optical system is one of the most common optical systems [318]. Similar to the matrix transformations in geometric optics, the geometric optical characteristics of the axisymmetric paraxial optical system can be described by a 2×2 ABCD-matrix [319]. Similarly, in wave optics, the transformation of a first-order optical system on the input field can also be described by Collins diffraction integral transformation equation based on ABCD-matrix [320]. Similar to the ABCD-matrix in matrix optics, the WDF of the original signal U after passing through a first-order optical system can be expressed as

$$W_{U_{1st}}(\mathbf{x}, \mathbf{u}) = W_U(A\mathbf{x} + B\mathbf{u}, C\mathbf{x} + D\mathbf{u}) \quad (183)$$

In phase space, the transformation between input and output coordinates is

$$\begin{bmatrix} \mathbf{x}' \\ \mathbf{u}' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} \quad (184)$$

where the ABCD-matrix is symplectic and the determinant equals unity. It can be seen that the five fundamental optical transformations described above can be regarded as special cases of the first-order optical system, and their corresponding ABCD-matrices are

$$\begin{bmatrix} 1 & -\lambda z \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ (\lambda f)^{-1} & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix} \quad (185)$$

The typical optical transformations in phase space are summarized in Fig. 42.

6.2.4. Wigner distribution representations of typical optical signals

We shall illustrate the concept of WDF by some typical examples from Fourier optics. For monochromatic deterministic fields, *i.e.*, it can be completely represented by a given complex amplitude function $U(\mathbf{x}) = a(\mathbf{x})e^{j\phi(\mathbf{x})}$. The WDFs of five typical coherent signals are shown in Fig. 43.

1) Point source

A monochromatic point source located at \mathbf{x}_0 can be expressed by the impulse signal $U(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_0)$. Its WDF takes the form

$$W(\mathbf{x}, \mathbf{u}) = \delta(\mathbf{x} - \mathbf{x}_0) \quad (186)$$

This is a 2D section of the 4D phase space in \mathbf{x} -plane. For the simplified 1D case [see Fig. 41(a)], it corresponds to a line perpendicular to the x -axis and at a distance \mathbf{x}_0 from the \mathbf{u} -axis in phase space.

2) Plane wave

A plane wave with a spatial frequency \mathbf{u}_0 can be described by $U(\mathbf{x}) = \exp(j2\pi\mathbf{u}_0\mathbf{x})$. Its WDF takes the form

$$W(\mathbf{x}, \mathbf{u}) = \delta(\mathbf{u} - \mathbf{u}_0) \quad (187)$$

This is a 2D section of the 4D phase space in \mathbf{u} -plane. For the 1D case [see Fig. 43(b)], it corresponds to a line perpendicular to the \mathbf{u} -axis and at a distance \mathbf{u}_0 from the \mathbf{x} -axis in phase space.

3) Spherical wave

A spherical wave under the paraxial approximation can be described by a quadratic-phase signal $U(\mathbf{x}) = \exp(j2\pi a\mathbf{x}^2)$. Its WDF takes the form

$$W(\mathbf{x}, \mathbf{u}) = \delta(\mathbf{u} - a\mathbf{x}) \quad (188)$$

For the 1D case, it corresponds to a straight line across the origin of phase space [see Fig. 43(c)].

4) Slow varying wave

For a smooth phase signal in the spatial domain, $U(\mathbf{x}) = \exp[j\phi(\mathbf{x})]$ where $\phi(\mathbf{x})$ is a smooth function of \mathbf{x} , its WDF takes the form [321]

$$W(\mathbf{x}, \mathbf{u}) \approx \delta\left(\mathbf{u} - \frac{1}{2\pi} \nabla \phi\right) \quad (189)$$

For more general slowly varying field $U(\mathbf{x}) = a(\mathbf{x})\exp[j\phi(\mathbf{x})]$ (the amplitude $a(\mathbf{x})$ is a nonuniform but smooth function of \mathbf{x}). Its WDF takes the form [208]

$$W(\mathbf{x}, \mathbf{u}) \approx I(\mathbf{x})\delta\left(\mathbf{u} - \frac{1}{2\pi} \nabla \phi\right) \quad (190)$$

More details about Eq. (190) will be introduced in Subsection 7.2.4. For the 1D case, it corresponds to a curve in phase space [see Fig. 43(d)]. The proof of Eq. (190) can be found in [208], and the required “slow varying” condition is strictly defined as

$$\phi\left(\mathbf{x} + \frac{\mathbf{x}'}{2}\right) - \phi\left(\mathbf{x} - \frac{\mathbf{x}'}{2}\right) \approx \mathbf{x}' \cdot \nabla \phi(\mathbf{x}) \quad (191)$$

$$a\left(\mathbf{x} + \frac{\mathbf{x}'}{2}\right)a\left(\mathbf{x} - \frac{\mathbf{x}'}{2}\right) \approx a^2(\mathbf{x}) = I(\mathbf{x}) \quad (192)$$

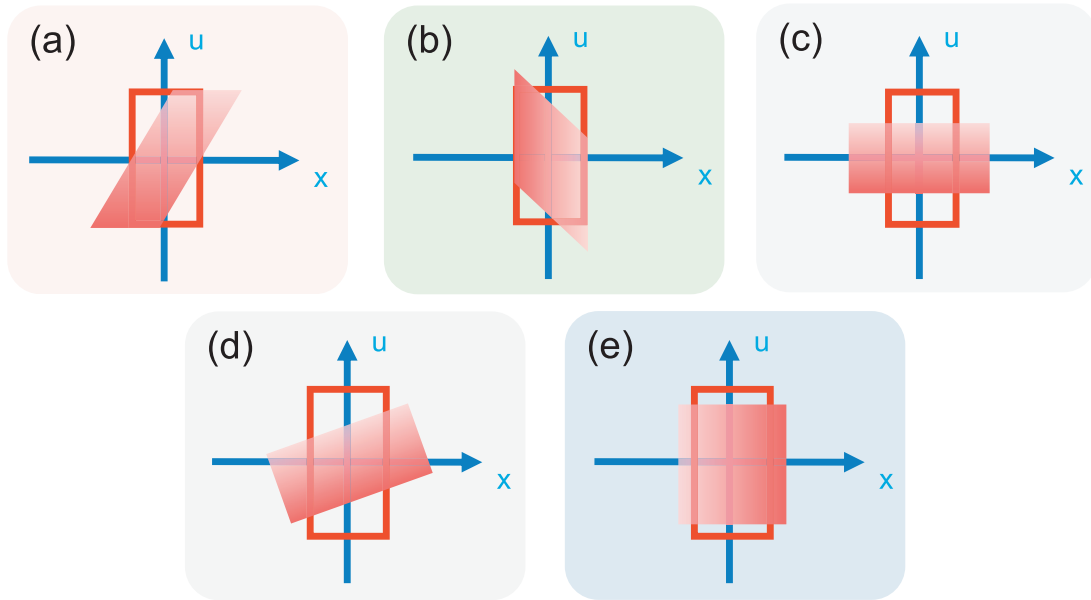


Fig. 42. Typical optical transformations in phase space. (a) Fresnel propagation; (b) Chirp modulation (lens); (c) Fourier transform; (d) fractional Fourier transform; (e) magnifier.

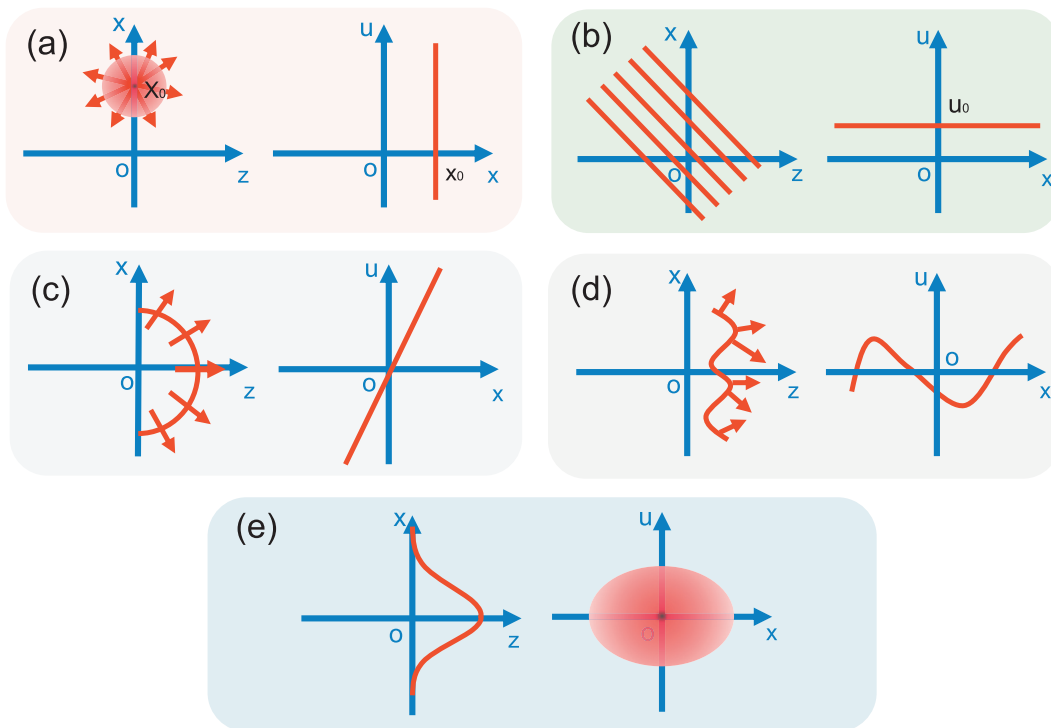


Fig. 43. WDFs of typical coherent signals. (a) Point source; (b) plane wave; (c) spherical wave; (d) slow-varying wave; (e) Gaussian beam.

5) Gaussian beam

The WDF of the Gaussian signal $U(x) = \exp\left\{-\frac{\pi}{\sigma^2}(x - x_0)^2\right\}$ is

$$W(x, u) = \exp\left\{-\left(\frac{\pi}{\sigma^2}(x - x_0)^2 + \frac{\sigma^2}{2\pi}u^2\right)\right\} \quad (193)$$

which is still a Gaussian signal in both x and u . For the 1D case, it corresponds to a 2D Gaussian function in phase space [see Fig. 43(e)].

For partially coherent fields, it cannot be described by the complex amplitude. Its statistical characteristics need to be represented by CSD

or WDF. Here, we introduce the WDFs of three typical partially coherent fields.

1) Spatially incoherent field

For spatially incoherent fields, the CSD can be expressed as [Eq. (145)]

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = S_\omega(\mathbf{x}_1)\delta(\mathbf{x}_1 - \mathbf{x}_2) \quad (194)$$

For quasi-monochromatic spatially incoherent fields

$$W(\mathbf{x}_1, \mathbf{x}_2) = I(\mathbf{x}_1)\delta(\mathbf{x}_1 - \mathbf{x}_2) \quad (195)$$

where the PSD $S_\omega(\mathbf{x})$ or intensity $I(\mathbf{x})$ is a nonnegative function. In the central differential space coordinate defined by Eq. (165), the CSD becomes

$$W_\omega\left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2}\right) = S_\omega(\mathbf{x})\delta(\mathbf{x}') \quad (196)$$

Its corresponding WDF is

$$W_\omega(\mathbf{x}, \mathbf{u}) = S_\omega(\mathbf{x}) \quad (197)$$

which is a function only of the space variable \mathbf{x} and that it does not depend on \mathbf{u} . This suggests that the power density of spatially incoherent field is independent with the spatial frequency (ray direction).

2) Spatially stationary field

The CSD of quasi-monochromatic fields satisfying the spatially stationary model can be expressed as [Eq. (163)]

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = S_0 g(\mathbf{x}_1 - \mathbf{x}_2) \quad (198)$$

The direction factor $g(\mathbf{x}_1 - \mathbf{x}_2)$ is actually the spectral CCF of the optical field. The intensity of the illumination is perfectly homogeneous. In the central differential coordinate system, the CSD becomes

$$W_\omega\left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2}\right) = S_0 g(\mathbf{x}') \quad (199)$$

Its corresponding WDF is

$$W_\omega(\mathbf{x}, \mathbf{u}) = S_0 \hat{g}(\mathbf{u}) \quad (200)$$

where $\hat{g}(\mathbf{u})$ is the Fourier transform of $g(\mathbf{x}')$. It can be seen that the WDF of spatially stationary fields is a function only of the spatial frequency variable \mathbf{u} . It has a form that is similar to the WDF of incoherent light [Eq. (197)] rotated by 90 degrees in phase space. According to the Fourier transform property of WDF [Eq. (179)], the spatially stationary field and the incoherent field essentially form a Fourier transform pair (far-field diffraction). The duality between incoherent light and spatially stationary light is, in fact, the Van Cittert-Zernike theorem [Eq. (151)]: the far-field diffraction of an incoherent light source creates a spatially stationary field.

3) Quasi-homogeneous field

The CSD of quasi-monochromatic fields satisfying the quasi-homogeneous model can be expressed as [Eq. (162)]

$$W_\omega(\mathbf{x}_1, \mathbf{x}_2) = S_\omega\left(\frac{\mathbf{x}_1 + \mathbf{x}_2}{2}\right)g(\mathbf{x}_1 - \mathbf{x}_2) \quad (201)$$

In the central differential coordinate system, the CSD becomes

$$W_\omega\left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2}\right) = S_\omega(\mathbf{x})g(\mathbf{x}') \quad (202)$$

Such quasi-homogeneous field can be locally considered as spatially stationary, having, however, a slowly varying intensity, i.e., S_ω is a relatively slow-varying signal compared with g . The WDF of quasi-homogeneous field takes the form

$$W_\omega(\mathbf{x}, \mathbf{u}) = S_\omega(\mathbf{x})\hat{g}(\mathbf{u}) \quad (203)$$

Both incoherent fields and spatially stationary fields are special cases of quasi-homogeneous fields: for spatial incoherent fields $\hat{g}(\mathbf{u}) = 1$ and for spatially stationary fields $I(\mathbf{x}) = S_0$.

6.2.5. Transport equation of Wigner distribution function

In Subsections 2.2.1 and Subsection 6.1.3, we learned that the propagation of coherent complex fields and MI/CSD both satisfy the Helmholtz equation [Eqs. (2) and (141)]. Under the paraxial approximation, Helmholtz equation can be further reduced to the paraxial wave equation [Eq. (4)]. As a new representation of coherent fields and partially coherent fields, WDF should also follow certain wave equation describing its propagation characteristics. Starting from the Helmholtz

equation and combining the properties of WDF under Liouville approximation (geometric optics approximation), we can deduce that the propagation of WDF follows the following *Liouville transport equation* [322]

$$\frac{\sqrt{k^2 - 4\pi^2|\mathbf{u}|^2}}{k} \frac{\partial W(\mathbf{x}, \mathbf{u})}{\partial z} = -\lambda \mathbf{u} \nabla_{\mathbf{x}} \cdot W(\mathbf{x}, \mathbf{u}) \quad (204)$$

where $\nabla_{\mathbf{x}} = \partial/\partial\mathbf{x} = (\partial_x, \partial_y)$. This equation can be solved analytically, and the solution takes the form

$$W_z(\mathbf{x}, \mathbf{u}) = W_0\left(\mathbf{x} - \frac{2\pi\mathbf{u}}{\sqrt{k^2 + 4\pi^2|\mathbf{x}|^2}}z, \mathbf{u}\right) \quad (205)$$

Eq. (205) describes the general propagation law of WDF in free space. Based on WDF, we can construct a 3D vector field $\mathbf{j}_r = [j_x, j_z]^T$, which is known as the *geometrical vector flux* [303,323]

$$j_x(\mathbf{x}) = \lambda \int \mathbf{u} W(\mathbf{x}, \mathbf{u}) d\mathbf{u} \quad (206)$$

$$j_z(\mathbf{x}) = \frac{1}{k} \int \sqrt{k^2 - 4\pi^2|\mathbf{u}|^2} W(\mathbf{x}, \mathbf{u}) d\mathbf{u} \quad (207)$$

It can be seen that the transverse flux j_x and the longitudinal flux j_z correspond to the integrals of the RHS and LHS of Liouville transport equation over the frequency variable \mathbf{u} , respectively. Geometrical vector flux is a radiometric quantity, which describes the light energy propagation in 3D space. In fact, the Liouville transport equation is an expression of the energy conservation law, which implies the geometrical vector flux has zero divergence.

Under the paraxial approximation $\sqrt{k^2 - 4\pi^2|\mathbf{u}|^2} \approx k$, the Liouville transport equation can be simplified as

$$\frac{\partial W(\mathbf{x}, \mathbf{u})}{\partial z} + \lambda \mathbf{u} \nabla_{\mathbf{x}} \cdot W(\mathbf{x}, \mathbf{u}) = 0 \quad (208)$$

This transport equation can again be solved explicitly, and the solution reads as

$$W_z(\mathbf{x}, \mathbf{u}) = W_0(\mathbf{x} - \lambda z\mathbf{u}, \mathbf{u}) \quad (209)$$

which is just the Fresnel diffraction property of WDF [Eq. (179)].

6.2.6. Coherence measurement

For monochromatic coherent fields, the WDF is defined based on the 2D complex amplitude [Eq. (167)], so the 4D phase-space representation is highly redundant. When the 2D intensity and phase distributions are known, the WDF can be directly calculated by Eq. (167). Therefore, for monochromatic coherent fields, the reconstruction of WDF essentially boils down to a phase recovery problem, which has been discussed in detail in previous sections.

It is more complicated when the field is not strictly coherent. Generally, the phase space WDF constitutes a rigorous and non-redundant representation for partially coherent fields. The 2D amplitude and phase are insufficient to determine a partially coherent field unambiguously. The complete characterization of the 4D coherence function (so-called *coherence measurement* or *coherence retrieval*) has always been an active research area. To fully characterize 4D partially coherent fields, one idea is to measure or reconstruct the two-point correlation function, such as MCF, CSD, or MI for quasi-monochromatic fields. After the two-point correlation function is obtained, the WDF can be calculated by definition [Eq. (164)]. For example, MCF can be directly measured through interferometry based on the contrast and displacement (phase) of the fringe pattern in Yang's double-slit interference experiment, as illustrated in Fig. 44(a). However, if we want to measure the 4D MCF of a 2D optical field, we need to traverse through the entire 4D space (2D by 2D), which is extremely time-consuming and inconvenient for practical implementation. Fortunately, in practical measurements, the principle of shearing interferometry can be used to create two overlapping copies of the test beam with a transversal, axial, or rotational shear

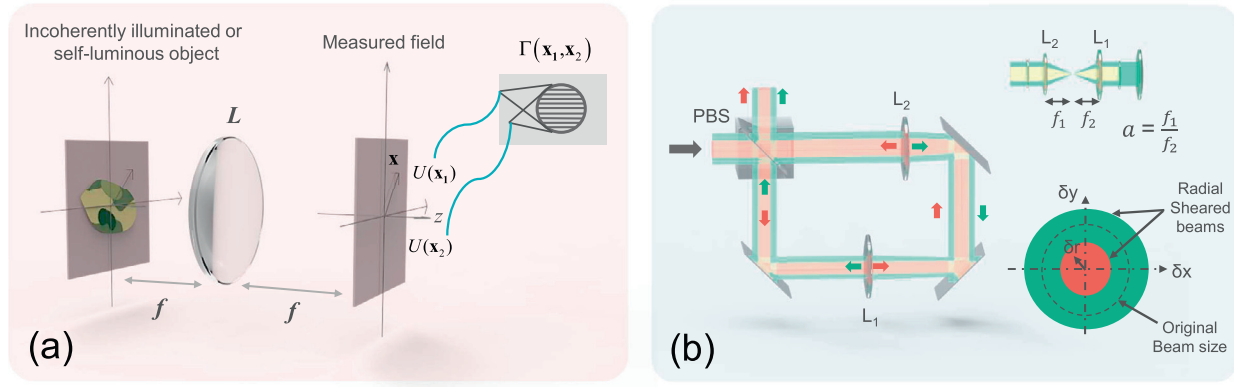


Fig. 44. Coherent measurements via interferometry. (a) Raster scan measurements based on Yang's double slit interference experiment; (b) parallel measurements based on a Sagnic radial shearing interferometer (PBS: Polarizing Beam Splitter).

based on Mach-Zehnder interferometer [324], Sagnic radial shearing interferometer [325,326], or rotational shearing interferometer [327] to achieve high-efficiency, parallel measurements of a large collection of point pairs, as illustrated in Fig. 44(b).

Another category of coherence measurement approaches is based on phase-space measurements, i.e., measuring or recovering the 4D WDF. The most well-known method is called *phase space tomography* (PST) [314,315]. By introducing asymmetric optical elements (e.g., cylindrical lens) into the field to be measured, and then collecting a large number of intensity distributions of the optical field at various propagation distances, the rotational projections of WDF at different angles can be obtained. Finally, the complete 4D WDF can be reconstructed in a way similar to the conventional CT.

Before introducing PST, let us first review the basic principle of the conventional CT [328]. For simplicity, we first consider a 2D function $f(x, y)$, and its projection can be represented as the integration along a straight line $l(s, \theta)$

$$R_f(s, \theta) = \iint f(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy = \int f(s \cos \theta + t \sin \theta, s \sin \theta - t \cos \theta) dt \quad (210)$$

The line equation $l(s, \theta)$ can be expressed as

$$x \cos \theta + y \sin \theta = s \quad (211)$$

or equivalently,

$$\begin{cases} x = s \cos \theta - t \sin \theta \\ y = s \sin \theta + t \cos \theta \end{cases} \quad (212)$$

where s represents the distance from the origin to the line, and θ represents the positive angle from the line to y -axis (or the positive angle from the projection plane to the x -axis). When θ and s are fixed, Eq. (211) represents a straight line with the angle θ between y -axis and the distance s from the origin. The projection of $f(x, y)$ onto the line can be obtained by Eq. (210). When θ is fixed and s is a variable, Eq. (211) represents a set of parallel lines with the angle θ . In this case, the 1D projection of the 2D function $f(x, y)$ along the line can be obtained by Eq. (210). If θ and s are both variables, the function $f(x, y)$ will be mapped to another 2D space (s, θ) to obtain projections in different directions, as illustrated in Fig. 45(c). Note that $R_f(s, \theta)$ is not defined in the polar coordinate system but on the surface of a semi-cylinder. Spreading the semi-cylinder surface onto a plane, we obtain an image as shown in Fig. 45(d). This image is also called a "Sinogram", because a point in (x, y) plane will be mapped into a sinusoid, as illustrated in Figs. 45(a) and 45(b). The mapping relation expressed in Eq. (210) forms a transform from the Cartesian coordinate (x, y) system into (s, θ) coordinate system, which is so-called *Radon transform* [329]. The inverse of Eq. (210) can

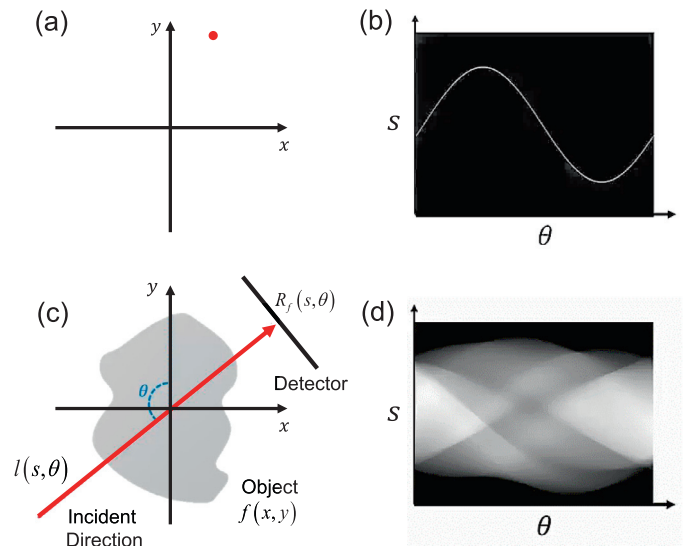


Fig. 45. Illustration of Radon transform. (a-b) One point in 2D space and the corresponding Radon transform; (c-d) the projection of a 2D object along a certain direction and the Radon transform of the 2D object.

be written as

$$f(x, y) = \iint \frac{\partial R_f(s, \theta)}{\partial s} \frac{1}{2\pi^2(x \cos \theta + y \sin \theta - s)} ds d\theta \quad (213)$$

Eq. (213) shows that after acquiring the projection $R_f(s, \theta)$ at each angle, the tomogram of the object can be reconstructed. Such a tomography method is called *inverse Radon transform*.

The mathematical mechanism of CT can also be understood by reinterpreting the problem in the spatial frequency domain. As the central theorem in classical CT, the *Fourier slice theorem* connects the projection data to the original function in Fourier space, which allows efficient CT reconstructions with FFT. Based on the definition of Fourier transform and variable substitution, it can be proven that the 1D Fourier transform $\hat{R}_f(\gamma, \theta)$ (γ corresponding to the spatial frequency s coordinate) of the 1D projection at angle θ is distributed along a straight line at the same angle in the Fourier space of the 2D object $f(x, y)$, as illustrated in Fig. 46.

$$\hat{R}_f(\gamma, \theta) = \hat{f}(u, v) \Big|_{v=\gamma \sin \theta}^{u=\gamma \cos \theta} = \hat{f}(\gamma \cos \theta, \gamma \sin \theta) = \hat{f}(\gamma, \theta) \quad (214)$$

where (γ, θ) is the polar coordinates in the frequency domain

$$\begin{cases} \gamma = \sqrt{u^2 + v^2} \\ \theta = \arctan \frac{v}{u} \end{cases} \quad (215)$$

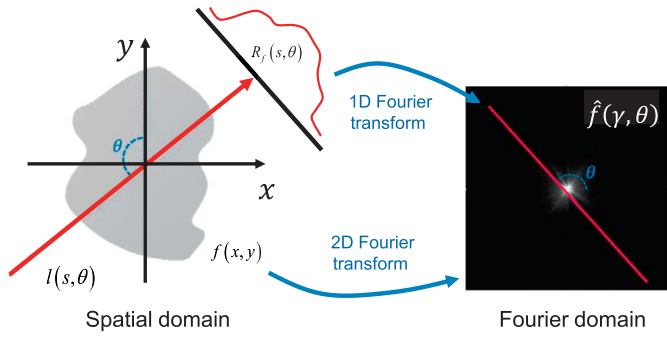


Fig. 46. Radon transform and the spatial and Fourier domain correspondence.

This theorem allows to access the whole 2D Fourier space of the object by collecting a set projection data $R_f(s, \theta)$ at different angles. Once $\hat{f}(\gamma, \theta)$ is completed, the object function can be reconstructed by a simple 2D inverse Fourier transform.

Next, we turn our attention to PST [314,315]. Based on the CT principle discussed earlier, PST recovers the WDF by successive projection measurements of the WDF at different rotation angles. According to the properties of WDF introduced in Subsection 6.2.2, for quasi-monochromatic fields, intensity is just a projection of WDF along x . In order to rotate the WDF, the simplest way is to take the Fourier transform (by far-field diffraction or introducing a lens) so that the WDF exchanges the space and spatial frequency variables, thus rotates $\pi/2$ in phase space. The intensity then corresponds to the projection of the WDF along u , so the projection at the angle $\pi/2$ is obtained. As represented by Eq. (150), the CSDs in the focal and back focal planes of a thin lens form a 4D Fourier transform pair

$$W(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\lambda^2 f^2} \iint W_{in}(\mathbf{x}'_1, \mathbf{x}'_2) \exp \left\{ \frac{2\pi}{\lambda f} [\mathbf{x}'_2 \cdot (\mathbf{x}_2 - \mathbf{x}_1) - \mathbf{x}'_1 \cdot (\mathbf{x}_2 - \mathbf{x}_1)] \right\} d\mathbf{x}'_1 d\mathbf{x}'_2 \quad (216)$$

which can be denoted as

$$W(\mathbf{x}_1, \mathbf{x}_2) = \hat{W}_{in}(\mathbf{u}_1, \mathbf{u}_2) \Big|_{\mathbf{u}_{1,2} = \pm \frac{\mathbf{x}_{1,2}}{\lambda f}} \quad (217)$$

Let $s = \lambda f$, according to the Fourier transform property, the WDFs in the input and output planes have the following relationship

$$W(\mathbf{x}, \mathbf{u}) = W_{in}(-s\mathbf{u}, s^{-1}\mathbf{x}) \quad (218)$$

Therefore, in the normalized coordinates $(\mathbf{x}_n, \mathbf{u}_n)$, where $\mathbf{x}_n = s^{-1/2}\mathbf{x}$, $\mathbf{u}_n = s^{1/2}\mathbf{u}$, the WDF rotates $\pi/2$. The intensities in the front and back focal planes of the lens are

$$I(\mathbf{x}) = \int W(\mathbf{x}, \mathbf{u}) d\mathbf{u} = \int W_{in}(-s\mathbf{u}, s^{-1}\mathbf{x}) d\mathbf{u} \quad (219)$$

Therefore, by introducing a lens, we can obtain the projection of the WDF at a different angle. In order to reconstruct the WDF, we need to collect more projections at various angles. Based on the fractional Fourier transform property of WDF, for 1D optical signals or axisymmetric optical fields, the corresponding fractional Fourier transform expressed in Eq. (180) directly corresponds to a θ -rotation of the original WDF $W_{in}(x, u)$ in phase space

$$W_{\mathcal{F}_\theta}(x, u) = W_{in}(x \cos \theta - u \sin \theta, u \cos \theta + x \sin \theta) \quad (220)$$

The corresponding intensity signal is just the Radon transform of the original WDF W_U at angle θ . Since WDF $W_{in}(x, u)$ and AF $A_{in}(u', x')$ form a Fourier transform pair [Eq. (169)], the Fourier slice theorem in phase space can be written as [330,331]

$$\begin{aligned} \hat{R}_{W_{in}}(\gamma, \theta) &= \mathcal{F} \{ W_{in}(x, u) \} \Big|_{u' = \gamma \cos \theta}^{u' = \gamma \cos \theta} \Big|_{x' = \gamma \sin \theta}^{x' = \gamma \sin \theta} \\ &= A_{in}(\gamma \cos \theta, \gamma \sin \theta) = A_{in}(\gamma, \theta) \end{aligned} \quad (221)$$

Fig. 47 illustrates the correspondence between WDF and AF. Therefore, for 1D signals or 2D axisymmetric fields, the reconstruction of WDF

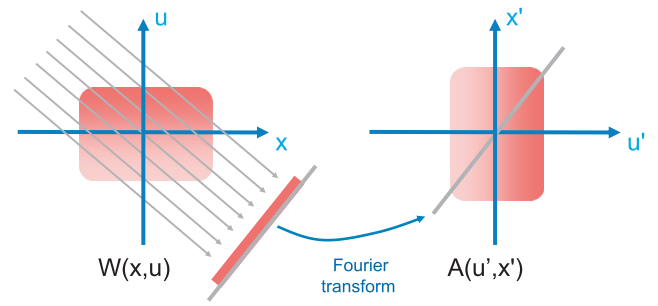


Fig. 47. Correspondence between Wigner distribution function and Ambiguity function.

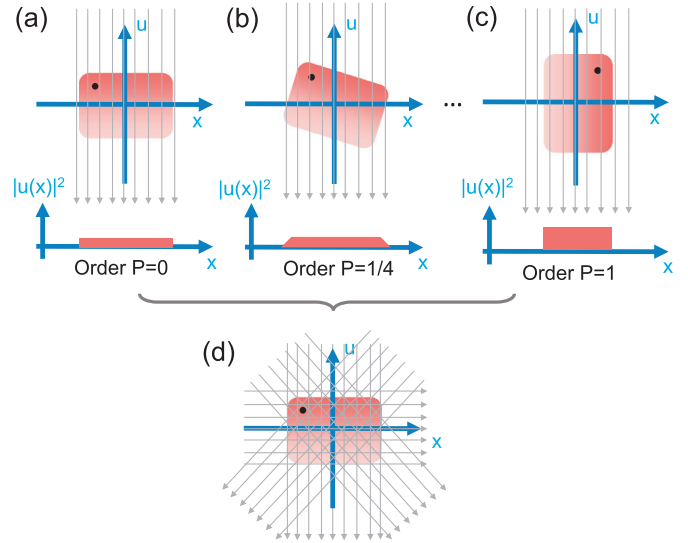


Fig. 48. Basic principle of phase space tomography. (a) Vertical projection (0-order fractional Fourier transform); (b) quarter rotation projection (1/4-order fractional Fourier transform); (c) 90-degree rotation projection (1-order fractional Fourier transform); (d) superposition of all projections from different angles.

by PST is completely consistent with the basic principle of traditional CT, as illustrated in Fig. 48.

In fact, not just limited to the fractional Fourier transform, we can also use Fresnel diffraction to shear the WDF [332,333]

$$W_{\Delta z}(x, u) = W_{in}(x - \lambda \Delta z u, u) \quad (222)$$

The corresponding intensity is $I(x) = \int W_{in}(x - \lambda \Delta z u, u) du$. Compared with Eq. (220), it can be found that the projection of WDF at angle θ is obtained “equivalently”, and the relation between θ and the propagation distance Δz is

$$\tan \theta = -\lambda \Delta z u \quad (223)$$

As shown in Fig. 49, although the transformations of fractional Fourier transform and Fresnel diffraction in phase space are different, they are equivalent for obtaining WDF projections. When Eq. (223) is satisfied, the projections obtained by fractional Fourier transform and Fresnel diffraction are both along the same line (P_1 and P_2 in Fig. 49). And the Fourier transform of the WDF projection obtained by Fresnel diffraction also corresponds to a 1D slice of the 2D object spectrum $A_{in}(u', -\lambda z u')$.

The case we discussed above is limited to 1D signal or 2D axisymmetric fields. However, it should be noted that for a general 2D optical field, its WDF is a 4D function. If we need to reconstruct the 4D WDF based on the CT principle, we need to “rotate” the 4D WDF over two independent phase space planes ($x - y$ and $u_x - u_y$). However, the fractional Fourier transform expressed by Eq. (180) takes the same order on

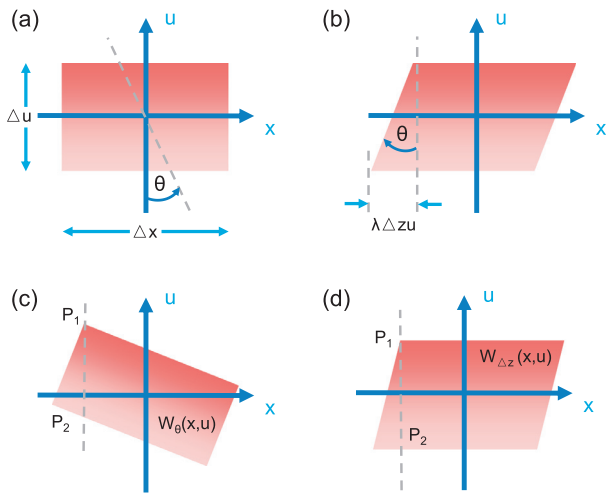


Fig. 49. Two different transformations of WDF for the implementation of phase space tomography. (a) WDF of the original signal; (b) WDF after Fresnel diffraction; (c) WDF after fractional Fourier transform; (d) the correspondence between Fresnel diffraction and fractional Fourier transform.

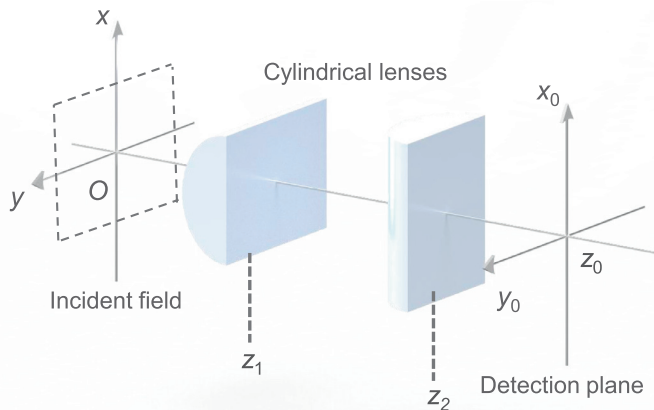


Fig. 50. Experimental arrangement of PST for recovering the 4D WDF of the 2D field at the plane of incidence O. A pair of cylindrical lenses oriented perpendicularly are used to introduce astigmatism to the measurement.

both axes $\mathbf{x} = (x, y)$, i.e., the rotation angles θ for both axes are identical. In fact, the fractional Fourier transform represented by the Eq. (180) is not the general form for 2D functions [310]. The transform kernel of fractional Fourier transform is a separable function in the Cartesian coordinate system, so its orders (rotation angles) for two axes x, y can assume different values. In optics, to realize asymmetric fractional Fourier transform for 2D optical fields, astigmatic optics should be introduced to break this symmetry in x and y dimensions, as illustrated in Fig. 50. By changing the focal lengths or the distances between the lenses, we can adjust θ_x, θ_y independently, so as to obtain the 2D rotational projections of the WDF. The 4D WDF of the input plane can be reconstructed by inverse Radon transform after the projection data of various angles are collected.

In addition to these “indirect” measurement methods based on interferometry and PST, we can also employ the property of the WDF as a “local spectrum (spectrogram)” or generalized radiance to probe the WDF directly. Two typical approaches are often used.

1) **Pinhole scanning:** the measured field is spatially localized by an aperture (usually a pinhole) scanning across \mathbf{x} . The corresponding 2D local spectrum can be recorded by far-field diffraction or based on the Fourier transform property of a thin lens. At a given location \mathbf{x}_0 , the captured local spectrum corresponds approximately to a spa-

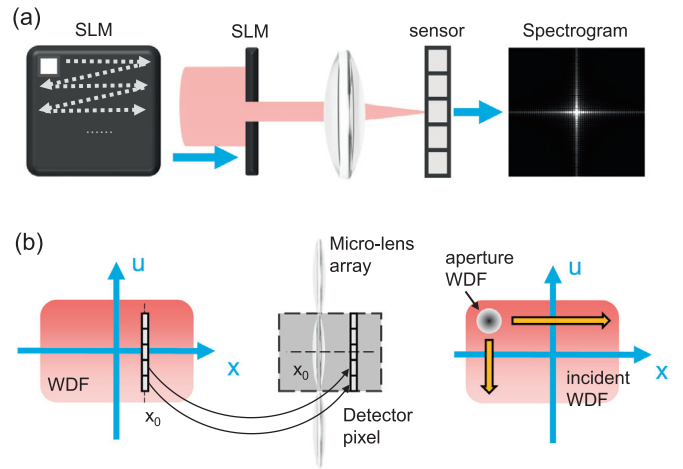


Fig. 51. Direct phase-space measurement methods. (a) Phase-space measurement based on pinhole scanning; (b) phase-space measurement based on a microlens array.

tial sampling of the WDF at \mathbf{x}_0 [334–336]. When the pinhole moves across the whole 2D space, the 4D WDF of the input field can be obtained, as illustrated in Fig. 51(a).

2) **Array measurements:** the experimental setup is quite similar to the Shack-Hartmann wavefront sensor [40–42] and light-field camera [337], as illustrated in Fig. 51(b). This method can be regarded as a “parallelized” version of the former pinhole scanning method, which allows single-shot acquisition (the intensity distribution behind each microlens corresponds to the local spectrum at different positions) [338,339]. However, the spatial resolution of the measurement is compromised, which only depends on the pitch of the microlens array. The basic principle of these two methods are quite similar: they both employ the approximate equivalence between WDF and the energy density function (radiance or light field in computer graphics) to realize the direct measurement of WDF.

Finally, it should be noted that the equivalence between WDF and radiance or light field can only be established under certain conditions (slowly varying coherent field or geometric optics approximation), so the local spectrum directly measured does not strictly correspond to the WDF. Strictly speaking, based on the convolution property of WDF, the spectrogram measured through a pinhole or a lens aperture corresponds to a smoothed version of the WDF [334,335].

$$W_s(\mathbf{x}, \mathbf{u}) = W_{in}(\mathbf{x}, \mathbf{u}) \otimes_{\mathbf{x}, \mathbf{u}} W_T(\mathbf{x}, \mathbf{u}) \quad (224)$$

where $\otimes_{\mathbf{x}, \mathbf{u}}$ represents the 4D convolution over both \mathbf{x} and \mathbf{u} , $W_{in}(\mathbf{x}, \mathbf{u})$ represents the WDF of input optical field and $W_T(\mathbf{x}, \mathbf{u})$ represents the WDF of aperture function. Although WDF may be negative, the spectrogram after convolution is always nonnegative and directly measurable. Obviously, if we want to make the measured spectrogram approach the true WDF, $W_T(\mathbf{x}, \mathbf{u})$ should tend to $\delta(\mathbf{x}, \mathbf{u})$. However, limited by Heisenberg’s uncertainty principle, it is physically impossible to shrink the support of WDF in both spatial and spatial-frequency domain. In order to alleviate this problem, a compromise between the blurring effects in the spatial and the spatial frequency domains should be made by appropriately designing the aperture function [335,336,340]. An alternative approach is to compensate for the aperture-induced blurring effect based on the WDF deconvolution method [341].

6.3. Image formation model under partially coherent illumination

In Section 4, we discussed the imaging formation for coherent imaging systems, where the source at the condenser aperture plane is as-

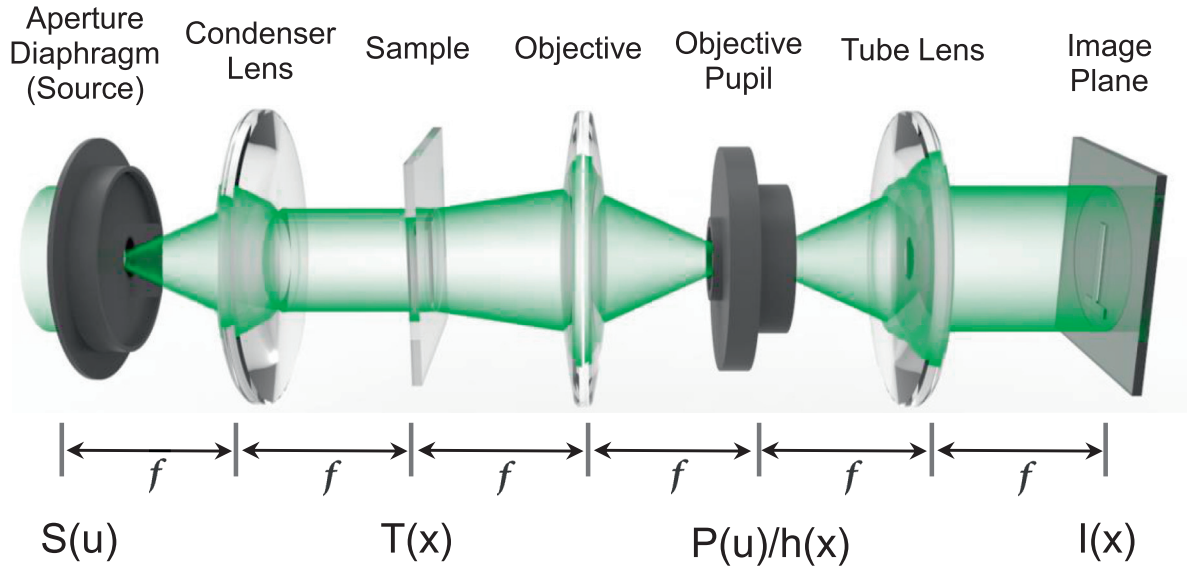


Fig. 52. Schematic diagram of a microscopic imaging system with partially incoherent illuminations. The condenser diaphragm can be regarded as an incoherent extended source with a finite size.

sumed to be an ideal on-axis point radiator with strict monochromaticity. For a practical optical microscopic imaging system, this condition is difficult to meet because the light source always has a certain bandwidth, and the condenser aperture always has a finite size. In other words, the illumination produced by the light source is neither perfect in temporal coherence nor spatial coherence. So in this section, we examine the effects of partial coherence on microscopic imaging. First, we ignore the limited aperture effect and discuss the influences of temporal coherence and spatial coherence on image formation. Then we take into account the aperture effect and derive the image formation model for a practical microscopic imaging system under partially coherent illuminations.

6.3.1. Ideal imaging model under partially incoherent illumination

In this section, we extend the coherent imaging condition discussed in Subsection 4.2, considering that the condenser aperture becomes incoherent extended source with a finite size [Eq. (145)]. Fig. 52 illustrates the corresponding optical configuration, where the light source (condenser aperture diaphragm) is placed at the front focal plane of the condenser lens, producing a highly homogeneous illumination field in the sample plane. Based on the Van Cittert-Zernike theorem, the CSD of the illumination just before the sample is given by the Fourier transform of the source intensity distribution $S(\mathbf{u})$ [Eq. (151)]

$$W_S(\mathbf{x}_1, \mathbf{x}_2) = W_S(\mathbf{x}_1 - \mathbf{x}_2) = \iint S(\mathbf{u}) e^{j2\pi\mathbf{u}\cdot(\mathbf{x}_1 - \mathbf{x}_2)} d\mathbf{u} \quad (225)$$

It only depends on the coordinate difference between the two points in the plane. In the following of this subsection, the formal description of the imaging process is limited to quasi-monochromatic light and unit ($1.0\times$) magnification. An extension to polychromatic illuminations and arbitrary magnification is straightforward.

When the illumination light transmits a thin object of complex amplitude $T(\mathbf{x}) = a(\mathbf{x}) \exp [j\phi(\mathbf{x})]$, the CSD just behind the object is given by

$$W_O(\mathbf{x}_1, \mathbf{x}_2) = W_S(\mathbf{x}_1, \mathbf{x}_2) T(\mathbf{x}_1) T^*(\mathbf{x}_2) \quad (226)$$

The CSD now depends on the complex amplitudes of the sample and the CSD of the illumination. The modulation of the optical field by the imaging system can be described in the spatial frequency domain based on the transfer function theory, as discussed in Subsection 5.1.3. After the transmitted optical field passing through the objective lens, the CSD in

the objective pupil is multiplied twice by the coherent transfer function $H(\mathbf{u})$ [Eq. (63)], and then followed by an inverse transform to finally obtain the mutual intensity in the image plane

$$W_I(\mathbf{x}_1, \mathbf{x}_2) = \iint \hat{W}_O(\mathbf{u}_1, \mathbf{u}_2) H(\mathbf{u}_1) H^*(\mathbf{u}_2) e^{j2\pi(\mathbf{u}_1\mathbf{x}_1 + \mathbf{u}_2\mathbf{x}_2)} d\mathbf{u}_1 d\mathbf{u}_2 \quad (227)$$

where $H(\mathbf{u}_1)H^*(\mathbf{u}_2)$ is the mutual coherent transfer function (discussed in Subsection 6.1.3). Eq. (227) suggests that for partially coherent imaging, the CSD is transferred linearly with the mutual coherent transfer function. It can also be written as a two-fold convolution of the sample CSD with the mutual PSF, defined as $h(\mathbf{x}_1)h^*(\mathbf{x}_2)$ (discussed in Subsection 6.1.3)

$$W_I(\mathbf{x}_1, \mathbf{x}_2) = \iint W_O(\mathbf{x}'_1, \mathbf{x}'_2) h(\mathbf{x}_1 - \mathbf{x}'_1) h^*(\mathbf{x}_2 - \mathbf{x}'_2) d\mathbf{x}'_1 d\mathbf{x}'_2 \quad (228)$$

This suggests that the CSD in the object space is blurred by the mutual PSF in both coordinates (\mathbf{x}_1 and \mathbf{x}_2). The image intensity $I(\mathbf{x})$ is given by the values “on the diagonal” of the CSD

$$I(\mathbf{x}) = W_I(\mathbf{x}, \mathbf{x}) = \iint W_S(\mathbf{x}_1, \mathbf{x}_2) T(\mathbf{x}_1) T^*(\mathbf{x}_2) h(\mathbf{x} - \mathbf{x}_1) h^*(\mathbf{x} - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \quad (229)$$

Substituting Eq. (225) into Eq. (229), the expression for the intensity can be simplified consequently

$$I(\mathbf{x}) = \int S(\mathbf{u}) \left| \int T(\mathbf{x}') h(\mathbf{x} - \mathbf{x}') e^{j2\pi\mathbf{u}\mathbf{x}'} d\mathbf{x}' \right|^2 d\mathbf{u} \equiv \int S(\mathbf{u}) I_{\mathbf{u}}(\mathbf{x}) d\mathbf{u} \quad (230)$$

Eq. (230) suggests that the intensity captured at the image plane can be interpreted as an incoherent superposition of the coherent partial images $I_{\mathbf{u}}(\mathbf{x})$ arising from all points of the incoherent source.

6.3.2. Effect of spatial coherence on image formation

In Subsection 4.2, we considered the case of plane wave coherent illumination, i.e., the light source at the condenser aperture plane is an on-axis point radiator $S(\mathbf{u}) = \delta(\mathbf{u})$. For such a coherent imaging case, the intensity in the image plane can be represented as [Eq. (60)]

$$I_0(\mathbf{x}) = \left| \int T(\mathbf{x}') h(\mathbf{x} - \mathbf{x}') d\mathbf{x}' \right|^2 = |T(\mathbf{x}) \otimes h(\mathbf{x})|^2 \quad (231)$$

Let us consider another extreme case: the source is infinitely extended such that the CSD of the illumination field degenerates to a δ

function.

$$W_S(\mathbf{x}_1, \mathbf{x}_2) = \iint C e^{j2\pi\mathbf{u}\cdot(\mathbf{x}_1-\mathbf{x}_2)} d\mathbf{u} = C\delta(\mathbf{x}_1 - \mathbf{x}_2) \quad (232)$$

Substituting Eq. (232) into Eq. (229) and neglecting the unimportant constant factor, we obtain

$$I(\mathbf{x}) = \iint \delta(\mathbf{x}_1 - \mathbf{x}_2) T(\mathbf{x}_1) T^*(\mathbf{x}_2) h(\mathbf{x} - \mathbf{x}_1) h^*(\mathbf{x} - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 = |T(\mathbf{x})|^2 \otimes |h(\mathbf{x})|^2 \quad (233)$$

Eqs. (232) and (233) represent two extreme cases for partially coherent imaging: completely coherent and incoherent. For partially coherent imaging, it should be somewhere in between: the extended source has a finite size. In order to reveal the influence of spatial coherence on the formed intensity image more clearly, the limited-aperture effect of the objective lens is neglected [the NA of the objective lens is infinite, i.e., $P(\mathbf{u}) \equiv 1$ in Eq. (63)]. When the object is imaged at a defocus distance Δz , the coherent transfer function degenerates into the angular spectrum transfer function for free-space propagation [Eq. (30)]. Under the paraxial approximation, the impulse response function can be further simplified as [Eq. (17)]

$$h_{\Delta z}(\mathbf{x}) = \frac{1}{j\lambda\Delta z} \exp(jk\Delta z) \exp\left\{\frac{j\pi}{\lambda\Delta z} |\mathbf{x}|^2\right\} \quad (234)$$

Substituting Eq. (234) into Eqs. (230) and (231), and ignoring the constant and phase factors that are independent of the transverse intensity

$$I(\mathbf{x}) = \int S(\mathbf{u}) \left| \int T(\mathbf{x}') \exp\left\{\frac{j\pi}{\lambda\Delta z} [|\mathbf{x}|^2 + |\mathbf{x}'|^2 - 2(\mathbf{x} - \lambda\Delta z\mathbf{u}) \cdot \mathbf{x}']\right\} d\mathbf{x}' \right|^2 d\mathbf{u} \quad (235)$$

$$I_0(\mathbf{x}) = \left| \int T(\mathbf{x}') \exp\left\{\frac{j\pi}{\lambda\Delta z} [|\mathbf{x}|^2 + |\mathbf{x}'|^2 - 2\mathbf{x} \cdot \mathbf{x}']\right\} d\mathbf{x}' \right|^2 \quad (236)$$

Adding an insignificant constant factor to the intensity in Eq. (235) and comparing it with Eq. (236) gives:

$$\begin{aligned} I(\mathbf{x}) &= \int S(\mathbf{u}) \left| \int T(\mathbf{x}') \exp\left\{\frac{j\pi}{\lambda\Delta z} [|\mathbf{x} - \lambda\Delta z\mathbf{u}|^2 + |\mathbf{x}'|^2 - 2(\mathbf{x} - \lambda\Delta z\mathbf{u}) \cdot \mathbf{x}']\right\} d\mathbf{x}' \right|^2 d\mathbf{u} \\ &= \int S(\mathbf{u}) \left| \int T(\mathbf{x}') \exp\left\{\frac{j\pi}{\lambda\Delta z} [|\mathbf{x} - \lambda\Delta z\mathbf{u} - \mathbf{x}'|^2]\right\} d\mathbf{x}' \right|^2 d\mathbf{u} \\ &= \int S(\mathbf{u}) I_0(\mathbf{x} - \lambda\Delta z\mathbf{u}) d\mathbf{u} \stackrel{\mathbf{u}=\frac{\Delta z\mathbf{x}}{f}}{=} S\left(\frac{\Delta z\mathbf{x}}{f}\right) \otimes I_0(\mathbf{x}) \end{aligned} \quad (237)$$

where $S(\Delta z\mathbf{x}/f)$ is the effective intensity distribution of the light source, and the size is proportional to the defocus distance Δz . The physical meaning of Eq. (237) can be explained by examining Fig. 53: when the object is illuminated by an oblique plane wave emitted from an off-axis point source, the image will be laterally shifted with a distance $\Delta z s/f$ that is proportional to the defocus distance Δz . Therefore, for an extended incoherent source (e.g., a homogeneous disk with a radius s , with the increase in Δz , the diffuse spot (defocused PSF) expands proportionally, creating a homogeneous circular disc with a radius of $\Delta z s/f$. The diffuse spots produced by different object points are superimposed, resulting in a blurred diffraction pattern (defocused intensity image).

In order to alleviate the blurring effect induced by the extended source, we can reduce the defocus distance Δz . When $\Delta z \rightarrow 0$, $S(\Delta z\mathbf{x}/f) \rightarrow \delta(\mathbf{u})$, and there will be no source blur. We can obtain an intensity distribution $I_0(\mathbf{x})$ that is completely consistent with the coherent case. However, as discussed in Subsection 4.2, the premise of TIE phase retrieval is to create phase contrast by defocusing, thus convert phase information into intensity information. The phase component of the object cannot be visualized without defocusing. Therefore, it is not difficult to understand that using extended sources (open up the condenser diaphragm) will attenuate the phase contrast in a microscopic image system. Note that the precondition for the establishment of Eq. (237) is that the aperture of the imaging system is infinite [$P(\mathbf{u}) \equiv 1$]. Eq. (237) may break down when considering the finite NA of the objective (see Subsection 6.3.4 for detailed discussions).

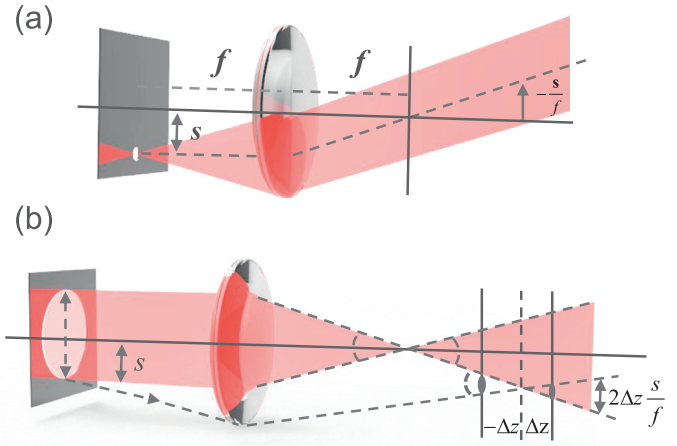


Fig. 53. Blurring effect due to spatial coherence of the light source [342]. (a) An off-axis point source produces tilted plane wave that travels with transverse vector direction $-s/f$; (b) intensity blurring resulting from the finite source size under defocus. Rays passing through a single point on the sample will produce a circular spot in the defocused planes located at $\pm\Delta z$ of diameter $2\Delta z s/f$.

6.3.3. Effect of temporal coherence on image formation

The temporal coherence of the light source is another important factor affecting the image formation. In current microscopes, the light source is usually a halogen lamp or a continuous spectrum LED. Neither of them can provide monochromatic illuminations like a laser so that the monochromatic assumption is no longer valid. As we know from Subsection 6.1, the temporal coherence of a light source depends on its PSD (spectral distribution). Assuming that the PSD of the light source is $S_\omega(\mathbf{u})$, since different wavelengths are uncorrelated, the total intensity distribution of the source is the integral of the PSD over all wavelengths

$$I_S(\mathbf{u}) = \int S_\omega(\mathbf{u}) d\omega \quad (238)$$

In the previous subsection, the light source is assumed to be a quasi-monochromatic extended source with an intensity distribution of $S(\mathbf{u})$. In this section, we assume that the light source has a certain spectral bandwidth with a PSD of $S_\omega(\mathbf{u})$. In this case, Eqs. (229) and (230) can be rewritten as

$$\begin{aligned} I(\mathbf{x}) &= \int W_\omega(\mathbf{x}, \mathbf{x}) d\omega \\ &= \iiint W_{S_\omega}(\mathbf{x}_1, \mathbf{x}_2) T_\omega(\mathbf{x}_1) T_\omega^*(\mathbf{x}_2) h_\omega(\mathbf{x} - \mathbf{x}_1) h_\omega^*(\mathbf{x} - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 d\omega \end{aligned} \quad (239)$$

and

$$\begin{aligned} I(\mathbf{x}) &= \iint S_\omega(\mathbf{u}) \left| \int T_\omega(\mathbf{x}') h_\omega(\mathbf{x} - \mathbf{x}') e^{j2\pi\mathbf{u}\cdot\mathbf{x}'} d\mathbf{x}' \right|^2 d\mathbf{u} d\omega \\ &\equiv \iint S_\omega(\mathbf{u}) I_{\mathbf{u},\omega}(\mathbf{x}) d\mathbf{u} d\omega \end{aligned} \quad (240)$$

Eqs. (239) and (240) indicate that the intensity distribution of the image is the superposition of the power density produced by different wavelength components. This conclusion seems straightforward because it only needs to integrate over all different wavelengths. However, it should be stressed that both the complex transmittance of the object and the PSF of the imaging system in Eqs. (239) and (240) are wavelength-dependent functions in a strict sense. More specifically, the wavelength-dependent RI induced sample dispersion and chromatic aberration often complicate accurate phase determination. Therefore, in practical phase imaging systems, the sample is often illuminated with quasi-monochromatic illumination by introducing a narrow bandpass filter

into the illumination path or using monochromatic LEDs as illumination sources. When it is necessary to use polychromatic illuminations, we have to impose certain assumptions on the sample: e.g., it should be non-dispersive and the absorption is wavelength-independent within the spectral range of the illumination (which are generally valid for unstained cells).

To better explain the influence of temporal coherence on image formation, let us turn our attention back to the complex impulse response function under Fresnel approximation

$$h_{\Delta z}(\mathbf{x}) = \frac{1}{j\lambda\Delta z} \exp(jk\Delta z) \exp\left\{\frac{j\pi}{\lambda\Delta z}|\mathbf{x}|^2\right\} \quad (241)$$

It is obvious that the impulse response function is related to the illumination wavelength. The first two terms on the RHS of Eq. (241) are constant and phase factors that are independent of the lateral coordinates. The third term $\exp\left\{\frac{j\pi}{\lambda\Delta z}|\mathbf{x}|^2\right\}$ is a quadratic phase factor which determines the intensity distribution. Close inspections reveal that the wavelength λ and Δz appear in pair, which suggests that a change in wavelength ($\lambda \rightarrow \lambda + \delta\lambda$) is equivalent to a change in propagation distance ($\Delta z \rightarrow \Delta z + \delta z$), where

$$\delta z = -\frac{\delta\lambda\Delta z}{\lambda + \delta\lambda} \quad (242)$$

Therefore, if polychromatic illuminations are used, the intensity finally captured at the defocused plane can be regarded as the superposition of many diffraction patterns of different wavelengths (or equivalently, at different defocus distances). Obviously, the superposition of mutually uncorrelated diffuse spots (defocused PSFs) at different propagation distances results in a blurred diffraction pattern. Therefore, similar to the lateral blurring effect due to the spatial coherence, the temporal coherence induced blurring results from the superposition of different diffuse spots at different axial propagation distances. Finally, it is worth noting that although the use of polychromatic illuminations attenuates the phase contrast in the defocused plane, the wavelength-dependent nature of the Fresnel diffraction can be wisely utilized to get rid of mechanical defocusing in TIE phase retrieval [191,212].

6.3.4. Transmission cross-coefficient (TCC) and weak object transfer function (WOTF)

In general, the image formation in microscopic imaging systems can be described by Fourier transform and linear filtering: a coherent imaging system is linear in complex amplitude [Eq. (231)], while an incoherent imaging system is linear in intensity [Eq. (232)]. However, for partially coherent imaging, the image formation is more complicated due to the nonlinear dependency of the image intensity on the object transmittance, light source, and imaging system. In Subsection 6.3.1, we gave the representations of the intensity distribution under partially coherent illuminations. Eq. (230) has a clear physical meaning: a partially coherent image can be regarded as the superposition of sub-images generated by different point sources. The image formation model described by Eq. (230) is the so-called Abbe's superposition method, which has a clear physical meaning: a partially coherent image can be regarded as the superposition of sub-images generated by different source points in the condenser aperture plane. It should be noted that the partially coherent imaging theory of a microscope was first established by Hopkins [247] from a different perspective in 1950s. Starting from Eq. (229) and combining Eq. (225), we can derive the following Fourier integral

$$I(\mathbf{x}) = \iint S(\mathbf{u})\hat{T}(\mathbf{u}_1)\hat{T}^*(\mathbf{u}_2)H(\mathbf{u} + \mathbf{u}_1)H^*(\mathbf{u} + \mathbf{u}_2)e^{j2\pi\mathbf{x}(\mathbf{u}_1 - \mathbf{u}_2)}d\mathbf{u}_1d\mathbf{u}_2d\mathbf{u} \quad (243)$$

Separating the contributions of the specimen and imaging system leads to the notion of the *transmission cross-coefficient* (TCC)

$$TCC(\mathbf{u}_1, \mathbf{u}_2) = \iint S(\mathbf{u})H(\mathbf{u} + \mathbf{u}_1)H^*(\mathbf{u} + \mathbf{u}_2)d\mathbf{u} \quad (244)$$

With the TCC, we can simplify Eq. (243) as follows

$$I(\mathbf{x}) = \iint \hat{T}(\mathbf{u}_1)\hat{T}^*(\mathbf{u}_2)TCC(\mathbf{u}_1, \mathbf{u}_2)e^{j2\pi\mathbf{x}(\mathbf{u}_1 - \mathbf{u}_2)}d\mathbf{u}_1d\mathbf{u}_2 \quad (245)$$

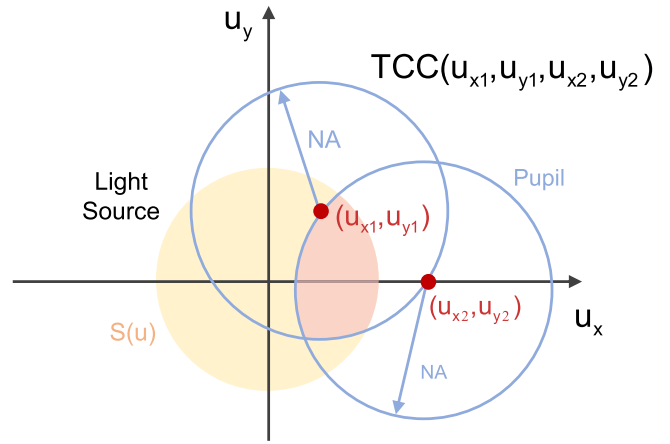


Fig. 54. Geometric interpretation of the transmission cross coefficient (TCC).

The image intensity is therefore computed by the linear superposition of the interference patterns of plane waves with spatial frequencies \mathbf{u}_1 and \mathbf{u}_2 [$e^{j2\pi\mathbf{x}(\mathbf{u}_1 - \mathbf{u}_2)}$], and amplitude given by the TCC and the object spectrum at \mathbf{u}_1 and \mathbf{u}_2 [$T(\mathbf{u}_1)T^*(\mathbf{u}_2)$]. In essence, TCC is a 2D convolution integral, which can be calculated by integrating the three functions over the overlapping area of the light source and the two shifted complex pupils, as illustrated in Fig. 54. The result of the integral at a given position (u_{x1}, u_{y1}) and (u_{x2}, u_{y2}) corresponds to one element in the 4D TCC matrix. When the two pupil functions shift across the (u_x, u_y) plane, the 4D TCC matrix can be constructed.

The image formation in partially coherent systems is not linear in either amplitude or intensity but is bilinear, which makes image restoration and phase recovery more complicated. To simplify the mathematical formulation, the weak object approximation is often assumed to linearize the phase retrieval problem, which was discussed under the condition of coherent imaging (see Subsection 4.3.1). In the following, we discuss the weak object approximation for partially coherent imaging. The complex transmittance of a weak object can be represented as [Eq. (77)]

$$T(\mathbf{x}) = a(\mathbf{x}) \exp[j\phi(\mathbf{x})] \begin{cases} \phi(\mathbf{x}) \ll 1 \\ \approx [a_0 + \Delta a(\mathbf{x})][1 + j\phi(\mathbf{x})] \\ \Delta a(\mathbf{x}) \ll a_0 \\ \approx a_0 + \Delta a(\mathbf{x}) + ja_0\phi(\mathbf{x}) \end{cases} \quad (246)$$

The corresponding Fourier transform can be expressed as

$$\hat{T}(\mathbf{u}) = a_0[\delta(\mathbf{u}) + \hat{\eta}(\mathbf{u}) + j\hat{\phi}(\mathbf{u})] \quad (247)$$

Then the *mutual spectrum* of the object can be written as

$$\hat{T}(\mathbf{u}_1)\hat{T}^*(\mathbf{u}_2) = a_0^2\delta(\mathbf{u}_1)\delta(\mathbf{u}_2) + a_0\delta(\mathbf{u}_2)[\hat{\eta}(\mathbf{u}_1) + j\hat{\phi}(\mathbf{u}_1)] + a_0\delta(\mathbf{u}_1)[\hat{\eta}(\mathbf{u}_2) + j\hat{\phi}(\mathbf{u}_2)] \quad (248)$$

Noted that the high order interference terms are neglected because the scattered light is weak, which just corresponds to the first order Born approximation commonly used in diffraction tomography (discussed in Section 8). Substituting Eq. (248) into Eq. (245) yields the intensity of the partially coherent image

$$I(\mathbf{x}) = a_0^2TCC(\mathbf{0}, \mathbf{0}) + 2a_0 \operatorname{Re} \left\{ \int TCC(\mathbf{u}, \mathbf{0})[\hat{\eta}(\mathbf{u}) + j\hat{\phi}(\mathbf{u})]e^{j2\pi\mathbf{x}\mathbf{u}}d\mathbf{u} \right\} \quad (249)$$

where the simple relation $TCC^*(\mathbf{0}, \mathbf{u}) = TCC(\mathbf{u}, \mathbf{0})$ is used (TCC is Hermitian symmetric). It now becomes obvious that the image contrast due to the absorption and phase are decoupled and linearized. In the following, we denote the $TCC(\mathbf{u}, \mathbf{0})$, the linear part of the TCC, as *weak object transfer function* (WOTF)

$$WOTF(\mathbf{u}) \equiv TCC(\mathbf{u}, \mathbf{0}) = \iint S(\mathbf{u}')H(\mathbf{u}' + \mathbf{u})H^*(\mathbf{u}')d\mathbf{u}' \quad (250)$$

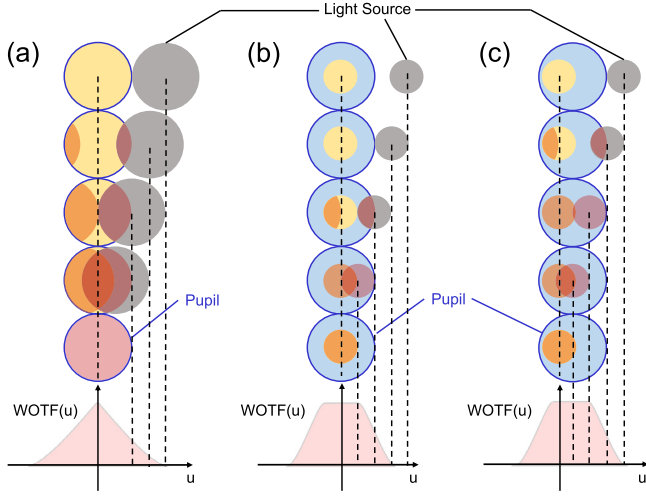


Fig. 55. Geometric interpretation of WOTF for three different illuminations. (a) Incoherent imaging (light source larger than or equal to the objective pupil); (b) for partially coherent imaging (light source smaller than the objective pupil); (c) partially coherent oblique illumination.

Since the absorption and phase of the object are both real functions, according to the properties of Fourier transform, the real part of $\hat{\eta}(\mathbf{u})$ is an even function, and the imaginary part of $\hat{\eta}(\mathbf{u})$ is an odd function. Due to the imaginary unit before the phase spectrum, the real part of $j\hat{\phi}(\mathbf{u})$ is an odd function, and the imaginary part of $j\hat{\phi}(\mathbf{u})$ is an even function.

For an aberration-free microscopic system with axisymmetric illumination and objective pupil, the coherent transfer function is the objective pupil function $H(\mathbf{u}) = P(\mathbf{u})$. For bright-field imaging, the circular condenser aperture diaphragm is enclosed by the objective pupil

$$WOTF(\mathbf{u}) \equiv TCC(\mathbf{u}, 0) = \iint P(\mathbf{u}')P(\mathbf{u}' + \mathbf{u})d\mathbf{u}' \quad (251)$$

Now the $WOTF(\mathbf{u})$ is always real and even, as illustrated in Fig. 55. When the real $WOTF(\mathbf{u})$ is multiplied with $j\hat{\phi}(\mathbf{u})$, the real part of the product becomes an odd function, and the imaginary part of the product becomes an even function. According to the symmetries of Fourier transform, the Fourier transform of the odd, real function is an odd, imaginary function, and the Fourier transform of an even, imaginary function is still an even, imaginary function. Therefore, the phase component is a purely imaginary function after Fourier transform, so it cannot be reflected in the intensity image. Therefore, if the microscope is in-focus, the phase does not manifest itself in the intensity image. In the same way, it can be found that the absorption component is a purely real function after Fourier transform, so only the absorption component can be observed in the in-focus image. For a general partially coherent imaging system, the odd-symmetric and even-symmetric components of the $WOTF(\mathbf{u})$ correspond to the contributions of amplitude and phase components in the intensity signal, respectively. We define them as the ATF and PTF for partially coherent imaging

$$I(\mathbf{x}) = a_0^2 TCC(\mathbf{0}, \mathbf{0}) + 2a_0 \text{Re} \left\{ \mathcal{F}^{-1} \left[H_A(\mathbf{u})\hat{\eta}(\mathbf{u}) + H_P(\mathbf{u})\hat{\phi}(\mathbf{u}) \right] \right\} \quad (252)$$

where

$$H_A(\mathbf{u}) = WOTF(\mathbf{u}) + WOTF^*(-\mathbf{u}) \quad (253)$$

$$H_P(\mathbf{u}) = WOTF^*(\mathbf{u}) - WOTF(-\mathbf{u}) \quad (254)$$

corresponding to the partially coherent ATF and partially coherent PTF. Therefore, in order to generate phase contrast signal, the symmetry of either source or pupil function has to be broken. If the illumination or the pupil function is asymmetric, the PTF $H_P(\mathbf{u})$ will be non-zero so that the phase component can manifest in the image. Such asymmetric

illumination or detection schemes are widely adopted in phase-contrast imaging, wavefront sensing, and QPI, e.g., differential phase-contrast (DPC) microscopy [343–345], pyramid wavefront sensor [44,346], and partitioned/programmable aperture microscopy (PAM) [347,348].

The situation discussed above is limited to an aberration-free optical system. In fact, another approach to break the realness of WOTF is to introduce aberration. Optical aberration produces an additional imaginary component in the WOTF, resulting in phase contrast. Defocusing is the most convenient way to introduce aberration into the optical system. When the imaging system is defocused with a distance Δz , the coherent transfer function becomes the product of the objective pupil function $P(\mathbf{u})$ and the angular spectrum transfer function [Eq. (30)]

$$H(\mathbf{u}) = P(\mathbf{u})H_{\Delta z}(\mathbf{u}) = P(\mathbf{u})e^{jk\Delta z\sqrt{1-\lambda^2|\mathbf{u}|^2}} \quad (255)$$

Substituting the complex pupil function into Eq. (250) results in a complex and even WOTF

$$\begin{aligned} WOTF(\mathbf{u}) &\equiv TCC(\mathbf{u}, 0) \\ &= \iint S(\mathbf{u}')P(\mathbf{u}')P(\mathbf{u}' + \mathbf{u})e^{jk\Delta z\left(-\sqrt{1-\lambda^2|\mathbf{u}'|^2} + \sqrt{1-\lambda^2|\mathbf{u}'+\mathbf{u}|^2}\right)}d\mathbf{u}' \end{aligned} \quad (256)$$

Similarly, according to the symmetries of Fourier transform, the transfer functions for the absorption and phase components of a weak object are then given by the real and imaginary components of the WOTF

$$H_A(\mathbf{u}) = \text{Re} [WOTF(\mathbf{u})] \quad (257)$$

$$H_P(\mathbf{u}) = -\text{Im} [WOTF(\mathbf{u})] \quad (258)$$

6.3.5. Phase transfer functions for coherent, incoherent, and partially coherent imaging

In Section 4, we know that the spatial coherence of the illumination in a microscope can be quantified by coherence parameter. The coherence parameter is defined as the ratio of the illumination NA to the objective NA [177,178,246,249], which is equivalent to the radius ratio of the condenser aperture to the objective aperture

$$s = \frac{NA_{ill}}{NA_{obj}} \quad (259)$$

Under coherent illuminations, the source degenerates into the ideal on-axis point radiator $S(\mathbf{u})=\delta(\mathbf{u})$, i.e., $s \rightarrow 0$. Then, the partially coherent WOTF degenerates into the complex pupil function

$$WOTF_{coh}(\mathbf{u}) = P(\mathbf{u})e^{jk\Delta z\left(\sqrt{1-\lambda^2|\mathbf{u}|^2}-1\right)} \quad (260)$$

and the real and the imaginary parts of the WOTF correspond to the ATF and PTF, respectively

$$H_A(\mathbf{u}) = P(\mathbf{u}) \cos \left[k\Delta z \left(\sqrt{1-\lambda^2|\mathbf{u}|^2} - 1 \right) \right] \quad (261)$$

$$H_P(\mathbf{u}) = -P(\mathbf{u}) \sin \left[k\Delta z \left(\sqrt{1-\lambda^2|\mathbf{u}|^2} - 1 \right) \right] \quad (262)$$

Under the paraxial approximation, $jk\Delta z\left(\sqrt{1-\lambda^2|\mathbf{u}|^2}-1\right) \approx -j\pi\lambda\Delta z|\mathbf{u}|^2$, and the transfer functions can be simplified as

$$H_A(\mathbf{u}) = P(\mathbf{u}) \cos (\pi\lambda\Delta z|\mathbf{u}|^2) \quad (263)$$

$$H_P(\mathbf{u}) = P(\mathbf{u}) \sin (\pi\lambda\Delta z|\mathbf{u}|^2) \quad (264)$$

These expressions are identical to Eqs. (92) and (93) discussed in Section 4. For weak defocusing $\Delta z \rightarrow 0$, the sine term can be further approximated by a parabolic function $\sin (\pi\lambda\Delta z|\mathbf{u}|^2) \approx \pi\lambda\Delta z|\mathbf{u}|^2$, which is a Laplacian in the Fourier space, corresponding to the PTF implied in TIE.

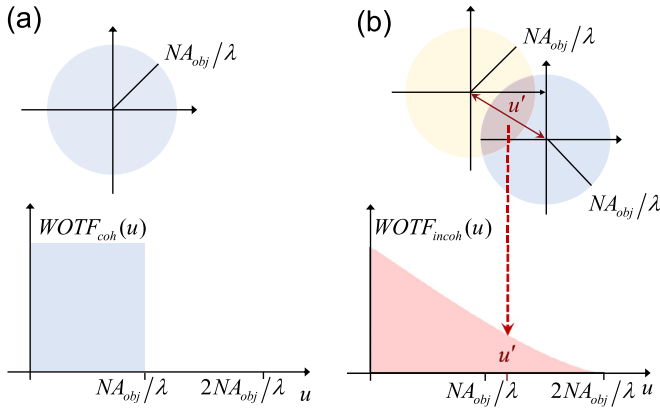


Fig. 56. Transfer functions of a rotation-symmetric system. (a) CTF in the coherent case; (b) OTF in the incoherent case.

Next we discuss the other extreme case, incoherent illumination. Strictly speaking, spatially incoherent illumination requires an infinitely extended source, *i.e.*, the coherence parameter $s \rightarrow \infty$ and the CSD reduces to a δ function [Eq. (232)]. However, when considering a weak object, as the illumination NA increases to the objective NA or even larger, $s = NA_{ill}/NA_{obj} \geq 1$, the circular condenser aperture is limited by the objective pupil. The form of WOTF will no longer change and become the autocorrelation of the coherent transfer function

$$WOTF_{incoh}(\mathbf{u}) = \iint H(\mathbf{u}' + \mathbf{u})H^*(\mathbf{u}')d\mathbf{u}' \quad (265)$$

At this time the system behaves the same as an incoherent microscope [142,178,249] [but note that it is not really incoherent (which requires $s \rightarrow \infty$) because a weak object is assumed in this case]. For an axisymmetric optical imaging system, the coherent transfer function $H(\mathbf{u})$ is an even function of \mathbf{u} , then $WOTF_{incoh}(\mathbf{u})$ is always real (even under defocusing), and its PTF is always 0. Hence the phase component of the specimen cannot be imaged. Consequently, it is impossible to perform phase retrieval under incoherent illumination.

For an ideal in-focus axisymmetric system with circular apertures (no pupil aberrations), the WOTF becomes a convolution of two circles of equal size (as shown in Fig. 56), which can be represented as

$$WOTF_{incoh}(\bar{\rho}) = \frac{2}{\pi} \left[\arccos\left(\frac{\bar{\rho}}{2}\right) - \frac{\bar{\rho}}{2} \sqrt{1 - \left(\frac{\bar{\rho}}{2}\right)^2} \right] \quad \bar{\rho} < 2 \quad (266)$$

where $\bar{\rho}$ is the normalized radial spatial coordinate with respect to the coherent resolution limit NA_{obj}/λ

$$\bar{\rho} = \frac{\lambda}{NA_{obj}} \rho \quad (267)$$

After normalization, the coherent diffraction limit NA_{obj}/λ corresponds to $\bar{\rho} = 1$. Fig. 57(a) shows the in-focus $WOTF_{incoh}(\bar{\rho})$, which is purely real-valued and can only image the absorption part. It can be seen that the absorption contrast gradually reduces as the increase in spatial frequency, with a 61% relative contrast loss at the coherent resolution limit (NA_{obj}/λ), and finally diminishes to zero at the incoherent diffraction limit ($2NA_{obj}/\lambda$). The contrast loss is even more severe when the system is defocused, as demonstrated by the $WOTF_{incoh}$ shown in Fig. 57(b).

From the above analysis, we know that under coherent illumination, phase contrast can be produced by defocusing, but imaging resolution is restricted to the coherent diffraction limit. Incoherent illumination provides higher imaging resolution (twice the coherent diffraction limit), but the phase component of the sample cannot be imaged. A compromise between the phase contrast and imaging resolution can be achieved by using partially incoherent illuminations with a reduced source dimension ($0 < s < 1$). Figs. 58 and 59 show the partially coherent ATF and PTF

for various coherent parameters and defocus distances. These results are numerically calculated from Eq. (250) directly, as a weighted area of overlap over $H^*(\mathbf{u}')$ and $H(\mathbf{u}'+\mathbf{u})$. For an ideal in-focus axisymmetric system with circular apertures, the WOTF can be represented analytically, which is just the overlapping area of two circles with different radii

$$WOTF_s(\bar{\rho}) = \begin{cases} \pi s^2 & 0 \leq \bar{\rho} \leq 1-s \\ s^2 \arccos\left(\frac{\bar{\rho}^2+s^2-1}{2\bar{\rho}s}\right) - \left(\frac{\bar{\rho}^2+s^2-1}{2\bar{\rho}}\right) \sqrt{s^2 - \left(\frac{\bar{\rho}^2+s^2-1}{2\bar{\rho}}\right)^2} & 1-s < \bar{\rho} < 1+s \\ +s^2 \arccos\left(\frac{\bar{\rho}^2-s^2+1}{2\bar{\rho}s}\right) - \left(\frac{\bar{\rho}^2-s^2+1}{2\bar{\rho}}\right) \sqrt{1 - \left(\frac{\bar{\rho}^2-s^2+1}{2\bar{\rho}}\right)^2} & 1+s \leq \bar{\rho} \leq 1+s \end{cases} \quad (268)$$

Fig. 57 gives the corresponding in-focus WOTFs for different coherent parameters. As can be seen, the cut-off frequency is extended to $1 + s$, which is resolution limit for partially coherent imaging. Although incoherent illumination has a higher resolution limit, details at a higher frequency are imaged with low contrast. On the other hand, for partially coherent imaging, a higher image contrast is achieved due to the better transfer at the low frequencies. For normal bright-field imaging (for absorptive sample), it has transpired to be a good compromise when the light source image fills approximately half area of the pupil ($s = 0.7 - 0.8$) for achieving both high imaging resolution and high image contrast (especially for low spatial frequencies).

Next, we focus on the partially coherent PTFs under defocusing. It can be seen from Fig. 58 that increasing coherence parameter diminishes the phase contrast considerably. This is in consistent with the well-known phenomenon that the condenser of a microscope must be stopped down to produce an appreciable contrast for phase information. But in this case, the associated reduction in resolution limit becomes a major issue, leading to a compromised imaging resolution. This suggests that the coherence parameter plays a very important role in phase imaging. Closing down the condenser diaphragm can improve the phase contrast at the expense of a reduced resolution limit. While opening up the condenser diaphragm improves the resolution limit, however, the response of the PTF is significantly attenuated. Therefore, in order to improve the quality of phase retrieval, it is necessary to select an appropriate coherence parameter and defocus distance to optimize the response of the PTF.

At the end of this section, let us consider the question raised at the beginning of this section. Teague [141] originally derived TIE under the assumption of complete coherence, *i.e.*, a monochromatic coherent field. However, in practical implementations, TIE seems to work well even when the illumination is not strictly coherent. In order to give a reasonable explanation for this phenomenon, let us recall the defocused coherent transfer function in the normalized coordinate [Eq. (76)]

$$H(\bar{\rho}) = P(\bar{\rho})e^{-\frac{1}{2}j\bar{z}\bar{\rho}^2} \quad (269)$$

where

$$\bar{\rho} = \frac{\rho}{NA_{obj}/\lambda} \quad (270)$$

and

$$\bar{z} = 2k \left(1 - \sqrt{1 - NA_{obj}^2} \right) z \approx k NA_{obj}^2 z \quad (271)$$

When analyzing the transfer function of a microscopic imaging system, the normalized coordinate is often adopted to simplify the evaluation of integrals. The scaling coefficients in Eqs. (73) and (74) are associated with the transverse and axial extension of the 3D OTF, which will be introduced in Section 8.

Substituting Eq. (269) into Eq. (250), we get the defocused WOTF under the paraxial approximation

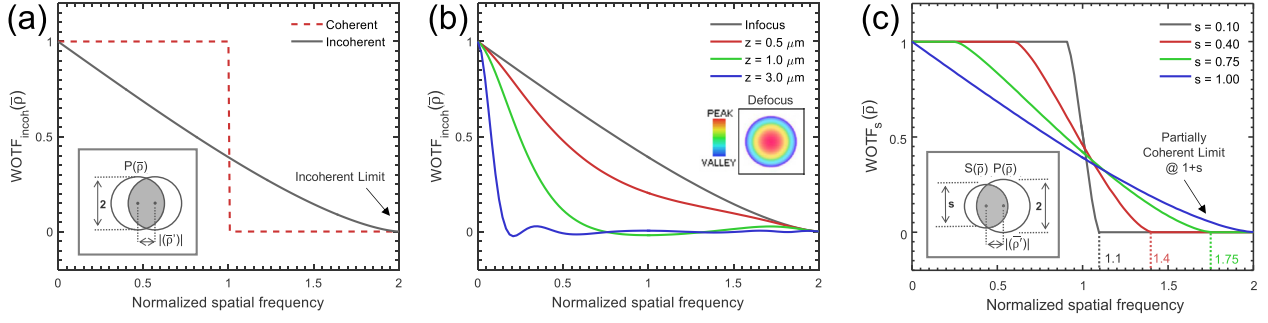


Fig. 57. Transfer functions under different spatial coherence. (a) In-focus $WOTF_{incoh}$, coherent parameters $s \geq 1$; (b) defocused $WOTF_{incoh}$, coherent parameters $s \geq 1$; (c) in-focus $WOTF_s$, $s = 0.1, 0.4, 0.75$ and 1.0 under different coherence parameters.

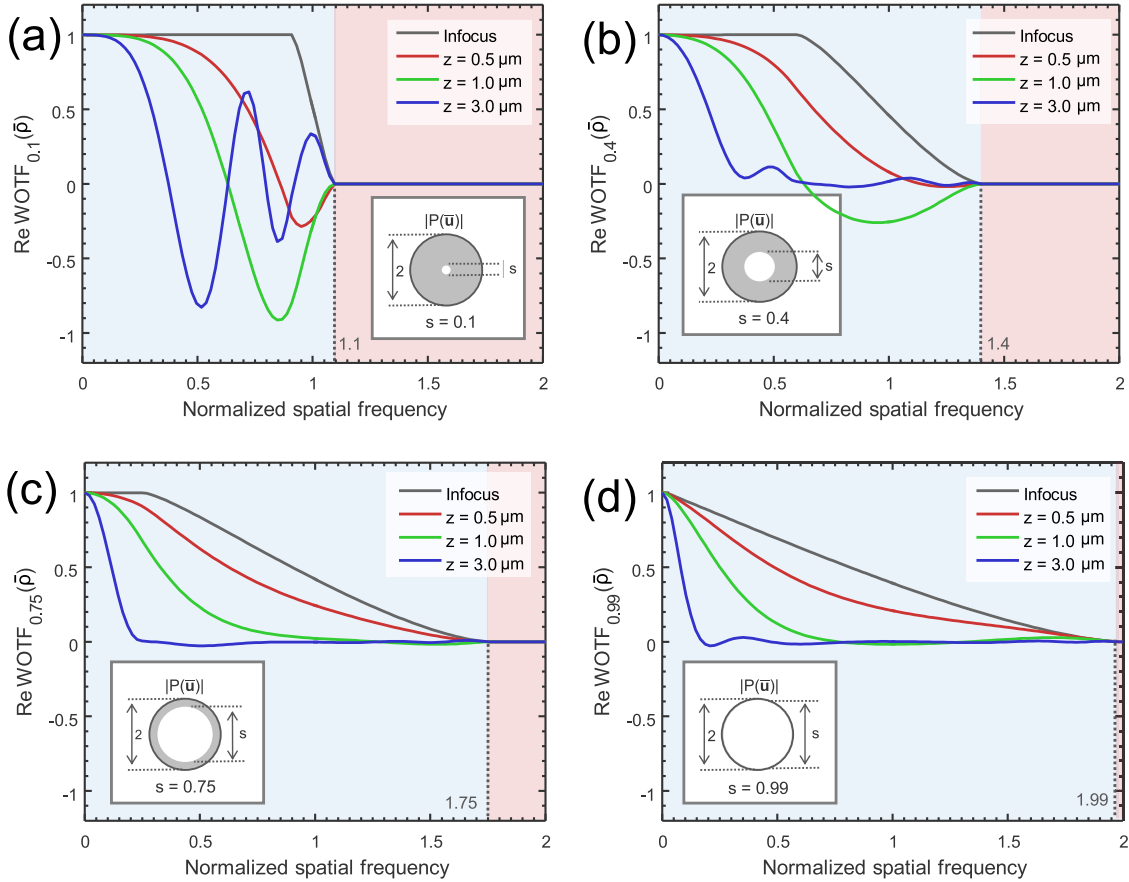


Fig. 58. The real part of $WOTF_s$ (ATF) for various coherent parameters and defocus distances ($NA_{obj}=0.8$, $\lambda = 550nm$, the spatial frequency coordinate is normalized against the coherent resolution limit NA_{obj}/λ). (a) $s = 0.1$; (b) $s = 0.4$; (c) $s = 0.75$; (d) $s = 0.99$.

$$WOTF_s(\bar{\rho}, \bar{z}) = \begin{cases} \int_{-(s+\bar{\rho}/2)}^{s-\bar{\rho}/2} 2\sqrt{s^2 - \left(\frac{\bar{\rho}}{2} + x\right)^2} \exp(j\bar{z}\bar{\rho}x) dx & 0 \leq \bar{\rho} < 1-s \\ \int_{-(1-s^2)/2\bar{\rho}}^{s-\bar{\rho}/2} 2\sqrt{s^2 - \left(\frac{\bar{\rho}}{2} + x\right)^2} \exp(j\bar{z}\bar{\rho}x) dx \\ + \int_{\bar{\rho}/2-1}^{-(1-s^2)/2\bar{\rho}} 2\sqrt{1 - \left(\frac{\bar{\rho}}{2} + x\right)^2} \exp(j\bar{z}\bar{\rho}x) dx & 1-s \leq \bar{\rho} \leq 1+s \end{cases} \quad (272)$$

Noted that this integral still cannot be evaluated analytically. But when the weak defocusing approximation is further introduced, i.e.

$$\Delta z \ll 1/\pi\lambda\rho^2 \quad (273)$$

the defocused coherent transfer function can be reduced to

$$H(\bar{\rho}) \approx 1 - \frac{1}{2}j\bar{z}\bar{\rho}^2 \quad (274)$$

The defocused WOTF (under paraxial and weak defocusing approximations) is obtained by substituting Eq. (274) into Eq. (250)

$$WOTF_s(\bar{\rho}, \bar{z}) = \begin{cases} \int_{-(s+\bar{\rho}/2)}^{s-\bar{\rho}/2} 2\sqrt{s^2 - \left(\frac{\bar{\rho}}{2} + x\right)^2} \left(1 - \frac{1}{2}j\bar{z}\bar{\rho}^2\right) dx & 0 \leq \bar{\rho} < 1-s \\ \int_{-(1-s^2)/2\bar{\rho}}^{s-\bar{\rho}/2} 2\sqrt{s^2 - \left(\frac{\bar{\rho}}{2} + x\right)^2} \left(1 - \frac{1}{2}j\bar{z}\bar{\rho}^2\right) dx \\ + \int_{\bar{\rho}/2-1}^{-(1-s^2)/2\bar{\rho}} 2\sqrt{1 - \left(\frac{\bar{\rho}}{2} + x\right)^2} \left(1 - \frac{1}{2}j\bar{z}\bar{\rho}^2\right) dx & 1-s \leq \bar{\rho} \leq 1+s \end{cases} \quad (275)$$

Now the integral can be solved explicitly. The real part is completely consistent with the in-focus ATF [Eq. (268)], and the imaginary part (i.e.,

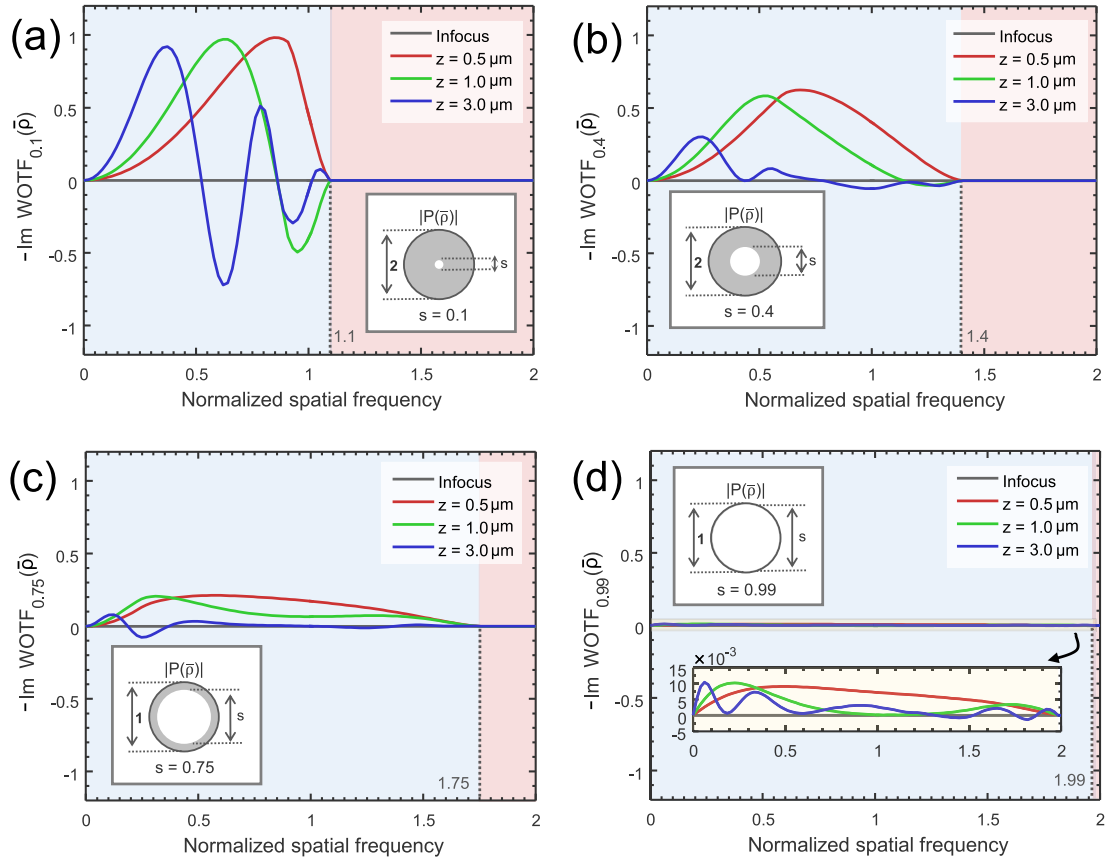


Fig. 59. The imaginary part of $WOTF_s$ (PTF) for various coherent parameters and defocus distances ($NA_{obj}=0.8$, $\lambda = 550nm$, the spatial frequency coordinate is normalized against the coherent resolution limit NA_{obj}/λ). (a) $s = 0.1$; (b) $s = 0.4$; (c) $s = 0.75$; (d) $s = 0.99$.

PTF) is given by

$$\begin{aligned}
 -\text{Im } WOTF_s(\bar{\rho}, \bar{z}) = & \begin{cases} \frac{1}{2} \pi \bar{z} \bar{\rho}^2 s^2 & 0 \leq \bar{\rho} \leq 1 - s \\ \frac{\bar{z} \bar{\rho}^2}{2} \left[s^2 \arccos \left(\frac{\bar{\rho}^2 + s^2 - 1}{2\bar{\rho}s} \right) - \arccos \left(\frac{\bar{\rho}^2 - s^2 + 1}{2\bar{\rho}} \right) \right] \\ + \frac{\bar{z}}{6\bar{\rho}} \left\{ \sqrt{s^2 - \left(\frac{\bar{\rho}^2 + s^2 - 1}{2\bar{\rho}} \right)^2} \left[s^2 \arccos \left(\frac{\bar{\rho}^2 + s^2 - 1}{2\bar{\rho}s} \right) - \arccos \left(\frac{\bar{\rho}^2 - s^2 + 1}{2\bar{\rho}} \right) \right] \right. \\ \left. \left[(1 - s^2)^2 - \frac{\bar{\rho}^2}{2} (1 + \bar{\rho}^2 + 7s^2) \right] \right\} & 1 - s \leq \bar{\rho} \leq 1 + s \\ + \sqrt{1 - \left(\frac{\bar{\rho}^2 - s^2 + 1}{2\bar{\rho}} \right)^2} \left[(1 - s^2)^2 - \frac{\bar{\rho}^2}{2} (1 + \bar{\rho}^2 + 7s^2) \right] & 1 - s \leq \bar{\rho} \leq 1 + s \end{cases} \quad (276)
 \end{aligned}$$

It can be found that when $0 \leq \bar{\rho} < 1 - s$, the PTF $\frac{1}{2} \pi \bar{z} \bar{\rho}^2 s^2$ is just a Laplacian in the Fourier domain. If we convert the normalized coordinate back to the original coordinate, and neglect the constant scaling factor (i.e., pupil area πs^2), $\frac{1}{2} \bar{z} \bar{\rho}^2 \rightarrow \pi \lambda z \rho^2$, it coincides with the TIE's PTF [Eq. (97)]. This suggests that under the paraxial approximation and weak defocusing, TIE (without considering the partial coherence) can still recover the phase information correctly within the low spatial frequency range of $0 \leq \bar{\rho} \leq 1 - s$. For spatial frequencies outside this range, $1 - s \leq \bar{\rho} \leq 1 + s$, the form of PTF become much more complicated. The amplitude of the PTF monotonically decays with the increase of ρ , which can be observed in Fig. 59 (see the red curves corresponding to $\Delta z = 0.5\mu m$).

In order to illustrate the relationship between WOTF and TIE more clearly, we show the ratio between the partially coherent weak defocused PTF and TIE's PTF in Fig. 60. It can be seen that the two PTFs are in complete agreement within the range of $0 \leq \bar{\rho} \leq 1 - s$, but begin to deviate when $\bar{\rho}$ becomes large. The divergence results from the overestimate of phase contrast in TIE due to the ignorance of limited aperture and partial coherence. Note that the trends of these curves are very similar to that of in-focus partially coherent ATFs [Fig. 57(c)], but

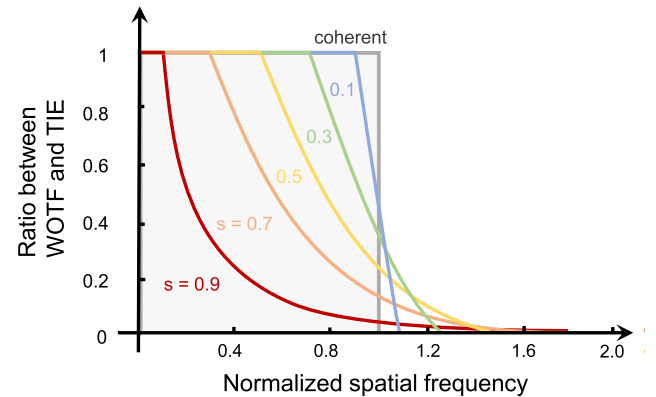


Fig. 60. The ratio between imaginary part of the WOTF for a weakly defocused, partially coherent bright-field microscope with coherent parameter s and the inverse Laplacian filter of TIE.

their meanings are completely different. One major difference is that when approaching incoherent illumination ($s \rightarrow 1$), the PTF gradually vanishes while the in-focus ATF still maintains a strong response.

7. Generalized TIE under partially coherent illuminations

No matter completely coherent fields, partially coherent fields, or completely incoherent fields, the intensity of the optical field is well-defined and directly accessible. Nevertheless, the “phase” is only well-defined for coherent fields. In a strict sense, any physically realizable light sources are not strictly coherent. Especially in adaptive opt-

tics [144–148], X-ray diffraction imaging [161,162], neutron radiography [164,165], and TEM [166,169,171–174], it is generally difficult to obtain highly coherent light sources so that TIE becomes a simple and effective alternative to interferometry. Even in the field of optical microscopy, partially coherent illumination is beneficial to enhance the imaging resolution, improve the image quality, and suppress the coherent noise. However, Teague's TIE, as well as his derivation, assumes a monochromatic, coherent beam [141]. The 2D complex amplitude function is no longer sufficient to fully describe partially coherent fields, and 4D MCF (or equivalently, CSD and WDF) should be used. Thus, Teague's TIE might encounter trouble when dealing with fields exhibiting non-negligible partial coherence. In this section, we analyze these issues and present the extensions of TIE for partially coherent fields. The challenges and opportunities brought by partially coherent illuminations for phase retrieval and QPI is discussed.

7.1. Extensions of TIE for partially coherent fields

The major obstacle in extending TIE for partially coherent fields arises from the fact that the partially coherent field does not have a well-defined phase since the field experiences statistical fluctuations over time. For partially coherent fields, the 2D complex amplitude is insufficient to fully describe the properties of the field. They require the 4D MCF or CSD of any two points in space, or equivalently being represented by WDF or AF in 4D phase space. In this case, the definition of phase should be clearly different from the one in a perfectly coherent field. As early as 1984, Streibl [142] used the 4D MI function to reformulate TIE in order to make it explicitly account for the partial coherence and demonstrated the possibility of using TIE for phase imaging under partially coherent illumination. Although he did not give the QPI results based on TIE (the numerical solution to TIE has not been proposed at that time), this work is still extremely important because he first pointed out that the phase retrieved by Teague's TIE can still be related to the phase of the object (not the field), provided that the intensity distribution of the primary is axis-symmetric about the optical axis. This pioneering work laid a preliminary theoretical foundation for the subsequent extensive applications of TIE in optical microscopy. In 1998, Paganin and Nugent [159] reinterpreted the "phase" of a partially coherent field as a scalar phase whose gradients become the normalized transverse time-averaged Poynting vector of the field. The significance of this work lies in the fact that it endows "phase" with a completely new and meaningful definition, providing a simple and reasonable physical basis for subsequent TIE phase retrieval under partially coherent illuminations. Unfortunately, this definition is only suitable for qualitative interpretation rather than quantitative analysis (because they did not give an explicit expression for the time-averaged Poynting vector of a partially coherent field). In 2004, Gureyev *et al.* [256] described an alternative interpretation with the generalized eikonal, based on the spectrum decomposition of a polychromatic field. He experimentally demonstrated that even under temporally partially coherent (polychromatic) illuminations, accurate phase retrieval can be obtained by TIE as long as the spectral distribution of the source is known. In 2016, Gureyev *et al.* [349] quantitatively analyzed the effect of partially coherent illumination on TIE phase retrieval based on the CSD analysis. They derived that the effects of both the spectral width and physical size of the source (Schell model) can be interpreted as a convolution on the reconstructed phase distribution, leading to phase blurring. For sources with a relatively high degree of coherence, the convolution effect is not obvious, so the phase distribution can still be accurately recovered. When the coherence of the light source is low, they also proposed to introduce one additional sample-free measurement and use deconvolution algorithm to compensate for the coherence-induced phase blur. In 2010, Zysk *et al.* [350] explicitly considered the spatially partial coherence with use of coherent mode decomposition, revealing that the phase recovered by TIE is a weighted average of the phases of all coherent modes. In 2014, Petrucci *et al.* [351] used the CSD representation to analyze TIE phase

imaging under partially coherent illuminations. They also proposed a double exposure method to compensate for the phase error induced by the spatially partial coherence. Their conclusions are consistent with the results of Streibl [142] and Gureyev *et al.* [352]. In 2015, Zuo *et al.* [208] derived the generalized TIE as an extension of the traditional TIE for partially coherent fields in phase space. In this subsection, we introduce two most representative extensions of TIE under partially coherent illuminations, *i.e.*, partially coherent TIE based on MI/CSD and generalized TIE based on WDF.

7.1.1. Partially coherent TIE based on MI/CSD

From Subsection 6.1.4, we know that the propagation of CSD satisfies the following pair of Helmholtz equations [Eqs. (141) and (142)]

$$\nabla_1^2 W_\omega(\mathbf{x}_1, \mathbf{x}_2) + k^2 W_\omega(\mathbf{x}_1, \mathbf{x}_2) = 0 \quad (277)$$

$$\nabla_2^2 W_\omega(\mathbf{x}_1, \mathbf{x}_2) + k^2 W_\omega(\mathbf{x}_1, \mathbf{x}_2) = 0 \quad (278)$$

The propagation of MI also satisfies similar Helmholtz equations. In the following, we adopt the CSD representation (note that the partially coherent TIE was originally derived by Streibl [142] based on MI in 1984. Then in 2013, Petrucci *et al.* [351] deduced the partially coherent TIE in differential coordinate system based on CSD). Under the paraxial approximation, the Helmholtz equations reduce to paraxial wave equations

$$\nabla_1^2 W_\omega(\mathbf{x}_1, \mathbf{x}_2) + 2jk \frac{\partial W_\omega(\mathbf{x}_1, \mathbf{x}_2)}{\partial z} = 0 \quad (279)$$

$$\nabla_2^2 W_\omega(\mathbf{x}_1, \mathbf{x}_2) - 2jk \frac{\partial W_\omega(\mathbf{x}_1, \mathbf{x}_2)}{\partial z} = 0 \quad (280)$$

In order to simplify Eqs. (279) and (280), the following differential coordinate system is defined

$$\mathbf{x} = \frac{\mathbf{x}_1 + \mathbf{x}_2}{2}; \mathbf{x}' = \mathbf{x}_1 - \mathbf{x}_2 \quad (281)$$

or equivalently,

$$\mathbf{x}_1 = \mathbf{x} + \frac{\mathbf{x}'}{2}; \mathbf{x}_2 = \mathbf{x} - \frac{\mathbf{x}'}{2} \quad (282)$$

In this case, it is not difficult to prove that $\nabla_1^2 = \nabla_x^2 + \nabla_{x'} \cdot \nabla_{x'}$ and $\frac{1}{4} \nabla_{x'}^2$, $\nabla_2^2 = \nabla_x^2 - \nabla_{x'} \cdot \nabla_{x'} + \frac{1}{4} \nabla_{x'}^2$, and we can get $\nabla_1^2 - \nabla_2^2 = 2 \nabla_{x'} \cdot \nabla_{x'}$. Then taking the difference between Eqs. (279) and (280) and simplify the result [142,351]

$$\frac{\partial}{\partial z} W_\omega \left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2} \right) = -\frac{j}{k} \nabla_{x'} \cdot \nabla_{x'} W_\omega \left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2} \right) \quad (283)$$

When $\mathbf{x}' \rightarrow 0$, the LSH of Eq. (283) is just the PSD, thus the following transport equation can be obtained

$$\frac{\partial S_\omega(\mathbf{x})}{\partial z} = -\frac{j}{k} \nabla_{x'} \cdot \nabla_{x'} W_\omega \left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2} \right) \Big|_{\mathbf{x}'=0} \quad (284)$$

Eq. (284) can be considered as the transport of spectrum equation for polychromatic fields. For a quasi-monochromatic field, the spectral density $S_\omega(\mathbf{x})$ just refers to the intensity $I(\mathbf{x})$, and the partially coherent TIE is obtained.

7.1.2. Generalized TIE in phase space

From Subsection 6.2.5, we know that the paraxial propagation of WDF obeys the Liouville transport equation [141]

$$\frac{\partial W_\omega(\mathbf{x}, \mathbf{u})}{\partial z} + \lambda \mathbf{u} \cdot \nabla_{\mathbf{x}} W_\omega(\mathbf{x}, \mathbf{u}) = 0 \quad (285)$$

and the solution takes the form

$$W_\omega(\mathbf{x}, \mathbf{u}, z) = W_\omega(\mathbf{x} - \lambda z \mathbf{u}, \mathbf{u}, 0) \quad (286)$$

where z is the propagation distance. In 2015, Zuo *et al.* [208] integrated on both sides of Eq. (285) over all spatial frequencies \mathbf{u} , and combining the definition of PSD given by Eq. (171), derived that

$$\frac{\partial S_\omega(\mathbf{x})}{\partial z} = -\nabla_{\mathbf{x}} \cdot \int \lambda \mathbf{u} W_\omega(\mathbf{x}, \mathbf{u}) d\mathbf{u} \quad (287)$$

Eq. (287) can be regarded as the transport of spectrum equation for polychromatic fields. It relates the longitudinal evolution rate of the PSD to the transverse divergence of the first frequency moment of the WDF. The time-averaged intensity of a partially coherent beam coincides with the integral of the PSD over all frequencies [256,350]. Hence, we integrate Eq. (287) over ω to obtain the generalized TIE (GTIE) [208]

$$\frac{\partial I(\mathbf{x})}{\partial z} = -\nabla_{\mathbf{x}} \cdot \iint \lambda \mathbf{u} W_{\omega}(\mathbf{x}, \mathbf{u}) d\mathbf{u} d\omega \quad (288)$$

Only the paraxial approximation is employed in deriving Eq. (288), thus it is general enough to cover various optical fields with arbitrary spatial and temporal coherence.

Now let us consider a few important special cases of the GTIE: first, when the field is quasi-monochromatic, *i.e.*, the field consists of almost a single frequency, the spectral density $S_{\omega}(\mathbf{x})$ is simply the intensity $I(\mathbf{x})$. In this case, the field can be regarded as almost completely temporally coherent. Thus, the transport of spectrum equation reduces to the GTIE for spatial partially coherent fields:

$$\frac{\partial I(\mathbf{x})}{\partial z} = -\lambda \nabla_{\mathbf{x}} \cdot \int \mathbf{u} W(\mathbf{x}, \mathbf{u}) d\mathbf{u}. \quad (289)$$

Note that although the temporal coherence of the illumination can be simply incorporated by the integral over all optical frequencies as in Eq. (288), it should be stressed that for dispersive samples, the inherent wavelength-dependent RI often complicates the accurate phase determination. For the remainder of this section, we will drop the spectral dependence ω (assuming quasi-monochromaticity) for simplicity, but it should be noted that for polychromatic fields, the WDF characterizes only a single spectral component of the whole field.

Similar to the traditional TIE, GTIE can also be derived from the energy conservation law of the electromagnetic field. In Subsection 6.2.5, we construct the 3D vector field $j_r = [j_x, j_z]^T$ [303,323], which is also known as the geometrical vector flux [Eqs. (206) and (207)]

$$j_x(\mathbf{u}) = \lambda \int \mathbf{u} W(\mathbf{x}, \mathbf{u}) d\mathbf{u} \quad (290)$$

$$j_z(\mathbf{x}) = \frac{1}{k} \int \sqrt{k^2 - 4\pi^2 |\mathbf{u}|^2} W(\mathbf{x}, \mathbf{u}) d\mathbf{u} \quad (291)$$

The energy conservation in free space implies the geometrical vector flux has zero divergence

$$\nabla \cdot j_r = 0 \quad (292)$$

Then the following continuity equation can be obtained

$$\frac{\partial j_z(\mathbf{x})}{\partial z} = -\nabla \cdot j_x(\mathbf{x}) \quad (293)$$

Under the paraxial approximation, $\sqrt{k^2 - 4\pi^2 |\mathbf{u}|^2} \approx k$. In this case, the axial energy flux reduces to the intensity, given by the space marginal of the WDF

$$j_z(\mathbf{x}) \approx \int W(\mathbf{x}, \mathbf{u}) d\mathbf{u} = I(\mathbf{x}) \quad (294)$$

Substituting Eq. (294) into Eq. (293), the GTIE for partially coherent fields is obtained [208]

$$\frac{\partial I(\mathbf{x})}{\partial z} = -\nabla \cdot j_x(\mathbf{x}) = -\lambda \nabla_{\mathbf{x}} \cdot \int \mathbf{u} W(\mathbf{x}, \mathbf{u}) d\mathbf{u} \quad (295)$$

7.1.3. Generalized definition of “phase” for partially coherent fields

In the previous subsection, we drop the spectral dependence ω by assuming quasi-monochromaticity. However, quasi-monochromatic fields are still not necessarily deterministic due to the statistical fluctuations over the spatial dimension. This randomness can be removed by further limiting the field to be completely spatially coherent as well. Then the field becomes deterministic and can be fully described by the 2D complex amplitude $U(\mathbf{x}) = \sqrt{I(\mathbf{x})} \exp[j\phi(\mathbf{x})]$, where $\phi(\mathbf{x})$ is the phase of the completely (both temporally and spatially) coherent field. From the space-frequency analysis perspective, the completely coherent field

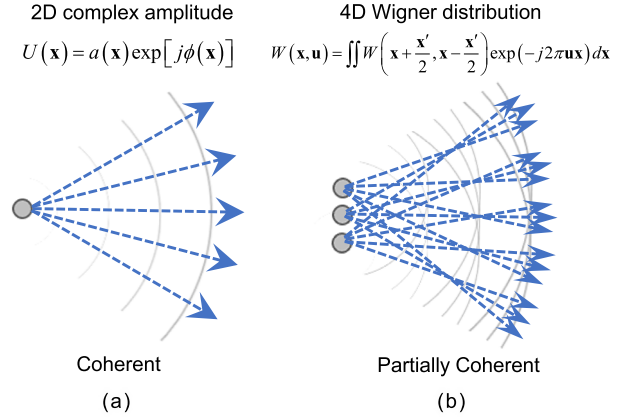


Fig. 61. Schematic of a simplistic view of coherent field and partially (spatially) coherent field. (a) For coherent field, the surface of the constant phase is interpreted as wavefronts with geometric light rays traveling normal to them. It is fully described by 2D complex amplitude; (b) partially coherent field needs 4D coherence function, like the Wigner distribution, to accurately characterize its properties, like its propagation and diffraction. In addition, a partially coherent field does not have a well-defined phase, but rather a statistical ensemble of phases (spatial frequencies, propagation directions) at every position in space.

can be regarded as a mono-component signal, and the first conditional frequency moment of WDF (instantaneous frequency) is related to the transverse phase gradient of the complex field [303,308]

$$\frac{\int \mathbf{u} W(\mathbf{x}, \mathbf{u}) d\mathbf{u}}{\int W(\mathbf{x}, \mathbf{u}) d\mathbf{u}} = \frac{1}{2\pi} \nabla_{\mathbf{x}} \phi(\mathbf{x}) \quad (296)$$

Substituting Eq. (296) into Eq. (295) leads to Teague’s TIE [141]

$$\frac{\partial I(\mathbf{x})}{\partial z} = -\frac{1}{k} \nabla_{\mathbf{x}} \cdot [I(\mathbf{x}) \nabla_{\mathbf{x}} \phi(\mathbf{x})] \quad (297)$$

As discussed earlier, the validity of Teague’s TIE is restricted to fully coherent fields, while GTIE, which is a generalized version of TIE in phase space, explicitly considers the coherence modes so that it can be applied to a much wider range of optical- and electron-beams.

One difficulty in extending GTIE to phase retrieval arises from the fact that the partially coherent field does not have a well-defined “phase” since the field experiences statistical fluctuations over time (Fig. 61). It can be seen from Eq. (296) that GTIE relates the axial intensity derivative with the transverse divergence of the conditional frequency moment of WDF, while the phase gradient (defined as “generalized phase”) of partially coherent fields is automatically connected with the conditional frequency moment of WDF (also known as the instantaneous frequency in the field of signal processing) [238]. It is not difficult to find that the phase-space representation on the left side of Eq. (296) is still valid for partially coherent fields, leading to a new meaningful and more general definition of “phase”. Here we refer the new “phase” defined by Eq. (296) as the generalized phase of partially coherent fields to distinguish it from its coherent counterpart. It can be seen from Eq. (296) that the generalized phase is a scalar potential whose gradient yields the conditional frequency moment of the WDF. It is clear from a distribution point of view that the quantity is the average spatial frequency at a particular location. Besides, the time-frequency joint description characteristic of WDF is very similar to the concept of light in geometric optics. It supposes that light is a description of the amplitude and propagation direction of light energy (Poynting vector). Thus, $W(\mathbf{x}, \mathbf{u})$ can be intuitively interpreted as the energy density of the ray travelling through the point \mathbf{x} and having a frequency (direction) \mathbf{u} . Eq. (286) exactly represents the geometric-optical behavior of a ray traveling through free space and Eq. (294) implies that the intensity at a point is simply the sum of the energy spreading over all possible directions. Similarly, the frequency moment of WDF [Eq. (296)], represents

the transversal ensemble/time-averaged flux vector (transversal time-averaged Poynting vector) [159]. However, it should be noted that the WDF is not a rigorous energy density function (radiance function) due to its possibility of negativity. But it is still very enlightening to use this connection between physical optics and geometric optics. For the cases when the diffraction effect is not obvious or even can be ignored, it provides an accurate and meaningful physical explanation for the light transport of partially coherent fields.

7.2. Phase retrieval under partially coherent illuminations

7.2.1. Phase retrieval based on GTIE

In this subsection, we explore the application of GTIE for phase retrieval under partially coherent illuminations. It should be emphasized that the major concern in such scenario is the well-defined phase shift introduced by the specimen, rather than the generalized phase of the partially coherent field itself. This leads to the natural choice of treating the contributions of the incident illumination and specimen separately by considering the transmitted field as a product of the illumination function $U_{in}(\mathbf{x})$ and the sample transmission function $T(\mathbf{x}) = a(\mathbf{x}) \exp[j\phi(\mathbf{x})]$, where $a(\mathbf{x})$ and $\phi(\mathbf{x})$ are the amplitude and the phase of the specimen. The CSD of the resultant field just leaving the object can be written as $W_{out}(\mathbf{x}_1, \mathbf{x}_2) = T(\mathbf{x}_1)T^*(\mathbf{x}_2)W_{in}(\mathbf{x}_1, \mathbf{x}_2)$. Substituting it into the definition of the WDF and using the convolution theorem of the Fourier transform, we can represent the overall WDF as a convolution of the object transmittance WDF, $W_T(\mathbf{x}, \mathbf{u})$, and the illumination WDF, $W_{in}(\mathbf{x}, \mathbf{u})$, over the spatial frequency variable \mathbf{u}

$$\begin{aligned} W_{out}(\mathbf{x}, \mathbf{u}) &= W_T(\mathbf{x}, \mathbf{u}) \otimes_{\mathbf{u}} W_{in}(\mathbf{x}, \mathbf{u}) \\ &= \int W_T(\mathbf{x}, \mathbf{u}') W_{in}(\mathbf{x}, \mathbf{u} - \mathbf{u}') d\mathbf{u}' \end{aligned} \quad (298)$$

Substituting Eq. (298) into the LHS of Eq. (296) and changing the order of integration, it can then be derived that the generalized phase of the transmitted field, $\hat{\phi}_{out}(\mathbf{x})$, should satisfy the following expression [208]

$$\frac{\int \mathbf{u} W_{out}(\mathbf{x}, \mathbf{u}) d\mathbf{u}}{\int W_{out}(\mathbf{x}, \mathbf{u}) d\mathbf{u}} = \frac{\int \mathbf{u} W_T(\mathbf{x}, \mathbf{u}) d\mathbf{u}}{\int W_T(\mathbf{x}, \mathbf{u}) d\mathbf{u}} + \frac{\int \mathbf{u} W_{in}(\mathbf{x}, \mathbf{u}) d\mathbf{u}}{\int W_{in}(\mathbf{x}, \mathbf{u}) d\mathbf{u}} \quad (299)$$

or equivalently,

$$\nabla_{\mathbf{x}} \hat{\phi}_{out}(\mathbf{x}) = \nabla_{\mathbf{x}} [\hat{\phi}_{in}(\mathbf{x}) + \phi(\mathbf{x})] \quad (300)$$

This representation shows the generalized phase accrues upon propagation through the object, behaving precisely as a conventionally defined phase. The total generalized phase is the sum of the phase of the object and the generalized phase of the incident illumination. In general, the determination of the object phase requires two independent measurements, performed with and without the presence of the specimen. The sample-free measurement is used to characterize $\hat{\phi}_{in}(\mathbf{x})$ of the incident beam and is subsequently subtracted from the total generalized phase $\hat{\phi}_{out}(\mathbf{x})$ to get the net phase introduced by the object only. However, if the illumination is chosen judiciously to directly nullify $\hat{\phi}_{in}(\mathbf{x})$, then

$$\int \mathbf{u} W_{in}(\mathbf{x}, \mathbf{u}) d\mathbf{u} = 0 \quad (301)$$

Eq. (301) is called “zero-moment condition”. When this condition is satisfied, the total generalized phase $\hat{\phi}_{out}(\mathbf{x})$ directly gives $\phi(\mathbf{x})$ and one single measurement is sufficient to retrieval the object phase even though the illumination is not fully coherent. Next, we will discuss two special cases that satisfy Eq. (301). Firstly, for completely coherent illumination, Eq. (301) simply means the wavefront is on-axis plane wave, and the waist of a Gaussian beam. For partially coherent spatially stationary illuminations [303] (discussed in Subsection 6.1.7), which will be generally true for the experimental arrangements in optical microscopy since they typically use Köhler illumination geometry, the spatially incoherent primary source (usually in the condenser aperture plane for an optical microscope) featured by the intensity distribution $S(\mathbf{x})$ and the positional CSD $W(\mathbf{x} + \mathbf{x}'/2, \mathbf{x} - \mathbf{x}'/2) = S(\mathbf{x})\delta(\mathbf{x}')$ is collimated by the

condenser (or simply propagated to the far field), producing the illumination WDF $W_{in}(\mathbf{x}, \mathbf{u})$ just before the object plane [Eq. (200)]

$$W_{in}(\mathbf{x}, \mathbf{u}) = S(\mathbf{u}) \quad (302)$$

Note that the above expression ignores the constant coordinate scaling factor associated with the Fourier transform pair, which is trivial when all computations are carried out in normalized units. Substituting Eq. (302) into Eq. (301) reveals that the primary source distribution must be symmetric about the optical axis, which corresponds to the case discussed by Streibl [142] and Petruccioli et al. [351].

Before proceeding further, we must emphasize that though GTIE is derived in the joint space-spatial frequency domain using WDF, here we do not intend directly to apply GTIE [Eq. (289)] or generalized phase [Eq. (296)] for phase retrieval because WDF is difficult to measure directly, as discussed in Subsection 6.2.6. The main point to be conveyed here is that phase retrieval can be realized by directly applying the original Teague’s TIE for partially coherent fields, by adopting the new GTIE and the generalized definition of “phase” that is valid for partially coherent fields. In other words, for completely coherent fields, if we take the axial intensity derivative and then solve Teague’s TIE, we obtain the phase of the field; for partially coherent fields, by following the same procedure, we obtain the generalized phase of the partially coherent field instead. Since the conditional frequency moments of WDF are additive [Eq. (299)], the gradient of the generalized phase also is additive [Eq. (300)]. Hence, the generalized phase of the transmitted field can be decomposed into the generalized phase of the incident illumination plus the phase shift introduced by the specimen. This decomposition is unique up to an additive constant that may float between the two components (constant phase is unimportant in phase retrieval and can be neglected). GTIE knows nothing about the object phase and the generalized phase of the illumination, and it only retrieves the generalized phase of the total beam passing through the object. However, our objective is to determine the well-defined phase shift introduced by the specimen rather than the “phase” of the illumination or the total transmitted partially coherent field. To solve this problem, one needs to either separate the two terms with two independent measurements explicitly, as suggested in [351,352]. Alternatively, we can directly nullify the “phase” of the illumination. As discussed above, for the coherent imaging, illumination with flat wavefront is required; and for spatially stationary illuminations, the primary source distribution must be symmetric about the optical axis. In fact, for the completely coherent case, such kind of treatment has been habitually adopted in TIE literature. For spatially stationary illuminations, it is also quite easy to realize in practice, e.g., the built-in Köhler illumination in a normal bright-field microscope (of course, the condenser aperture must be properly centered on the optical axis, which is usually clearly explained in the microscope operating manual).

7.2.2. Phase retrieval based on the definition of “generalized phase”

As mentioned above, the monochromatic coherent optical field can be represented by a 2D complex amplitude function, so its 4D phase-space representation is highly redundant. For a slowly varying coherent optical field, the redundancy of phase space becomes more obvious. As shown in Eq. (190), the signal only fills a 2D section in phase space, and the WDF is always strictly greater than 0, i.e., the non-negativity is guaranteed. Therefore, for a slowly-varying coherent field, the diffraction effect can be approximately ignored, and the WDF is equivalent to the energy density function (radiation brightness) of the optical field. When spatial coordinates and spatial frequency coordinates are defined, $W(\mathbf{x}, \mathbf{u})$ represented by Eq. (190) indicates that the light (energy flow) passing through the point only propagates along one direction, which is determined by the phase gradient (normal line). This property enables phase measurement can be achieved by measuring the light direction, which is just the basic idea of the Shake-Hartmann wavefront sensor mentioned in Section 1 [40–42]. Fig. 62 shows a 1D diagram of WDF and field distribution of a smooth coherent wavefront (spherical wave).

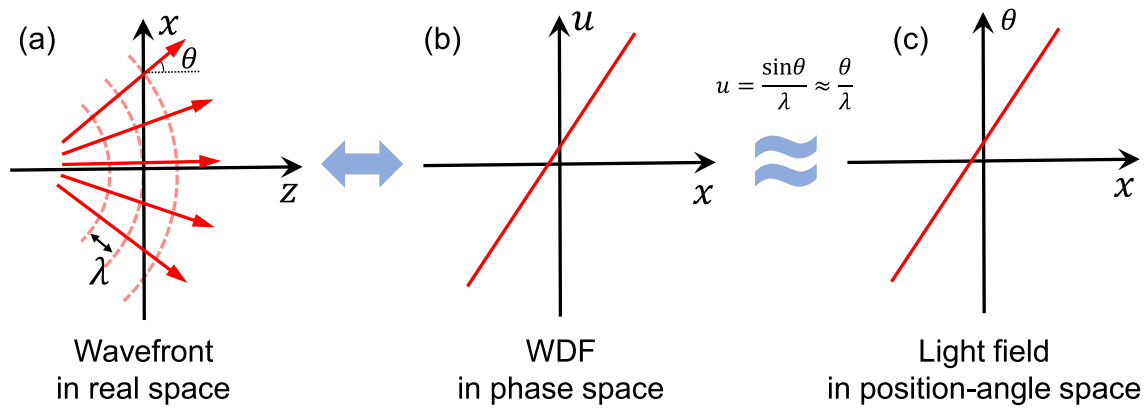


Fig. 62. The WDF and light field of a smooth coherent wavefront. Phase is represented as the localized spatial frequency (instantaneous frequency) in the WDF representation. Rays travel perpendicularly to the wavefront (phase normal). (a) Wavefront in real space; (b) WDF in phase space; (c) light field in position-angle space.

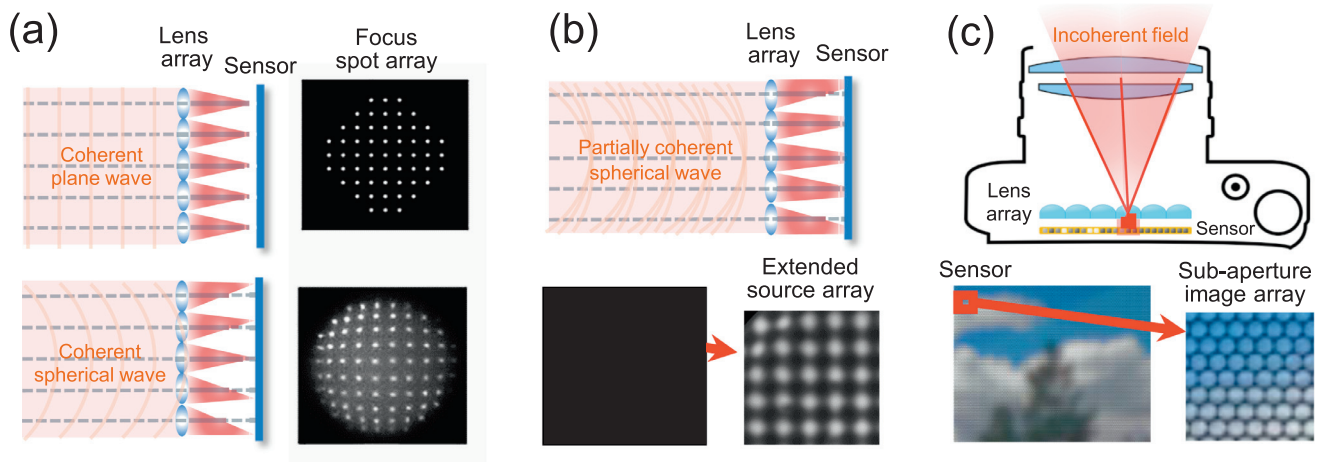


Fig. 63. Principle of the Shack-Hartmann sensor and light-field camera. (a) For coherent field, the Shack-Hartmann sensor forms a focus spot array sensor signal; (b) for a partially coherent field, the Shack-Hartmann sensor forms an extended source array sensor signal; (c) for incoherent imaging, the light-field camera produces a 2D sub-aperture image array.

It can be found that the direction of light propagation is perpendicular to the wavefront (phase gradient direction). Under the paraxial approximation, the relation between the spatial frequency in the WDF and the light propagation angle in the light field can be simply summarized as $\theta \approx \lambda u$, where θ represents the inclined angle of the light along the optical axis. Obviously, since the signal only fills a 2D section in phase space, each microlens array behind the Hartmann wavefront sensor can only approximately collect a highly localized δ function with different offset [Fig. 62(a)], which reflects the high redundancy of the slowly varying coherent optical field in the spatial frequency domain.

It is more complicated when the field is not strictly coherent. Generally, the phase-space WDF constitutes a rigorous and non-redundant representation for partially coherent fields. From the perspective of geometric optics, for each point on the beam, light rays (energy flow) no longer propagate in only one direction. Instead, they fan out to create a 2D distribution, *i.e.*, a 2D sub-aperture image is created behind each microlens in the Hartmann wavefront sensor [Fig. 63(b)], which accounts for the higher dimensionality of a partially coherent field. If the field exhibits significant spatial incoherence, the phase-space negativity and oscillations smooth out, and the WDF again approaches the radiance or the light field. In the computer graphics community, the light-field camera, as a counterpart of the Shack-Hartmann wavefront sensor, also allows joint measurement of the spatial and directional distribution of the incoherent light field [337] [Fig. 63(c)]. The “light field” usually

refers to as a collection of light, represented by $L(\mathbf{x}, \theta)$, where \mathbf{x} is the spatial location, and θ is the angular distribution of all light rays passing through a given point in space. Under the geometric-optics approximation, the WDF is equivalent to radiance [300,301] or light field $L(\mathbf{x}, \theta)$ [$L(\mathbf{x}, \theta) \approx W(\mathbf{x}, \lambda u)$] [335]. Since the intensity and angular distribution of all rays are recorded by light-field imaging, ray-tracing technique can be used to reconstruct synthetic photographs, estimate depth, and change focus or viewing perspectives [353]. This is similar to the case when the 4D coherence function of a partially coherent optical field is retrieved, it allows to control the light propagation and manipulate the optical field numerically in a computer. However, it significantly sacrifices spatial resolution as compared to conventional imaging systems.

As described in Subsection 6.2.6, it is difficult to measure the WDF directly because of its high dimensionality, so it is not recommended to directly use the definition of the generalized phase [Eq. (296)] for phase retrieval. But it must be admitted that it is a very intuitive and interesting approach for phase measurement. As is shown by Eq. (296), the phase of the field (regardless of its state of coherence), which is a scalar potential whose gradient yields the conditional frequency moment of the WDF, can be retrieved from TIE with a minimum of two closely spaced intensity measurements. From Section 6, the 4D WDF can be approximately acquired directly by using a microlens array (the key component of a Shack-Hartmann sensor and light-field camera). Applying the geometric-optics approximation $L(\mathbf{x}, \theta) \approx W(\mathbf{x}, \lambda u)$, we can

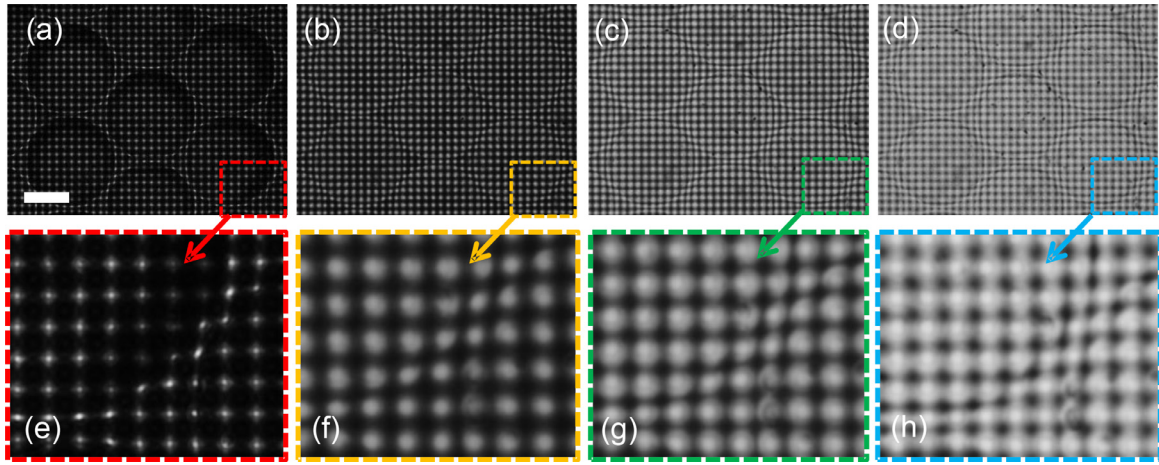


Fig. 64. Images captured with a light-field microscope with different illumination NAs . (a) $NA_{ill} = 0.05$; (b) $NA_{ill} = 0.15$; (c) $NA_{ill} = 0.2$; (d) $NA_{ill} = 0.25$; scale bar length is $50 \mu m$.

describe the phase in terms of the light field [Eq. (296)] can be changed to

$$\frac{\int \theta L(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}}{\int L(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}} = k^{-1} \nabla \phi(\mathbf{x}) \quad (303)$$

the quantity on the left side is just the centroid of the light field, *i.e.*, the average direction of light at one given position. Eq. (303) clearly reveals that a standard TIE measurement can provide important (though not complete) information of the light field, at least its angular marginal and first angular moment. Conversely, it also tells us that the phase gradient can be easily recovered from the 4D light field by a simple centroid detection scheme. This is similar to the standard procedure in the Shack-Hartmann method [40–42].

A simple experiment is presented here to verify the correctness of Eq. (303). The experiment is based on a light field microscope built upon a conventional microscope (Olympus BX41) with a microlens array (pitch $150 \mu m$, ROC $10.518 mm$) inserted in the intermediate image plane just before the camera sensor]. The sample is a plano-convex microlens array (pitch $100 \mu m$), which is imaged with a $20\times$ objective with $NA = 0.4$. Four light field images with the condenser NA from 0.05 to 0.25 are recorded, as shown in Fig. 64. From the enlarged images, we can clearly see the intensity changing corresponding to each lenslet, from a focus spot array to a 2D sub-aperture image array. The centroid for each sub-image is calculated, followed by an integration to reconstruct the phase, as shown in Fig. 65. The results confirm that the phase can be extracted from the light field, even though the illumination is not completely coherent. Note that when the condenser aperture is open up to $NA_{ill} = 0.25$, the sub-aperture images are too large to make them overlap, preventing reliable centroid detection. This results in artifacts in the final retrieved phase image, as shown in Fig. 65(d). Nevertheless, due to limited number of microlens arrays and the size of CCD, the spatial resolution of the final reconstructed phase is quite low. These experimental results prove the correctness of Eq. (303), showing that the quantitative phase information can be directly extracted from the light field (or WDF) based on the definition of “generalized phase”. However, such an approach for phase retrieval is not recommended due to the low spatial resolution and the experimental complexity as compared to the full-resolution, defocus-based TIE techniques.

7.2.3. Influence of the limited aperture of imaging system

Another critical assumption made by TIE is ideal imaging, which is not fulfilled for a practical imaging system. In fact, what we measure is the phase of the field in the image plane, not the phase of the object itself, especially when the pupil of the imaging system is insufficient to transmit all spatial frequencies of interest of the object. In this case,

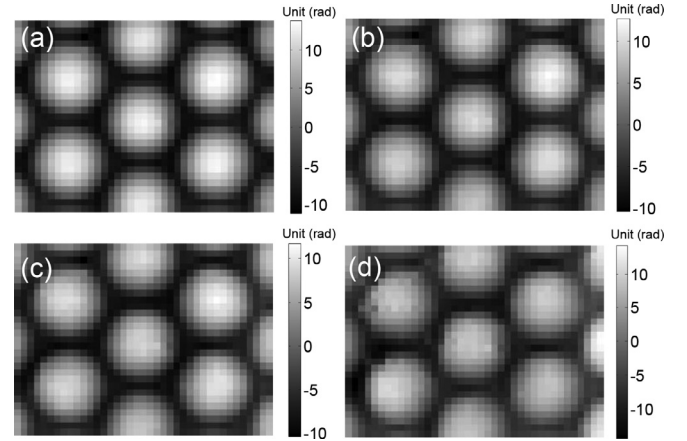


Fig. 65. Phases reconstructed from the light fields with different illumination NAs . (a) $NA_{cond} = 0.05$; (b) $NA_{cond} = 0.15$; (c) $NA_{cond} = 0.2$; (d) $NA_{cond} = 0.25$; scale bar length is $50 \mu m$.

understanding and quantifying the effect of the imaging system appears particularly important. Consider a practical imaging system with a finite aperture, the CSD in the image plane can be written as

$$W_{image}(\mathbf{x}_1, \mathbf{x}_2) = W_{out}(\mathbf{x}_1, \mathbf{x}_2) \otimes_{\mathbf{x}_1, \mathbf{x}_2} h(\mathbf{x}_1, \mathbf{x}_2) \quad (304)$$

The mutual PSF $h(\mathbf{x}_1, \mathbf{x}_2)$ is defined as

$$h(\mathbf{x}_1, \mathbf{x}_2) = h(\mathbf{x}_1) h^*(\mathbf{x}_2) \quad (305)$$

where $h(\mathbf{x})$ is the coherent PSF of the imaging system. Based on the convolution property of WDF, the WDF in the image plane can be written as

$$\begin{aligned} W_{image}(\mathbf{x}, \mathbf{u}) &= \int \Gamma_{image} \left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2} \right) \exp(-j2\pi \mathbf{u} \mathbf{x}') d\mathbf{x}' \\ &= W_{out}(\mathbf{x}, \mathbf{u}) \otimes_{\mathbf{x}} W_{psf}(\mathbf{x}, \mathbf{u}) \\ &= W_T(\mathbf{x}, \mathbf{u}) \otimes_{\mathbf{u}} W_{in}(\mathbf{x}, \mathbf{u}) \otimes_{\mathbf{x}} W_{psf}(\mathbf{x}, \mathbf{u}) \end{aligned} \quad (306)$$

It is seen that the effect of the imaging system is equivalent to convolving the WDF of imaging PSF over the spatial variable \mathbf{x} . More importantly, $W_{psf}(\mathbf{x}, \mathbf{u})$ is zero when \mathbf{u} falls outside of the pupil (in most cases the pupil function is equal to a circ-function, *i.e.*, $P(\mathbf{u}) = 1$, $|\mathbf{u}| \leq u_{NA}$ and $P(\mathbf{u}) = 0$, $|\mathbf{u}| > u_{NA}$), which means all WDF components outside the pupil will be dumped by the imaging system. From the definition of the generalized function Eq. (296), TIE retrieves the conditional frequency

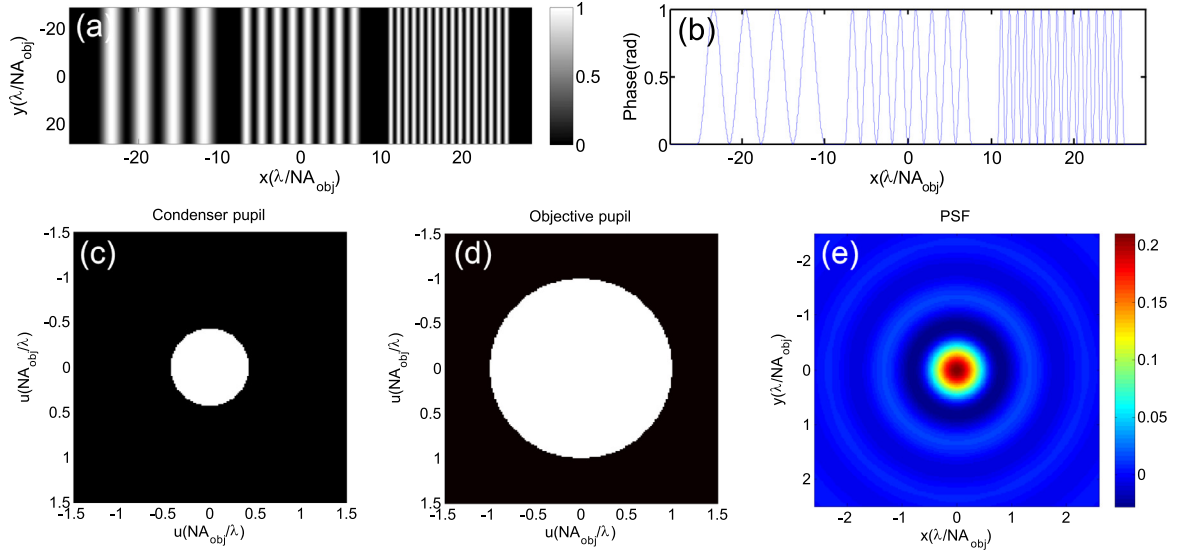


Fig. 66. Numerical simulation on a pure phase sinusoidal grating. (a) Phase distribution of the simulated object; (b) 1D profile of the phase function; (c) Condenser pupil function. (d) Objective pupil function; (e) The coherent PSF of the imaging system. All axes are expressed in the normalized units.

moment of the WDF as the phase gradient, so the reconstructed phase in the image plane is not coincident with the true phase of the object in general [expect for the case of perfect imaging, *i.e.*, $P(\mathbf{u}) = 1$, and the $W_{psf}(\mathbf{x}, \mathbf{u})$ reduces to $\delta(\mathbf{x})$]. Due to the bilinear nature of image formation in partially coherent imaging systems, such kind of phase discrepancy is difficult to analyze or compensate directly. However, we consider a slowly varying specimen under spatially stationary illumination, Eq. (306) can be further simplified as [208]

$$W_{image}(\mathbf{x}, \mathbf{u}) \approx I(\mathbf{x})S\left[\mathbf{u} - \frac{1}{2\pi}\nabla_{\mathbf{x}}\phi(\mathbf{x})\right]|P(\mathbf{u})|^2 \quad (307)$$

Thus, the reconstructed gradient of the generalized phase in the image plane is the frequency centroid of the overlapping area of the shifted primary source and the pupil function. Without considering the effect of the imaging system, the phase gradient is just the centroid of the shifted primary source. As long as the source distribution is symmetric concerning the optical axis, the phase can be accurately retrieved, regardless of the source size (spatial coherence of the illumination). However, in a practical imaging system, it is necessary to give higher importance to the illumination coherence, because the size of the light source has a significant influence on the imaging. Though decreasing the source size does help improve the phase retrieval accuracy (better linear transfer for lower phase gradient), it will compromise the resolution limit. Furthermore, a certain degree of illumination coherence is necessary. For incoherent imaging (the source size is larger than the pupil), the real frequency centroid corresponding to the object phase gradient can never be correctly identified by TIE due to the pupil cut-off. For partially coherent imaging (the source size smaller than the pupil), the imaging system induced phase distortion still exists but can be further compensated, which is discussed in Subsection 7.2.4.

We present a series of numerical simulations to better illustrate the above-mentioned theories. As shown in Fig. 66(a), the simulated object is a pure phase sinusoidal grating with three different periods, $3 \mu\text{m}$, $1.5 \mu\text{m}$ and $0.75 \mu\text{m}$ respectively. To visualize the phase-space quantities more conveniently, the sinusoidal grating is represented by a 1D signal, as shown in Fig. 66(b). The object is illuminated by Köhler illumination with a circular condenser aperture $NA_{ill} = 0.3$, and the illumination intensity uniformly distributed over the aperture plane. The central wavelength of the quasi-monochromatic illumination is $\lambda = 550\text{nm}$. The object is imaged with an objective with $NA_{obj} = 0.7$, thus the coherent diffraction limit of the system is $0.7863 \mu\text{m}$. All computations are

performed in normalized units of λ/NA_{obj} for the space coordinate and NA_{obj}/λ for the spatial frequency coordinate. As shown in Fig. 66(c) and (d), in normalized coordinates, the radii of the $S(\mathbf{u})$ and $P(\mathbf{u})$ are s and 1 , respectively, where s is the coherence parameter. The inverse Fourier transform of the objective pupil gives the coherent PSF of the imaging system, which is shown in Fig. 66(e).

Fig. 67 (a) shows the WDF $W_T(\mathbf{x}, \mathbf{u})$ of the specimen computed from the object transmittance. According to Eq. (296), the phase derivative obtained through TIE is equal to the first conditional frequency moment of the WDF. To verify the accuracy of the phase retrieved by TIE, we compared its derivative with the ideal phase derivative calculated from the original phase function, shown in the normalized range from $[-1, 1]$ [Fig. 67(b)]. The perfect match between the two curves indicates that the phase can reliably be retrieved by TIE for the completely coherent case. Next, we examine the case when the specimen is illuminated by the partially coherent Köhler illumination, but assuming perfect imaging conditions. According to Eq. (298), the WDF at the object plane $W_{out}(\mathbf{x}, \mathbf{u})$ can be calculated by convoluting the WDF of the object transmittance WDF $W_T(\mathbf{x}, \mathbf{u})$ with the illumination WDF $S(\mathbf{u})$, as illustrated in Fig. 67(c). Compared with Fig. 67(a), the blurring of WDF along the frequency dimension is clearly seen in Fig. 67(d). However, due to the symmetry of the condenser aperture $S(\mathbf{u})$, this blurring does not change the frequency centroid of the object WDF. Thus, the phase derivative can still be accurately retrieved, as verified by Fig. 67(e).

Next, the effect of the imaging system is considered. The WDF of the imaging PSF, $W_{psf}(\mathbf{x}, \mathbf{u})$, is shown in Fig. 67(f), and according to Eq. (306) the image plane $W_{image}(\mathbf{x}, \mathbf{u})$ can be calculated by convoluting $W_{out}(\mathbf{x}, \mathbf{u})$ with $W_{psf}(\mathbf{x}, \mathbf{u})$ along the x -direction, as shown in Fig. 67(g). The imaging system removes all the WDF component falling outside of the pupil, which in turn causes blurring in the retrieved phase derivative, as shown in Fig. 67(h). The imaging PSF greatly reduces the phase contrast of the $0.75 \mu\text{m}$ grating, but the phase structure for lower spatial frequencies is less affected. It is instructive to further examine the effect of the illumination coherence (*i.e.*, the effect of gradually changing the condenser aperture) when the imaging PSF is considered. This is also consistent with our theoretical analysis.

Based on GTIE, Zuo *et al.* [208] quantitatively analyzed the image formation and phase retrieval under partially coherent illuminations. Through theoretical analysis and simulations, the following two conclusions were obtained. First, when the primary light source is symmetric about the optical axis, the phase of the object under partially coherent

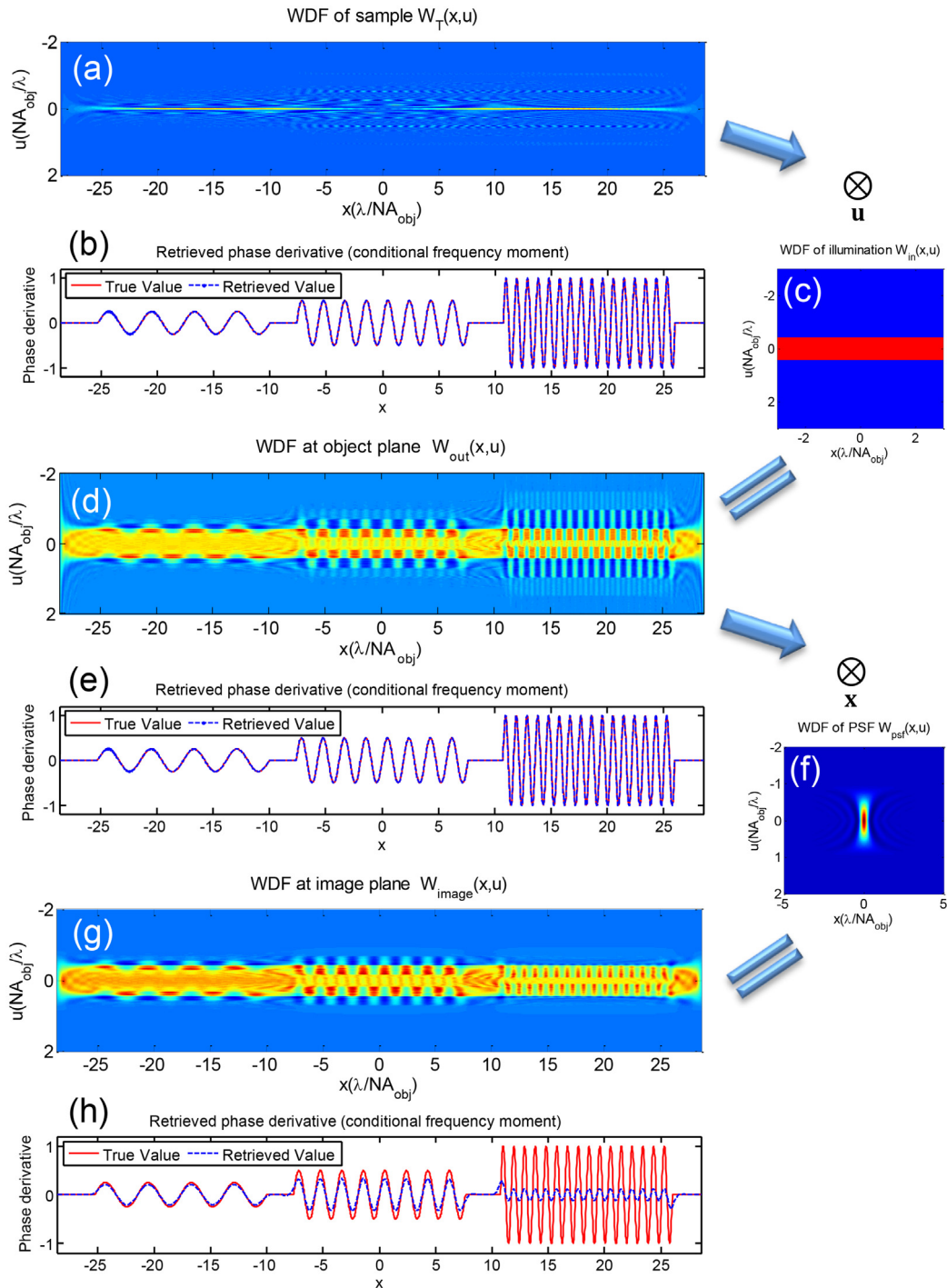


Fig. 67. Phase-space description of the effects of illumination and imaging system on TIE phase retrieval. See the text for details. (a) WDF of the specimen; (b) retrieved phase derivative from (a); (c) WDF of the illumination; (d) WDF of the field in object plane; (e) retrieved phase derivative from (d); (f) WDF of the PSF of the imaging system; (g) WDF of the field in image plane; (h) retrieved phase derivative from (g).

illuminations can be recovered by TIE through one-time measurement. While when the primary source is not axisymmetric, the generalized phase of the illumination can be measured by TIE (without the sample), and then its influence can be compensated by subtracting the generalized phase of the illumination from the measured total phase (with the sample). Secondly, although TIE does not impose any requirements on illumination coherence, for a practical imaging system, a certain level of spatial coherence is indispensable (an extreme case is when $s \geq 1$, the

phase effect completely vanishes), and narrowing down the condenser aperture a bit ($s = 0.3 \sim 0.5$) is indeed conducive to accurate phase retrieval.

7.2.4. Phase gradient transfer function and coherent error compensation

In GTIE, the reconstructed generalized phase gradient just corresponds to the centroid of the shifted primary source. Without considering the effect of lens aperture, reliable phase retrieval can be achieved as

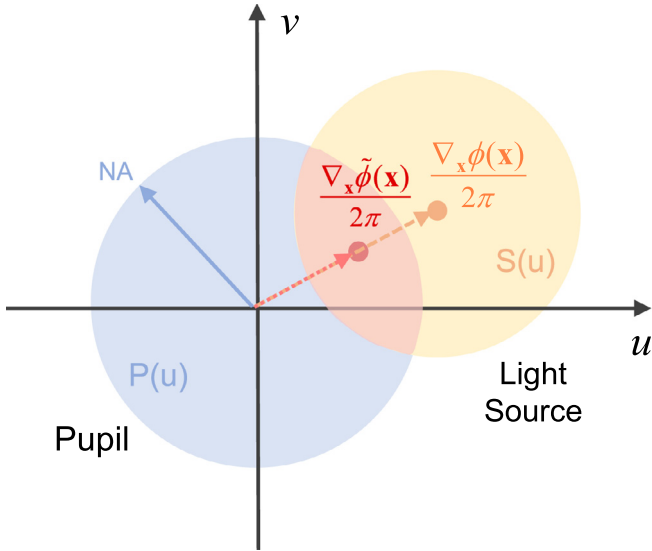


Fig. 68. Geometric interpretation of phase gradient recovered by TIE.

long as the primary source is symmetrical about the optical axis. However, for practical imaging systems, the frequency moment of WDF in the image plane is deviated due to the cut-off effect of the objective pupil, as shown in Fig. 68. The relation between the phase gradient recovered by TIE $\nabla_x \tilde{\phi}(\mathbf{x})$ and the true phase gradient of the object $\nabla_x \phi(\mathbf{x})$ can be represented as

$$\begin{aligned} \frac{\nabla_x \tilde{\phi}(\mathbf{x})}{2\pi} &= \frac{\int \mathbf{u} W_{image}(\mathbf{x}, \mathbf{u}) d\mathbf{u}}{\int W_{image}(\mathbf{x}, \mathbf{u}) d\mathbf{u}} \\ &= \frac{\int \mathbf{u} S[\mathbf{u} - \nabla_x \phi(\mathbf{x})/2\pi] |P(\mathbf{u})|^2 d\mathbf{u}}{\int S[\mathbf{u} - \nabla_x \phi(\mathbf{x})/2\pi] |P(\mathbf{u})|^2 d\mathbf{u}} \end{aligned} \quad (308)$$

To quantitatively characterize the attenuation effect of the imaging system on different phase gradient (corresponding to different spatial frequency) components, we can define a transfer function about the phase gradient, so-called *phase gradient transfer function* (PGTF) [343,354], which is the ratio between the measured phase gradient in the image plane and the ideal phase gradient of the object

$$\begin{aligned} PGTF &= \frac{\nabla_x \tilde{\phi}(\mathbf{x})}{\nabla_x \phi(\mathbf{x})} \\ &= \frac{1}{\nabla_x \phi(\mathbf{x})} \frac{\int \mathbf{u} S[\mathbf{u} - \nabla_x \phi(\mathbf{x})/2\pi] |P(\mathbf{u})|^2 d\mathbf{u}}{\int S[\mathbf{u} - \nabla_x \phi(\mathbf{x})/2\pi] |P(\mathbf{u})|^2 d\mathbf{u}} \end{aligned} \quad (309)$$

Without considering the lens aperture effect ($|P(\mathbf{u})| = 1$), the PGTF is always 1 and the phase gradient can be recovered correctly. However, when the objective aperture is taken into account, the estimated gradient will be smaller than the true value because of the source is truncated by the pupil. The analytical expressions of PGTF for an ax-symmetric imaging system was derived by Sheppard et al. [354] under different illumination conditions. Fig. 69 shows the PGTFs corresponding to different coherence parameters. When $s = 1$, it can be deduced from the geometric relationship between the light source and the pupil function that the estimated gradient is always half of the true phase gradient (PGTF is always 0.5). When $s < 1$, the low-gradient (low-frequency) components corresponding to $0 \sim s$ are consistent with the real value, while the high-gradient (high-frequency) components are underestimated. Meanwhile, it can be observed that there is a trade-off between the resolution and accuracy of the phase reconstruction. Large coherent parameters provides higher cut-off frequencies at the expense of lower PGTF responses. In contrast, a small coherence parameter can provide unbiased estimates for low-gradient phase components,

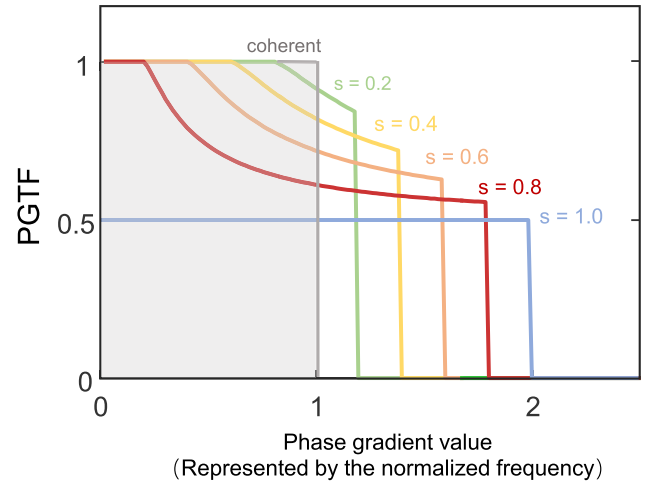


Fig. 69. The phase gradient transfer function under different coherent parameters. Note that the abscissa represents the phase gradient value, but the coordinate unit is the normalized spatial frequency.

but the imaging resolution is compromised. Note that the PGTF shares some similarities with the ratio between the partially coherent WOTF and TIE under weak defocusing [Eq. (60)]. However, they are essentially quite different: PGTF is defined in the spatial domain and represents the transfer characteristics of the spatial gradient of the phase function. In contrast, WOTF and PTF are defined in the frequency domain and represent the transfer characteristics of different frequency components of the phase function. Their establishment conditions are also different: PGTF is based on the slowly varying object approximation, while WOTF is based on the weak object approximation. When $s = 1$, the WOTF becomes the incoherent OTF, and the phase information completely vanishes, while the PGTF still maintains half of the response.

As suggested by Zuo et al. [208,356], the PGTF can be used to compensate the phase error induced by the spatial coherence of the illumination. A lookup table (LUT) can be built based on PGTF to calibrate the phase gradient value recovered by TIE. The basic idea is similar to deconvolution, but note that the compensation is performed on gradient domain rather than Fourier domain. The entries of the LUT is the phase gradient value recovered by solving TIE, and the output of the LUT is the corresponding true phase gradient. The compensated phase gradient is then numerically integrated to obtain the phase that is free from coherence-induced error [356].

7.3. Phase retrieval based on geometric-flow speckle tracking

7.3.1. Phase retrieval under coherent speckle illuminations

“Speckle” is a pattern generated when coherent light or partially coherent light illuminates an object consisting of randomly distributed scatterers [357]. Although speckle should be usually avoided in laser display [358] and coherent optical imaging [359], it has very important applications in many fields, such as speckle imaging in astronomy [360], ESPI in rough surface stress measurement [361], and dynamic speckle imaging in biomedical research [362]. Recently, speckle illumination has been introduced into the field of phase retrieval and QPI. In 2018, Paganin et al. [355] proposed a *geometric-flow speckle tracking* (GFST) method for X-ray phase-contrast imaging based on the concept of geometric energy flow. As shown in Fig. 70, a paraxial forward propagating beam illuminates a thin object and then traverses a short distance z to the planar image sensor. $I_R(\mathbf{x})$ indicates the intensity of the reference speckle, i.e., the image taken in the absence of the object, where \mathbf{x} is transverse coordinates in the plane perpendicular to the optical axis z . The image in the presence of the sample is represented

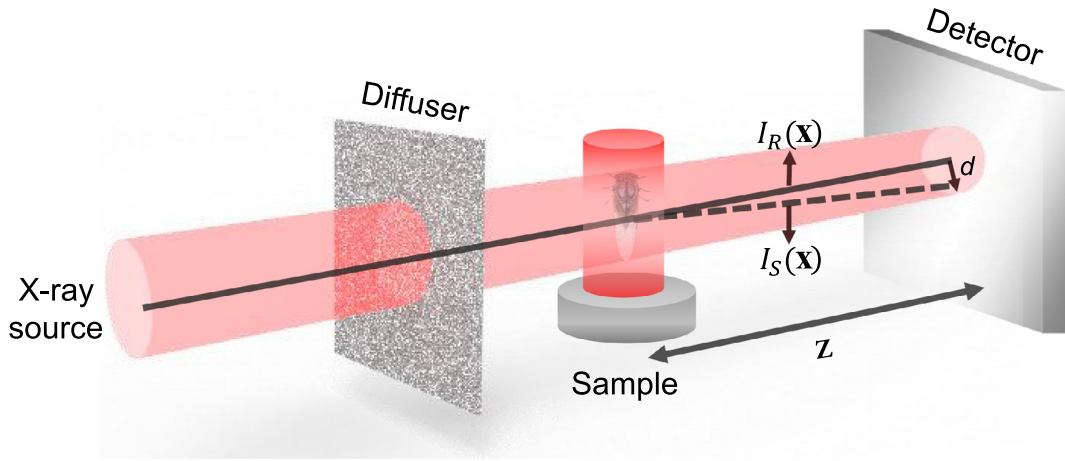


Fig. 70. Experimental setup of the geometric-flow speckle tracking method [355].

as $I_S(\mathbf{x})$. It is a resultant graph of distorted $I_R(\mathbf{x})$ due to the presence of the measured sample, whose non-planar phase geometrically distorts the reference speckles $I_R(\mathbf{x})$.

Assuming that the illumination is an ideal coherent plane wave, the phase of the sample with different surface slopes will induce translation of the speckle pattern, and the speckle displacement vector field $\nabla_{\mathbf{x}}d(\mathbf{x})$ is proportional to the phase gradient $\nabla_{\mathbf{x}}\phi(\mathbf{x})$ [355]

$$\nabla_{\mathbf{x}}d(\mathbf{x}) = \frac{z}{k} \nabla_{\mathbf{x}}\phi(\mathbf{x}) \quad (310)$$

where z is the distance between the sample and the detector. Therefore, based on the relation between $d(\mathbf{x})$ and $\phi(\mathbf{x})$, the phase information of the sample can be obtained by simply calculating the displacement vector field $\nabla_{\mathbf{x}}d(\mathbf{x})$ of the speckle pattern. In the GFST method, the flow is defined as a conserved current associated with the deformation of the reference speckles induced by the phase of the sample. By representing the transverse flow associated with deforming $I_S(\mathbf{x})$ into $I_R(\mathbf{x})$ as the gradient of an auxiliary scalar potential function $\nabla_{\mathbf{x}}\Lambda = I_R D_{\mathbf{x}}$, the displacement field $D_{\mathbf{x}}$ can be obtained as

$$D_{\mathbf{x}}(\mathbf{x}) = \frac{j}{I_R(\mathbf{x})} \mathcal{F}^{-1} \left((u, v) \left\{ \frac{\mathcal{F}[I_R(\mathbf{x}) - I_S(\mathbf{x})]}{u^2 + v^2} \right\} \right) \quad (311)$$

Assuming the field of the flow to be irrotational, the displacement field can be described as the gradient of the scalar potential $d_{\mathbf{x}}(\mathbf{x})$

$$D_{\mathbf{x}}(\mathbf{x}) \equiv (D_x(\mathbf{x}), D_y(\mathbf{x})) \approx \nabla_{\mathbf{x}}d(\mathbf{x}) \quad (312)$$

Thus, the final formula for reconstructing the phase shift $\phi(\mathbf{x})$ based on GFST is

$$\phi(\mathbf{x}) = \nabla_{\mathbf{x}}^{-1} \left(\frac{D_{\mathbf{x}}(\mathbf{x}) \cdot \mathbf{k}}{z} \right) = \frac{k}{z} \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}[(\mathbf{x}_0 + j\mathbf{y}_0) \cdot D_{\mathbf{x}}(\mathbf{x})]}{ju - v} \right\} \quad (313)$$

where \mathbf{x}_0 and \mathbf{y}_0 are unit vectors in x and y directions, respectively. Compare with the FFT-based solver of TIE [Eq. (50)], it can be found that GFST and TIE are essentially based on the same principle [356]. Both of them retrieve the phase information by solving the resultant Poisson equations with the aid of scalar potentials, or so-called Teague's auxiliary function.

7.3.2. Phase retrieval based on partially coherent speckle illuminations

The idea of GFST is not just limited to X-ray diffraction imaging, but also suitable for quantitative phase microscopy under partially coherent illuminations. Recently, Lu *et al.* [356] designed a QPI camera with a weak diffuser (QPICWD) that is compatible with conventional microscope platforms, as shown in Fig. 71. By combining the definition of the generalized phase based on the energy flow (or current density), the GFST was extended to QPI under partially coherent illuminations. As described in Subsection 7.2.2, for each point on the partial coherent fields, light rays (energy flow) no longer propagate in only one direction.

Instead, they fan out to create a 2D distribution. For partially coherent fields, the distortion of the speckle field is a result of the statistical average of many rays passing through each spatial position and deflected by the phase of the object. This is consistent with the generalized phase, and the speckle displacement vector field $\nabla_{\mathbf{x}}d(\mathbf{x})$ is proportional to the generalized phase gradient $\nabla_{\mathbf{x}}\phi(\mathbf{x})$, *i.e.*, the first-order conditional moment of the WDF [356]. The high-resolution speckle distortion field can be obtained by solving TIE, and then the high-resolution phase distribution can be reconstructed by phase integration. However, similar to the case discussed in Subsection 7.2.4, the phase we measured is the generalized phase of the "image" instead of the true phase of the object. Lu *et al.* [356] analyzed the effect of illumination coherence on the resolution and accuracy of phase retrieval, revealing that the sample's phase can be reliably reconstructed when the coherence parameter is between 0.3~0.5. This is consistent with the case of the uniform partially coherent Köhler illumination discussed in Subsection 7.2.3. Similarly, we can also establish a LUT based on PGTF to compensate for the phase blurring induced by the illumination spatial coherence [363].

7.4. Computational light-field imaging based on GTIE

For the TIE phase retrieval, it requires one in-focus image and one additional defocused image to recover the phase information, which seems plausible for the amount of information. Because the complex amplitude is a 2D function, using the 2D defocused intensity information in exchange for the 2D phase information appears "conservative" in the amount of information. Thus, it is obviously highly redundant to represent a 2D coherent field in the 4D phase space. For a slowly varying object, in Subsection 6.2.4, we mentioned that the redundancy of phase space will become more obvious because the signal only occupies a 2D section in phase space [Eq. (190)] [208]

$$W(\mathbf{x}, \mathbf{u}) = I(\mathbf{x}) \delta \left[\mathbf{u} - \frac{1}{2\pi} \nabla \phi(\mathbf{x}) \right] \quad (314)$$

The WDF now becomes nonnegative and takes on all the properties of radiance. It clearly describes that the geometrical light ray at a single position travels only along a single direction described by the phase normal (coincides with the direction of the Poynting vector). It also tells that the total amount of light carried by each ray is described by the intensity of the field. This is an advantageous feature to allow phase measurement simply by measuring the directions of rays, *e.g.*, the Shack-Hartmann sensor [41].

The situation becomes more complex when the optical field is not strictly coherent. Generally, the phase-space WDF constitutes a rigorous and non-redundant description for partially coherent fields. The knowledge of amplitude and (generalized) phase are insufficient to determine

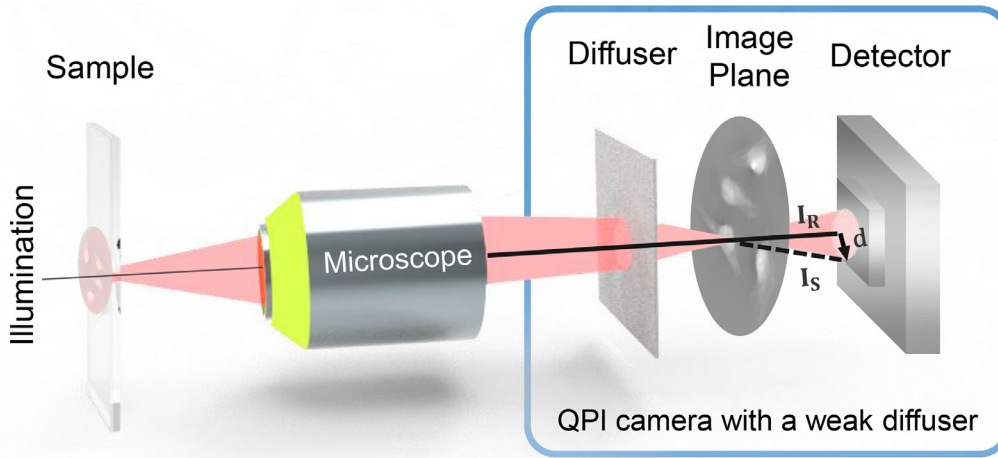


Fig. 71. The schematic diagram of the quantitative phase microscope based on the weak diffuser camera [356].

the full field unambiguously. The measurement and retrieval of the complete 4D coherence function (or equivalent, WDF and AF) are much more complicated (see Subsection 6.2.6 for details). From the geometric-optics perspective, for each point on the beam there exists many rays with different directions; they fan out to make a 2D distribution, which accounts for the higher dimensionality of the partially coherent field. The light-field camera, as a counterpart of the Shack-Hartmann sensor in the computer graphics community, allows joint measurement of the spatial and directional distribution of light [337]. In the optics community, light field is often known as “radiance” in radiometry [300,301]. As early as in 1968, Walther *et al.* [300] introduced the generalized radiance, an equivalent of WDF, to lay a solid wave-optics foundation for radiometry. In 2009, Zhang *et al.* [335] further clarified the equivalence of light field and WDF under geometric-optics approximation, *i.e.*, $L(\mathbf{x}, \boldsymbol{\theta}) \approx W(\mathbf{x}, \lambda \mathbf{u})$ [as shown in Figs. 62 (b) and (c)]. As discussed in Subsection 6.2.6, the light-field camera directly captures the intensity and angular distributions of all light rays, which can also be considered as a coherence measurement technique. However, it requires elaborate optical setups (Fig. 49) and significantly sacrifices spatial resolution (traded for angular resolution) as compared to conventional imaging techniques.

Although TIE cannot recover the complete 4D light field, it does provide important information about the light field. In interferometry, the phase information is completely encoded into the fringe patterns, so the coherent sources are required to produce observable fringes. However, TIE recovers the phase based on intensity propagation, and the intensity is always directly measurable, regardless of the degree of coherence of the optical field. This property makes TIE much easier to implement. For partially coherent optical fields, the phase obtained by TIE is the generalized phase [Eq. (296)] whose gradient is the first-order conditional frequency moment of WDF. Under geometric-optics approximation, the WDF is equivalent to the light field $L(\mathbf{x}, \boldsymbol{\theta}) \approx W(\mathbf{x}, \lambda \mathbf{u})$, and the definition of generalized phase then becomes [208]

$$\frac{\int \boldsymbol{\theta} L(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}}{\int L(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}} = k^{-1} \nabla \phi(\mathbf{x}) \quad (315)$$

The LHS of Eq. (315) is the centroid of the light field, *i.e.*, the weighted average of the directions of light rays passing through a certain point in space. Based on Eq. (315), two conclusions can be drawn [208]. First, the 4D light field includes the 2D phase information, so that the phase gradient can be easily recovered by localizing the centroid of each sub-aperture image. This is similar to the standard procedure in the Shack-Hartmann method [40–42]. Second, TIE cannot recover the entire 4D light field, but the conditional frequency moment (centroid) of the light field can be obtained. In addition, under certain simplified conditions

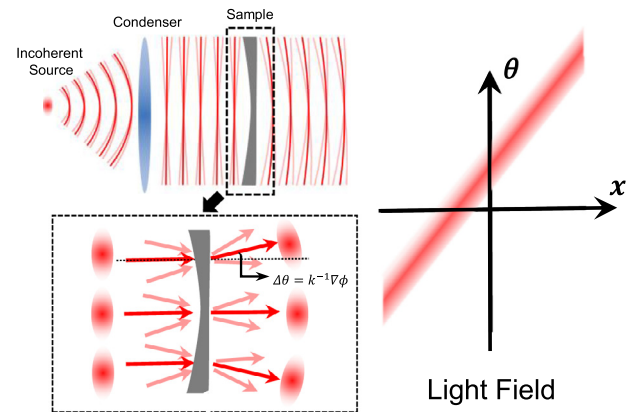


Fig. 72. Light field representation of a slowly varying object under spatially stationary illumination [208]. The sample exhibits angle-shift invariance: at each location, the direction of each incident ray shifts by the amount of object phase gradient.

(slowly varying objects under spatially stationary illuminations), the 4D light field is highly redundant (as shown in Fig. 72, the specimen can be regarded as a “spreadless” or “angular shift-invariant system”: it does not change the angular spread of the incident rays, which is fully determined by the source intensity distribution. On the premise that the light source distribution is known, the 4D light field can be fully characterized by TIE [208].

Under the geometric-optics approximation, $L(\mathbf{x}, \boldsymbol{\theta}) \approx W(\mathbf{x}, \lambda \mathbf{u})$, the physical picture behind Eq. (307) becomes quite clear. The angular distribution of the light field just leaving the object is determined by the source intensity distribution shifted by the amount of the phase gradient of the object. The imaging system only allows rays with the angles within the pupil to pass and blocks the rest with larger angles. The intensity finally captured in the image plane is the sum of all rays passing through the imaging system, determined by the overlapping area of the shifted primary source and the pupil function

$$I_{image}(\mathbf{x}) = I(\mathbf{x}) \int \mathbf{u} S \left[\mathbf{u} - \frac{1}{2\pi} \nabla_{\mathbf{x}} \phi(\mathbf{x}) \right] |P(\mathbf{u})|^2 d\mathbf{u} \quad (316)$$

Eq. (316) can be used to recreate 2D images of the sample from arbitrary perspectives: with the retrieved phase gradient $\nabla_{\mathbf{x}} \phi(\mathbf{x})$, one can simply synthesize different views through Eq. (316) by shifting the position of the primary source artificially. Compared with the method presented in paper [364] (so-called pinhole renderings as in traditional light-field imaging), which constructs the 4D light field first and then extracts its

2D slices as perspective-shifted 2D images, Eq. (316) employs no empirical assumptions and gives a more physically meaningful way for high-resolution view synthesis with the effect of the imaging system taken into account.

In fact, the first attempt to realize computational light-field imaging based on TIE is the “light field moment imaging” (LMI) proposed by Orth and Crozier [364] in 2013. They found that, with use of only a pair of images exhibiting slight defocusing, the first angular moment of the light field can be retrieved by solving a partially differential equation (they did not notice that the equation is precisely TIE). With the moment retrieved, the perspective-shifted views of the original scene can be synthesized. The “moment” in LMI just means that the method can only recover the first angular moment, rather than the complete information of the light field. To realize light-field imaging, Orth and Crozier [364] assumed that the angular distribution of the light field is Gaussian so that the missing light field data can be predicted. The empirical Gaussian angular distribution assumption seems not physically founded, but it does provides satisfactory parallax effect in real experiments. In 2014, Zuo et al. [207] commented that the LMI is actually associated with TIE at the geometric optics limit. Thus, all numerical solutions and axial intensity derivative estimation algorithms in TIE can be directly applied to the LMI, e.g., Liu et al. [365] used the multi-plane high-order finite-difference method to improve the accuracy of axial intensity derivative estimation, providing higher SNR and better visual effect for LMI.

7.5. Phase retrieval based on WOTF deconvolution

As demonstrated in previous subsections, although the definition of phase for partially coherent fields is different from its coherent counterpart, TIE can still be used to recover the phase of the object under partially coherent illuminations. However, it should be noted that GTIE is also based on weak defocusing approximation. However, in a practical microscopic imaging system, a larger defocusing distance is often preferred for better phase contrast, so the effect of defocusing distance on phase reconstruction cannot be simply ignored. In Section 6, we quantitatively analyzed TIE’s PTFs under coherent and partially coherent illuminations (Fig. 59). Compare TIE’s PTF and the WOTF under partially coherent illuminations, it is found that they only overlap at low frequencies. As the degree of coherence decreases, the mismatching becomes more obvious. Therefore, when the phase is directly reconstructed by TIE under partially coherent illuminations, the high-frequency information will be excessively attenuated, resulting in the loss of high-frequency details. To solve this problem, we can use weak object approximation to achieve the linearization between intensity and phase. This method is very similar to the coherent case, the only difference is that the coherent PTF in Subsection 4.3 should be replaced by the partial coherent WOTF obtained in Subsection 6.3.5, and then followed by a deconvolution corresponding to the WOTF to achieve more accurate phase reconstruction. Another major advantage of this method is that the range of linearization can be extended from weak defocusing (near Fresnel zone) to an arbitrary defocus distance.

As discussed in Section 6.3.5, the real and imaginary parts of WOTF corresponds to the ATF $H_A(\mathbf{u})$ and PTF $H_P(\mathbf{u})$, respectively. For an defocused axisymmetric imaging system, it can be found that $H_A(\mathbf{u})$ is an even function of the defocus distance Δz , $H_P(\mathbf{u})$ is an odd function of Δz . If we take two defocused images at opposite and equal defocus distances $\pm \Delta z$ along the optical axis, and calculate their normalized difference in the Fourier domain

$$\frac{\hat{I}_{\Delta z}(\mathbf{u}) - \hat{I}_{-\Delta z}(\mathbf{u})}{4\hat{I}_0(\mathbf{u})} = -\text{Im}[WOTF(\mathbf{u})]\hat{\phi}(\mathbf{u}) \quad (317)$$

where $\hat{I}_0(\mathbf{u}) = a_0^2 TCC(0, 0)$ is the Fourier transform of the intensity distribution at the in-focus plane. $-\text{Im}[WOTF(\mathbf{u})]$ is the PTF of the partially coherent imaging system, $\hat{\phi}(\mathbf{u})$ is the spectrum of the phase to be retrieved. It should be noted that the amplitude information is canceled

in the intensity difference, producing only phase contrast, and the phase retrieval can be realized based on the WOTF deconvolution.

It should be noted that the response curve of the partially coherent PTF $H_P(\mathbf{u})$ tends to be gradually oscillatory as the increase in defocus distance, leading to several zero-crossings (see Fig. 59). This phenomenon also occurs under coherent conditions. The zero-crossings make the deconvolution ill-conditioned, causing severe reconstruction artifacts. In order to avoid the zero-crossings, we can reduce the defocused distance to make $\Delta z \rightarrow 0$ [256,257], thereby mitigating the oscillation of the PTF at high frequencies. However, it can be seen from Fig. 59, when the defocus distance is reduced, the phase contrast is also attenuated, especially for low spatial frequencies, making it difficult to obtain high-SNR phase reconstructions [254,258]. Another solution is to capture more intensity images at multiple defocus distances to synthesize the PTF, just like the coherent case [200,256,259,260]. In this way, more spatial frequencies can be more uniformly covered and zero-crossings can be avoided in the synthesized PTF, so as to reduce the influence of noise and improve the accuracy of phase reconstruction. For example, in 2014, Jenkins et al. [366] proposed a multi-plane partially coherent phase retrieval approach termed multifilter phase imaging (MFPI), which is an extension of the OSF method proposed by Zuo et al. [205] in the case of coherent illumination. MFPI replaces the CTFs in OSF with the partially coherent PTFs and recalculates the cut-off frequencies of the SGDFs with different orders (Fig. 73). In 2015, Jenkins et al. [367] proposed another multi-plane partially coherent phase retrieval method based on WOTF deconvolution. The weighting functions for the intensities at different defocus distances are optimized based on the linear least-square method. It is worth mentioning that the SGDF and the least-square WOTF deconvolution methods share similar ideas. SGDF is essentially a least-square approach in axial derivative estimation (SGDF is the convolution form of the least-square fitting) [205]. The phase reconstructed from each SGDF is calculated based on all the intensity measurements. In the least-square WOTF deconvolution method, one phase reconstruction is based on the intensity difference between a pair of symmetrically defocused images, so it does not make full use of all intensity measurements. Recently, Bao et al. [368] compares and contrasts these two approaches in detail, revealing that the SGDF approach is more accurate than the least-square WOTF deconvolution method but is also slower in computation.

7.6. Resolution enhancement based on coherence engineering

In Subsection 6.3.5, we learned that for a coherent imaging system, the defocus PTF has a strong response, i.e., the captured intensity image can provide a relatively high phase contrast. Nevertheless, the imaging resolution is limited to the coherent diffraction limit. Partial coherent imaging extends the maximum achievable imaging resolution beyond the coherent diffraction limit. The resolution limit of partially coherent imaging is determined by the sum of the NA of the objective lens and NA of the illumination. As the coherence parameter increases, the theoretical imaging resolution also improves. However, it also causes a significant reduction of the response of the PTF, exacerbating the ill-posedness of WOTF deconvolution. Therefore, for a conventional bright-field microscope with a circular condenser diaphragm, in order to achieve a compromise between the imaging resolution and phase contrast, the coherence parameter should be generally set between 0.3 and 0.5 [208]. Although multi-plane approaches utilizing intensity measurements at both small and large defoci allow the response of PTF to be optimized over a wider range of spatial frequency, the noise-to-resolution tradeoff is still not fundamentally solved. As is predicted by the WOTF analysis (Fig. 59), the phase contrast progressively vanishes as the illumination NA approaches the objective NA, suggesting the phase information can hardly be transferred into intensity via defocusing when illumination NA is large. This poses a fundamental obstacle to improving the resolution of TIE phase imaging up to twice of coherent diffraction limit (diffraction limit of incoherent imaging). It should also be mentioned that although synthetic aperture techniques via oblique [369] or struc-

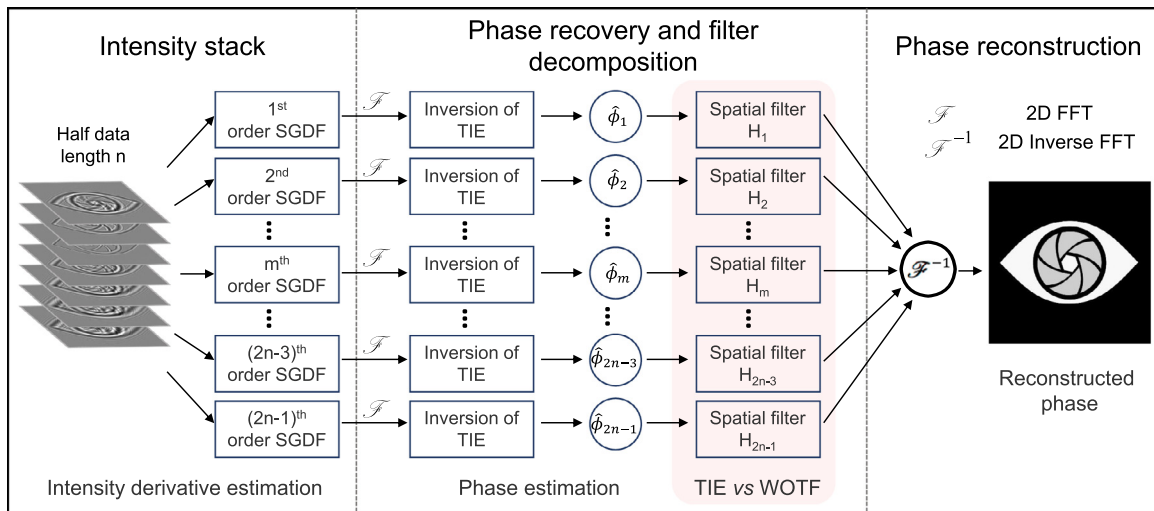


Fig. 73. The block diagram of quantitative phase recovery method based on OFS with multi-plane TIE under partially coherent illuminations.

tured illumination [370,371] have been demonstrated to be possible solutions to enhance the phase imaging resolution of TIE, most of them require relatively complicated optical systems which are not typically available to most bio/pathologists, prohibiting their widespread use in biological and medical science.

In the TIE phase retrieval, the distribution of the PTF is determined by the defocus distance and the coherence parameter. Nevertheless, there is another very important adjustable parameter that is not considered in our previous discussions: the shape of the illumination aperture. The condenser aperture of conventional microscopes is roughly circular in shape, and the spatial coherence of the illumination (coherence parameter s) can be tuned by changing the radius of condenser diaphragm. However, the shape of the condenser aperture is not just limited to circular. Zernike's phase-contrast microscope is a very good example: a specially designed annular diaphragm, which is matched in diameter and optically conjugate to an internal phase plate residing in the objective rear focal plane, is placed in the condenser front focal plane to modulate the illumination aperture. Recently, it has been found that coherence engineering by changing the illumination aperture is generally more effective than changing the defocus distance for optimizing the WOTF [209,210,342,372]. In 2017, Zuo *et al.* [209] suggests to replace the conventional circular illumination aperture with an annular one

$$S(\vec{\rho}) = \begin{cases} 1 & s_1 \leq |\vec{\rho}| \leq s_2 \\ 0 & |\vec{\rho}| < s_1, |\vec{\rho}| > s_2 \end{cases} \quad (318)$$

where s_1 and s_2 are the normalized inner and outer diameters of the annular aperture, respectively. The corresponding PTFs (the imaginary part of the WOTF) with different combinations of s_1 and s_2 can be calculated based on Eq. (250), which are shown in Figs. 74 and 75. Fig. 74 shows the PTFs of annular illuminations with fixed annulus width but different inner and outer diameters. It can be found that when the annular illumination has a smaller inner diameter, the response amplitude of the PTF is large, but the cut-off frequency of the PTF is still limited to near the coherent diffraction limit. However, as the inner and outer diameters of the annular illumination increase, the cut-off frequency of PTF is extended to the incoherent diffraction limit, and the overall curve is gradually moved from the first quadrant to the forth quadrant. This result indicates that an annular illumination source with a maximum outer diameter and inner diameter not only provides non-coherent diffraction-limited imaging resolution but also provides a strong amplitude response in the cut-off frequency passband. In Fig. 75, we further demonstrate the effect of varying the thickness of the annulus by fixing the NA of the outer circles to be 1 ($s_2 = 1$) and only changing the thickness of the annulus ($s_1 = \Delta s$ from 0 to 1). As might be

expected, the phase contrast is reduced as the annulus width increases. When $\Delta s \rightarrow 1$, the phase contrast finally goes to zero, which is just the incoherent case of the circular illumination. It is also shown that the cut-off frequency of the WOTF is reduced from 2 to $2 - \Delta s$ with the increase of the annulus width. From the results shown in Figs. 74 and 75, it can be concluded that we should choose the diameter of the annulus to be equal to that of the objective pupil, and make its thickness as small as possible to optimize both phase contrast and imaging resolution.

In Fig. 76, we compare the magnitudes of the PTFs of the annular illuminations ($\Delta s = 0.01$ and $\Delta s = 0.1$) and circular illuminations ($s = 0.1, 0.75, 0.99$) under weak defocusing ($\Delta z = 0.5 \mu\text{m}$). It can be observed that the annular illumination provides strong responses inside the passband as well as a spatial frequency cut-off near the incoherent diffraction limit ($1.99NA_{obj}$ for $\Delta s = 0.01$, $1.9NA_{obj}$ for $\Delta s = 0.1$). Compared with a conventional microscope with a wide-open circular aperture ($s = 0.75$), the total phase contrast (the area enclosed by the PTF curve and the frequency-axis) provided by the annular illumination is more than doubled (2.35 times for $\Delta s = 0.01$, 2.11 times for $\Delta s = 0.1$). Not only the frequency coverage is extended, but the response in both low and high spatial frequencies is significantly enhanced. The total phase contrast provided by the annular illumination is comparable to that of a conventional microscope with nearly coherent ($s = 0.1$) illumination (91% for $\Delta s = 0.01$, 82% for $\Delta s = 0.1$), but the response is much smoother and more extensive. The spatial frequency cut-off is almost doubled, and the phase contrast of low spatial frequencies is significantly increased. These results demonstrate that replacing the conventional circular aperture with an annular one provides a convenient way to optimize the WOTF for achieving a broadband frequency coverage and enhanced response in both low- and high-frequencies. Moreover, the resulting PTF contains no deep dips and zero-crossings in its pass-band, which removes the ill-posedness of the WOTF inversion. It is expected to achieve high-quality phase reconstruction and overcome the noise-resolution tradeoff in TIE under partially coherent illuminations.

In Fig. 77, we compare the phase retrieval results of annular illumination TIE and circular illumination TIE based on simulations. The Siemens star image is used as an example phase object [shown in Fig. 77(a)] which is defined on a grid with 256×256 pixels with a pixel size of $0.13 \mu\text{m} \times 0.13 \mu\text{m}$. The wavelength of the illumination is 550nm , and the NA_{obj} is 0.80. For such an imaging configuration, the best phase imaging resolution can be achieved is 344nm ($\lambda/2NA_{obj}$), which is also shown in Fig. 77(a). To simulate the noise effect, each defocused image is corrupted by Gaussian noise with a standard deviation of 0.01. Fig. 77(b) compares the defocused images, and the phase retrieval re-

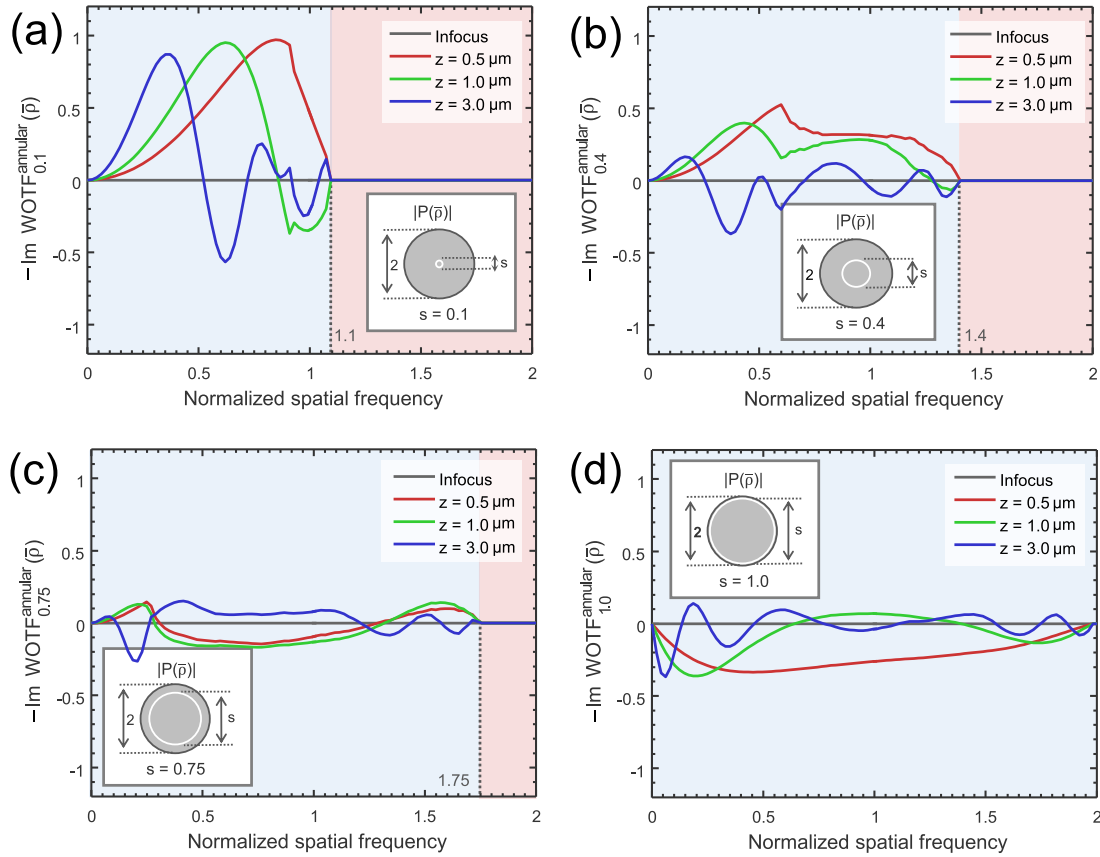


Fig. 74. Phase transfer function of annular illumination apertures with fixed annulus width but different inner and outer diameters at different defocus distances ($NA_{obj} = 0.8$, $\lambda = 550\text{nm}$, the spatial frequency coordinate is normalized against the coherent resolution limit NA_{obj}/λ). (a) $s = 0.1$; (b) $s = 0.4$; (c) $s = 0.75$; (d) $s = 1.0$.

sults of different illumination settings for a small defocus distance ($\Delta z = 0.5\mu\text{m}$). The metric used to measure the accuracy of phase retrieval is given by the root mean square error (RMSE), which quantifies the overall difference between the true phase and the retrieved phase. For the case of circular illumination, the overall phase contrast reduces with the increase in coherent parameter s , which is in coincidence with the WOTF analysis (Fig. 59). The poor response at low-spatial frequencies leads to cloud-like artifacts superimposed on the reconstructed phases. Besides, the phase imaging resolution is improved by opening up the condenser diaphragm (increasing the coherence parameter s). However, for the case of nearly matched illumination ($s = 0.99$), the washout in phase contrast prevents any recognizable phase information to be reconstructed, leading to significant artifacts and a very large RMSE. The phase contrast, especially for low-frequency components can be significantly enhanced by using the annular illumination. The Siemens star appears dark in the defocused image, demonstrating the negative phase contrast as predicted by the theory. The strong phase contrast is finally converted to the quantitative phase images by WOTF inversion, resulting in high-quality reconstructions with a uniform background and improved resolution. The RMSE values for the AI-TIE are comparable with the conventional two-distance TIE approaches and significantly lower than the case when only a single defocus distance is used. Besides, the theoretical resolution for AI-TIE is improved to $1.9NA_{obj}$ for $\Delta s = 0.1$ and $1.99NA_{obj}$ for $\Delta s = 0.01$, which approaches to the incoherent limit [RHS of Fig. 77(a)]. The above simulation results suggest that use of annular illumination matching the objective NA allows for high-quality, low-noise phase reconstruction with a lateral resolution close to the incoherent diffraction limit, providing significant resolution improvement over its circular alternatives.

The above theoretical analysis and experimental results suggest that reshaping of illumination source provides new possibilities to enhance the imaging resolution and improve low-frequency performance of TIE imaging. However, the choice of annular aperture was empirically designed based on intuitive criteria related to the shape of WOTF. It is still unclear whether the annular illumination is the best choice for TIE phase imaging. Due to the complicated form of WOTF and high degree of freedom of the aperture function, solving for an optimum source pattern analytically seems quite challenging. In 2018, Li *et al.* [211] developed a numerical scheme to optimize the illumination pattern based on a combined quantitative criterion for evaluating the “goodness” of an aperture. In order to make the size of the solution searching space tractable, only binary-coded axis-symmetric illumination patterns are considered. Note that the axis-symmetric illumination satisfies the “zero-moment condition” [Eq. (301)], which is a precondition for unbiased TIE phase retrieval under partially coherent illuminations [208]. As shown in Fig. 78, the light source is divided radially, and the pupil of the circular incoherent lighting source is separated into many concentric annuli with equal spacing. The concentric annuli are represented by a 12-bit binary number, and all the illumination source patterns can be indexed by a decimal value of the corresponding binary number.

The PTFs corresponding to different source patterns are calculated. Fig. 79 shows five typical illumination patterns and their corresponding PTFs. The properties of PTF directly determines phase imaging performance, *i.e.*, the cut-off frequency determines the imaging resolution, the number of zero-crossings indicates the ill-posedness of WOTF deconvolution, and the area enclosed by the PTF and the coordinate axis represents the total phase contrast (the information-bearing portion of the image). Therefore, the PTFs are evaluated based on a combined quanti-

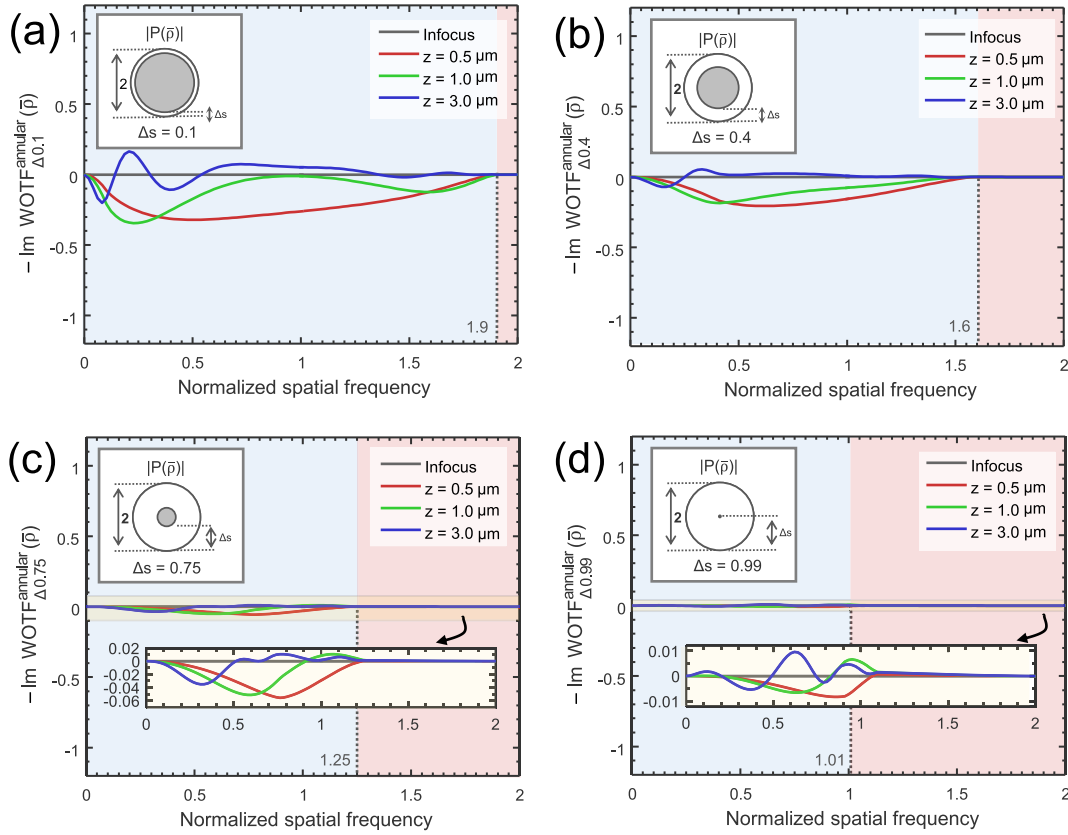


Fig. 75. Phase transfer function of different defocus distances under annular illumination with maximum annular outer diameter but different inner ($NA_{obj} = 0.8$, $\lambda = 550nm$, the spatial frequency coordinate is normalized against the coherent resolution limit NA_{obj}/λ). (a) $\Delta s = 0.1$; (b) $\Delta s = 0.4$; (c) $\Delta s = 0.75$; (d) $\Delta s = 0.99$.

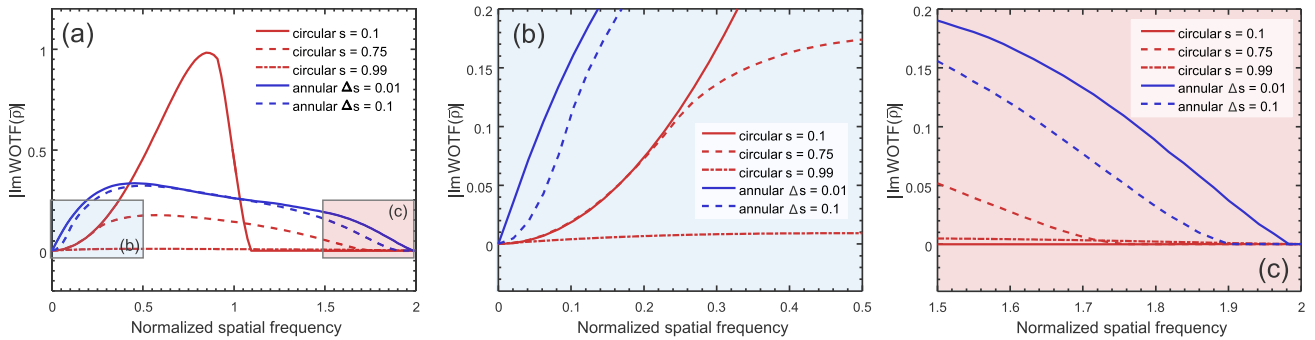


Fig. 76. Magnitude comparison of the phase transfer functions of the annular illuminations ($s = 0.1$ and $s = 0.01$) and circular illuminations ($s = 0.1, 0.75, 0.99$) when the defocus distance is $0.5 \mu m$. (a) Phase transfer function curves of different illuminations; (b) enlarged curve of low frequency component (blue boxed region); (c) enlarged curve of high frequency component (red boxed region).

tative criterion. First, the cut-off frequency of the PTF must reach twice the NA of the objective, which means the final imaging resolution is extended to the incoherent diffraction limit. Then, the PTF does not intersect with the coordinate axis within the cut-off frequency, *i.e.*, there is no zero-crossing in the PTF. Finally, the area enclosed by the PTF should be maximized to guarantee the best SNR. By comparing the PTFs based on the combined criterion, the optimal illumination pattern was identified, which is just a thin annulus matching the NA of the objective, as we mentioned above.

The optimality of the annular illumination is experimentally verified through experiments, as shown in Fig. 80. In order to control the illumination patterns flexibly, the light source of a conventional bright-field microscope is replaced by a programmable LED array. Fig. 80 shows

the illumination patterns, corresponding PTFs, axial intensity derivatives and Fourier spectra, and final reconstructed phases under different illumination patterns. It can be seen that the PTF corresponding to the matched annular illumination provides the highest imaging resolution and strongest phase contrast, especially for low-frequency components. Thus, the corresponding reconstructed phase has the best contrast, resolution, and SNR. Finally, it is worth mentioning that the annular illumination scheme and the multi-plane TIE approaches are not contradictory in principle, and they can be combined to gain complementary advantages. For example, the OSF method and least-square deconvolution method can be extended to annular illumination by simply replacing the coherent CTFs to WOTFs under annular illuminations [373].

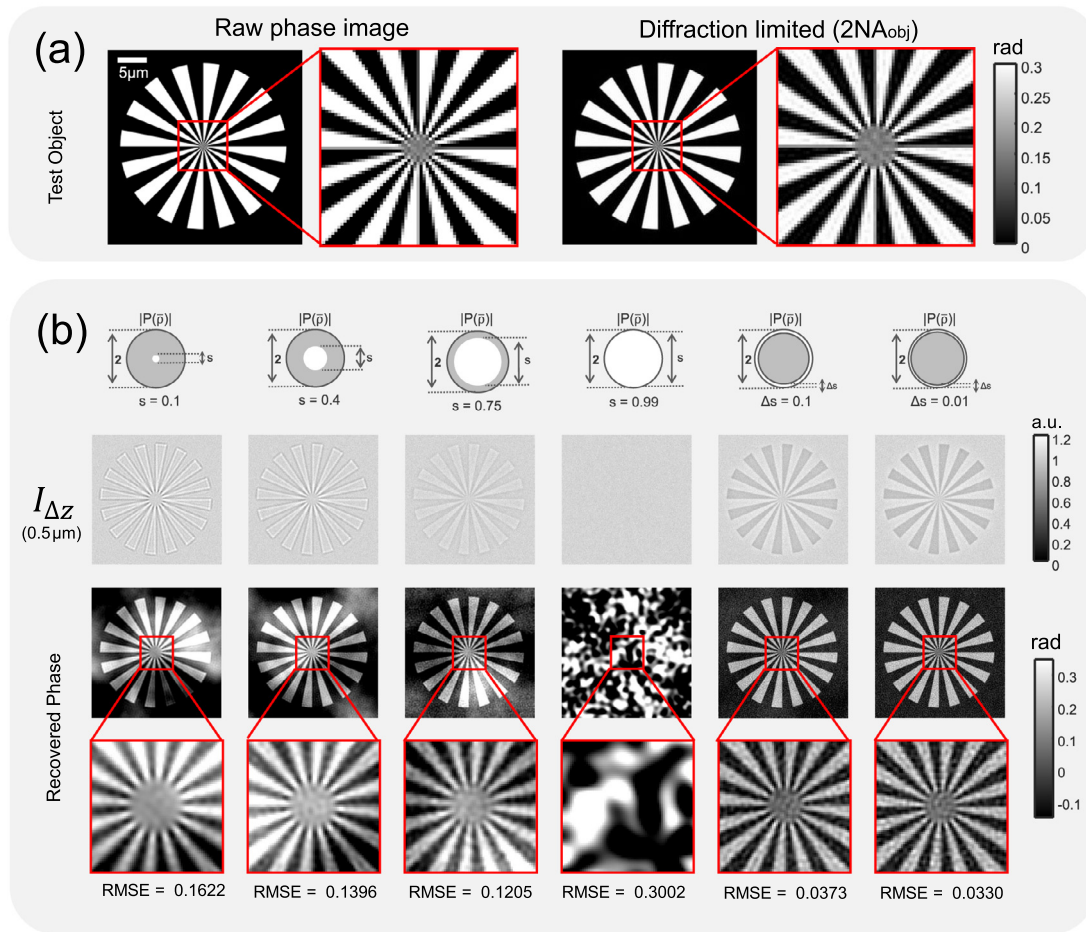


Fig. 77. Comparison between annular illumination TIE and circular illumination TIE. (a) The raw Siemens star image and the corresponding best diffraction limited image can be achieved based on the simulation parameter; (b) Comparison of over-defocus images and reconstruction results of different illumination settings for a small defocus distance ($\Delta z = 0.5 \mu\text{m}$).

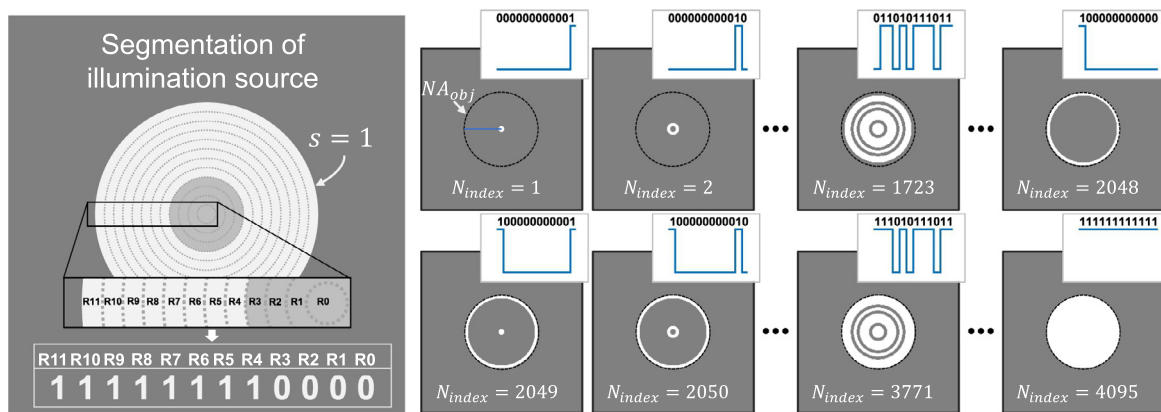


Fig. 78. Illumination source encoding scheme. The circular incoherent source is divided into 12 concentric annuli, which can be indexed by a 12-bit binary number. All possible illumination source patterns can be uniquely indexed by a decimal value of the corresponding binary number.

8. 3D phase imaging under partially coherent illuminations

When we talk about QPI techniques, it is usually assumed that the measured samples are 2D (thin) objects, which can be represented as a 2D complex transmittance function composed of absorption component and phase component. The complex amplitude distribution of the transmitted optical field is the product of the complex amplitude of the incident optical field and that of the object. However, the phase delay in-

duced by the object is actually the axial projection (accumulation) of its 3D RI distribution onto a 2D plane (commonly known as 2.5D imaging), which is an integral along the light propagation direction instead of real 3D information [374–377]. This problem can be effectively overcome by optical diffraction tomography (ODT) techniques [375,376,378,379], which enables high-resolution real 3D (lateral + axial) imaging of the 3D sample by recovering its 3D RI distribution. Conventionally, this method needs to combine the phase measurement technique (e.g., digi-

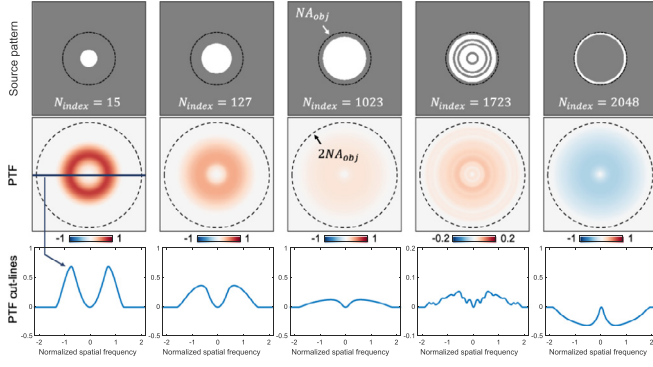


Fig. 79. The phase transfer functions and corresponding 1D cross-sections for 5 different illumination patterns.

tal holography or phase retrieval) with the CT technique: first obtains a set of quantitative phase projections by rotating the object or changing the illumination directions, and then reconstructs the 3D RI distribution based on the CT principle. In recent years, “*transport of intensity diffraction tomography*” (TIDT) has gradually emerged as a new diffraction tomography technique based on the principle of non-interferometric, intensity only measurement inspired by TIE [214,380,381]. Compared with traditional ODT techniques, this method does not require coherent illumination and interferometric measurement. It only needs to capture the through-focus intensity images across the object, and then uses the image deconvolution algorithm to directly retrieve the 3D RI distribution of the object. Thus, it can effectively bypass the difficulties associated with the traditional interferometric diffraction tomography techniques, such as interferometric measurements, object rotation, and beam scanning. This section is devoted to the study of diffraction tomography and TIDT.

8.1. 3D Fourier spectrum and Ewald sphere for coherent fields

In Subsection 2.2, we have learned that any 2D deterministic coherent fields can be decomposed into a superposition of plane waves with different propagation directions, which is known as the angular spectrum representation. Generalizing this rule into 3D space, the complex amplitude distribution of the stationary coherent optical field is assumed to be $U(x, y, z)$, and its Fourier spectrum in 3D space can be connected by the following 3D Fourier transform

$$U(x, y, z) = \iiint \hat{U}_{3D}(u_x, u_y, u_z) e^{j2\pi(u_x x + u_y y + u_z z)} du_x du_y du_z \quad (319)$$

where the exponential primitive $e^{j2\pi(u_x x + u_y y + u_z z)}$ represents a plane wave $\exp(j\mathbf{k} \cdot \mathbf{r})$ propagating in 3D space. Considering an instantaneous shot of a plane wave propagating in 3D space, wave peaks and troughs can be observed, as shown in Fig. 81. The distance from the peak to the trough corresponds to the wavelength of the light wave (when the light propagates in non vacuum, it is the ratio of light wave to RI of medium). The frequency vector is perpendicular to the wavefronts of constant phase, and is thus parallel to the optical direction cosine vector ($\cos \alpha$, $\cos \beta$, $\cos \gamma$), so

$$u_x = \frac{n \cos \alpha}{\lambda}, u_y = \frac{n \cos \beta}{\lambda}, u_z = \frac{n \cos \gamma}{\lambda} \quad (320)$$

where n is the RI of the medium. Due to the fact that the frequency vector is of length n/k , for any frequency vectors, the three frequency components (u_x , u_y , u_z) are not completely independent in the 3D Fourier space and are related by the following formula

$$\sqrt{u_x^2 + u_y^2 + u_z^2} = \frac{n}{\lambda} \quad (321)$$

Eq. (321) suggests that the frequency vectors of plane waves in different directions (spatial frequency) are the same, and they are all located on

a sphere in 3D Fourier space with the radius of $\frac{n}{\lambda}$, the so-called “*Ewald sphere*”. In particular, if the lateral spatial frequency component (u_x , u_y) of the optical field is determined, the z component of the spatial frequency vector can be represented by

$$u_z = \pm \sqrt{\left(\frac{n}{\lambda}\right)^2 - u_x^2 - u_y^2} \quad (322)$$

According to the angular spectrum theory introduced in Subsection 2.2.2, once the complex amplitude of the coherent field $U(x, y, 0)$ in a certain plane is known (without loss of generality, we consider $z = 0$), the complex field $U(x, y, z)$ at an arbitrary distance of z can be determined

$$\begin{aligned} \hat{U}_{2D}(u_x, u_y, z) &= \hat{U}_{2D}(u_x, u_y, 0) e^{jkz \sqrt{1 - (\lambda u_x)^2 - (\lambda u_y)^2}} \\ &= \hat{U}_{2D}(u_x, u_y, 0) e^{j2\pi z \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}} \end{aligned} \quad (323)$$

Taking Fourier transform on z of both sides of Eq. (323) yields

$$\begin{aligned} c\hat{U}_{3D}(u_x, u_y, u_z) &= \int \hat{U}_{2D}(u_x, u_y, 0) e^{j2\pi z \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}} e^{-j2\pi z u_z} dz \\ &= \int \hat{U}_{2D}(u_x, u_y, 0) e^{-j2\pi z \left(u_z - \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}\right)} dz \\ &= \hat{U}_{2D}(u_x, u_y, 0) \delta\left(u_z - \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}\right) \end{aligned} \quad (324)$$

Eq. (324) shows that the coherent optical field is highly redundant in 3D Fourier space, and non-zero values can only be taken on the Ewald sphere. $\delta\left(u_z - \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}\right)$ denotes the projection of a 2D Fourier spectrum onto the Ewald hemisphere surface. The direction of the hemisphere depends on the light propagation direction with respect to z -axis. Therefore, the 2D Fourier spectrum in $z = 0$ plane actually contains all the information about the 3D Fourier spectrum of 3D optical field

$$\hat{U}_{2D}(u_x, u_y, 0) = \hat{U}_{3D}(u_x, u_y, u_z) \Big|_{u_z = \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}} \quad (325)$$

Similarly, the 2D Fourier spectrum in the defocused plane Δz can also be associated with the 3D Fourier spectrum

$$\begin{aligned} \hat{U}_{2D}(u_x, u_y, \Delta z) &= \hat{U}_{3D}\left(u_x, u_y, \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}\right) e^{j2\pi \Delta z \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}} \\ &= \hat{U}_{3D}(u_x, u_y, u_z) e^{j2\pi u_z \Delta z} \end{aligned} \quad (326)$$

8.2. 3D coherent transfer function and generalized aperture

In Section 4, we have learned that the coherent imaging system is linear about the complex amplitude, which is completely determined by its coherent transfer function (the defocused pupil function) [Eq. (63)]. For axisymmetric optical system, the defocused pupil function is presented by Eq. (70)

$$\begin{aligned} H(\rho) &= P(\rho) e^{jk\Delta z \sqrt{1 - \lambda^2 \rho^2}} \\ P(\rho) &= \text{circ}\left(\frac{\rho}{NA/\lambda}\right) = \begin{cases} 1 & \rho \leq \frac{NA}{\lambda} \\ 0 & \text{else} \end{cases} \end{aligned} \quad (327)$$

where $\rho = \sqrt{u_x^2 + u_y^2}$ is the radial spatial frequency, $P(\rho)$ is the objective pupil function with the cut-off frequency of $\frac{NA}{\lambda}$. According to Hankel transform, the defocused PSF of coherent imaging system can be denoted as

$$h(r, z) = \int_{\rho} P(\rho) e^{jkz \sqrt{1 - \lambda^2 \rho^2}} J_0(2\pi r \rho) 2\pi \rho d\rho \quad (328)$$

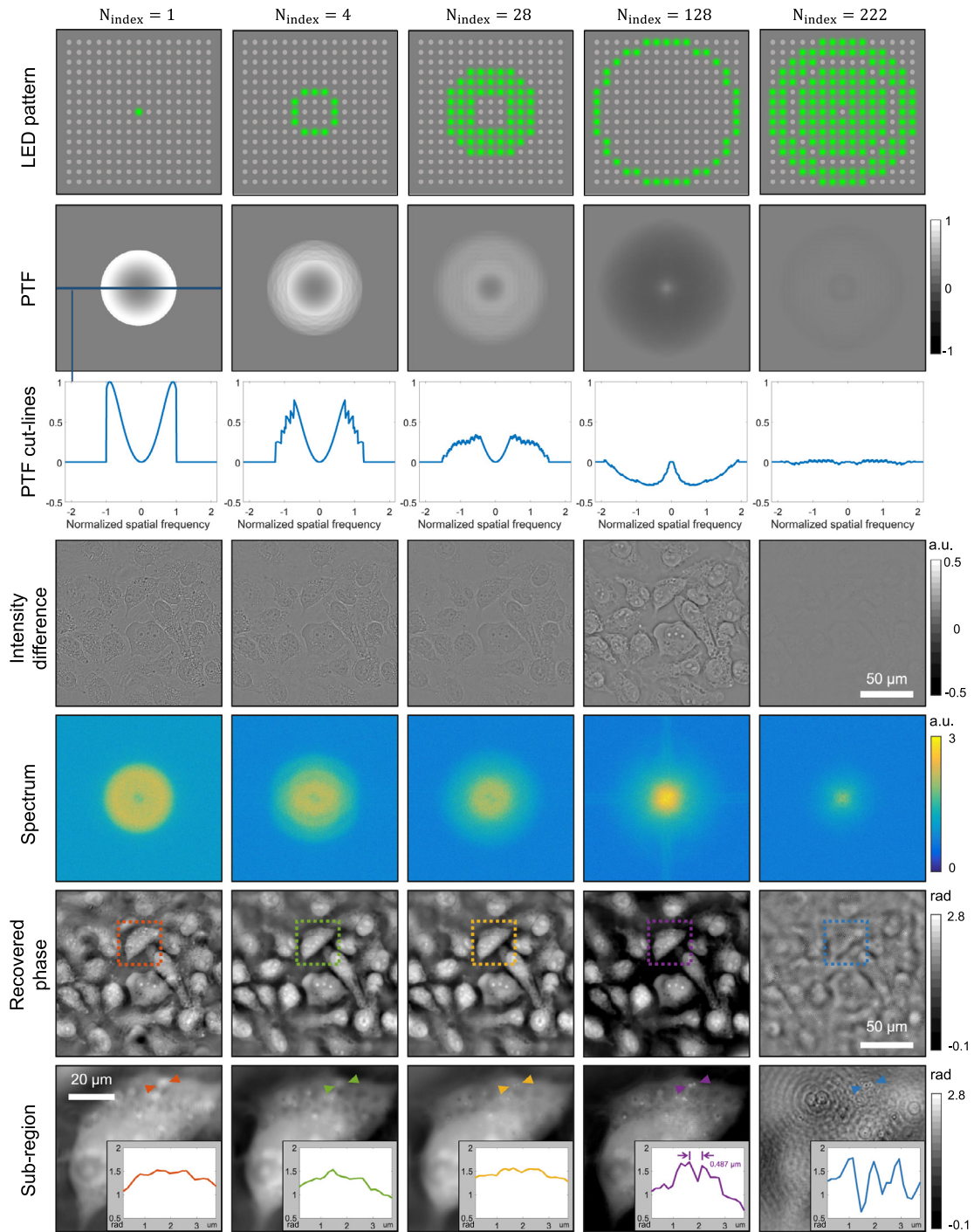


Fig. 80. The illumination patterns, phase transfer functions, axial intensity derivatives and Fourier spectra, and final reconstructed phases under different illumination patterns.

Note here we explicitly write the defocused PSF as a function of r and z in polar coordinates. This is because Eq. (328) actually represents the 3D PSF of the imaging system, which describes the 3D complex amplitude distribution in the image space formed by an ideal point source. Taking Fourier transform about z on both sides of Eq. (328), the 3D coherent transfer function of the coherent optical field can be obtained

$$H(\rho, \eta) = \int P(\rho) e^{jkz\sqrt{1-\lambda^2\rho^2}} e^{-j2\pi zl} dz = P(\rho) \delta\left(l - \sqrt{\left(\frac{1}{\lambda}\right)^2 - \rho^2}\right) \quad (329)$$

Eq. (328) is quite similar to Eq. (324), which indicates that the 3D coherent transfer function of the coherent optical field can only take non-zero value on the Ewald sphere in the 3D Fourier space (which is understandable, because the 3D coherent transfer function itself is the Fourier transform of the 3D PSF, and the 3D PSF is actually a coherent field defined in 3D space). $\delta\left(\eta - \sqrt{\left(\frac{1}{\lambda}\right)^2 - \rho^2}\right)$ represents the projection of the 2D pupil function $P(\rho)$ (coherent transfer function) onto the 3D Ewald sphere surface. The projection direction depends on the light propagation direction with reference to z -axis.

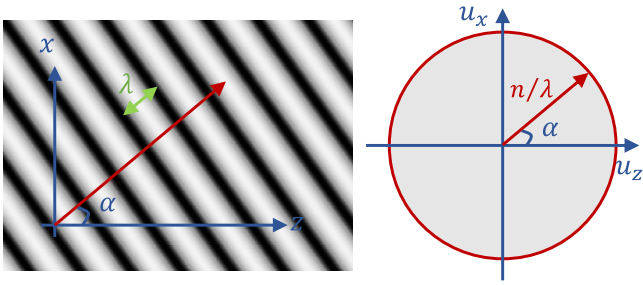


Fig. 81. Description of a monochromatic plane wave in the 3D spatial and frequency space.

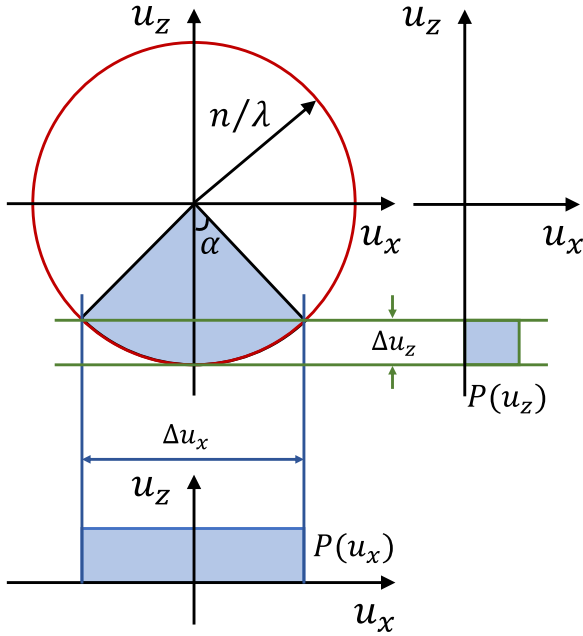


Fig. 82. Generalized aperture $P(\mathbf{u})$ as a 3D frequency spectrum limited by the aperture angle on the Ewald sphere.

Note that in the above derivation, we assume that the light wave propagates in vacuum. For the case that the light wave propagates in the medium (e.g., when an oil immersion objective lens is used), the wavelength in Eq. (328) should be replaced by the equivalent wavelength of the light wave in the medium λ/n , where n is the RI of the medium. In addition, wave vector $(k_x, k_y, k_z) = 2\pi(u_x, u_y, u_z)$ and the spatial frequency vector (u_x, u_y, u_z) differ only by a factor of 2π and are therefore used synonymously. The radius of Ewald sphere in the 3D wave vector space is thus $|\mathbf{k}| = n \frac{2\pi}{\lambda}$ (see Subsection 8.3 for details).

The aperture of the imaging system determines the maximum transverse frequency of the spectrum, and thus, the lateral resolution limit. For coherent imaging, the diffraction limit of the lateral resolution is $\frac{NA}{\lambda}$. However, when imaging thick objects, the limited aperture also affects the diffraction limit of axial resolution (i.e., depth of field). As shown in Fig. 82, the 2D complex pupil function is projected onto the 3D spherical surface, resulting in a segment of the Ewald sphere surface. According to McCutchen [250], this segment of the Ewald sphere is called the *generalized aperture* (i.e., 3D coherent transfer function), and its inverse Fourier transform corresponds to the 3D PSF of the imaging system. The generalized aperture is equivalent to a band-pass filter in 3D Fourier space, whose lateral cut-off frequency corresponds to the coherent diffraction limit $\frac{NA}{\lambda}$, and the axial cut-off frequency is $\frac{n\sqrt{n^2 - NA^2}}{\lambda}$. The aperture angle α denotes the largest cone of wave vectors that can pass through the imaging lens. It should be noted that under the paraxial approximation, the axial cut-off frequency of the generalized aperture

just corresponds to the normalized coefficient of the axial coordinate used in Subsection 6.3.5.

8.3. Scattering potential representation and approximate conditions for 3D objects

In the 2D theory of optical imaging, as described in Sections 4, a thin object is usually represented as a 2D complex transmittance function composed of absorption component and phase component $T(x, y) = a(x, y) \exp[j\phi(x, y)]$. The complex amplitude distribution of the transmitted field is given by the product of the complex amplitude of the incident illumination and the complex transmittance of the object. But when we image a 3D object, how to model the object and describe its interaction with the incident illumination? A straightforward idea is to extend the 2D definition, defining the thick object as a 3D complex transmittance function $T(x, y, z)$ so that the complex amplitude distribution of the transmitted optical field can also be represented as the product of the incident field and the object. This is the main content to be discussed in this subsection.

Generally, the 3D phase imaging of thick object is also known as ODT, which is first proposed by Wolf in 1969 [375]. Considering a 3D sample with RI distribution $n(\mathbf{r})$ illuminated by a monochromatic plane wave travelling in 3D space (x, y, z) , and the RI distribution of the surrounding medium of the sample is n_m . Note that $n(\mathbf{r})$ is a complex distribution, whose real and imaginary parts represent the RI and absorption components of the sample, respectively. Since the RI of the object is different from that of the surrounding medium, the light propagation and scattering obey the following inhomogeneous wave equation

$$[\nabla^2 + k^2(\mathbf{r})]U(\mathbf{r}) = 0 \tag{330}$$

Note that the difference between Eq. (330) and Helmholtz equation [Eq. (2)] in free space (in homogeneous medium) is that the wave number $k(\mathbf{r}) = k_0[n_m + \Delta n(\mathbf{r})]$ is a variable related to the spatial distribution of the RI. Here $\Delta n(\mathbf{r}) = n(\mathbf{r}) - n_m$, k_0 is the wave vector of light wave in free space. $U(\mathbf{r})$ represents the complex amplitude distribution of the total 3D optical field in 3D space. Expanding the LHS of Eq. (330), we obtain

$$(\nabla^2 + k_m^2)U(\mathbf{r}) = -f(\mathbf{r})U(\mathbf{r}) \tag{331}$$

where $k_m = k_0 n_m$ is the wave number in the medium. $f(\mathbf{r})$ is defined as the scattering potential function of the sample

$$f(\mathbf{r}) = k_0^2 [n(\mathbf{r})^2 - n_m^2] \tag{332}$$

Obviously, the scattering potential outside the sample is 0. In fact, the scattering potential of the sample is equivalent to the complex RI distribution, so $f(\mathbf{r})$ is often called the object function. The total field $U(\mathbf{r})$ can be written as the superposition of the incident (un-scattered) field $U_{in}(\mathbf{r})$, and the scattered field $U_s(\mathbf{r})$

$$U(\mathbf{r}) = U_{in}(\mathbf{r}) + U_s(\mathbf{r}) \tag{333}$$

while the incident field satisfies the Helmholtz equation in homogeneous medium

$$(\nabla^2 + k_m^2)U_{in}(\mathbf{r}) = 0 \tag{334}$$

Substituting Eq. (333) into Eq. (331) and using Eq. (334), the scattered field satisfies the following equation

$$(\nabla^2 + k_m^2)U_s(\mathbf{r}) = -f(\mathbf{r})U(\mathbf{r}) \tag{335}$$

The partial differential equation expressed in Eq. (335) cannot be directly solved, but based on Green's method, its solution can be expressed as the following integral form

$$U_s(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') U(\mathbf{r}') d\mathbf{r}' = [f(\mathbf{r}') U(\mathbf{r}')] \otimes G(\mathbf{r}) \tag{336}$$

where $G(\mathbf{r})$ is the 3D Green's function corresponding to Helmholtz equation in homogeneous medium, which is a outgoing spherical wave originates from the point of observation \mathbf{r}

$$(\nabla^2 + k_m^2)G(\mathbf{r}) = -\delta(\mathbf{r}) \quad (337)$$

$$G(\mathbf{r}) = \frac{\exp(jk_m|\mathbf{r}|)}{4\pi|\mathbf{r}|} \quad (338)$$

The convolution integral [Eq. (336)] suggests to decompose the source term $f(\mathbf{r})U(\mathbf{r})$, *i.e.*, the product of the total field and the object function, into the shifted and weighted sum of a delta function $\delta(\mathbf{r})$. And the total scattered field is the superposition of the scattered fields generated by these point sources $\delta(\mathbf{r})$. We can further separate the source function into two terms, which respectively represent the contributions of the incident field and the coupled correlation of the scattered field itself

$$U_s(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')U_{in}(\mathbf{r}')d\mathbf{r}' + \int G(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')U_s(\mathbf{r}')d\mathbf{r}' \quad (339)$$

Though Eq. (339) provides a linearized relation between the scattered field and the scattering potential of the object, it is still quite challenging to invert it directly since $U_s(\mathbf{r})$ appears in the both sides of Eq. (339). However, if we assume that the light is scattered only once by the object, *i.e.*, the light is deflected only by a single interaction, a linearized relation between the resultant the first-order scattered field, U_{s1} , and the scattering potential of the object can be obtained

$$U_{s1}(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')U_{in}(\mathbf{r}')d\mathbf{r}' \quad (340)$$

$U_{s1}(\mathbf{r})$ represents the scattered field produced by the single scattering between the incident field and the object, which is only a portion of the total scattered field

$$U_B(\mathbf{r}) = U_s(\mathbf{r}) - \int G(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')U_s(\mathbf{r}')d\mathbf{r}' \quad (341)$$

Eq. (340) suggests that the first-order scattered field of the entire object is the superposition of elemental scattered fields of all points that constitute the object, *i.e.*, convolution of the Green's function by the product of the object function and the incident field. As is clear here, the solution requires the first-order scattered field $U_{s1}(\mathbf{r})$ either to be a measurable quantity or can be obtained by other means. Two approximations are often used to determine $U_{s1}(\mathbf{r})$:

1) First-order Born approximation [374,375]

When the RI values of the sample and the medium are quite close, the light scattering should be very weak. In this case, it can be assumed that the scattered field is negligible compared with the incident field $U_{in} \gg U_s$ so that the second term in Eq. (341) can be ignored

$$U_s(\mathbf{r}) \approx U_{s1}(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')U_{in}(\mathbf{r}')d\mathbf{r}' \quad (342)$$

Eq. (342) suggests that under the first-order Born approximation, the total scattered field is approximated by the first-order Born scattered field, and the high-order scattered field [the second term in Eq. (339)] is ignored. In this case, there is a linear relationship between the object function and the total scattered field. It should be noted that the first-order Born approximation assumes that the objects are weak scattering. Numerical simulation verified that only if the absorption and total phase delay introduced by the object is small (typically less than π ,) the first-order Born approximation can produce accurate result. When the sample is large or its RI is much higher than that of the surrounding medium, the first-order Born approximation will be no longer valid. In this case, the first-order Rytov approximation is preferred.

2) First-order Rytov approximation [374,382,383]

The first-order Rytov approximation assumes that the total field has a complex phase function related to the scattered field, *i.e.*, $U(\mathbf{r}) = \exp[\phi(\mathbf{r})]$ and $U_{in}(\mathbf{r}) = \exp[\phi_{in}(\mathbf{r})]$. The complex phase of the total field is the sum of the complex phase of the incident field and the complex phase of the scattered field

$$\phi(\mathbf{r}) = \phi_{in}(\mathbf{r}) + \phi_s(\mathbf{r}) \quad (343)$$

Note that the complex phase of the scattered field $\phi_s(\mathbf{x})$ here is different from the "phase" in the conventional sense. Instead, it represents the change of the complex phase of the incident field induced by the object. Since $U_s(\mathbf{r}) = U(\mathbf{r}) - U_{in}(\mathbf{r})$, substituting it into Eq. (343), the scattered optical field can be expressed

$$\begin{aligned} U_s(\mathbf{r}) &= \exp[\phi_{in}(\mathbf{r}) + \phi_s(\mathbf{r})] - \exp[\phi_{in}(\mathbf{r})] \\ &= \exp[\phi_{in}(\mathbf{r})]\{\exp[\phi_s(\mathbf{r})] - 1\} \\ &= U_{in}(\mathbf{r})\{\exp[\phi_s(\mathbf{r})] - 1\} \end{aligned} \quad (344)$$

Then, the complex phase of the scattered field can be represented as follows

$$\phi_s(\mathbf{r}) = \ln\left(\frac{U_s(\mathbf{r})}{U_{in}(\mathbf{r})} + 1\right) = \ln\left(\frac{U(\mathbf{r})}{U_{in}(\mathbf{r})}\right) \quad (345)$$

Based on above expressions, it can be proven that the solution to the inverse scattering problem [Eq. (335)] can be written as the following integral form

$$U_{in}(\mathbf{r})\phi_s(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')\left[U_{in}(\mathbf{r}') + |\nabla\phi_s(\mathbf{r}')|^2\right]d\mathbf{r}' \quad (346)$$

Similarly, the first-order Rytov approximation is introduced to make this problem linearized and solvable. When the phase gradient introduced by the object is small so that the following slowly varying condition is satisfied

$$n_\delta \gg |\nabla\phi_s|^2\left(\frac{\lambda}{2\pi}\right)^2 \quad (347)$$

where n_δ is the RI variation of the object within the range of one wavelength. Then, the contribution of the complex phase gradient term in Eq. (346) can be safely ignored for small n_δ , and Eq. (346) can be simplified as

$$U_{in}(\mathbf{r})\phi_s(\mathbf{r}) \approx U_B(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')U_{in}(\mathbf{r}')d\mathbf{r}' \quad (348)$$

It can be found that the complex phase of the scattered field can be interpreted as the modulation to the incident field introduced by the object $\phi_{s1}(\mathbf{x}) \approx U_{in}(\mathbf{r})/U_B(\mathbf{r})$. The difference between Eq. (348) and Eq. (342) is only that the approximation imposed on the first-order scattered field $U_{s1}(\mathbf{r})$. It is not difficult to prove that Rytov approximation can be reduced to Born approximation when the scattered field is weak or the total phase delay induced by the object is small [374]. Since the validity of the Rytov approximation is not dependent on the absolute phase change introduced by the sample, but on the gradient of the RI within the sample [374,384], this makes the Rytov approximation more suitable to thick biological samples than the Born approximation [374,384–386].

8.4. Fourier diffraction theorem

In the previous subsection, we learned that the 3D object is modeled as the scattering potential satisfying the differential equation of Eq. (335). Under Born or Rytov approximation, this equation can be solved linearly, and the solution is represented as the integral form expressed by Eq. (340). In this subsection, we will discuss the physical meaning behind Eq. (340) and how to use it to realize ODT for 3D objects. First, rewriting the RHS of Eq. (340) into the convolution form

$$U_{s1}(\mathbf{r}) = G(\mathbf{r}) \otimes [f(\mathbf{r})U_{in}(\mathbf{r})] \quad (349)$$

Taking Fourier transform on the both sides of Eq. (349) yields

$$\hat{U}_{s1}(\mathbf{k}) = \hat{G}(\mathbf{k})\{\hat{f}(\mathbf{k}) \otimes \hat{U}_{in}(\mathbf{k})\} \quad (350)$$

Here we use the wave vector $\mathbf{k} = (k_x, k_y, k_z)$ to represent the frequency coordinates in 3D Fourier space, and its modulus is the wave number in the medium $|\mathbf{k}| = k_m$. In general, the incident field is a plane wave and can be expressed as $U_{in}(\mathbf{r}) = \exp(j\mathbf{k}_i \cdot \mathbf{r})$, where \mathbf{k}_i is the incident wave vector of plane wave. Since the image sensor is a 2D detector array, we

can only measure the field distribution on a plane. As we discussed at the beginning of this section, the Fourier spectrum of the 3D coherent optical field is highly redundant, and the complex amplitude distribution on a single plane is sufficient to fully characterize the whole 3D field. It should be remembered that Green's function in Eq. (350) serves as a probing function which originates from the point \mathbf{r} under consideration, inside the volume. We first implement the 3D angular spectrum decomposition (Fourier transform) of Green's function [374], and then take the 1D inverse Fourier transform with respect to z . Finally, assuming that the detection plane is located at $z = z_D$, and only the forward propagation component of the scattered field can be detected (neglecting the back-scattering), we can get the following expression

$$\hat{U}_{s1}(\mathbf{k}_\perp, z_D) = -j \frac{e^{jk_z z_D}}{k'_z} \hat{f}(\mathbf{k}_\perp - \mathbf{k}_{i\perp}, k'_z - k_{iz}) \Big|_{z_D \geq 0} \quad (351)$$

Here the transverse frequency coordinate is defined as $\mathbf{k}_\perp = (k_x, k_y)$, and we define a new quantity $k'_z = \sqrt{k_m^2 - k_x^2 - k_y^2}$, which is a propagation constant and only related to the transverse frequency k_x, k_y , to distinguish from the axial frequency coordinate k_z . $\hat{U}_{s1}(\mathbf{k}_\perp, z_D)$ is the 2D Fourier spectrum $\hat{U}_{s1}(k_x, k_y, z_D)$ of the first-order scattered field at the detection plane $z = z_D$. The exponential term $e^{jk_z z_D}$ just corresponds to the angular spectrum propagation kernel for a coherent complex field [according to Eq. (326)], which accounts for the coordinate shift in the z direction. Note that this exponential term will automatically vanish if the measurement is performed at the nominal "in-focus" plane ($z = 0$). Because the Fourier spectrum of the 3D coherent field can only assume non-zero values on the Ewald sphere and the object function on the RHS is also defined in the 3D Fourier space, in order to show the physical meaning behind Eq. (351) more clearly, we can also explicitly write the LHS of Eq. (351) as a 3D function, which corresponds to the projection of the 2D spectrum onto the 3D Ewald sphere

$$\hat{U}_{s1}(\mathbf{k}_{s1}, z_D) = j \frac{\exp(jk_z z_D)}{k_z} \hat{f}(\mathbf{k}_{s1} - \mathbf{k}_i) \quad (352)$$

Eq. (352) is known as the Fourier diffraction theorem [374,375], which relates the first-order scattered field at the detection plane ($z = z_D$) to the scattering potential in the Fourier space. Here $\mathbf{k}_{s1} = (k_x, k_y, k_z = \sqrt{k_m^2 - k_x^2 - k_y^2})$ represents the 3D spatial frequency coordinate of the first-order scattered field. It should be noted that the the first-order scattered field is defined on the Ewald sphere, and k_z is actually dependent on (k_x, k_y) . $\hat{f}(\mathbf{k}_{s1} - \mathbf{k}_i)$ on the RHS represents that the spectrum of the scattered field \mathbf{k}_{s1} is given by the object frequency spectrum shifted by the frequency of the incident field \mathbf{k}_i

$$\mathbf{g} = \mathbf{k}_{s1} - \mathbf{k}_i \quad (353)$$

A consequence of Eq. (353) is the well-known Laue equation [220] with \mathbf{k}_{s1} being the scattered and \mathbf{k}_i being the incident wave vector. The Laue equation results from the the assumption of single scattering. According to the Laue equation the scattered wave vector \mathbf{k}_{s1} can be determined from the sum of an incident wave vector \mathbf{k}_i and a wave vector \mathbf{g} from the scattering object. In the far-field approximation, i.e., if evanescent waves can be neglected, the transfer function of free space has also to be considered. This means that only the frequencies \mathbf{k}_{s1} on the Ewald sphere can propagate over distances large in comparison to the wavelength. Fig. 83 illustrates this in a graphical way. \mathbf{k}_{s1} represents the Ewald sphere of the first-order scattered field, \mathbf{k}_i is the incident vector. The propagating contributions to the scattered field are limited to vectors on the Ewald sphere. The wave vector \mathbf{g} is the accessible frequency component of the object, which is a shifted Ewald sphere passing through the origin with center $-\mathbf{k}_i$. Of course, here we ideally assume that we can capture the transmitted and reflected first-order scattered field from both sides of the object. Due to the technical layout, as for conventional microscopic imaging in transmissive geometry, the reflected parts are mostly lost,

and only forward scattering can be received [see red circle in Fig. 83(a)]. In addition, we can change the direction of the wave vector \mathbf{k}_i by varying the illumination angle. If the object is illuminated from all possible directions in 3D space (360° surround), it can be imagined that the maximum frequency coverage of the object becomes a sphere with a radius of $2k_m$. We call this large sphere *Ewald limiting sphere* in the object space (which has a doubled radius compared with the Ewald sphere of the coherent field), as illustrated in Fig. 83(c) [387].

8.5. 3D diffraction tomographic imaging under coherent illumination

The Fourier diffraction theorem [Eq. (352)] provides a powerful tool for solving the inverse scattering problem and recovering the quantitative 3D RI distribution. When the sample is illuminated by a plane wave from a certain direction, the complex amplitude of the forward transmission component of the first-order scattered field is acquired. Taking Fourier transform, we obtain a portion of object spectrum on the Ewald sphere. The complex amplitude of the incident plane wave illumination $U_{in}(\mathbf{r})$ can be calibrated or predefined, and the complex amplitude of the total optical field in a certain plane $U(\mathbf{r})$ is measured via interferometry or phase retrieval approaches. Thus, the first-order scattered field can be determined based on two approximations described in the previous subsection

$$U_B(\mathbf{r}) \approx \begin{cases} U_s(\mathbf{r}) = U(\mathbf{r}) - U_{in}(\mathbf{r}) & \text{Born} \\ U_{in}(\mathbf{r})\phi_s(\mathbf{r}) = U_{in}(\mathbf{r}) \ln\left(\frac{U(\mathbf{r})}{U_{in}(\mathbf{r})}\right) & \text{Rytov} \end{cases} \quad (354)$$

After obtaining $U_{s1}(\mathbf{r})$, its planar 2D Fourier spectrum $\hat{U}_{s1}(\mathbf{k})$ is projected onto a semi-spherical surface according to the Laue equation as depicted in Fig. 83. By changing the incident angle of the illumination \mathbf{k}_i , the resultant first-order scattered field can access different regions of the 3D Fourier spectrum of the object, as illustrated in Fig. 85. With a series of angle-dependent complex amplitude measurements, a certain portion of the 3D Fourier spectrum of the object can be completed, and an estimate of $\hat{f}(\mathbf{k})$ can be obtained. Finally, an inverse 3D Fourier transform of the resultant Fourier spectrum produces the 3D complex RI tomogram of a sample. When the incident illumination covers all possible directions ($\pm 90^\circ$), the frequency coverage of the scattering potential is shown in Fig. 85(a). It can be found that only a limited region of the spectrum can be accessed, which results from the shifted Ewald sphere through the origin with center $-\mathbf{k}_i$ rotating $\pm 360^\circ$ along the z axis. Although the extension in the lateral direction reaches the border of Ewald limiting sphere, a missing cone appears in the axial direction.

Additional limitations of the frequency coverage emerge from limited aperture angles or numerical apertures of the illumination optics and the objective, respectively. For on-axis plane wave illumination, the accessible object frequency just corresponds to the support of the generalized aperture, with a lateral frequency extent $\Delta u_x = \frac{NA}{\lambda}$ and an axial extent $\Delta u_z = \frac{n - \sqrt{n^2 - NA^2}}{\lambda}$, as shown in Fig. 82. The aperture angle corresponding to the largest cone of wave vectors that can be illuminated and detected is $\text{arc sin}\left(\frac{NA}{n}\right)$. For matched illumination, i.e., the maximum illumination angle allowed is limited to the maximum collection NA, a torus-shaped structure in 3D frequency space is formed [Figs. 85(b) and (d)]. Only those object frequencies can be transferred as a result of the limited illumination and objective apertures. The resulting accessible volume in frequency space represents the 3D transfer function of tomographic imaging. It describes the possible object frequencies contributing to the image formation, and can be used to assess both lateral and axial resolution limit quantitatively. Incomplete illumination angle leads to larger missing cone region, as shown in Fig. 85(c). The missing cone will reduce of the resolution of the reconstructed RI, especially the axial resolution [376,378]. For partially coherent optical imaging, the object is illuminated by a limited angular spectrum. Only the information for a finite angular range is gathered. So, besides the low-pass filtering by the imaging system, the finite illumination aperture has also to be taken into account, which will be discussed later in Subsection 8.8.

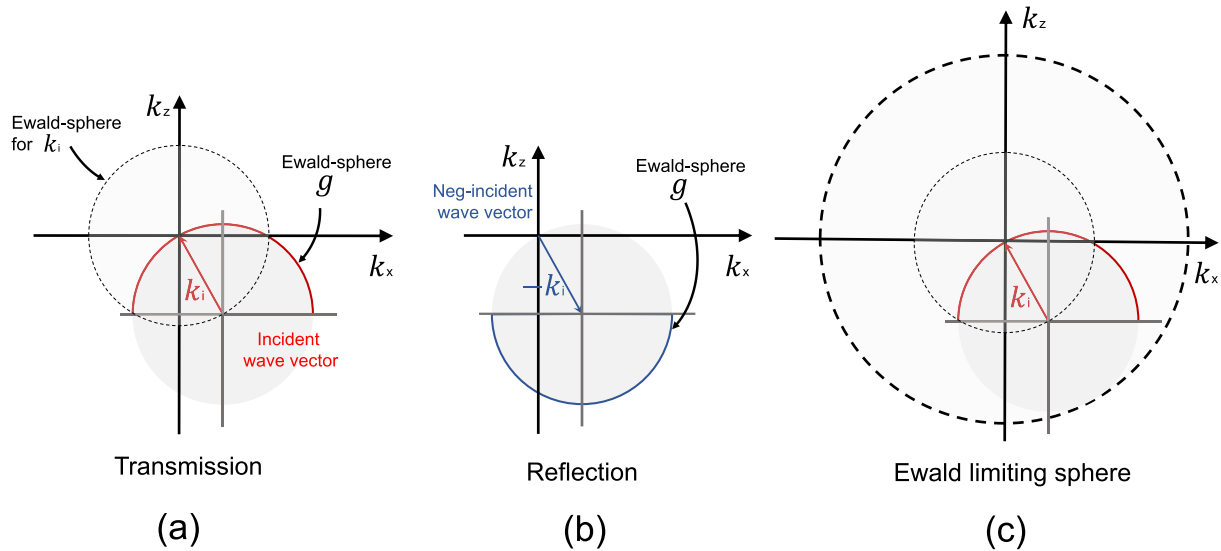


Fig. 83. Fourier-diffraction projection theorem. (a)-(b) Laue construction for determining a scattering wave vector under transmission and reflection; (c) the corresponding Ewald limiting sphere under illuminations at all possible angles.

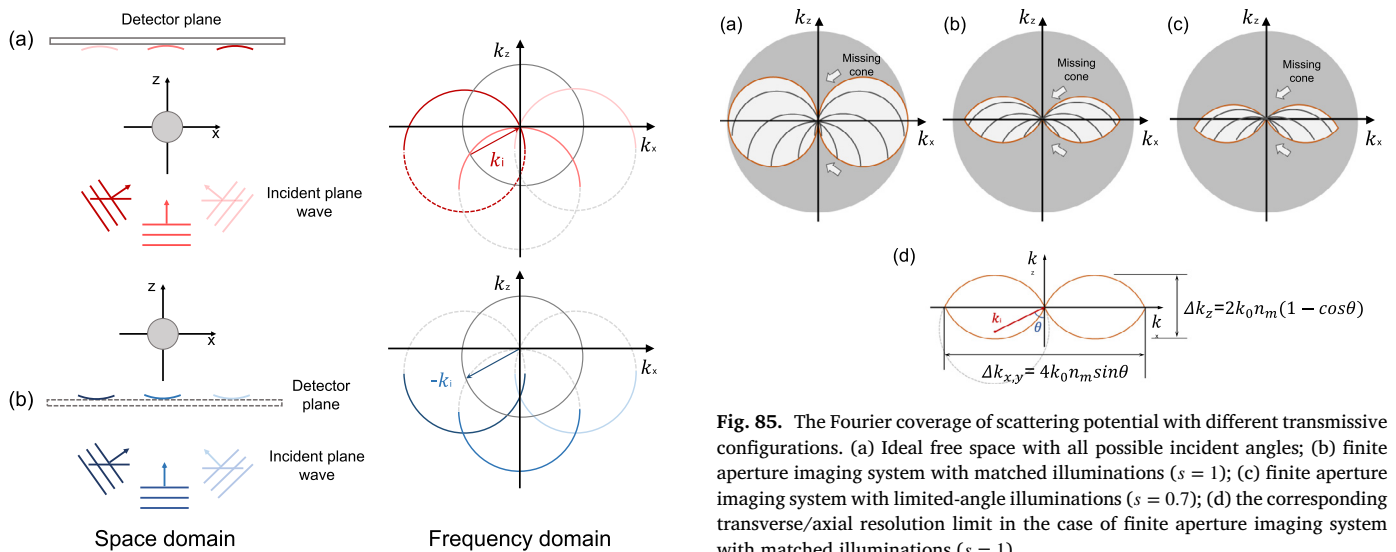


Fig. 84. Illustration of the Fourier diffraction theorem under transmission and reflection configurations. (a) Transmission configuration; (b) reflection configuration.

Besides changing the illumination angle, we can also rotate the sample to achieve larger spectrum coverage, which is similar to the case of conventional CT imaging. In this case, the illumination is fixed with normal incidence with respect to the detector $\mathbf{k} = (0, 0, k_m)$. Without considering the aperture effect, the spectrum information once acquired corresponds to a shifted hemisphere with vertex at the origin, as illustrated in Fig. 86(a). The sample rotation is equivalent to spectrum rotation. If the object is rotated about y-axis, the hemisphere can sweep around the object spectrum along the sample-rotating axis (y), filling a horn torus-like shape in Fourier space, as shown in Fig. 86(b). Under this setting, the maximum extension of the swept spectrum in k_x and k_z plane is $\sqrt{2}k_m$, as shown in Fig. 86(c). But notice that rotating the sample only about one axis still results in missing-cone problem [376], as illustrated in Fig. 86(b). This results in directional blurring along the

Fig. 85. The Fourier coverage of scattering potential with different transmissive configurations. (a) Ideal free space with all possible incident angles; (b) finite aperture imaging system with matched illuminations ($s = 1$); (c) finite aperture imaging system with limited-angle illuminations ($s = 0.7$); (d) the corresponding transverse/axial resolution limit in the case of finite aperture imaging system with matched illuminations ($s = 1$).

axis of rotation (y). In order to alleviate the missing cone problem, iterative non-negativity constraint can be applied to computationally fill the missing information [384]. In addition, in the limit of small wavelengths $\lambda \rightarrow 0$ (e.g., X-ray), the radius of the Ewald sphere $k_m \rightarrow \infty$, the Fourier diffraction theorem converges to the Fourier slice theorem discussed in Subsection 6.2.6 [374].

8.6. Different forms of 3D coherent transfer functions

In the 2D Fourier theory of optical imaging, as described in Section 4, the diffracted spectrum is given by the convolution of the incident spectrum with the object spectrum. As it turns out in the 3D treatment of optical imaging, the reason for this is the linearity of diffraction in frequency space, e.g., approximated by the first-order Born or Rytov approximation. When the effect of aperture is considered, and the incident illumination is assumed to be perpendicular to the detection plane,

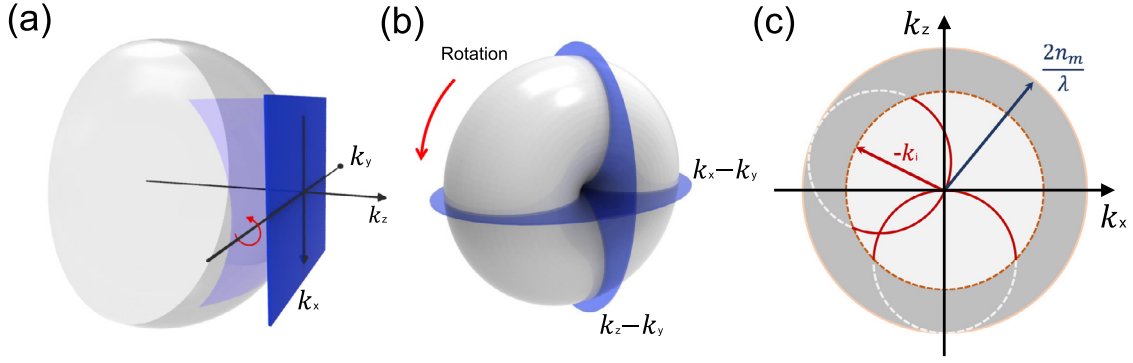


Fig. 86. The Fourier coverage of scattering potential with sample rotation. (a) Information on the hemisphere surface can be obtained with a single illumination direction; The surface of the sphere is oriented along z axis in the case of normal incidence; (b) rotating the object As a result of sample rotation, the hemisphere sweep around the object spectrum along the sample-rotating axis (y), filling a horn torus-like shape inside the 3D spectrum of the object; (c) the $k_x - k_z$ cross-section of the region swept by the hemisphere.

Eq. (352) can be rewritten as

$$\begin{aligned} \hat{U}_{s1}(u_x, u_y, u_z) &= j \frac{H(u_x, u_y)}{2\pi\sqrt{\left(\frac{n_m}{\lambda}\right)^2 - u_x^2 - u_y^2}} \hat{f}(u_x, u_y, u_z) \otimes \hat{U}_{in}(u_x, u_y, u_z) \\ &= j \frac{H(u_x, u_y)}{2\pi\sqrt{\left(\frac{n_m}{\lambda}\right)^2 - u_x^2 - u_y^2}} \hat{f}(u_x, u_y, u_z) \delta\left(u_z - \sqrt{\left(\frac{n_m}{\lambda}\right)^2 - u_x^2 - u_y^2}\right) \end{aligned} \quad (355)$$

$H(u_x, u_y)$ is the 2D coherent transfer function with the limited aperture function and propagation factor included [Eq. (63)]. It is plain to see that the scattered spectrum is given by the convolution of the incident spectrum with the 3D frequency spectrum of the object, and there is a linear relationship between the object function and the first-order scattered field $\hat{U}_{s1}(u_x, u_y, u_z) = H_{3D}(u_x, u_y, u_z) \cdot \hat{f}(u_x, u_y, u_z)$, and they are related by the following 3D transfer function

$$H_{3D}(u_x, u_y, u_z) = \frac{H(u_x, u_y)}{\sqrt{\left(\frac{n_m}{\lambda}\right)^2 - u_x^2 - u_y^2}} \cdot \delta\left(u_z - \sqrt{\left(\frac{n_m}{\lambda}\right)^2 - u_x^2 - u_y^2}\right) \quad (356)$$

Note that we ignore the unimportant constant factor and imaginary unit j in Eq. (356) [the extra j appears because the (real) amplitude and (imaginary) phase information become to the (imaginary) absorption and (real) RI in the scattering potential representation)]. Comparing Eq. (356) with Eq. (329), it can be seen that the 3D coherent transfer function derived from the Fourier diffraction theorem is not coincident with that derived directly from the angular spectrum theory in Subsection 8.2 due to the additional factor $u_z = \sqrt{\left(\frac{n_m}{\lambda}\right)^2 - u_x^2 - u_y^2}$. Therefore, according to the Fourier diffraction theorem, the 3D coherent transfer function is not the direct and uniform projection of the 2D coherent transfer function onto the 3D Ewald sphere, but is given by the weighted projection related to u_z . Fig. 87 illustrates the transformation from 3D to 2D coherent transfer function representation. A homogeneous amplitude on the Ewald sphere, corresponding to, e.g., a spherical wave, leads to an amplitude distribution $1/\cos \alpha$, increasing with the aperture angle α .

The main reason for this discrepancy stems from the selection of Green's function: When deriving the Fourier diffraction theorem, the Green's function used is a ideal diverging spherical wave with uniform amplitude. When only forward scattering is considered, the Green's function is equivalent to the impulse response function [Eq. (13)] with the RS-II type obliquity factor [$K(\theta) = 1$]. The Green's function is isotropic because it is just a half-space spherical wave. However, for the angular spectrum diffraction formula [Eq. (30)], the Green's function

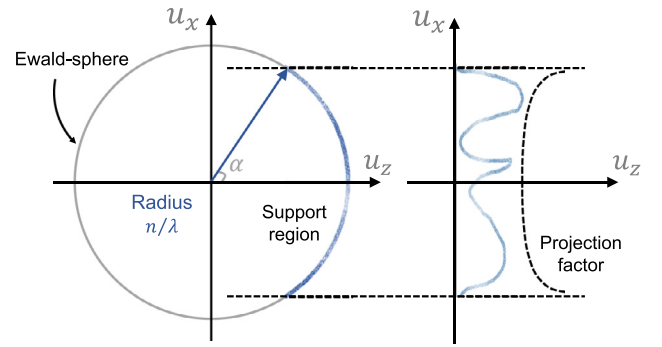


Fig. 87. The projection of 3D coherent transfer function onto the 2D plane.

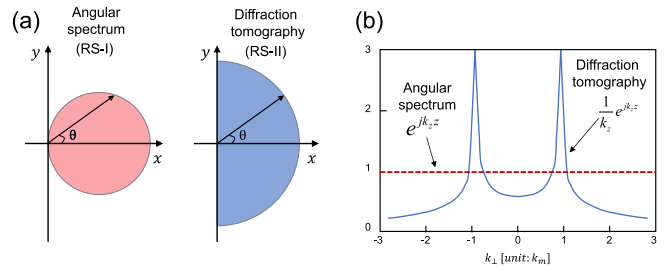


Fig. 88. The comparison of Green's functions used in angular spectrum propagation and diffraction tomography. (a) Green's functions corresponding to angular spectrum propagation and diffraction tomography; (b) the Green's function of diffraction tomography has singular points in frequency domain (infinite gain).

corresponds to the impulse response function [Eq. (13)] with the RS-I type obliquity factor [$K(\theta) = \cos \theta$] [2,5]. Note that the propagation angle θ is a function of transverse coordinates in both original plane and diffraction plane. There is no field radiated along the transverse plane, as shown in Fig. 88(a). The 2D Fourier transform of the Green's function of the angular spectrum diffraction is

$$\mathcal{F}_{x,y} \left\{ \frac{\exp(jk_m|\mathbf{r}|)}{j\lambda|\mathbf{r}|} \cos \theta \right\} = \mathcal{F}_{x,y} \left\{ \frac{\partial \exp(jk_m|\mathbf{r}|)}{\partial z} \frac{1}{j\lambda|\mathbf{r}|} \right\} = \exp(jk_z z) \quad (357)$$

The exponential is simply the phase shift of the corresponding plane wave for propagation over a distance z . It can be found that the additional k_z in the denominator is canceled due to the cosine obliquity factor, which converges to the angle spectrum diffraction theory. In contrast, for the diffraction tomography representation, there is an extra k_z

Table 7
Apodization functions of different design conditions
($\rho = \cos \theta / \lambda$)

| Design conditions | Apodization function |
|----------------------|--|
| Sine condition | $P(\theta) = P(\rho)\sqrt{\cos \theta}$ |
| Herschel condition | $P(\theta) = P(\rho)$ |
| Lagrangian condition | $P(\theta) = P(\rho)\sqrt{\theta/\sin \theta}$ |
| Helmholtz Condition | $P(\theta) = P(\rho)\left(1/\sqrt{\cos \theta}\right)^3$ |

in the denominator. The magnitude of the Fourier transform shows a singularity at $k_z = 0$, i.e., at a spatial frequency of $k_\perp = k_m$ [Fig. 88(b)]. It stems from the plane wave component propagating perpendicular to z-axis (cf. the angular spectrum of a spherical wave).

Though both widely used among the literatures, the 3D coherent transfer functions obtained from two different derivations are inconsistent with each other. Under the paraxial approximation $\cos \theta \approx 1$, the weighting factor in Eq. (358) vanishes, and then the two forms are unified. But in the cases of high-resolution 3D imaging with high-NA illumination and objective, the effect of u_z cannot be ignored. How to distinguish the correct one for a given situation is still a controversial issue [388,389]. In the literature on the theory of image formation in the confocal fluorescence microscope, the 3D coherent transfer function with a large NA is often evaluated based on Debye integral [250,252,390,391]. In the diffraction computation according to the Debye approximation, each point on the aperture plane is taken as the source of a plane wave with amplitude, phase and polarization prescribed by this point. The final image field is the superposition of plane wavelets within the geometric cone (solid angle) limited by the aperture. For the low-NA case, the pupil function of objective is just the generalized aperture function [Eq. (329)]. For larger NAs ($NA > 0.75$), the effect of the apodization function become pronounced and must be included in the imaging model. The apodization function $P(\theta)$ is not equal to the pupil function $P(\rho)$ for a high NA objective: pupil function $P(\rho)$ represents the ray density in the transverse plane while the apodization function $P(\theta)$ gives the ray density over the converging wavefront [76,359]. Apodization functions depend on the transmission coefficients at various interfaces of an imaging system and on spatial filters inserted in the path of an imaging system. The relationship of an apodization function to the corresponding pupil function can be represented as a weighted projection $g(\theta)$:

$$H(\rho, \eta) = P(\rho)g(\theta)\delta\left(\eta - \sqrt{\left(\frac{1}{\lambda}\right)^2 - \rho^2}\right) \quad (358)$$

The commonly used apodization functions are provided in Table 7. For a commercial objective, the sine condition is usually obeyed in design processes so that a perfect image of a thin object can be obtained within the field of view of an objective. Such an imaging system exhibits 2D transverse space invariance, which is called the aplanatic imaging system. $P(\theta) = P(\rho)\cos \theta$ indicates that the ray density over the angle of convergence is decreased with the aperture angle. The Herschel condition $P(\theta) = 1$ is also called the uniform angular condition. The apodization function is completely consistent with the generalized aperture function [Eq. (329)]. In other words, the ray density is constant over the range of the angle of convergence. Under the Helmholtz condition, the apodization function increases with the aperture angle, which is similar to the trend of the 3D coherent transfer function of Fourier diffraction theorem [Eq. (356)]. However, the Helmholtz condition is generally only available for reflective lenses, and the design of most microscopic objective follows the sine condition. In the rest part of this tutorial, the 3D coherent transfer function in Eq. (329) (Herschel condition) is used for simplicity in the derivation of the 3D partially coherent transfer functions. But it should be noted that in high-NA microscopy applications, the sine condition should be considered to produce more physically meaningful results.

8.7. 3D partially coherent transfer functions, 6D TCC and 3D WOTF

In the previous subsections, we have discussed the following linear relationship between the object function and the first-order scattered field in 3D coherent imaging

$$U_{s1}(\mathbf{r}) = \int h(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')U_{in}(\mathbf{r}')d\mathbf{r}' \quad (359)$$

where $h(\mathbf{r})$ is the 3D coherent PSF, whose Fourier transform is the generalized aperture $H(\rho, l) = \mathcal{F}\{h(\mathbf{r})\}$. In order to simplify the derivation, we introduce the first-order Born approximation $U_{s1}(\mathbf{r}) \approx U_s(\mathbf{r})$, the total field can be expressed as

$$\begin{aligned} U(\mathbf{r}) &= U_{in}(\mathbf{r}) + U_s(\mathbf{r}) \\ &= U_{in}(\mathbf{r}) + \int h(\mathbf{r} - \mathbf{r}')f(\mathbf{r}')U_{in}(\mathbf{r}')d\mathbf{r}' \\ &= \int h(\mathbf{r} - \mathbf{r}')T(\mathbf{r}')U_{in}(\mathbf{r}')d\mathbf{r}' \end{aligned} \quad (360)$$

where $T(\mathbf{r}) = 1 + f(\mathbf{r}) = 1 + k_0^2[n(\mathbf{r})^2 - n_m^2]$ is defined as the 3D complex transmittance function of the object. Based on this complex transmittance expression, the 3D coherent imaging and 2D coherent imaging can be formally unified. In the following, we follow the path mirroring that in Subsection 6.3 to derive the 3D transfer functions under partially coherent illuminations. When an 3D object is illuminated by a partially coherent field with the CSD defined as $W_S(\mathbf{r}_1, \mathbf{r}_2)$, the CSD of the transmitted field can be expressed as

$$W_O(\mathbf{r}_1, \mathbf{r}_2) = W_S(\mathbf{r}_1, \mathbf{r}_2)T(\mathbf{r}_1)T^*(\mathbf{r}_2) \quad (361)$$

According to Van Cittert-Zernike theorem, the CSD of the illumination before reaching the object can be written as

$$W_S(\mathbf{r}_1, \mathbf{r}_2) = W_S(\mathbf{r}_1 - \mathbf{r}_2) = \iint S(\mathbf{u})e^{i2\pi\mathbf{u}\cdot(\mathbf{r}_1 - \mathbf{r}_2)}d\mathbf{u} \quad (362)$$

Note that here \mathbf{u} represents the 3D frequency vector (ρ, η). Similar to the generalized aperture, $S(\mathbf{u})$ refers to the incoherent *generalized source* defined on the Eward sphere (in 6f imaging system the source is located at the conjugated plane of the objective pupil)

$$S(\rho, \eta) = S(\rho)\delta\left(\eta - \sqrt{\left(\frac{1}{\lambda}\right)^2 - \rho^2}\right) \quad (363)$$

After 6D Fourier transform, the transmitted light field needs to multiply the 3D coherent transfer function $H(\mathbf{u})$ [Eq. (358)] twice at the pupil plane, and then the CSD in the image space can be obtained after inverse Fourier transform

$$W_I(\mathbf{r}_1, \mathbf{r}_2) = \iint \hat{W}_O(\mathbf{u}_1, \mathbf{u}_2)H(\mathbf{u}_1)H^*(\mathbf{u}_2)e^{j2\pi(\mathbf{u}_1\mathbf{r}_1 + \mathbf{u}_2\mathbf{r}_2)}d\mathbf{u}_1d\mathbf{u}_2 \quad (364)$$

in which $H(\mathbf{u}_1)H^*(\mathbf{u}_2)$ is the 3D mutual coherent transfer function. Equivalently, it can also be written in the convolution form in the spatial domain

$$W_I(\mathbf{r}_1, \mathbf{r}_2) = \iint W_O(\mathbf{r}'_1, \mathbf{r}'_2)h(\mathbf{r}_1 - \mathbf{r}'_1)h^*(\mathbf{r}_2 - \mathbf{r}'_2)d\mathbf{r}'_1d\mathbf{r}'_2 \quad (365)$$

where $h(\mathbf{r}_1)h^*(\mathbf{r}_2)$ is the 3D mutual PSF. The image intensity that we can captured is given by the values “on the diagonal” of the CSD function

$$I(\mathbf{r}) = W_I(\mathbf{r}, \mathbf{r}) = \iint W_S(\mathbf{r}_1, \mathbf{r}_2)T(\mathbf{r}_1)T_S^*(\mathbf{r}_2)h(\mathbf{r} - \mathbf{r}_1)h^*(\mathbf{r} - \mathbf{r}_2)d\mathbf{r}_1d\mathbf{r}_2 \quad (366)$$

This formula can be further simplified by substituting Eq. (362) into Eq. (366)

$$I(\mathbf{r}) = \int S(\mathbf{u}) \left| \int T(\mathbf{r}') h(\mathbf{r} - \mathbf{r}') e^{j2\pi\mathbf{u}\mathbf{r}'} d\mathbf{r}' \right|^2 d\mathbf{u} \\ \equiv \int S(\mathbf{u}) I_{\mathbf{u}}(\mathbf{r}) d\mathbf{u} \quad (367)$$

Eq. (367) suggests that the final intensity that can be captured at the image plane can be regarded as the intensity (incoherence) superposition of coherent partial images $I_{\mathbf{u}}(\mathbf{r})$ arising from all points over the incoherent generalized source function. This is similar to the 2D case discussed in Subsection 6.3.1. Substituting Eq. (362) into Eq. (367) and rewriting it in the form of Fourier integral gives

$$I(\mathbf{r}) = \iiint S(\mathbf{u}) \hat{T}(\mathbf{u}_1) \hat{T}^*(\mathbf{u}_2) H(\mathbf{u} + \mathbf{u}_1) H^*(\mathbf{u} + \mathbf{u}_2) e^{j2\pi\mathbf{r}(\mathbf{u}_1 - \mathbf{u}_2)} d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u} \quad (368)$$

Separating the contribution of the specimen and imaging system leads to the notion of the 3D TCC

$$TCC(\mathbf{u}_1, \mathbf{u}_2) = \iint S(\mathbf{u}) H(\mathbf{u} + \mathbf{u}_1) H^*(\mathbf{u} + \mathbf{u}_2) d\mathbf{u} \quad (369)$$

Noted that the 3D TCC is actually a 6D function. Based on the 3D TCC, Eq. (368) can be written as

$$I(\mathbf{r}) = \iint \hat{T}(\mathbf{u}_1) \hat{T}^*(\mathbf{u}_2) TCC(\mathbf{u}_1, \mathbf{u}_2) e^{j2\pi\mathbf{r}(\mathbf{u}_1 - \mathbf{u}_2)} d\mathbf{u}_1 d\mathbf{u}_2 \quad (370)$$

The above expressions are similar to those derived from 2D partial coherence imaging, except that the dimension of the TCC increases from four to six. It can be calculated from the overlapping region of the generalized source and two shifted generalized aperture (three hemispherical shells). However, the calculation requires prohibitively large computation time for the 6D integral of the TCC. Considering the real part and the imaginary part of the 3D complex transmission function separately, *i.e.*, the phase component $P(\mathbf{r})$ and the absorption component $A(\mathbf{r})$ (note that the real part of the 3D object function corresponds to the phase component, instead of absorption as in the 2D case)

$$T(\mathbf{r}) = 1 + f(\mathbf{r}) = 1 + k_0^2 [n(\mathbf{r})^2 - n_m^2] \\ = 1 + P(\mathbf{r}) + jA(\mathbf{r}) \quad (371)$$

Under the first-order Born approximation, the self-interference of the scattered light can be neglected, and the mutual spectrum of the object can be approximated as

$$\hat{T}(\mathbf{u}_1) \hat{T}^*(\mathbf{u}_2) = \delta(\mathbf{u}_1) \delta(\mathbf{u}_2) + \delta(\mathbf{u}_2) [\hat{P}(\mathbf{u}_1) + j\hat{A}(\mathbf{u}_1)] + \\ \delta(\mathbf{u}_1) [\hat{P}(\mathbf{u}_2) - j\hat{A}(\mathbf{u}_2)] \quad (372)$$

Substituting Eq. (372) into Eq. (371) and using the Hermitian symmetry property of TCC yields the intensity of the partially coherent image for a weak object

$$I(\mathbf{r}) = TCC(\mathbf{0}, \mathbf{0}) + 2 \operatorname{Re} \left\{ \int TCC(\mathbf{u}, \mathbf{0}) [\hat{P}(\mathbf{u}) + j\hat{A}(\mathbf{u})] e^{j2\pi\mathbf{r}\mathbf{u}} d\mathbf{u} \right\} \quad (373)$$

where $TCC(\mathbf{u}, \mathbf{0})$ is the linear part of the TCC, which is the 3D WOTF

$$WOTF(\mathbf{u}) \equiv TCC(\mathbf{u}, \mathbf{0}) = \iint S(\mathbf{u}') H(\mathbf{u}' + \mathbf{u}) H^*(\mathbf{u}') d\mathbf{u}' \quad (374)$$

Taking Fourier transform on Eq. (373), and separating the contributions of phase component $P(\mathbf{r})$ and absorption component $A(\mathbf{r})$ gives

$$\hat{I}(\mathbf{u}) = I_0 \delta(\mathbf{u}) + H_A(\mathbf{u}) \hat{A}(\mathbf{u}) + H_P(\mathbf{u}) \hat{P}(\mathbf{u}) \quad (375)$$

where I_0 is the background intensity. It now becomes obvious that the image contrast due to the absorption and phase are decoupled and linearized. Because the intensity is always real, the real and even component of $WOTF(\mathbf{u})$ corresponds to the partially coherent 3D ATF $H_A(\mathbf{u})$, and the imaginary and odd component of $WOTF(\mathbf{u})$ corresponds to the 3D PTF $H_P(\mathbf{u})$ of the object, respectively.

$$H_A(\mathbf{u}) = WOTF(\mathbf{u}) + WOTF^*(-\mathbf{u}) \quad (376)$$

$$H_P(\mathbf{u}) = WOTF^*(\mathbf{u}) - WOTF(-\mathbf{u}) \quad (377)$$

Note that here 3D ATF and 3D PTF is then given by the real and imaginary components of 3D WOTF respectively, which is now in coincidence with the 2D case. The reason for this lies in the additional imaginary unit in the projection of the 2D aperture onto the 3D Eward sphere, as described in Eq. (355).

8.8. 3D PTFs for coherent, incoherent, and partially coherent imaging

Next, we analyze the 3D transfer functions of an axisymmetric optical system under different illumination conditions. First, under coherent illumination (coherence parameter $s \rightarrow 0$), the light source degenerates into an ideal on-axis point radiator

$$S(\rho, \eta) = \delta(\rho) \delta\left(\eta - \frac{1}{\lambda}\right) \quad (378)$$

where $\rho = \sqrt{u_x^2 + u_y^2}$ represents the radial spatial frequency coordinate and η represents the axial spatial frequency coordinate. Taking Eq. (378) into the 3D WOTF

$$WOTF(\rho, \eta) = H\left(\rho, \eta + \frac{1}{\lambda}\right) \quad (379)$$

It is shown that under coherent illumination, the 3D WOTF just corresponds to the shifted generalized aperture, *i.e.*, the shifted 3D coherent transfer function $H(\mathbf{u})$ [Eq. (358)] passing through the origin. This is not difficult to understand because the linearization of 3D imaging is derived based on the first-order Born approximation (*e.g.*, the Fourier diffraction theorem), indicating that the 3D coherent transfer function implies the “weak object” approximation. Ideally, the generalized aperture is a real function, the even part and the odd part of the 3D WOTF correspond to the ATF and PTF, respectively.

$$H_A(\rho, \eta) = H\left(\rho, \eta + \frac{1}{\lambda}\right) + H\left(\rho, -\eta - \frac{1}{\lambda}\right) \quad (380)$$

$$H_P(\rho, \eta) = H\left(\rho, \eta + \frac{1}{\lambda}\right) - H\left(\rho, -\eta - \frac{1}{\lambda}\right) \quad (381)$$

They are two back-to-back Eward hemispherical shells intersecting at the origin. The 3D ATF is even symmetric about the η -axis with the maximum value at the origin; while the 3D phase transfer function is odd symmetric about the η -axis with zero at the origin, indicating that a homogeneous object produces no phase contrast.

In the case of incoherent illumination, the generalized source is equal to or even larger than the generalized aperture $s = NA_{ill}/NA_{obj} \geq 1$, and the 3D WOTF becomes the autocorrelation of 3D coherent transfer function

$$WOTF_{\text{incoh}}(\mathbf{u}) = \iint H(\mathbf{u}' + \mathbf{u}) H^*(\mathbf{u}') d\mathbf{u}' \quad (382)$$

Note that it is always a real function (even under defocusing), which means the PTF is always 0. When the apodization function in the generalized aperture/3D coherent transfer function $H(\mathbf{u})$ [Eq. (358)] is ignored (satisfying Herschel condition), the transfer function can be expressed as [391]

$$WOTF_{\text{incoh}}(\bar{\rho}, \bar{\eta}) = \frac{4}{\pi K} \arccos \left[\frac{1}{M} \left(\frac{2 \cos \alpha}{|\bar{\eta}|} + 1 \right) \right] \quad (383)$$

where $K = \sqrt{\bar{\rho}^2 + \bar{\eta}^2}$, $M = \frac{2|\bar{\rho}|}{K|\bar{\eta}|} \sqrt{1 - \frac{K^2}{4}}$, and α is the aperture angle of the imaging system. Note here the lateral and axial frequencies $\bar{\rho}, \bar{\eta}$ are both normalized by $1/\lambda$. Fig. 89 shows the formation of the 3D incoherent OTF. Because of the nature of the autocorrelation function, the support of the 3D incoherent OTF is a torus-shaped region of Fourier space. The maximum accessible object frequencies in free space correspond to $\Delta u_x = \frac{NA}{\lambda}$ and $\Delta u_z = \frac{1 - \sqrt{1 - NA^2}}{\lambda}$, which gives a reminder of the normalized coordinate system used in Subsection 4.2 and

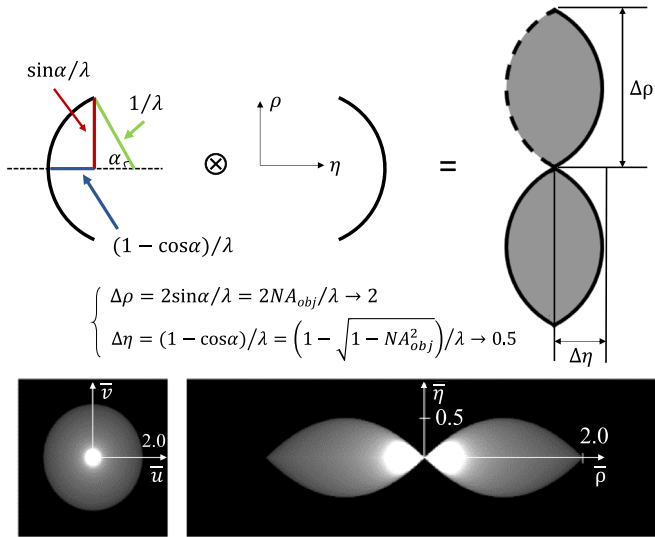


Fig. 89. Transverse and axial cross-sections of the 3D incoherent WOTF, represented in normalized coordinates.

Subsection 6.3.4. It can be found the normalization coefficients are associated with the maximum extent of the 3D incoherent OTF. And the lateral and axial extents become 0.5 and 2 in the normalized coordinate system $(\bar{\rho}, \bar{\eta})$, as clearly illustrated in Fig. 89.

Under partially coherent illuminations, the transfer functions of an axial-symmetric aberration-free imaging system are given by

$$H_A(\bar{\rho}, \bar{\eta}) = WOTF(\bar{\rho}, \bar{\eta}) + WOTF(\bar{\rho}, -\bar{\eta}) \quad (384)$$

$$H_P(\bar{\rho}, \bar{\eta}) = WOTF(\bar{\rho}, \bar{\eta}) - WOTF(\bar{\rho}, -\bar{\eta}) \quad (385)$$

where $WOTF(\bar{\rho}, \bar{\eta})$ is represented as

$$WOTF(\bar{\rho}, \bar{\eta}) = \frac{\lambda}{2\pi} \frac{\bar{\rho}^2 \sigma}{K^2 \bar{\eta}} \sqrt{1 - \frac{K^2}{4} - \frac{K^2 \sigma^2}{\bar{\eta}^2}} + \left(\sqrt{1 - \frac{K^2}{4} - \frac{\bar{\eta}^2}{2K}} \right) \arccos \left(\frac{K\sigma}{\bar{\eta} \sqrt{1 - K^2/4}} \right) \quad (386)$$

when $0 < \bar{\rho} < \bar{\rho}_P - \bar{\rho}_S$:

$$\sigma = \frac{\bar{\eta}}{\bar{\rho}} \left(\frac{\bar{\eta}}{2} - \sqrt{1 - \bar{\rho}_S^2} \right), \text{ if } \sqrt{1 - \bar{\rho}_S^2} - \sqrt{1 - (\bar{\rho}_P - \bar{\rho}_S)^2} \leq \bar{\eta} \leq \sqrt{1 - \bar{\rho}_S^2} - \sqrt{1 - (\bar{\rho}_P + \bar{\rho}_S)^2} \quad (387)$$

when $0 < \bar{\rho} < \bar{\rho}_P - \bar{\rho}_S$:

$$\sigma = \frac{\bar{\eta}}{\bar{\rho}} \left(\frac{\bar{\eta}}{2} - \sqrt{1 - \bar{\rho}_S^2} \right), \text{ if } \sqrt{1 - \bar{\rho}_S^2} - \sqrt{1 - (\bar{\rho}_S - \bar{\rho})^2} \leq \bar{\eta} \leq \sqrt{1 - \bar{\rho}_S^2} - \sqrt{1 - \bar{\rho}_P^2}$$

$$\sigma = \frac{\bar{\eta}}{\bar{\rho}} \left(-\frac{\bar{\eta}}{2} - \sqrt{1 - \bar{\rho}_P^2} \right), \text{ if } \sqrt{1 - \bar{\rho}_S^2} - \sqrt{1 - \bar{\rho}_P^2} \leq \bar{\eta} \leq \sqrt{1 - (\bar{\rho}_P - \bar{\rho})^2} - \sqrt{1 - \bar{\rho}_S^2} \quad (388)$$

where $\bar{\rho}_S$ and $\bar{\rho}_P$ are both normalized by $1/\lambda$ so that $\bar{\rho}_S = NA_{ill}$, $\bar{\rho}_P = NA_{obj}$. In fact, the partially coherent 3D WOTF was first derived by Streibl [178] under the paraxial approximation as early as 1985, and then it was further summarized and applied to 3D microscopic imaging by Sheppard [179] based on the TCC representation. The expression of the transfer function under the paraxial approximation is much simpler. Until recently, the analytical expression of the 3D WOTF under the non-paraxial condition was derived, and has been successfully applied to the high-resolution TIDT [214,381,388].

Finally, it should be noted that we can directly obtain the 2D defocused transfer functions from the 3D transfer functions. The Fourier transform of 3D intensity and the 2D intensity at different depths can be related by taking the Fourier transform about z [177]

$$I_z(\mathbf{r}) = \int I(\mathbf{u}) e^{-j2\pi z \eta} d\eta \quad (389)$$

Expanding the exponential term in Eq. (389) based on Euler formula, we can get the following relationship between 3D transfer functions and defocused 2D transfer functions [217]

$$H_A(\rho, z) = \int H_A(\rho, \eta) \cos(2\pi z \eta) d\eta \quad (390)$$

$$H_P(\rho, z) = \int H_P(\rho, \eta) \sin(2\pi z \eta) d\eta \quad (391)$$

It can be verified that, without considering the apodization function in the generalized aperture/3D coherent transfer function (satisfying the Herschel condition), the 3D ATF and PTF derived in this subsection can be directly converted to the 2D defocused ATF and PTF derived in Subsection 4.3 [Eqs. (92) and (93)].

8.9. Diffraction tomographic (3D phase) imaging under partially coherent illuminations

As introduced at the beginning of this section, the widely applied ODT techniques [375,376,378,379] usually require specially designed holographic microscopes working with a coherent light source that sequentially illuminates the object from different sides by using a scanning module [384,392–394] or rotating the object mechanically [395,396]. However, such coherent ODT techniques recover the 3D RI by assembling numerous sections of the object spectrum obtained for each illumination directions. Moreover, coherent ODT techniques suffer from speckle noise and parasitic interference due to the coherent illumination. In this subsection, we will discuss another way to achieve 3D diffraction tomography by extending the idea of TIE, namely TIDT. Similar to QPI techniques based on TIE, TIDT does not require coherent illumination and interferometry, it only captures the through-focus intensity images and directly recover the 3D RI distribution of the object by 3D deconvolution based on the 3D PTF. The basic idea of TIDT is similar to the widefield 3D fluorescence deconvolution microscopy, except that the 3D transfer function used here is the 3D partially coherent PTF, instead of incoherent 3D OTF [178]. It can be viewed as an extension of TIE phase retrieval technique, which generalizes the idea of 2D “transport of intensity at a 2D plane” into “transport of intensity through a 3D volume” to realize diffraction tomography. Compared with the object rotation or beam scanning, the data acquisition based on the axial scanning not only greatly simplifies the system configuration, but also is completely compatible with conventional wide-field transmission microscope platforms.

Fig. 90 illustrate the difference between the traditional coherent ODT and the partially coherent TIDT techniques. Traditional coherent ODT techniques recover the 3D RI by assembling numerous of sections of the object spectrum obtained for each incident angle. With a series of angle-dependent complex amplitude measurements, a certain portion of the object spectrum can be filled, which produces an estimate of 3D complex RI of the sample [Fig. 90(a)]. In contrast to the coherent ODT, TIDT

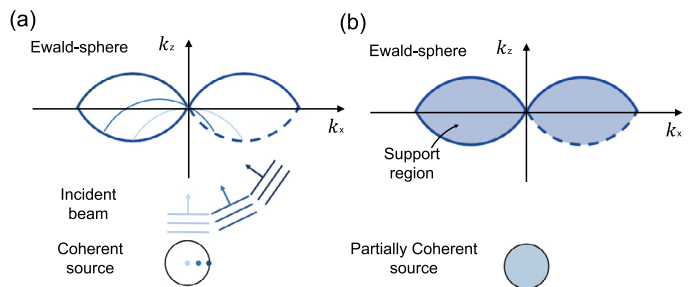


Fig. 90. Comparison of two different diffraction tomography strategies under coherent illumination (the conventional ODT) and partially coherent illumination (TIDT). (a) Coherent illumination; (b) partially coherent illumination.

can be regarded as its “parallelized” version due to the use of partially coherent illumination. The sample is simultaneously illuminated by a quasi-monochromatic and spatially incoherent extended source from all possible directions allowed by the condenser aperture of the wide-field microscope [Fig. 90(b)]. The raw data is the through-focus intensity stack obtained by optical refocusing (axial scanning). Thus, TIDT is inherently faster and more efficient in terms of spectrum coverage and, in practice, simplifies the measurement and the reconstruction process. An additional advantage is that the low spatial coherent illumination avoids speckle noise, which plagues coherent ODT and other coherent imaging methods. It should be noted that though the fundamentals of TIDT were established by Streibl [178] more than thirty years ago, only recently its experimental implementation has been demonstrated [214,380,381].

In Subsection 8.7 and Subsection 8.8, we have learned that for weakly scattering object under partially coherent illumination, the 3D Fourier transform of the stack of intensity distributions at the image space is given by

$$\hat{I}(\mathbf{u}) = I_0 \delta(\mathbf{u}) + H_A(\mathbf{u})\hat{A}(\mathbf{u}) + H_P(\mathbf{u})\hat{P}(\mathbf{u}) \quad (392)$$

It should be noted that in Eq. (392), the phase (RI) information and the intensity (absorption) information of the object are coupled together, *i.e.*, both phase contrast and amplitude contrast contribute to the measured intensity. In order to recover the quantitative RI information for objects with nonnegligible absorption, we need to decouple it from the contribution arising from the intensity information. There are four feasible solutions:

- 1) **Recording an additional set of intensity stacks with opposing illumination or via object rotation over 180°:** the symmetries of 3D ATF and PTF allow the phase component to be separated by subtraction of the two intensity stacks recorded with opposing illumination or via object rotation over 180° because the absorption contrast is an even function about each scatterer. This is similar to TIE phase retrieval based on central finite difference with the defocus distances of the two images equal and opposite relative to the central in-focus image.
- 2) **Collecting two or more datasets of the same object with different pupil functions:** since the corresponding 3D transfer functions are different, it should be possible to separate the phase contrast and amplitude contrast by solving the corresponding linear equations [178].
- 3) **Assuming the phase-attenuation duality analogous to 2D phase imaging:** the absorption is assumed to be proportional to RI $\hat{P}(\mathbf{u}) \approx \varepsilon \hat{A}(\mathbf{u})$, then Eq. (392) can be simplified as [380,381]

$$\hat{I}(\mathbf{u}) = B\delta(\mathbf{u}) + \hat{P}(\mathbf{u})[H_P(\mathbf{u}) + \varepsilon H_A(\mathbf{u})] \quad (393)$$

where ε is a constant parameter, which is usually empirically adjusted to optimize the reconstruction quality [381]. When the sample is a pure phase object, $\varepsilon = 0$, and the 3D RI reconstruction boils down to the following one-step 3D deconvolution

$$\hat{P}(\mathbf{u}) = \frac{\hat{I}(\mathbf{u})H_P^*(\mathbf{u})}{|H_P(\mathbf{u})|^2 + \beta} \quad (394)$$

where β is the Tikhonv regularization parameter determined by the SNR of the raw data. The support and response of the 3D PTF are directly associated with the resolution and SNR of the 3D RI reconstruction. According to Eq. (385), we can analyze the 3D PTF distribution of a perfectly aligned microscope with rotationally symmetric illumination. Figs. 91(a)-(d) display the $\bar{\rho} - \bar{\eta}$ sections of the 3D PTF corresponding to an objective with $NA = 1.4$ and circular illumination aperture with different coherence parameters $s = 0.3, 0.6$ and 0.9 . As can be seen, smaller coherence parameters provide better low-frequency response at an expense of smaller spectral coverage, leading to compromised lateral and axial resolution. As the coherence parameter increases, the spectral coverage is extended, and the response at both low and high frequencies

also gradually increases. If the coherence parameter further increases to ~ 1 , the spectral coverage reaches the maximum (incoherent diffraction limit). However, phase contrast progressively vanishes, suggesting that the phase information can hardly be transferred into intensity when the illumination NA is large. In other words, the intensity image gives no phase contrast under incoherent illumination. Thus, there is an inherent tradeoff between phase contrast and imaging resolution in TIDT with conventional circular illumination apertures. Note that these observations are analogous to the case of the 2D QPI based on the TIE phase retrieval. Smaller coherence coefficients correspond to stronger phase contrast especially for low spatial frequencies, but the resolution is limited. Larger coherence coefficients provide higher theoretical resolution limit, while the poor phase contrast compromises the SNR for phase reconstruction. Therefore, considering both of these two aspects, s should be set between 0.4 and 0.7 to guarantee better 3D phase imaging performance. It should be also noticed that, regardless of the illumination aperture, there always exists a missing cone frequency region around the $\bar{\eta}$ axis, which is a well-known common problem of the ODT techniques for wide-field transmission configuration.

There are two possible strategies to overcome the above issue and improve the performance of TIDT. One is to optimize the source distribution and improve the homogeneity of the PTF [397]. A similar approach has been used to optimize the illumination pattern for the TIE phase retrieval [211], as discussed in Subsection 7.6. It has been found that Gaussian illumination is better than traditional uniform circular illumination for its higher homogeneity and less low-value regions in the 3D PTF [397]. The other strategy is to exploit the illumination diversity and OFS. For a single illumination aperture, the strong frequency response and large spectral coverage of the PTF are difficult to be achieved simultaneously. Therefore, similar to the case of 2D QPI, we can incorporate multiple intensity stacks with different illumination apertures to achieve a synthesized 3D PTF, resulting in high-quality 3D RI reconstruction with improved resolution and SNR. Note that unlike the multi-plane TIE, the axial scanning is already included in the data acquisition, so “multi-plane” is impossible and is meaningless for TIDT, leaving only the illumination aperture is adjustable. So in order to obtain a more complete frequency coverage and balanced transfer function, Li *et al.* [214] proposed a multi-frequency combination scheme that incorporates three intensity stacks captured under three illumination apertures including an annular aperture and two circular apertures with different coherence parameters. The optimal frequency components of 3D PTFs corresponding to multiple illumination apertures are combined together by utilizing linear least-squares method, and a more accurate 3D reconstruction result with imaging resolution up to the incoherent diffraction limit can be obtained. Inspired by the resolution enhancement for 2D QPI in Subsection 7.6, the annular illumination aperture is introduced to not only achieve the maximum allowable lateral and axial resolution limit, but also maintain strong amplitude response of the 3D PTF, preventing phase contrast vanishing under large circular illumination aperture [Fig. 91(e)]. Fig. 92 gives the flowchart of the multi-aperture TIDT based on multi-frequency combination. The Fourier spectra of the corresponding three intensity stacks are taken as $\hat{I}_1(\mathbf{u})$, $\hat{I}_2(\mathbf{u})$ and $\hat{I}_3(\mathbf{u})$ under these three illumination apertures, and the corresponding PTFs are $H_{P1}(\mathbf{u})$, $H_{P2}(\mathbf{u})$ and $H_{P3}(\mathbf{u})$. The synthesized PTF can be expressed as

$$P(\mathbf{r}) = \mathcal{F}^{-1} \left(\frac{\hat{I}_1(\mathbf{u})}{H_{P1}(\mathbf{u})} \varepsilon_1 + \frac{\hat{I}_2(\mathbf{u})}{H_{P2}(\mathbf{u})} \varepsilon_2 + \frac{\hat{I}_3(\mathbf{u})}{H_{P3}(\mathbf{u})} \varepsilon_3 \right) \quad (395)$$

The weighting coefficients of the frequency component corresponding to each pupil can be obtained by the following least squares method

$$\varepsilon_i = \frac{H_{P_i}^*(\mathbf{u})H_{P_i}(\mathbf{u})}{|H_{P1}(\mathbf{u})|^2 + |H_{P2}(\mathbf{u})|^2 + |H_{P3}(\mathbf{u})|^2} \quad (396)$$

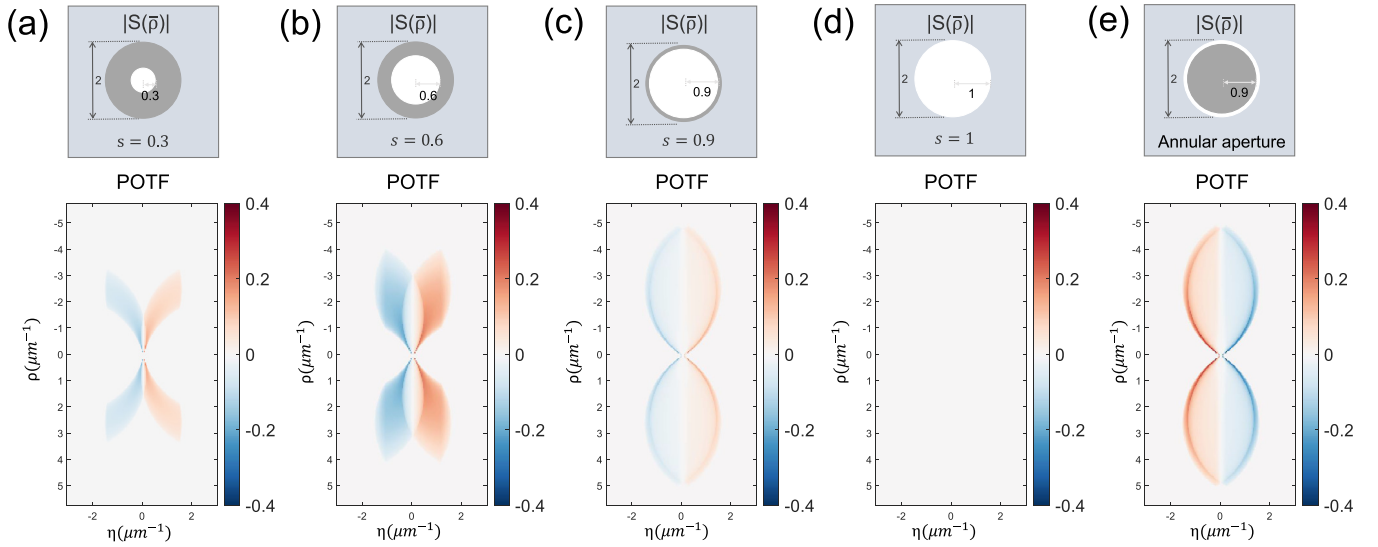


Fig. 91. 2D cross-sections of 3D PTFs for different (circular and annular) illumination apertures with different coherence parameters [214].

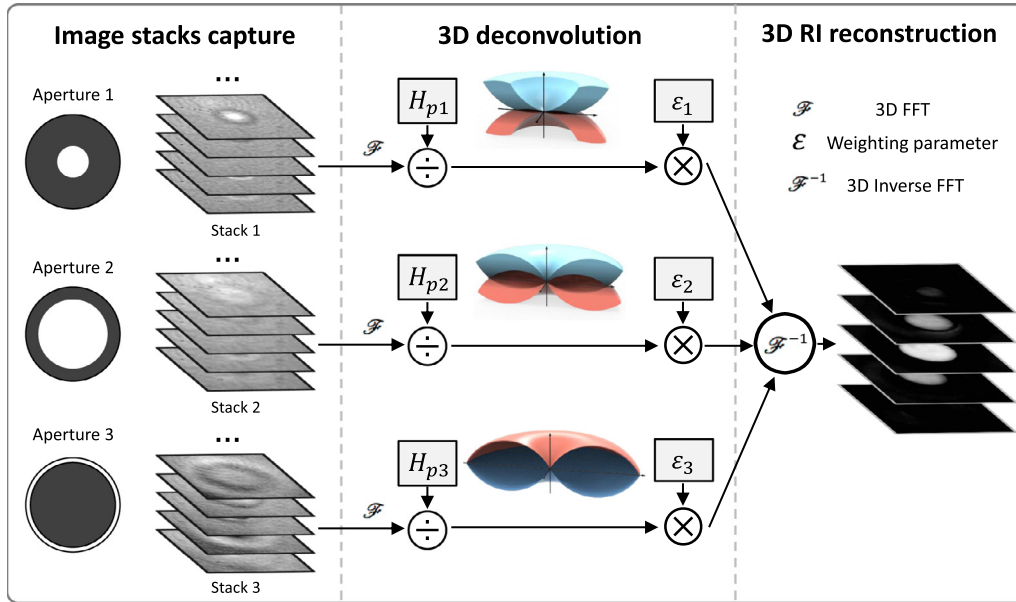


Fig. 92. Flowchart of the multi-aperture TIDT method based on multi-frequency combination [214].

Substituting Eq. (396) into the Eq. (395), the phase component can be finally reconstructed as follows

$$P(\mathbf{r}) = \mathcal{F}^{-1} \left\{ \frac{\hat{I}_1(\mathbf{u})H_{p1}^*(\mathbf{u}) + \hat{I}_2(\mathbf{u})H_{p2}^*(\mathbf{u}) + \hat{I}_3(\mathbf{u})H_{p3}^*(\mathbf{u})}{|H_{p1}(\mathbf{u})|^2 + |H_{p2}(\mathbf{u})|^2 + |H_{p3}(\mathbf{u})|^2 + \alpha} \right\} \quad (397)$$

Before closing this section, we need to further clarify the term “3D phase imaging”. It is generally known that the “phase” refers to the optical path difference introduced by the non-uniform spatial distribution of the RI or thickness of the object. It implies the meanings of “projection” and “accumulation”, which is defined on a 2D plane instead of in a 3D volume. So the definition of “phase” is only valid for 2D thin object, otherwise it can only be used to represent the optical path integral of a 3D thick object along the light propagation direction. Consequently, the term “3D phase imaging” is not rigorous. In this tutorial, the term “3D phase imaging” is only adopted as a colloquial alternative to “3D RI diffraction tomography”.

9. Applications of TIE in optical imaging and microscopy

In this section, we present optical configurations and representative applications of TIE with an emphasis on optical imaging, metrology, and microscopy. In Subsection 9.1, we present the conventional experimental setups for the TIE phase retrieval. In these setups, in order to acquire multiple intensity images at different defocus distances, either the sample or the camera needs to be translated along the optical axis, necessitating manual adjustment or mechanic devices. In Subsection 9.2, we present several improved setups of TIE, which eliminate the mechanical motion for through-focus intensity acquisition, extending TIE to dynamic phase imaging applications. In Subsection 9.3, we present the optical configurations for TIE based diffraction tomography. Finally, in Subsection 9.4, we review the typical applications of TIE in the field of optical imaging and microscopy, and present some representative experimental results.

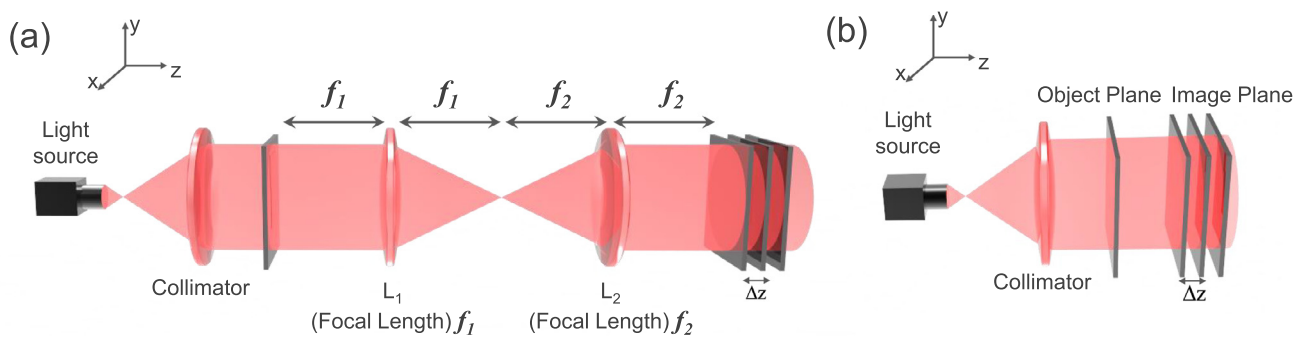


Fig. 93. Conventional optical configurations for the TIE phase imaging. (a) $4f$ system-based configuration; (b) lensless configuration.

9.1. Basic optical configurations for TIE phase imaging

The TIE phase retrieval method needs to collect intensity images at multiple defocus planes. In order to capture these intensity images, it is often necessary to use a $4f$ system, as illustrated in Fig. 93(a). Since the object and image planes in the $4f$ system are strictly conjugated, the intensity information at defocused planes can be obtained by either moving the object or the camera. These two methods are essentially equivalent, but when considering lateral and axial magnifications, the object-space defocus distance is scaled by a factor of f_2^2/f_1^2 when the defocusing is achieved by moving the camera in the image space. For an optical microscope with an infinity-corrected optical system, the optical path is equivalent to Fig. 93(a). The collimator of Fig. 93(a) is equivalent to the condenser lens in the Köhler illumination, while the lenses L_1 and L_2 correspond to objective and tube lens in the infinity-corrected microscope, and f_2/f_1 is the magnification of the microscope. Therefore, an infinity-corrected microscope can be directly used for TIE retrieval without any modifications. Besides, TIE can also adopt the lensless imaging setup shown in Fig. 93(b), in which the intensity images are all taken at the out-of-focus plane. So the phase reconstructed by TIE is also located at the defocused plane, and additional numerical back-propagation or digital refocusing algorithms are required to reconstruct the complex amplitude distribution of the sample at the object plane.

9.2. Advanced optical configurations for dynamic TIE phase imaging

In traditional optical setups of TIE, one needs to move the object or camera manually or mechanically when capturing the intensity images, which inevitably reduces the speed of data acquisition, precluding dynamic, high-speed, and real-time measurement applications. To address this problem, numerous advanced optical configurations for dynamic TIE phase imaging have been proposed in recent years. Their common feature is to avoid mechanical motion during the image acquisition process. For example, Blanchard *et al.* [398] placed a quadratically distorted grating in front of the imaging lens, creating three laterally shifted images corresponding to different defocus distances in a single image plane [Fig. 94(a)]. Waller *et al.* [191] utilized the chromatic aberration inherent in the microscope to obtain three intensity images at different axial distances from a single color image. Waller *et al.* [190] also proposed single-shot TIE optical configurations based on a multiplexed volume hologram to laterally separate images from different focal planes, enabling acquisition of intensity images at different depths from a single exposure [Fig. 94(b)]. Almoro *et al.* [193] realized non-mechanical image defocusing by introducing a SLM in the Fourier plane of the $4f$ imaging system [Fig. 94(c)]. Gorthi and Schonbrun [194] applied TIE to flow cytometry and realized the automatic collection of through-focus intensity images by tilting the microfluidic channel vertically [Fig. 94(d)]. Martino *et al.* [399] developed a single-shot TIE setup based on a beam splitter and several mirrors, permitting recording two intensity images at different depth simultaneously [Fig. 94(e)].

Not just limited to independent imaging systems, TIE is also fully compatible with an off-the-shelf infinity-corrected microscope. Dynamic QPI can also be achieved by introducing additional optics to the camera port of the microscope. Zuo *et al.* [216,217] proposed two microscope-based dynamic TIE phase imaging systems. The tunable lens based TIE (TL-TIE) system introduces an *electrically tunable lens* (ETL) and a $4f$ system into a conventional bright-field microscope. The TL-TIE permits high-speed axial scanning to realize the non-mechanical focus control with constant magnification for high-speed TIE phase imaging [217] [Fig. 95(a)]. The *single-shot quantitative phase microscopy* (SQPM) system realizes the programmable digital defocusing based on an SLM. Based on a Michelson-like optical architecture attached to an unmodified inverted bright-field microscope, two laterally separated images from different depths can be obtained simultaneously by a single camera exposure, enabling TIE phase imaging at the camera limited speed [216] [Fig. 95(b)].

9.3. Optical configurations for TIE phase tomography

Although TIE has been successfully used for phase retrieval and QPI, the fact that phase is a product of thickness and average RI can potentially result in inaccurate interpretations of QPI data when imaging complex 3D structures. In order to gain more accurate morphological information, diffraction tomography is often desired. Compared with TIE phase imaging, the optical configurations for diffraction tomography based on TIE are relatively complicated, and there are two typical solutions. The first category of techniques is based on traditional CT or ODT models, which obtains multiple quantitative phase distributions by sample rotating or beam scanning. Different from traditional diffraction tomography based on interferometry or digital holography, the complex amplitude of the field under each illumination angle is retrieved by TIE in a noninterferometric manner. In 2000, Barty *et al.* [175] combined TIE phase retrieval with sample rotation based on a traditional microscope platform. Two different optical fibers were fixed with the tip of a syringe under the control of a rotating motor [Fig. 96(a)]. The RI profiles of the fibers were reconstructed by the inverse Radon transform. In 2011, based on similar ideas and experimental setups, Lee *et al.* [400] achieved 3D tomography of a glass diamond fixed on a rotating stage based on the back filtered projection method. Subsequently, Nguyen *et al.* [401,402] measured the 3D RI distributions of fibers and gemstone microspheres fixed on a rotating turret based on the TL-TIE [217] and SQPM systems [216]. In those early works, the 3D reconstruction physical model is similar to what was used in x-ray CT, which assumed phase measurement to be an integration of RI along the projection angle and ignored the significance of the optical diffraction effect, thus limiting the earlier TPM systems for objects with simple structures. In biology imaging, this inaccurate tomography model can significantly affect the 3D reconstruction resolution, especially for cells that are much thicker than the depth of field of the imaging systems. In 2015, Zuo *et al.* [212] proposed a lensless diffraction tomographic platform using only an image sensor and a programmable color LED array. By illuminating

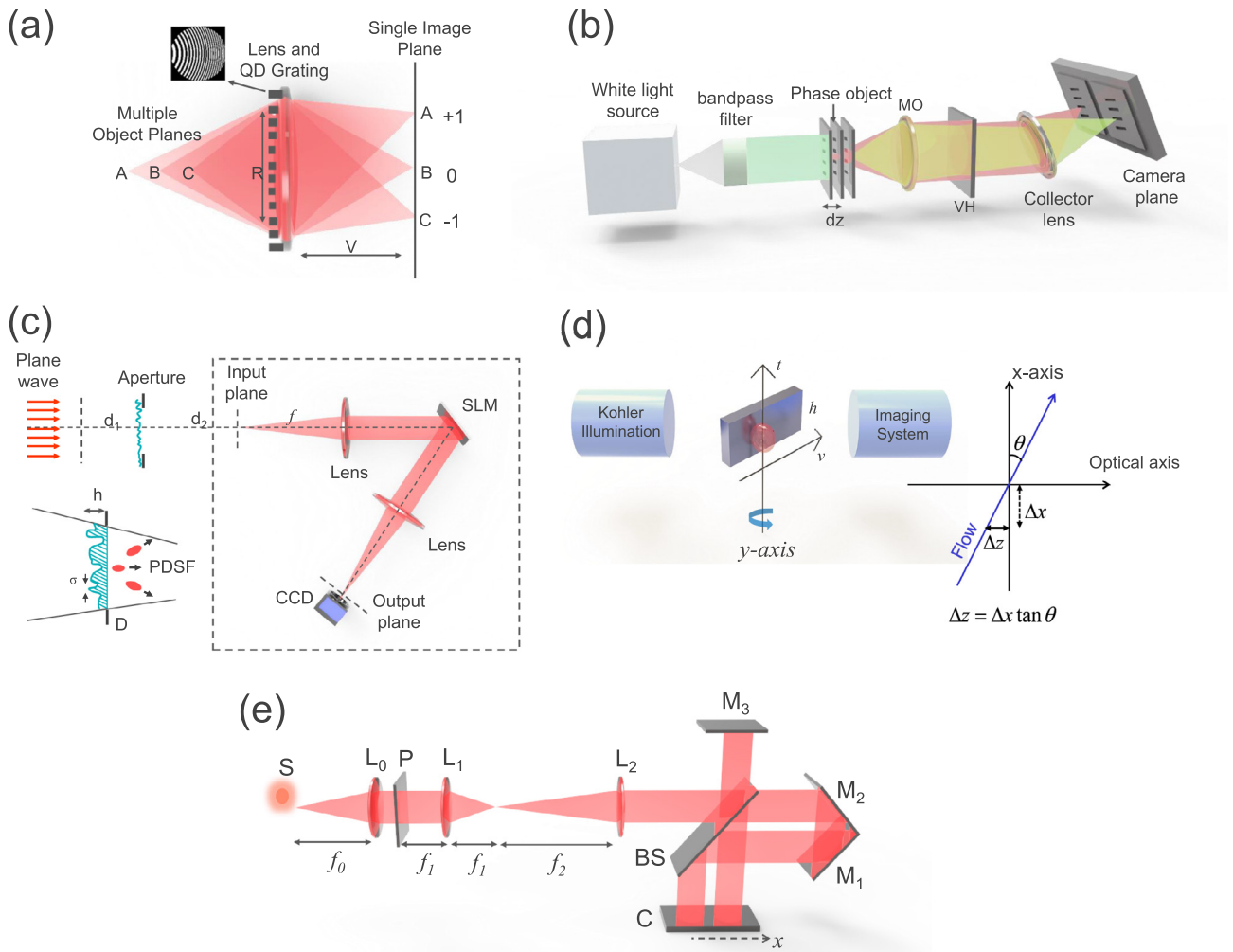


Fig. 94. Several improved optical configurations for dynamic TIE phase imaging. (a) optical configuration based on quadratically distorted grating [398]; (b) optical configuration based on volume holography [190]; (c) optical configuration based on spatial light modulator [193]; (d) optical configuration based on flow cytometry [194]; (e) optical configuration based on beam splitter and multiple reflections [399].

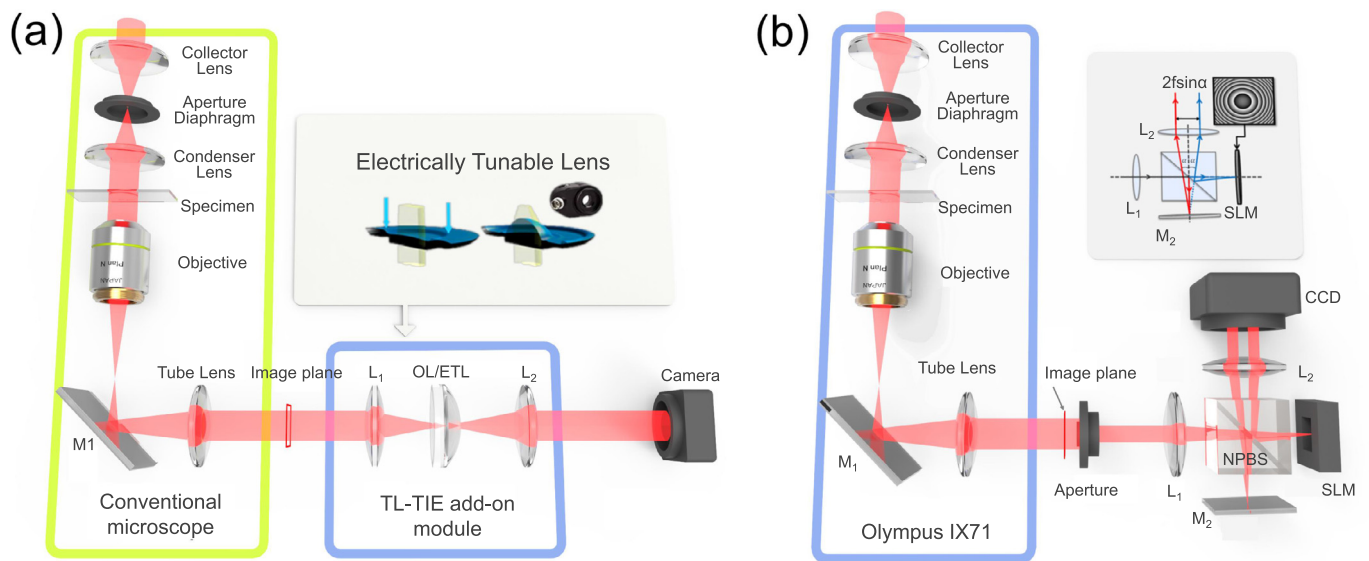


Fig. 95. Two optical configurations for dynamic TIE phase imaging, which can be implemented as a simple add-on module to a conventional off-the-shelf microscope. (a) electrically tunable lens based TIE microscopic system (TL-TIE) [217]; (b) single-shot TIE system based on a SLM (SQPM) [216].

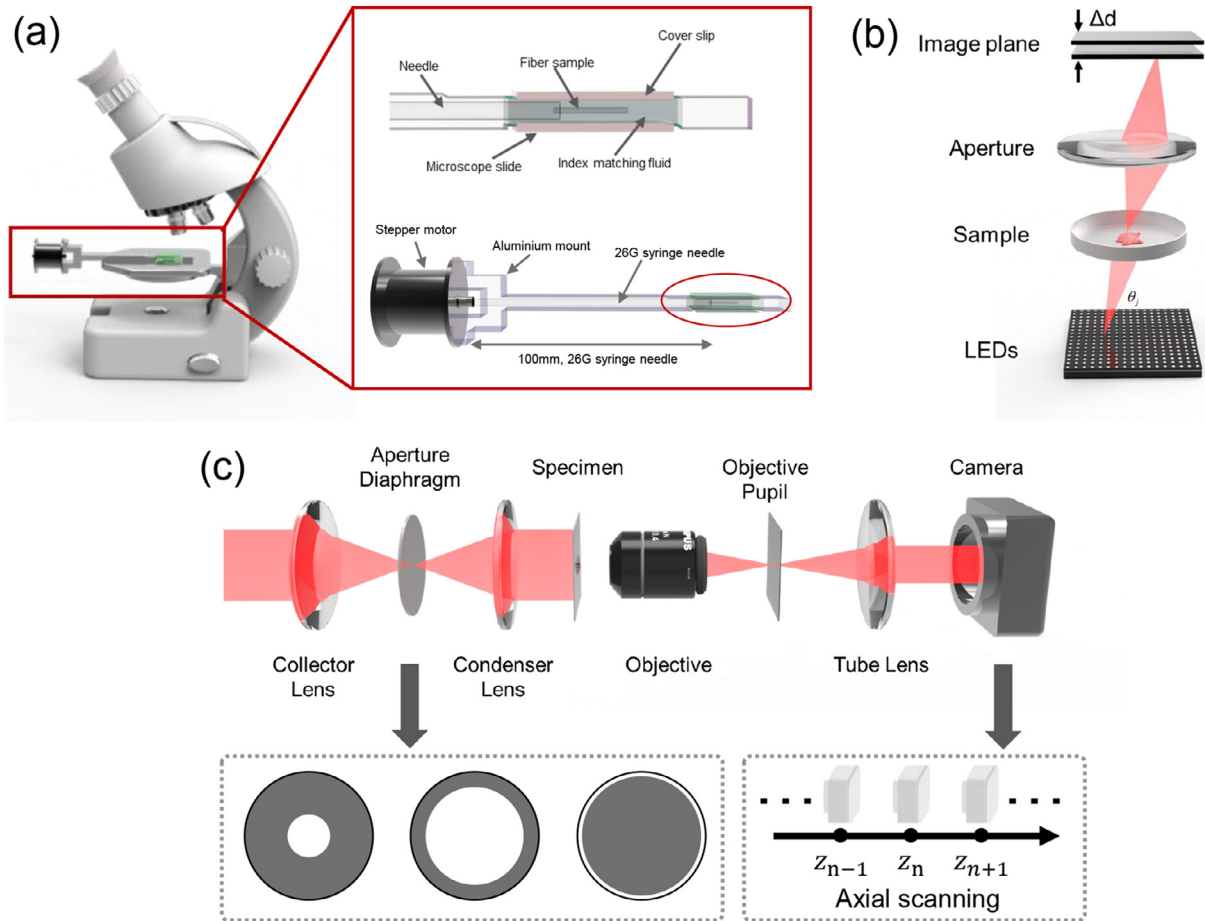


Fig. 96. Optical configurations for 3D diffraction tomography based on TIE. (a) Tomographic system in which the sample is fixed by a needle tip and rotated by a motorized stage [175]; (b) TIE diffraction tomographic system based on multi-angle illuminations with an LED array [213]; (c) TIE diffraction tomographic system based on intensity axial scanning with an off-the-shelf microscope [214].

the sample at multiple wavelengths, the quantitative phase of the specimen can be retrieved by TIE. Based on the wavelength dependence of Fresnel propagation, this approach treats the diffraction patterns generated by RGB three-color illuminations as intensity images at three different defocus distances. Furthermore, by illuminating the sample sequentially with different LEDs across the full array, the complex fields of the object at different illumination angles can be combined into the 3D RI distribution based on the ODT model under the first Rytov approximation. In 2015, Jenkins *et al.* [380] combined the object rotation and axial scanning to realize 3D tomographic reconstruction of the RI distributions of different types of optical fibers. In 2017, Li *et al.* [213] proposed a TIE tomography system based on an off-the-shelf microscope by replacing the Köhler illumination system of the traditional bright-field microscope with a programmable LED array [Fig. 96(b)]. The intensity images of the sample are captured in two different defocused planes at each illumination angle, and the corresponding phase distributions are recovered by TIE. Finally, the 3D RI information of the object can be reconstructed based on the Fourier diffraction theorem. Different from the interferometric tomographic system where the incident laser illumination is scanned with the help of galvanometer mirrors, the LED-array-based configurations are inherently much simpler and achieve improved imaging stability and speed.

Since 2017, TIDT has gradually emerged as a new diffraction tomography technique and attracted much attention of researchers. TIDT is a partially coherent ODT approach, which only needs to capture the through-focus intensity images across the object, and then uses the image deconvolution algorithm to directly retrieve the 3D RI distribution

of the object. Thus, it effectively bypasses the difficulties associated with the traditional interferometric diffraction tomography techniques, such as interferometric measurements, object rotation, and beam scanning. The hardware setup is fully compatible with conventional bright-field microscope systems. Soto *et al.* [381] realized the TIDT based on the TL-TIE system [217]. An electrically tunable lens is employed to axially scan the sample in high speed. The resultant through-focus stack is deconvoluted based on 3D WOTF to recover the 3D RI distribution of the sample. In 2018, Li *et al.* [214] combined TIDT with illumination coherence engineering. Three sets of through-focus intensity images are captured under an annular aperture and two circular apertures with different coherence parameters [Fig. 96(c)]. By changing the condenser aperture diaphragms of a traditional bright-field microscope, the 3D WOTFs corresponding to different illumination apertures are combined to obtain a synthesized frequency response, achieving high-quality, low-noise 3D reconstructions with imaging resolution up to the incoherent diffraction limit.

9.4. Applications of TIE in optical imaging and microscopy

Similar to many other phase retrieval algorithms, TIE was mainly oriented to adaptive optics [144–148], X-ray diffraction imaging [161,162], STEM [166,169,171,172], neutron radiography [164,165], transmission electron microscopy *etc.*. In recent years, as the TIE technique advances and matures, it has been increasingly applied to the optical imaging and microscopy of the visible light band. Due to the space

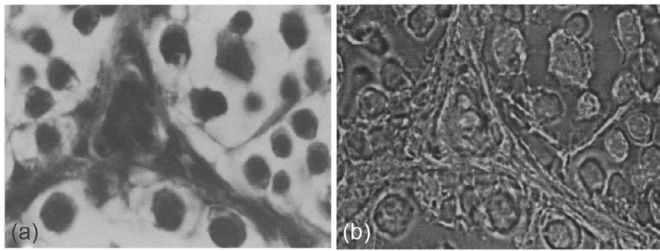


Fig. 97. TIE phase-contrast imaging of testicle of a rat [142]. (a) bright-field intensity image; (b) axial intensity derivative distribution.

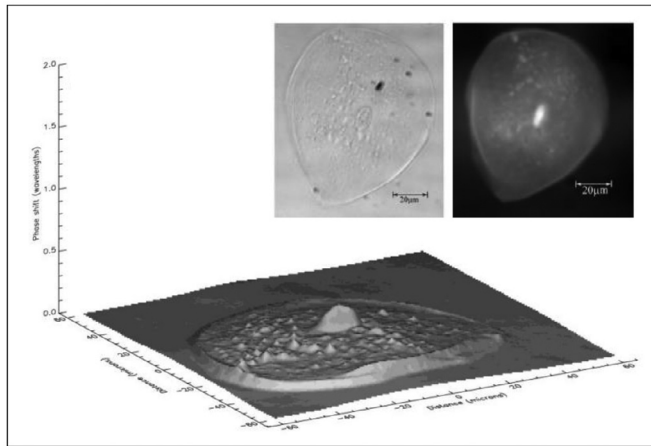


Fig. 98. Quantitative phase imaging of an unstained cheek cell [175]. The insets show the DIC image of the cell and the recovered quantitative phase image by TIE.

limitation, this section reviews several representative applications of TIE, focus particularly on optical imaging, metrology, and microscopy.

9.4.1. QPI for bio-applications

The applications of TIE to optical microscopy can be traced back to 1984 (only about one year after Teague deduced TIE). Streibl [142] briefly proved that TIE is applicable to phase-contrast imaging of transparent phase objects under partially coherent Köhler illumination of a off-the-shelf microscope. He demonstrated the phase gradient enhancement effect of unstained mouse testis cells (Fig. 97). Since the numerical solution to TIE has not been proposed at that time, he only presented the axial intensity derivative image rather than the quantitative phase reconstruction result. Streibl believed that the axial intensity derivative image could effectively highlight phase details that are difficult to be observed in the bright-field image. In 1998, Barty *et al.* [175] briefly reported the QPI results of cheek cells and optical fibers based on TIE (Fig. 98). Strictly speaking, it was the debut of TIE in the field of quantitative phase microscopy. They also emphasized that the phase recovered by TIE is continuous, and no phase unwrapping is needed. The beautiful experimental results presented by Barty *et al.* has opened the door for the subsequent applications of TIE in biological and biomedical imaging. In 2004, Curl *et al.* [180,181] used TIE to quantitatively detect the morphology and growth rate of cells, indicating that quantitative phase information is very conducive to data processing and quantitative analysis, such as cell segmentation and counting. In 2005, Ross *et al.* [182] utilized TIE to enhance the imaging contrast of unstained cells under microbeam irradiations. In 2007, Dragomir *et al.* [183] successfully applied TIE to the quantitative measurement of the birefringence effect of unstained isolated cardiomyocytes.

In 2010, Kou *et al.* [188] demonstrated that TIE can be directly implemented based on an off-the-shelf DIC microscope to realize QPI, and successfully obtained the quantitative phase images of human cheek cells.

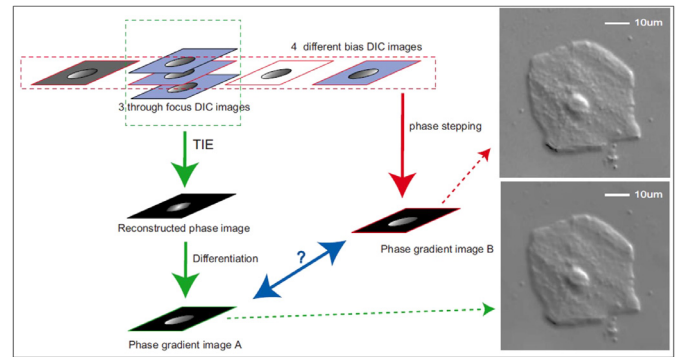


Fig. 99. Quantitative phase imaging of an unstained cheek cell [192]. The insets show the phase-gradient image from quantitative phase reconstructed by TIE and phase-shifting DIC, respectively.

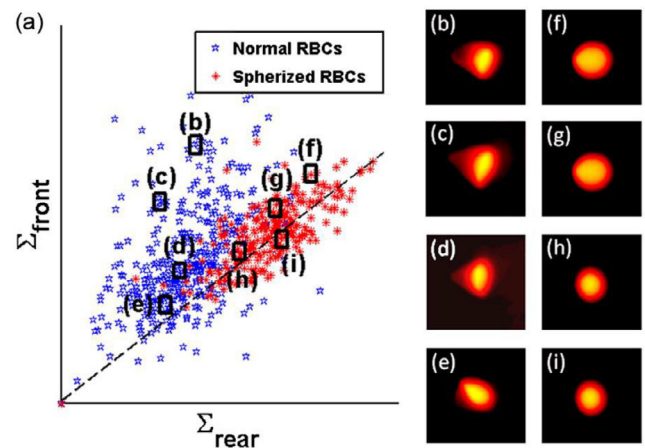


Fig. 100. Quantitative phase imaging and RBC screening [403]. (a) Scatterplot of RBCs subjected to isotonic and hypotonic solutions; (b)-(i) phase images in the gated regions marked in (a); Normal RBCs (b)-(e) in flow took parachute like or slipper like shapes; Spherized RBCs (f)-(i) remained spherical even in flow.

The phase reconstruction results were in good agreement with those obtained by the polarization phase-shifting technique (Fig. 99). Waller *et al.* proposed single-shot TIE optical configurations based on volume holography [190] and chromatic aberration [191], enabling acquisition of intensity images at different defocus distances from a single exposure. Based on these setups, they presented the QPI results of deformable micromirror arrays and HeLa cells. In 2011, Kou *et al.* [192] put forward the WOTF deconvolution method for QPI under partially coherent illuminations and quantitatively compared it with the conventional TIE. The QPI results of human *Ascaris* are very similar to those obtained by TIE, which is understandable because the low-frequency component of the WOTF under weak defocusing is a Laplacian in the frequency domain (TIE's PTF). In 2012, Gorthi and Schonbrun [194] first applied TIE to flow cytometry and realized the automatic collection of through-focus intensity images by tilting the microfluidic channel vertically. The high-throughput sorting of red blood cells was demonstrated based on morphological information extracted from their quantitative phase images. In the same year, Phillips *et al.* [404] measured the RI, dry mass, volume, and density of red blood cells based on TIE with use of a traditional bright-field microscope (though they claimed that they obtained 3D RI distribution of the cell, their multi-plane TIE tomography method is not rigorous in principle). Subsequently, Phillips *et al.* [403] further applied their method for circulating tumor cell (CTC) screening in an ovarian cancer patient: the TIE phase retrieval was employed to measure cellular dry mass, and the reconstructed DIC images were used to measure cellular volume. They found that high-definition CTCs were more

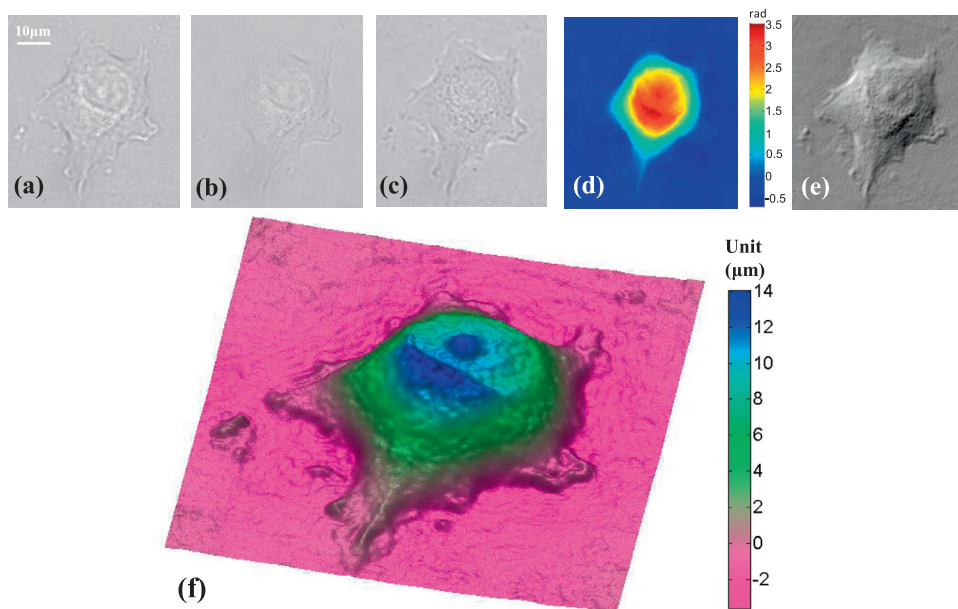


Fig. 101. Quantitative phase imaging of an individual MCF-7 cell based on TL-TIE [217]. (a-c) Intensity images with defocus distance -2.5 , 0 , $2.5 \mu\text{m}$; (d) phase map recovered by TIE; (e) digitally simulated DIC image from (d); (f) 3D pseudo-color rendering of the cell thickness.

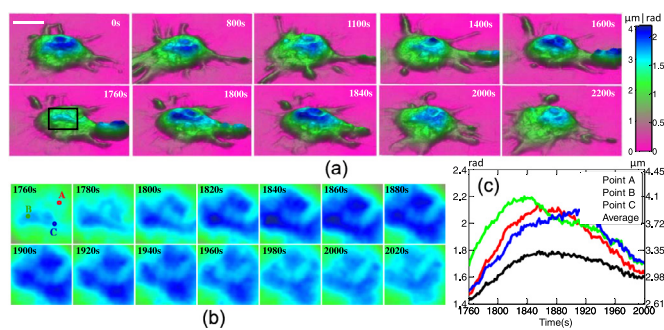


Fig. 102. Dynamic TIE phase imaging of macrophage phagocytosis [216]. (a) Color-coded phase profiles at different stages of phagocytosis; (b) phase maps of the nuclear region of the macrophage—the black square in (a)—during the internalization stage of phagocytosis; (c) phase/thickness variation with time of three points, indicated by the dots in (b): (A) red, (B) green, and (C) blue, and the average of the whole square region (bottom, black curve). Scale bar: $10 \mu\text{m}$.

massive than leukocytes, which provides key insights into the fluid dynamics of cancer, and may provide the rationale for strategies to isolate, monitor or target CTCs based on their physical properties.

In 2013, Zuo *et al.* [217] demonstrated the dynamic QPI results of a live MCF-7 breast cancer cell based on the TL-TIE system. The high-resolution quantitative phase images revealed the subcellular dynamics, *i.e.*, the nucleus inside the cell, ruffles on the membrane, and the lamellipodia in the cell periphery (Fig. 101). In the same year, Zuo *et al.* [216] investigated of chemical-induced apoptosis and the dynamic phagocytosis process of RAW264.7 macrophages (Fig. 102) based on SQPM. The high-resolution dynamic 3D QPI videos of macrophage phagocytosis were presented, suggesting TIE is a simple and promising approach for label-free imaging and quantitative study of cell dynamics at subcellular resolution. In 2016, Bostan *et al.* [275] used TIE to recover the quantitative phase images of HeLa cells. The measured phases were consistent with those obtained by digital holographic microscopy. They also pointed out that quantitative phase images are extremely convenient for subsequent cell segmentation and counting (Fig. 103). In 2016, Li *et al.* [215] realized multi-modal microscopic imaging based on TIE based on an off-the-shelf bright-field microscope, realizing digi-

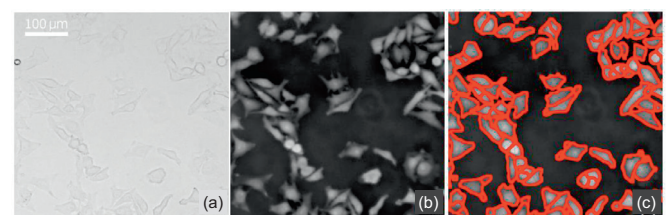


Fig. 103. Delineation of HeLa cells [275]. (a) In-focus bright-field image; (b) phase image obtained by using TIE reconstruction; and (c) segmentation result with watershed algorithm.

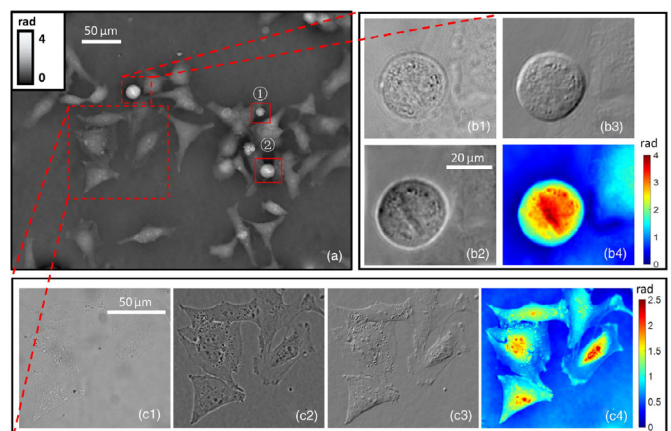


Fig. 104. Multi-modal imaging of live HeLa cells [406]. (a) Quantitative phase map of the whole FOV. (b-c) experimental results for bright-field intensity, phase contrast, DIC, and quantitative phase images of mitotic cell and interphase cells.

tal ZPC, DIC, QPI, and light-field imaging of HeLa cells under division (Fig. 104). It should be mentioned that a similar multi-modal TIE imaging approach was previously presented by Paganin *et al.* [405] in lensless X-ray diffraction imaging.

In 2017, Zuo *et al.* [209] employed the annular illumination based TIE for long-term time-lapse imaging of HeLa cell dividing in culture.

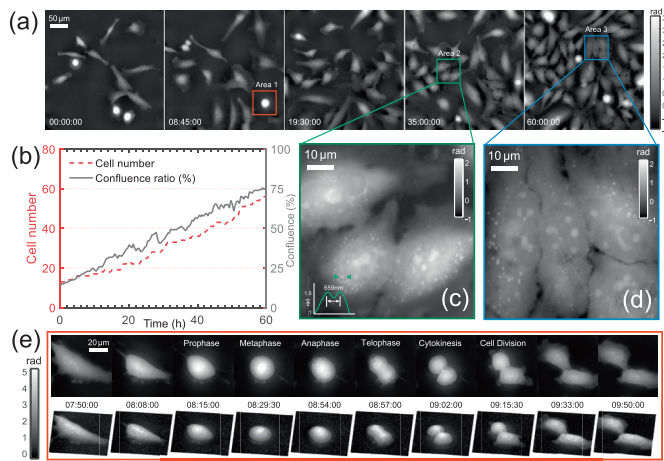


Fig. 105. Time-lapse TIE phase imaging of HeLa cell division over a long period (60 h) [209]. (a) Representative quantitative phase images at different time points; (b) the change of cell number and confluence ratio over the culture passage period; (c) and (d) show the magnified views corresponding to two regions of interest (Area 2 and Area 3); (e) 10 selected time-lapse phase images and the corresponding 3D renderings showing the morphological features of a dividing cell (Area 1) at different stages of mitosis.

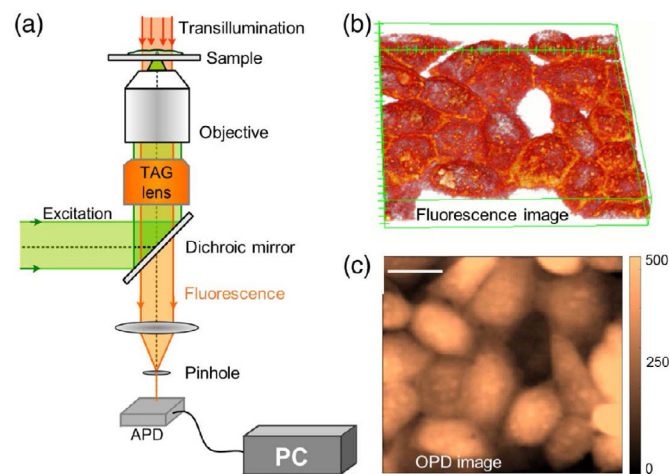


Fig. 106. Dual-mode confocal fluorescence and TIE phase imaging in a confocal laser scanning microscope [407]. (a) Schematic depiction of the optical setup; (b) exemplary fluorescence; and (c) optical path difference images measured with this setup. Scale bar, 20 μm ; color bar labels, nm.

The high-resolution quantitative phase images allow for assessing the progression of confluence ratio over 60 hours (Fig. 105). Zheng *et al.* [408,409] introduced TIE into Coherent Anti-Stokes Raman Scattering (CARS) microscopy. The Several CARS images at different defocused planes perpendicular to the propagation direction were recorded to reconstruct a quantitative phase map based on TIE, which allows for non-resonant background suppression in CARS microscopy. In 2018, Li *et al.* [406] realized dynamic TIE phase imaging of live human osteoblasts based on a modified SQPM system. Liao *et al.* [410] combined the TIE phase retrieval technique with a self-developed dual-LED-based auto-focusing microscope, achieving the whole slide QPI of an unstained mouse kidney section. Zheng *et al.* [407] proposed the dual-mode phase and fluorescence imaging in a confocal laser scanning microscope by combining the confocal microscopy with TIE phase imaging (Fig. 106). The limited depth of field was extended about sixfold by axial scanning with a tunable acoustic gradient index of refraction lens, making TIE compatible with confocal microscopy. The dual-mode images with

pixel-to-pixel correspondence enable the standalone determination of the RI of live cells.

In 2019, Li *et al.* [411] quantitatively characterized the cell morphology and cytoskeletal dynamics of living osteoblastic cells based on TIE. The changes in mass transport are visualized based on phase correlation imaging [412]. Bian *et al.* [413] proposed a miniaturized portable microscope to assess the inhomogeneous optical properties of tissues based on the quantitative phase maps obtained by TIE. Rajput *et al.* [414] demonstrated that the quantitative phase recovered by TIE could be used for non-scanning 3D fluorescence imaging. Though the propagation of incoherent fields does not obey the Fresnel diffraction law, and the phase of an incoherent field is not well-defined, their experimental results on neurite and cells did show certain feasibility. In 2020, Mandula *et al.* [415] combined TIE with multi-color fluorescence for long-term dual-mode imaging of dividing HeLa cells based on a standard bright-field microscope platform. The changes of the focus are achieved with specifically introduced chromatic aberration in the imaging system, allowing motion-free image acquisition. Wittkopp *et al.* [416] compared the QPI results of resolution target and check cells obtained by DHM and TIE based on a custom-built microscope system, revealing quantitative agreement between the two methods in both static and live samples.

9.4.2. Phase tomography for bio-applications

The key difference between phase tomography and QPI is the ability to obtain the “true 3D” RI profile of a biological sample. In 2015, Zuo *et al.* [212] demonstrated the lensless quantitative phase microscopy and diffraction tomographic imaging of fertilized eggs of *Parascaris equorum* based a lensless compact on-chip platform with a programmable color LED array (Fig. 108). By illuminating the sample with a multi-wavelength LED, the quantitative phase of the specimen can be retrieved by TIE. With the quantitative phase retrieved, the diffracted object field can be digitally refocused through back propagation without using any lenses. Furthermore, by illuminating the sample sequentially with different LEDs across the full array, the complex fields of the same object from different illumination angles can be obtained. Finally, the 3D RI distribution of the sample is reconstructed based on the ODT model. The miniaturized system offers high quality, depth-resolved images with high resolution (lateral resolution 3.72 μm , axial resolution 5 μm) and across a wide FOV (24 mm^2). In 2017, Soto *et al.* [381] reconstructed the 3D RI distribution of human RBCs based on TIDT [Fig. 107(a)-(c)]. This technique was subsequently applied to 3D quantitative RI reconstruction of dynamic biological samples by Rodrigo *et al.* [417]. They demonstrated 3D RI visualization of living bacteria both freely swimming and optically manipulated by using freestyle laser traps allowing for their trapping and transport along 3D trajectories. In 2017, Li *et al.* [214] demonstrated quantitative RI imaging based on TIDT under multiple illumination apertures. Experimental results verified that the use of multiple illumination apertures provides a broader frequency coverage and higher frequency responses. Fig. 107(d)-(e) show 3D RI rendering results of a unstained *Pandorina morum* algae and a HeLa cell reconstructed by Li *et al.* [214]. In 2019, Ma *et al.* [418] presented a simplified TIDT approach based on spatially-incoherent annular illumination for weakly absorbing samples, e.g., tissues and diatoms. A 3D gradient operation was adopted to remove the background, leaving only the 3D skeleton structure of the weakly absorbing sample visible in the images. Li *et al.* [419] proposed a high-speed *in vitro* intensity diffraction tomographic technique utilizing annular illumination to rapidly characterize large-volume 3D RI distributions with native bacteria, and live *Caenorhabditis elegans* specimens. Soto *et al.* [420] reported the use of TIDT for quantitative analysis of unstained DH82 cell line infected with *Leishmania infantum*. The experimental results demonstrated that the infected DH82 cells exhibit a higher RI than healthy samples.

9.4.3. Optical metrology

As a quantitative phase measurement technique in nature, TIE was also widely used in the field of optical metrology. As early as in 1998,

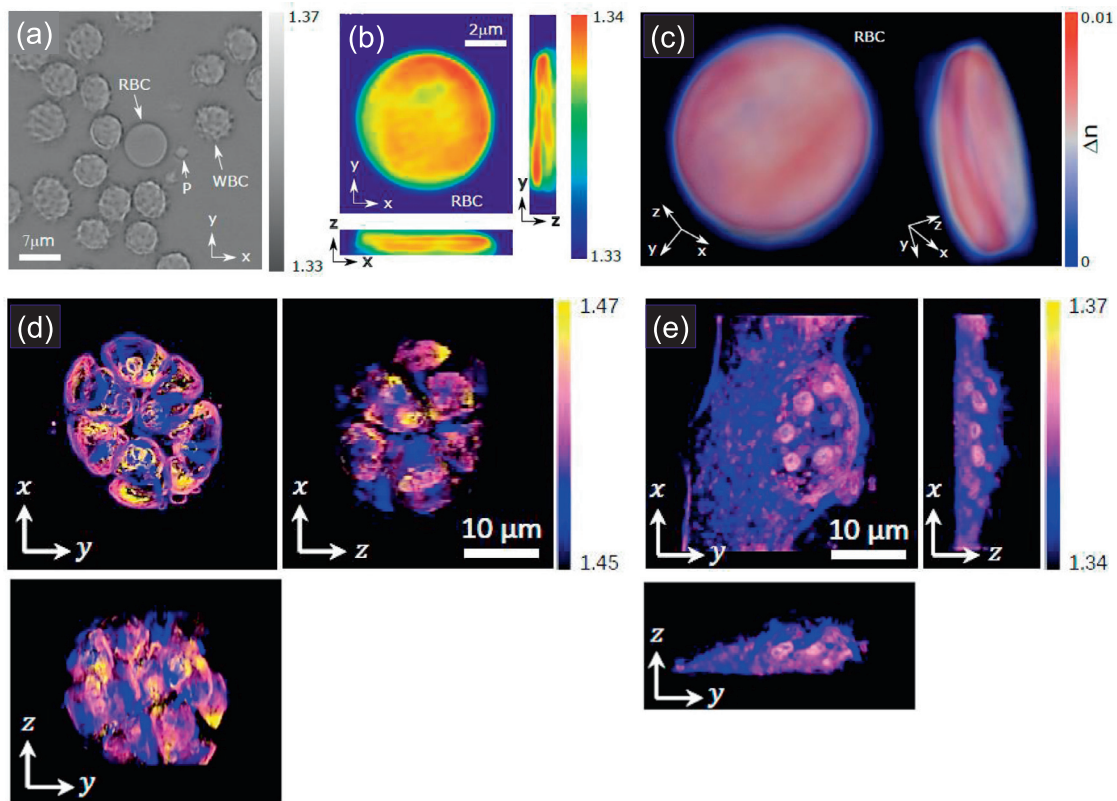


Fig. 107. 3D Tomographic reconstruction of several biological samples. (a)-(c) bright-field image, RI slices, and 3D rendering of RI distribution of human blood cell sample [381]; (d)-(e) 3D rendering of RI tomograms of a *Pandorina morum algae* and a HeLa cell [214].

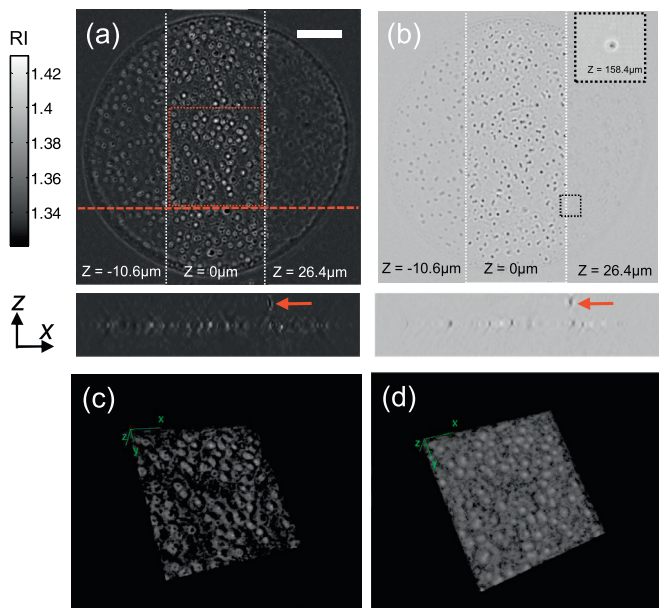


Fig. 108. 3D tomographic imaging of a slice of the uterus of *Parascaris equorum* [212]. (a) RI; (b) absorption sections at different depth. The second row shows the corresponding $x - z$ views of 3D stacks (scale ratio of z to x axis is 1:2). The arrows point out the dust particle located at a higher layer (at $z = 168.4\mu\text{m}$); (c) 3D rendering of the RI; (d) 3D rendering of the absorption distribution. Scale bar: $400\mu\text{m}$.

Barty *et al.* [175] successfully applied TIE to characterize the phase profiles of optical fibers immersed in RI-matching medium in a conventional optical microscope platform. In 2000, Barty *et al.* [176] extended this

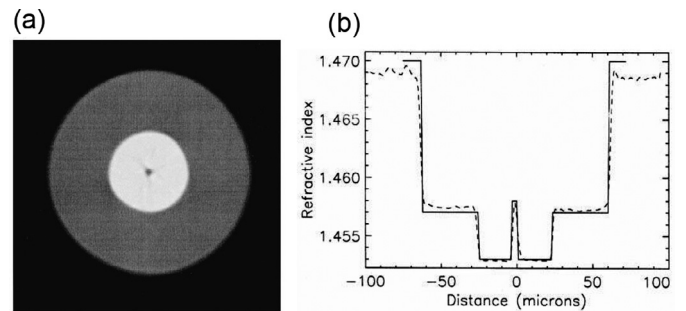


Fig. 109. Quantitative phase tomography of a single-mode fiber [176]. (a) 2D slice through the 3D reconstruction perpendicular to the rotation axis; (b) line profile through the middle of the reconstructed 3D distribution (dashed line) compared to the known RI distribution (solid line).

method to phase tomography. They measured the phase distribution of the sample at various angles by rotating the object, and 3D RI distribution of various optical fibers (single-mode, multi-mode) can be synthesized according to the inverse Radon transform (Fig. 109). In 2002, based on the axisymmetric assumption of fiber, Roberts *et al.* [184] obtained the radial RI distribution of the fiber only by performing inverse Abel transform on the phase distribution measured by TIE. The experimental results are consistent with the exponential profile obtained by the commercial profiler, and the relative error is lower than 0.0005. Due to the simplicity and effectiveness, this method has also attracted extensive attention and been further improved [185,187,421]. In 2007, Dorner and Zuegel [186] used TIE to quantitatively characterize the topological variations induced by the magnetorheological finishing process on laser rods. They successfully applied the technique to the Omega EP laser device (OMEGA EP Laser Facility). In 2010, Almoro *et al.* [69] used

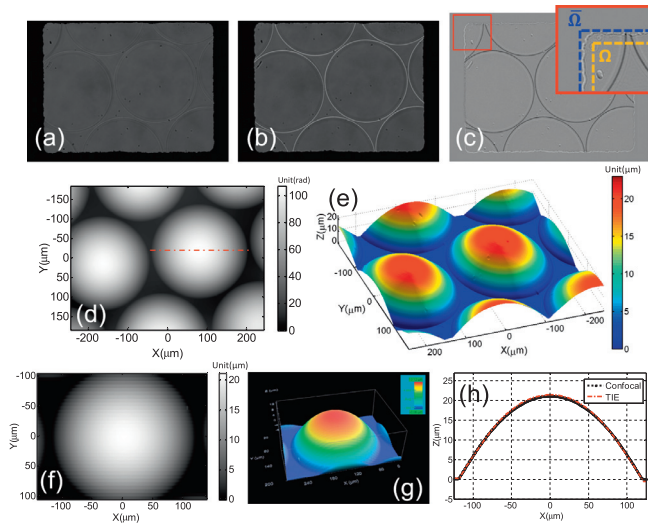


Fig. 110. Characterization of a plano-convex quartz microlens array (pitch $250\mu\text{m}$) [202]. (a) In-focus intensity; (b) defocused intensity distribution ($\Delta z = -550\mu\text{m}$); (c) axial intensity derivative, the inset shows the enlarged boxed region; (d) retrieved phase; (e) Rendered surface plot; (f) confocal microscopic result; (g) 3D topography by confocal microscopy; (h) Comparison of the line profiles of single lens.

speckle field illumination to enhance the phase contrast of the defocused intensity images for smooth objects. They demonstrated that this technique could effectively improve the phase measurement accuracy for lens characterization. In the same year, Shomali *et al.* [422] proposed the idea of applying TIE to aspheric optics testing. The traditional method is based on “null test”, *e.g.*, by using a wavefront-matching diffractive optical element (DOE) or tilted wavefront interferometry, resulting in extremely complex and expensive optical configurations. TIE only needs to capture several intensity images of the sample at different defocus distances to recover the absolute phase of the complex surface directly. It effectively bypasses the phase aliasing problem arising from the over-dense fringes around large wavefront curvature in traditional interferometry.

In 2014, Zuo *et al.* [202] also extended TIE to the characterization of micro-optical components. With use of the DCT-based TIE solver, different types of micro-optical components, *e.g.*, microlens arrays, cylindrical microlenses, Fresnel lenses were accurately characterized without any boundary artifacts which is common for FFT-based solver (the periodic boundary condition is not satisfied). Fig. 110 shows the measurement results for a plano-convex quartz microlens array with a pitch of $250\mu\text{m}$. It can be seen that the curvature radius recovered by the TIE method is $346.7\mu\text{m}$, and the result obtained by confocal microscopy is $350.4\mu\text{m}$, both these two results are consistent with the standard reference value ($350\mu\text{m}$) provided by the manufacturer. In 2015, Jenkins *et al.* [380] combined the object rotation and axial scanning to realize 3D tomographic reconstruction of the RI distributions of different types of optical fibers, including single-mode, polarization-maintaining, and photonic-crystal fibers as well as an azimuthally varying CO₂-laser-induced long-period fiber grating period as test phase objects. In 2017, Pan *et al.* [423] reconstructed the complex amplitude distribution of the laser beam based on TIE, and then used the angular spectrum propagation method to obtain the beam intensity distribution of the optical field in the arbitrary cross-section along the propagation direction, and finally calculated the M^2 quality factor of the beam. Experimental results of He-Ne and high-power fiber laser sources indicated that this method could obtain a more accurate M^2 quality factor, and the measured values coincide with the results obtained by the beam propagation analyzer. In 2019, Bunsen and Tateyama [424] used TIE to detect the phase distribution of the signal beam of holographic data storage (HDS),

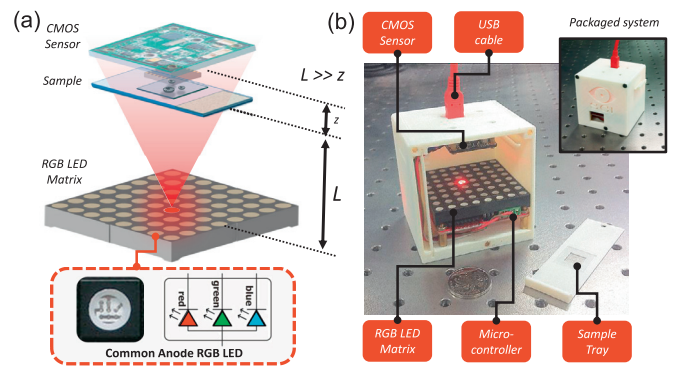


Fig. 111. Lensless TIE phase imaging system with multi-wavelength LED illuminations [212]. (a) Schematics explaining the principle of lensless imaging; (b) photograph of the system.

contributing to miniaturization of the optical system and improvement of the vibration tolerance of HDS.

9.4.4. Lensless imaging

On-chip lensless holographic microscopy is a new computational imaging technology in which the sample is placed directly on the surface of the image sensor, and the resultant diffraction pattern under quasi-coherent illumination is recorded. The image is reconstructed by numerical back propagation once the phase of the diffraction field can be retrieved. Lensfree holographic microscopy was originally proposed in 2009 by the Ozcan group of UCLA [425]. Compared with conventional lens-based microscopic systems, the lensless imaging configuration enjoys two distinct advantages: first, the imaging magnification is close to 1 due to the small distance between the object and the detector (generally less than 1mm), *i.e.*, the imaging FOV is completely determined by the size of the photosensitive area of the image sensor; second, the effective NA of the imaging system is close to 1 due to the small sample-to-sensor distance, *i.e.*, almost forward scattering light can be acquired by the image sensor. Initially, the lensless on-chip microscopic technique is mostly based on iterative phase retrieval algorithms, which recover the phase distribution based on the assumption of sparse samples [70,425] or multi-distance intensity measurements [426,427], and then the intensity and phase distribution at the object plane can be reconstructed by the angular spectrum back propagation algorithm. In 2015, Zuo *et al.* [212] developed a miniaturized lensless TIE tomographic microscope based on a programmable LED array (Fig. 111). With the interchangeability of illumination wavelength and propagation distance in the Fresnel domain, the quantitative phase of the specimen can be retrieved through TIE with multi-wavelength LED illuminations. By illuminating the sample sequentially with different LEDs across the full array, the complex fields of same object from different illumination angles can be mapped in 3D Fourier space, and the tomogram of the sample can be reconstructed. This was the first demonstration of TIE for lensless phase microscopy and diffraction tomography.

After TIE techniques gained considerable attention in microscopic imaging, it was also gradually introduced into the field of lensless microscopy. Since only 2-3 intensity images are sufficient to recover the phase distribution, TIE can be used to create an initial estimate of the phase distribution quickly and conveniently [212,429,433]. The initial estimate is then optimized by iterative phase retrieval to improve its spatial resolution and compensate for the phase discrepancy due to paraxial approximation and Teague’s assumption. On the other hand, lensless holographic microscopy shifts the problem of the limited SBP of conventional microscopic systems to the limited sampling spacing of image sensors. In principle, the key factor limiting the imaging resolution of lensless microscopy is the high-frequency signal aliasing in-

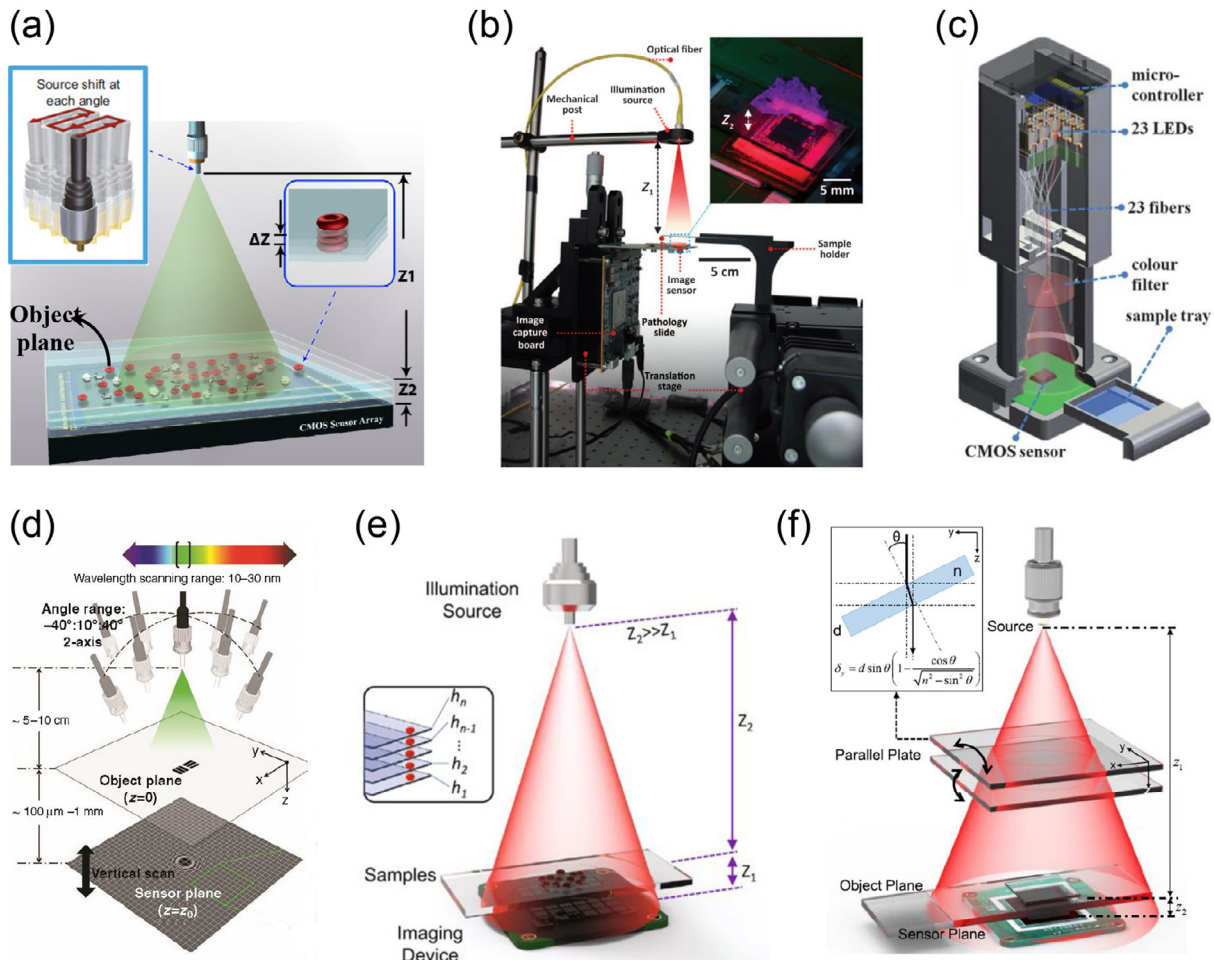


Fig. 112. Schematic diagrams of the pixel super-resolution based lensfree on-chip imaging setup. (a) Sub-pixel shifting of illumination source [428]; (b) 2D horizontal sub-pixel sensor motion [429]; (c) fiber-optic array based source scanning [430]; (d) illumination wavelength scanning [431]; (e) axial scanning with multiple sample-to-sensor distances [426]; (f) active source micro-scanning using parallel plates [432].

duced by undersampling, rather than the loss in high-frequency details induced by the limited aperture [434]. Therefore, a lot of research works in this field focused on “pixel super-resolution”, which reduces the effective pixel size through computational approaches so as to realize “sub-pixel imaging”. Representative methods include: sub-pixel shifting of illumination sources [428] [Fig. 112(a)], sub-pixel lateral motion of image sensor [429] [Fig. 112(b)], fiber-optic array-based source scanning [430] [Fig. 112(c)], illumination wavelength scanning [431] [Fig. 112(d)], axial scanning with multiple sample-to-sensor distances [426] [Fig. 112(e)], etc.. In 2018, Zhang *et al.* [432] proposed an active sub-pixel micro-scanning scheme based on two parallel plates [Fig. 112(f)], which realized high-precision controllable sub-pixel scanning with low-cost mechanical components. By combining TIE and GS phase retrieval methods, the Nyquist-Shannon sampling resolution limit imposed by the pixel size of the imaging sensor was effectively surpassed, and the high-resolution wide-FOV dynamic QPI can be achieved (effective imaging $NA \approx 0.5$).

9.4.5. Light-field imaging

In Section 7, we discussed that conditional frequency moment (centroid) of the light field could be recovered by TIE. The moment represents the average direction of the light rays passing through each spatial point. In 2013, Orth and Crozier [364] reported a computational light-field imaging method termed LMI to extract the first angular moment of light field. They derived a continuity equation to reconstruct the light field with only a pair of intensity images exhibiting a slight defocusing.

The “moment” of “light field moment imaging” means that this method can only recover the first angular moment of the light field rather than the complete 4D light field information. In order to “fill in” the missing data, Orth and Crozier [364] assumed that the angular distribution of the light field is Gaussian. Although this assumption lacks physical evidence, their experimental results indeed exhibit good visual effects. In 2014, Zuo *et al.* [207] commented that LMI could be associated with TIE under the geometric-optics limit, so all numerical solutions and axial intensity derivative estimation algorithms in TIE can be directly applied to LMI. For example, Liu *et al.* [365] employed the multi-plane high-order finite-difference method to improve the accuracy of axial intensity derivative estimation, thus reduced the noise and artifacts in the reconstructed light field. In 2015, Zuo *et al.* [208] derived the GTIE for partially coherent fields in phase space based on the transport equation of WDF. Under the geometric-optics limit, WDF is equivalent to light field [335], and TIE can be associated with the light-field imaging. Zuo *et al.* [208] deduced that under certain conditions (*e.g.*, slow-varying objects under spatially stationary illuminations), the 4D light field is highly redundant, and the complete 4D light field can be fully recovered by TIE with the pre-knowledge of the source intensity distribution. In 2018, Wang *et al.* [435] presented an efficient data acquisition approach for LMI with precise focal plane sweeping controlled by a SLM rather than mechanical movement. Chen *et al.* [436] presented an overview of 3D imaging from depth measurements, and provided a summary of the connections between the ray-based and wavefront-based 3D imaging techniques.

Table 8
Comparison of digital holography and transport of intensity equation

| Terms | Digital holography | Transport of intensity equation |
|-----------------------------|---|---|
| Data formulation | Interferometric | Non-interferometric |
| Beam coherence requirement | High | Low, can be partially coherent |
| Raw data type | Hologram (interferogram) | Through-focus intensity images (diffraction patterns) |
| Numerical processing method | Phase demodulation and numerical propagation | Solving partially differential equation |
| Main noise type | Spatial coherent noise (speckle) | Low-frequency noise (cloud-like artifacts) |
| Resolution limit | CCD size and pixel resolution (lensless geometry), coherent diffraction limit (lens-based geometry) | Diffraction limit of partially coherent imaging (illumination aperture, numerical aperture) |
| Dynamic measurement | Off-axis geometry | Need specific configurations |
| Phase recovered range | Wrapped to $[-\pi, \pi)$ (need phase unwrapping) | Continuous (no unwrapping required) |

9.4.6. Other applications

In addition to the above-mentioned imaging and measurement applications, TIE was recently extended to become a spatial phase unwrapping algorithm. Since TIE can directly obtain a continuous phase without phase wrapping, it seems that it can effectively bypass the phase unwrapping step, as is common for interferometric phase measurement methods. Based on this characteristic, in 2013, Zuo *et al.* [277] introduced TIE into digital holography. Three intensity images with slightly defocusing were numerically calculated based on Fresnel diffraction from the complex amplitude reconstructed by digital holography. By solving TIE with artifactually created intensity images, the continuous phase can be recovered. Zuo *et al.* [277] also pointed out that this method can also be used as a spatial phase unwrapping algorithm by assuming a uniform intensity distribution. It should be noted that no phase unwrapping is a major advantage of the TIE phase retrieval, but it assumes that the phase is a single-valued continuous function, which does not fundamentally address the 2π ambiguity problem in traditional phase measurement techniques. In 2016, Pandey *et al.* [437] proposed a similar phase unwrapping method based on TIE and demonstrated that this method is superior to the traditional path-dependent Goldstein method. Subsequently, Zuo *et al.* [438] pointed out that TIE under uniform intensity is a Poisson equation, and TIE phase unwrapping is essentially equivalent to the standard least-squares phase unwrapping algorithm. In 2017, Martinez-Carranza *et al.* [439] further improved TIE phase unwrapping algorithm. The idea of the least-squares phase unwrapping algorithm was used to calculate the Laplacian of the wrapped phase efficiently, the time-consuming computation associated with the wave propagation can be avoided, and the phase reconstruction accuracy can be effectively improved. In 2018, Zhou *et al.* [440] experimentally demonstrated that the continuous phase demodulated from a digital hologram by TIE is less noisy than the one obtained by conventional holographic reconstruction. In 2019, Cheng *et al.* [441] proposed a dual-wavelength phase unwrapping algorithm based on TIE. Zhao *et al.* [442] used the iterative DCT TIE solver to correct the phase unwrapping error iteratively, which improves the robustness of the TIE phase unwrapping under noisy conditions. In addition to phase unwrapping, TIE has also been applied to optical encryption [443–446], and microscopic imaging in miniaturized mobile phone platforms [447,448].

10. Conclusions and future directions

The development and fulfillment of TIE marks an important step for phase retrieval and QPI technologies: the transition from interferometric to non-interferometric, from iterative to deterministic, from phase recovery to light-field imaging, and from completely coherent to partially coherent or even incoherent. This tutorial has reviewed the current status and latest developments of TIE, including its basic principles, numerical solutions, axial intensity derivative estimation, partially coherent imaging, 3D diffraction tomography, and its representative applications in optical imaging, metrology, and microscopy. TIE provides an advanced and invaluable label-free tool for biological imaging and biomedical research. It does not require strictly spatially and temporally coherent illuminations, making it fully compatible with any off-the-shelf

bright-field microscopes. Thanks to the built-in Köhler illumination optics and aberration-corrected objective lens, high-quality quantitative phase images with diffraction-limited resolution can be easily obtained with no worries about the parasitic interference and speckle noise. It can be easily combined with fluorescence techniques to gain molecular specificity and thus provides a wider window to investigate biological processes. In addition, its single-beam (common-path) configuration is highly stable, and therefore this technique is inherently insensitive to environmental vibrations and other disturbances, eliminating the need for bulky vibration isolation platforms. Furthermore, continuous phase distributions are directly recovered so that associated difficulties in the 2D phase unwrapping can be efficiently bypassed. The above-mentioned advantages of TIE along with the rapid developments in related areas provide us substantial grounds for believing that TIE now has become a competitive alternative to its conventional interferometric counterparts and occupied an important place in the fields of optical imaging and metrology (a brief comparison between the digital holography and TIE is given in Table 8 [449]).

Though TIE has shown its great potential in the field of optical imaging and microscopy, there are still several important theoretical and technical issues deserve to be further studied in the future:

- 1) The applicability of TIE in the presence of zero intensity and or phase vortex. This has long been a controversial issue since the inception of TIE. In 2001, Allen *et al.* [167] proposed to address this problem by iterative phase retrieval, which requires to record an additional intensity image at a large defocus distance. In 2013, Lubk *et al.* [450] pointed out the mistake in Allen *et al.*'s derivation and suggested a new “exhaustive testing” method to rule out the intensity ambiguities induced by phase vortex. However, in practical experiments, it is very difficult to accurately localize all the singular points (zero intensities) inside the domain due to the noise and other disturbances.
- 2) The “deterministic” of the phase (RI) retrieval depends on certain approximations. Deterministic phase retrieval relies on the linearization of the relationship between the intensity and phase information. However, the establishment of linearization generally depends on certain approximations. For example, the traditional TIE is based on paraxial approximation and the weak defocusing approximation. The CTF or mixed transfer function methods require the sample to satisfy the weak object approximation or slowly varying object approximation. 3D diffraction tomography based on TIE is only valid for weak scattering samples under first-order Born approximation, *etc.*. However, in practical applications, these approximations are generally not easy to be satisfied strictly.
- 3) The quantitiveness of phase (RI) recovery remains debatable. Although TIE has proven to be a quantitative phase recovery method, according to the reported results among the present literature, the phase measurement accuracy is still not very high, especially when compared with the interferometric methods. In fact, although TIE greatly relaxes the stringent requirements on the light source and image system of interferometry, it imposes stricter requirements for precise calibrations of axial defocus distance, accurate acquisition of

boundary conditions, and correct implementation of numerical solutions. These factors become a bottleneck of TIE for high-precision phase measurement in practical engineering applications.

In addition to the above three challenging issues, the following research directions, we would like to highlight the following future research directions.

- 1) **Super-resolved QPI based on TIE.** At present, the spatial resolution of the phase reconstructed by TIE is still limited to the diffraction limit of the imaging system (coherent diffraction limit under coherent illuminations, and incoherent diffraction limit under the matched annular illumination [209]). To the best of our knowledge, super-resolved (beyond the incoherent diffraction limit) QPI based on TIE has not been reported so far, which is an important but challenging direction for future research.
- 2) **4D light-field retrieval based on “transport of intensity”.** For general partially coherent fields, it is difficult to retrieve its complete reconstruction by TIE due to its inherent high dimensionality (4D). But if we can further incorporate some additional information, such as combining the low-resolution field camera (Shack-Hartmann) measurement with the high-resolution through-focus intensity stack, the amount of information should be sufficient for complete 4D light-field retrieval, or equivalently, WDF/AF retrieval at full image resolution.
- 3) **Generalized phase retrieval beyond “transport of intensity”.** In TIE, the phase contrast is produced by defocusing (free-space propagation), thereby converting the invisible phase information into visible intensity signal. Nevertheless, the way to create the phase contrast is not just limited to defocusing. Many aberration functions, such as astigmatism, spherical aberration, and Zernike phase annulus, can create phase contrast even at the in-focus plane, which may provide more possibilities for designing completely new deterministic phase retrieval methods beyond the “Transport of intensity”.
- 4) **“Deep learning” enabled TIE phase retrieval.** The paraxial approximation, weak defocusing approximation, boundary artifacts, phase discrepancy, low-frequency noise, and high-frequency phase blurring are typical limiting factors of TIE, which would probably be addressed by the recent deep learning technologies [451–455]. Different from the traditional TIE, which first “establishes the forward mathematical model” and then “solves the corresponding inverse problem”, deep learning can directly establish the complex nonlinear relationship between the input (defocused intensity images) and the ideal output (quantitative phase distribution) with the help of a large amount of training data, thus opening new avenues for circumventing the limitations of the conventional TIE.

The four research directions mentioned above are mainly at the theoretical level. Of course, developing TIE microscopic imaging systems towards more integrated, more miniaturized, more easy-to-use, higher performance, and higher reliability is also an important direction for future research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Redistribution and use of MATLAB source code must retain the copyright notice, list of conditions, the disclaimer from the authors.

CRedit authorship contribution statement

Chao Zuo: Conceptualization, Writing - original draft, Software, Data curation, Visualization, Supervision, Project administration, Funding acquisition. **Jiaji Li:** Investigation, Writing - review & editing. **Jiasong Sun:** Investigation, Writing - review & editing. **Yao Fan:** Data

curation, Writing - review & editing. **Jialin Zhang:** Software, Writing - review & editing. **Linpeng Lu:** Writing - review & editing. **Runnan Zhang:** Writing - review & editing. **Bowen Wang:** Visualization. **Lei Huang:** Writing - review & editing. **Qian Chen:** Resources, Supervision, Project administration, Funding acquisition.

Acknowledgments

The authors thank J. Ding for help with the schematic drawing. This work was supported by the [National Natural Science Foundation of China \(61722506\)](#), Leading Technology of Jiangsu Basic Research Plan (BK20192003), Final Assembly “13th Five-Year Plan” Advanced Research Project of China (30102070102), Equipment Advanced Research Fund of China (61404150202), National Defense Science and Technology Foundation of China (2019-JCJQ-JJ-381), Outstanding Youth Foundation of Jiangsu Province (BK20 170034), The [Key Research and Development Program of Jiangsu Province \(BE2017162\)](#), “333 Engineering” Research Project of Jiangsu Province (BRA2016407), [Fundamental Research Funds for the Central Universities \(30920032101\)](#), and Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (30918 01410411).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.optlaseng.2020.106187](https://doi.org/10.1016/j.optlaseng.2020.106187)

References

- [1] Cowley JM. *Diffraction physics*. Elsevier; 1995.
- [2] Goodman JW. *Introduction to Fourier Optics*. Roberts and Company Publishers; 2005.
- [3] Griffiths D.J. *Introduction to electrodynamics*. 2005.
- [4] Hecht E. *Optics*. 4 edition. Reading, Mass: Addison-Wesley; 2001. ISBN 978-0-8053-8566-3.
- [5] Born M, Wolf E, Bhatia AB, Clemmow PC, Gabor D, Stokes AR, Taylor AM, Wayman PA, Wilcock WL. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 7th edition. Cambridge ; New York: Cambridge University Press; 1999. ISBN 978-0-521-64222-4.
- [6] High-speed camera. 2017. Page Version ID: 786672531;
- [7] Oppenheim AV, Lim JS. The importance of phase in signals. *Proceedings of the IEEE* 1981;69(5):529–41.
- [8] Giloh H, Sedat JW. Fluorescence microscopy: reduced photobleaching of rhodamine and fluorescein protein conjugates by n-propyl gallate. *Science* 1982;217(4566):1252–5.
- [9] Webb RH. Confocal optical microscopy. *Rep Prog Phys* 1996;59(3):427.
- [10] Diaspro A, et al. *Confocal and two-photon microscopy: foundations, applications, and advances*, vol 1. Wiley-Liss New York; 2002.
- [11] Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J Microsc* 2000;198(Pt 2):82–7.
- [12] Hell SW, Wichmann J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt Lett* 1994;19(11):780–2.
- [13] Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 2006;313(5793):1642–5.
- [14] Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 2006;3:793–6. doi:10.1038/nmeth929.
- [15] Stephens DJ, Allan VJ. Light microscopy techniques for live cell imaging. *Science* 2003;300(5616):82–6.
- [16] Zernike F. Phase contrast, a new method for the microscopic observation of transparent objects. *Physica* 1942;9:686–98.
- [17] Nomarski G. Differential microinterferometer with polarized waves; 1955. 9S–11S.
- [18] Popescu G. *Quantitative Phase Imaging of Cells and Tissues*. McGraw-Hill Professional; 2011.
- [19] Park Y, Depeursinge C, Popescu G. Quantitative phase imaging in biomedicine. *Nat Photonics* 2018;12(10):578–89.
- [20] Abramovici A, Althouse WE, Drever RW, Grsel Y, Kawamura S, Raab FJ, Shoemaker D, Sievers L, Spero RE, Thorne KS, Vogt RE, Weiss R, Whitcomb SE, Zucker ME. Ligo: The laser interferometer gravitational-wave observatory. *Science* 1992;256(5055):325–33.
- [21] Collaboration LS, Collaboration V. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys Rev Lett* 2016;116(6):061102.
- [22] Lekberg O. Electronic speckle pattern interferometry. *Phys in Tech* 1980;11(1):16.
- [23] Wang W-C, Hwang C-H, Lin S-Y. Vibration measurement by the time-averaged electronic speckle pattern interferometry methods. *Appl Optics* 1996;35(22):4502–9.

- [24] Schwarz CJ, Kuznetsova Y, Brueck SRJ. Imaging interferometric microscopy. *Opt Lett* 2003;28(16):1424–6.
- [25] Popescu G, Ikeda T, Dasari RR, Feld MS. Diffraction phase microscopy for quantifying cell structure and dynamics. *Opt Lett* 2006;31(6):775–7.
- [26] Kuznetsova Y, Neumann A, Brueck SRJ. Imaging interferometric microscopy—approaching the linear systems limits of optical resolution. *Opt Express* 2007;15(11):6651–63.
- [27] Cuhe E, Bevilacqua F, Depeursinge C. Digital holography for quantitative phase-contrast imaging. *Opt Lett* 1999a;24(5):291–3.
- [28] Cuhe E, Marquet P, Depeursinge C. Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms. *Appl Optics* 1999b;38(34):6994–7001.
- [29] Cuhe E, Marquet P, Depeursinge C. Spatial filtering for zero-order and twin-image elimination in digital off-axis holography. *Appl Optics* 2000;39(23):4070–5.
- [30] Schnars U, Jptner WPO. Digital recording and numerical reconstruction of holograms. *Meas Sci Technol* 2002;13(9):R85.
- [31] Schnars U, Jueptner W. Digital holography: digital hologram recording, numerical reconstruction, and related techniques. Springer Science & Business Media; 2005.
- [32] Marquet P, Rappaz B, Magistretti PJ, Cuhe E, Emery Y, Colomb T, Depeursinge C. Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy. *Opt Lett* 2005;30(5):468–70.
- [33] Kemper B, Langehanenberg P, Von Bally G. Digital holographic microscopy: A new method for surface analysis and marker-free dynamic life cell imaging. *Optik Photonik* 2007;2(2):41–4.
- [34] Kemper B, Bally Gv. Digital holographic microscopy for live cell applications and technical inspection. *Appl Optics* 2008;47(4):A52–61.
- [35] Kim MK. Digital holographic microscopy. In: *Digital Holographic Microscopy*. Springer; 2011. p. 149–90.
- [36] Wang Z, Millet L, Mir M, Ding H, Unarunotai S, Rogers J, Gillette MU, Popescu G. Spatial light interference microscopy (SLIM). *Opt Express* 2011;19(2):1016–26.
- [37] Bhaduri B, Pham H, Mir M, Popescu G. Diffraction phase microscopy with white light. *Opt Lett* 2012;37(6):1094–6.
- [38] Bon P, Maucort G, Wattellier B, Monneret S. Quadriwave lateral shearing interferometry for quantitative phase microscopy of living cells. *Opt Express* 2009;17(15):13080–94.
- [39] Girshovitz P, Shaked NT. Compact and portable low-coherence interferometer with off-axis geometry for quantitative phase microscopy and nanoscopy. *Opt Express* 2013;21(5):5701–14.
- [40] Shack RV. Production and use of a lecticular hartmann screen. *J Opt Soc Am* 1971;61:656–61.
- [41] Hartmann J. Bemerkungen über den bau und die justirung von spektrographen. *Zt Instrumentenk* 1990;20(47):17–27.
- [42] Platt BC, Shack R. History and principles of shack-hartmann wavefront sensing. *J Refract Surg* 2001;17(5):S573–7.
- [43] Ragazzoni R. Pupil plane wavefront sensing with an oscillating prism. *J Mod Opt* 1996;43(2):289–93.
- [44] Esposito S, Riccardi A. Pyramid wavefront sensor behavior in partial correction adaptive optic systems. *Astron Astrophys* 2001;369(2):L9–L12.
- [45] Ragazzoni R, Diolaiti E, Vernet E. A pyramid wavefront sensor with no dynamic modulation. *Opt Commun* 2002;208(1–3):51–60.
- [46] Neil MAA, Booth MJ, Wilson T. New modal wave-front sensor: a theoretical analysis. *J Opt Soc Am A* 2000;17:1098–107.
- [47] Booth MJ. Wave front sensor-less adaptive optics: a model-based approach using sphere packings. *Opt Express* 2006;14:1339–52.
- [48] Schfer B, Mann K. Determination of beam parameters and coherence properties of laser radiation by use of an extended hartmann-shack wave-front sensor. *Appl Optics* 2002;41:2809–17.
- [49] Schfer B, Lbbecke M, Mann K. Hartmann-shack wave front measurements for real time determination of laser beam propagation parameters. *Rev Sci Instrum* 2006;77:053103.
- [50] Pfund J, Lindlein N, Schwider J, Burow R, Blmel T, Ellsner KE. Absolute sphericity measurement: a comparative study of the use of interferometry and a shack-hartmann sensor. *Opt Lett* 1998;23:742–4.
- [51] Greivenkamp JE, Smith DG, Gappinger RO, Williby GA. Optical testing using shack-hartmann wavefront sensors. In: *Optical Engineering for Sensing and Nanotechnology* (ICOSN 2001), vol. 4416. International Society for Optics and Photonics; 2001. p. 260–3.
- [52] Dayton D, Goglewski J, Pierson B, Spielbusch B. Atmospheric structure function measurements with a Shack-Hartmann wave-front sensor. *Opt Lett* 1992;17:1737–9.
- [53] Ricklin JC, Davidson FM. Atmospheric turbulence effects on a partially coherent Gaussian beam: implications for free-space laser communication. *J Opt Soc Am A* 2002;19:1794–802.
- [54] Booth MJ. Adaptive optics in microscopy. *Philos Trans R Soc Lond Math Phys Eng Sci* 2007;365:2829–43.
- [55] Cha JW, Ballesta J, So PTC. Shack-hartmann wavefront-sensor-based adaptive optics system for multiphoton microscopy. *J Biomed Opt* 2010;15:046022–046022–10.
- [56] Liang J, Grimm B, Goelz S, Bille JF. Objective measurement of wave aberrations of the human eye with the use of a hartmann-shack wave-front sensor. *J Opt Soc Am A* 1994;11:1949–57.
- [57] Moreno-Barruso E, Navarro R. Laser ray tracing versus hartmann-shack sensor for measuring optical aberrations in the human eye. *J Opt Soc Am A* 2000;17:974–85.
- [58] Kohnen T, Koch DD, et al. Cataract and refractive surgery. Springer; 2006.
- [59] Gerchberg RW. Phase determination from image and diffraction plane pictures in the electron microscope. *Optik* 1971;34:275–84.
- [60] Gerchberg RW, Saxton WO. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik (Jena)* 1972;35:237.
- [61] Miao J, Charalambous P, Kirz J, Sayre D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* 1999;400(6742):342–4.
- [62] Zuo JM, Vartanyants I, Gao M, Zhang R, Nagahara LA. Atomic resolution imaging of a carbon nanotube from diffraction intensities. *Science* 2003;300(5624):1419–21.
- [63] Marchesini S, He H, Chapman HN, Hau-Riege SP, Noy A, Howells MR, Weierstall U, Spence JCH. X-ray image reconstruction from a diffraction pattern alone. *Phys Rev B* 2003;68:140101.
- [64] Eisebitt S, Lning J, Schlotter WF, Lrgen M, Hellwig O, Eberhardt W, Sthr J. Lensless imaging of magnetic nanostructures by X-ray spectro-holography. *Nature* 2004;432:885–8.
- [65] Gonsalves RA, Chidlaw R. Wavefront sensing by phase retrieval. In: *Applications of Digital Image Processing III*, vol. 207. International Society for Optics and Photonics; 1979. p. 32–9.
- [66] Guyon O. Limits of adaptive optics for high-contrast imaging. *ApJ* 2005;629(1):592.
- [67] Pedrini G, Osten W, Zhang Y. Wave-front reconstruction from a sequence of interferograms recorded at different planes. *Opt Lett* 2005;30(8):833–5.
- [68] Anand A, Pedrini G, Osten W, Almoró P. Wavefront sensing with random amplitude mask and phase retrieval. *Opt Lett* 2007;32(11):1584–6.
- [69] Almoró PF, Pedrini G, Gundu PN, Osten W, Hanson SG. Phase microscopy of technical and biological samples through random phase modulation with a diffuser. *Opt Lett* 2010;35(7):1028–30.
- [70] Mudanyali O, Tseng D, Oh C, Serhan OI, Sencan I, Bishara W, Oztoprak C, Seo S, Khademhosseini B, Ozcan A. Compact, light-weight and cost-effective microscope based on lensless incoherent holography for telemedicine applications. *Lab Chip* 2010;10(11):1417–28.
- [71] Tseng D, Mudanyali O, Oztoprak C, Serhan OI, Sencan I, Yaglidere O, Ozcan A. Lensfree microscopy on a cellphone. *Lab Chip* 2010;10(14):1787–92.
- [72] Wackerman CC, Yagle AE. Use of fourier domain real-plane zeros to overcome a phase retrieval stagnation. *J Opt Soc Am A* 1991;8(12):1898–904.
- [73] Lu G, Zhang Z, Yu FTS, Tanone A. Pendulum iterative algorithm for phase retrieval from modulus data. *Opt Eng* 1994;33(2):548–55.
- [74] Takajo H, Takahashi T, Kawanami H, Ueda R. Numerical investigation of the iterative phase-retrieval stagnation problem: territories of convergence objects and holes in their boundaries. *J Opt Soc Am A* 1997;14(12):3175–87.
- [75] Misell DL. A method for the solution of the phase problem in electron microscopy. *J Phys D: Appl Phys* 1973;6(1):L6.
- [76] Allen LJ, Oxley MP. Phase retrieval from series of images obtained by defocus variation. *Opt Commun* 2001;199:65–75.
- [77] Zhang Y, Pedrini G, Osten W, Tiziani HJ. Whole optical wave field reconstruction from double or multi in-line holograms by phase retrieval algorithm. *Opt Express* 2003;11(24):3234–41.
- [78] Bao P, Zhang F, Pedrini G, Osten W. Phase retrieval using multiple illumination wavelengths. *Opt Lett* 2008;33(4):309–11.
- [79] Zhang F, Pedrini G, Osten W. Phase retrieval of arbitrary complex-valued fields through aperture-plane modulation. *Phys Rev A* 2007;75(4):043805.
- [80] Fannjiang A, Liao W. Phase retrieval with random phase illumination. *J Opt Soc Am A* 2012;29(9):1847–59.
- [81] Almoró P, Pedrini G, Osten W. Complete wavefront reconstruction using sequential intensity measurements of a volume speckle field. *Appl Optics* 2006;45(34):8596–605.
- [82] Gao P, Pedrini G, Zuo C, Osten W. Phase retrieval using spatially modulated illumination. *Opt Lett* 2014;39(12):3615.
- [83] Bates RHT. Uniqueness of solutions to two-dimensional fourier phase problems for localized and positive images. *Computer Vision, Graphics, and Image Processing* 1984;25(2):205–17.
- [84] Fienup JR, Wackerman CC. Phase-retrieval stagnation problems and solutions. *J Opt Soc Am A* 1986;3(11):1897–907.
- [85] Seldin JH, Fienup JR. Numerical investigation of the uniqueness of phase retrieval. *J Opt Soc Am A* 1990;7(3):412–27.
- [86] Guizar-Sicairos M, Fienup JR. Understanding the twin-image problem in phase retrieval. *J Opt Soc Am A* 2012;29(11):2367–75.
- [87] Fienup JR. Phase retrieval algorithms: a comparison. *Appl Optics* 1982;21(15):2758–69.
- [88] Fienup JR. Reconstruction of an object from the modulus of its Fourier transform. *Opt Lett* 1978;3(1):27–9.
- [89] Bauschke HH, Combettes PL, Luke DR. Hybrid projection-reflection method for phase retrieval. *J Opt Soc Am A* 2003;20(6):1025–34.
- [90] Elser V. Phase retrieval by iterated projections. *J Opt Soc Am A* 2003;20(1):40–55.
- [91] Luke DR. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems* 2005;21(1):37–50.
- [92] Oszlányi G, Sütő A. Ab initio structure solution by charge flipping. *Acta Crystallogr A* 2004;60(2):134–41.
- [93] Marchesini S. Invited article: A unified evaluation of iterative projection algorithms for phase retrieval. *Rev Sci Instrum* 2007;78(1):011301.
- [94] Hegerl R, Hoppe W. Dynamische Theorie der Kristallstrukturanalyse durch Elektronenbeugung im inhomogenen Primärstrahlwellenfeld. *Berichte der Bunsengesellschaft für physikalische Chemie* 1970;74(11):1148–54.
- [95] Faulkner HML, Rodenburg JM. Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm. *Phys Rev Lett* 2004;93(2):023903.

- [96] Faulkner HML, Rodenburg JM. Error tolerance of an iterative phase retrieval algorithm for moveable illumination microscopy. *Ultramicroscopy* 2005;103(2):153–64.
- [97] Guizar-Sicairos M, Fienup JR. Phase retrieval with transverse translation diversity: a nonlinear optimization approach. *Opt Express* 2008;16(10):7264–78.
- [98] Thibault P, Dierolf M, Menzel A, Bunk O, David C, Pfeiffer F. High-resolution scanning x-ray diffraction microscopy. *Science* 2008;321(5887):379–82.
- [99] Maiden AM, Rodenburg JM. An improved Ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy* 2009;109(10):1256–62.
- [100] Thibault P, Dierolf M, Bunk O, Menzel A, Pfeiffer F. Probe retrieval in Ptychographic coherent diffractive imaging. *Ultramicroscopy* 2009;109(4):338–43.
- [101] Thibault P, Guizar-Sicairos M. Maximum-likelihood refinement for coherent diffractive imaging. *New J Phys* 2012;14(6):063004.
- [102] Maiden A, Johnson D, Li P. Further improvements to the Ptychographical iterative engine. *Optica* 2017;4(7):736–45.
- [103] Maiden AM, Humphry MJ, Sarahan MC, Kraus B, Rodenburg JM. An annealing algorithm to correct positioning errors in Ptychography. *Ultramicroscopy* 2012a;120:64–72.
- [104] Beckers M, Senkbeil T, Gorniak T, Giewekemeyer K, Salditt T, Rosenhahn A. Drift correction in Ptychographic diffractive imaging. *Ultramicroscopy* 2013;126:44–7.
- [105] Zhang F, Peterson I, Vila-Comamala J, Diaz A, Berenguer F, Bean R, Chen B, Menzel A, Robinson IK, Rodenburg JM. Translation position determination in Ptychographic coherent diffraction imaging. *Opt Express* 2013;21(11):13592.
- [106] Thibault P, Menzel A. Reconstructing state mixtures from diffraction measurements. *Nature* 2013;494(7435):68–71.
- [107] Batey DJ, Claus D, Rodenburg JM. Information multiplexing in Ptychography. *Ultramicroscopy* 2014;138:13–21.
- [108] Clark JN, Huang X, Harder RJ, Robinson IK. Dynamic imaging using Ptychography. *Phys Rev Lett* 2014;112(11):113901.
- [109] Karl R, Bevis C, Lopez-Rios R, Reichanadter J, Gardner D, Porter C, Shanblatt E, Tanksalvala M, Mancini GF, Murnane M, Kapteyn H, Adams D. Spatial, spectral, and polarization multiplexed Ptychography. *Opt Express* 2015;23(23):30250.
- [110] Maiden AM, Humphry MJ, Zhang F, Rodenburg JM. Superresolution imaging via Ptychography. *J Opt Soc Am A* 2011;28(4):604–12.
- [111] Humphry MJ, Kraus B, Hurst AC, Maiden AM, Rodenburg JM. Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging. *Nat Commun* 2012;3:730.
- [112] Stockmar M, Cloetens P, Zanette I, Enders B, Dierolf M, Pfeiffer F, Thibault P. Near-field Ptychography: phase retrieval for inline holography using a structured illumination. *Sci Rep* 2013;3(1).
- [113] Takahashi Y, Suzuki A, Furutaku S, Yamauchi K, Kohmura Y, Ishikawa T. High-resolution and high-sensitivity phase-contrast imaging by focused hard x-ray Ptychography with a spatial filter. *Appl Phys Lett* 2013;102(9):094102.
- [114] Maiden AM, Humphry MJ, Rodenburg JM. Ptychographic transmission microscopy in three dimensions using a multi-slice approach. *J Opt Soc Am A* 2012b;29(8):1606–14.
- [115] Godden TM, Suman R, Humphry MJ, Rodenburg JM, Maiden AM. Ptychographic microscope for three-dimensional imaging. *Opt Express* 2014;22(10):12513.
- [116] Suzuki A, Furutaku S, Shimomura K, Yamauchi K, Kohmura Y, Ishikawa T, Takahashi Y. High-resolution multislice x-ray Ptychography of extended thick objects. *Phys Rev Lett* 2014;112(5).
- [117] Shimomura K, Suzuki A, Hirose M, Takahashi Y. Precession x-ray Ptychography with multislice approach. *Phys Rev B* 2015;91(21).
- [118] Rodenburg JM, Hurst AC, Cullis AG, Dobson BR, Pfeiffer F, Bunk O, David C, Jefimovs K, Johnson I. Hard-x-ray lensless imaging of extended objects. *Phys Rev Lett* 2007a;98(3):034801.
- [119] Giewekemeyer K, Thibault P, Kalbfleisch S, Beerlink A, Kewish CM, Dierolf M, Pfeiffer F, Salditt T. Quantitative biological imaging by Ptychographic x-ray diffraction microscopy. *Proceedings of the National Academy of Sciences* 2010;107(2):529–34.
- [120] Maiden AM, Morrison GR, Kaulich B, Gianoncelli A, Rodenburg JM. Soft X-ray spectromicroscopy using Ptychography with randomly phased illumination. *Nat Commun* 2013;4:1669.
- [121] Rodenburg JM, Hurst AC, Cullis AG. Transmission microscopy without lenses for objects of unlimited size. *Ultramicroscopy* 2007b;107(2-3):227–31.
- [122] Hue F, Rodenburg JM, Maiden AM, Sweeney F, Midgley PA. Wave-front phase retrieval in transmission electron microscopy via Ptychography. *Phys Rev B* 2010;82(12).
- [123] Hue F, Rodenburg JM, Maiden AM, Midgley PA. Extended Ptychography in the transmission electron microscope: Possibilities and limitations. *Ultramicroscopy* 2011;111(8):1117–23.
- [124] Brady GR, Guizar-Sicairos M, Fienup JR. Optical wavefront measurement using phase retrieval with transverse translation diversity. *Opt Express* 2009;17(2):624–39.
- [125] Maiden AM, Rodenburg JM, Humphry MJ. Optical Ptychography: a practical implementation with useful resolution. *Opt Lett* 2010;35(15):2585–7.
- [126] Marrison J, Rty L, Marriott P, O'Toole P. Ptychography - a label free, high-contrast imaging technique for live cells using quantitative phase information. *Sci Rep* 2013;3(1).
- [127] Zheng G, Horstmeyer R, Yang C. Wide-field, high-resolution Fourier Ptychographic microscopy. *Nat Photonics* 2013;7(9):739–45.
- [128] Ou X, Zheng G, Yang C. Embedded pupil function recovery for Fourier Ptychographic microscopy. *Opt Express* 2014;22(5):4960.
- [129] Sun J, Chen Q, Zhang Y, Zuo C. Efficient positional misalignment correction method for Fourier Ptychographic microscopy. *Biomed Opt Express* 2016a;7(4):1336.
- [130] Yeh L-H, Dong J, Zhong J, Tian L, Chen M, Tang G, Soltanolkotabi M, Waller L. Experimental robustness of Fourier Ptychography phase retrieval algorithms. *Opt Express* 2015;23(26):33214.
- [131] Dong S, Shiradkar R, Nanda P, Zheng G. Spectral multiplexing and coherent-state decomposition in Fourier Ptychographic imaging. *Biomed Opt Express* 2014;5(6):1757.
- [132] Tian L, Li X, Ramchandran K, Waller L. Multiplexed coded illumination for Fourier Ptychography with an LED array microscope. *Biomed Opt Express* 2014;5(7):2376–89.
- [133] Sun J, Chen Q, Zhang Y, Zuo C. Sampling criteria for Fourier Ptychographic microscopy in object space and frequency space. *Opt Express* 2016b;24(14):15765.
- [134] Li P, Batey DJ, Edo TB, Rodenburg JM. Separation of three-dimensional scattering effects in tilt-series Fourier Ptychography. *Ultramicroscopy* 2015;158:1–7.
- [135] Tian L, Waller L. 3d intensity and phase imaging from light field measurements in an LED array microscope. *Optica* 2015a;2(2):104.
- [136] Horstmeyer R, Chen RY, Ou X, Ames B, Tropp JA, Yang C. Solving Ptychography with a convex relaxation. *New J Phys* 2015;17(5):053044.
- [137] Zuo C, Sun J, Chen Q. Adaptive step-size strategy for noise-robust Fourier Ptychographic microscopy. *Opt Express* 2016a;24(18):20724.
- [138] Sun J, Zuo C, Zhang L, Chen Q. Resolution-enhanced Fourier Ptychographic microscopy based on high-numerical-aperture illuminations. *Sci Rep* 2017;7(1):1187.
- [139] Williams GJ, Quiney HM, Peele AG, Nugent KA. Coherent diffractive imaging and partial coherence. *Phys Rev B* 2007;75:104102. doi:10.1103/PhysRevB.75.104102.
- [140] Teague MR. Irradiance moments: their propagation and use for unique retrieval of phase. *J Opt Soc Am* 1982;72(9):1199–209.
- [141] Teague MR. Deterministic phase retrieval: a green's function solution. *J Opt Soc Am A* 1983;73(11):1434–41.
- [142] Streibl N. Phase imaging by the transport equation of intensity. *Opt Commun* 1984;49(1):6–10.
- [143] Ichikawa K, Lohmann AW, Takeda M. Phase retrieval based on the irradiance transport equation and the Fourier transform method: experiments. *Appl Optics* 1988;27(16):3433–6.
- [144] Roddier F. Wavefront sensing and the irradiance transport equation. *Appl Optics* 1990;29(10):1402–3.
- [145] Roddier F, Roddier C, Roddier N. Curvature Sensing: A New Wavefront Sensing Method. In: Morris GM, editor. *Statistical Optics*, vol. 0976. SPIE; 1988. p. 203–9. International Society for Optics and Photonics
- [146] Roddier F. Curvature sensing and compensation: a new concept in adaptive optics. *Appl Optics* 1988;27(7):1223–5.
- [147] Roddier NA. Algorithms for wavefront reconstruction out of curvature sensing data. In: *Active and Adaptive Optical Systems*, vol. 1542. International Society for Optics and Photonics; 1991. p. 120–9.
- [148] Roddier C, Roddier F. Wave-front reconstruction from defocused images and the testing of ground-based optical telescopes. *J Opt Soc Am A* 1993;10(11):2277–87.
- [149] Roddier F, Roddier C. Wavefront reconstruction using iterative Fourier transforms. *Appl Optics* 1991;30:1325–7.
- [150] Roddier F, Northcott M, Graves JE. A simple low-order adaptive optics system for near-infrared applications. *Publ Astron Soc Pac* 1991;103(659):131.
- [151] Roddier FJ, Anuskiewicz J, Graves JE, Northcott MJ, Roddier CA. Adaptive optics at the university of hawaii i: current performance at the telescope. In: *Adaptive Optics in Astronomy*, vol. 2201. International Society for Optics and Photonics; 1994. p. 2–9.
- [152] Graves JE, Northcott MJ, Roddier FJ, Roddier CA, Close LM. First light for hokupa'a: 36-element curvature AO system at UH. In: *Adaptive Optical System Technologies*, vol. 3353. International Society for Optics and Photonics; 1998. p. 34–43.
- [153] Arsenault R, Alonso J, Bonnet H, Brynneel J, Delabre B, Donaldson R, Dupuy C, Fedrigo E, Farinato J, Hubin NN, et al. Macao-vlti: an adaptive optics system for the eso interferometer. In: *Adaptive Optical System Technologies II*, vol. 4839. International Society for Optics and Photonics; 2003. p. 174–85.
- [154] Hayano Y, Takami H, Guyon O, Oya S, Hattori M, Saito Y, Watanabe M, Murakami N, Minowa Y, Ito M, et al. Current status of the laser guide star adaptive optics system for Subaru telescope. In: *Adaptive Optics Systems*, vol. 7015. International Society for Optics and Photonics; 2008. p. 701510.
- [155] Gureyev TE, Roberts A, Nugent KA. Partially coherent fields, the transport-of-intensity equation, and phase uniqueness. *J Opt Soc Am A, JOS A* 1995a;12(9):1942–6.
- [156] Gureyev TE, Roberts A, Nugent KA. Phase retrieval with the transport-of-intensity equation: matrix solution with use of Zernike polynomials. *J Opt Soc Am A* 1995b;12(9):1932–41.
- [157] Gureyev TE, Nugent KA. Phase retrieval with the transport-of-intensity equation. II. orthogonal series solution for nonuniform illumination. *J Opt Soc Am A* 1996;13(8):1670–82.
- [158] Gureyev TE, Nugent KA. Rapid quantitative phase imaging using the transport of intensity equation. *Opt Commun* 1997;133(1):339–46.
- [159] Paganin D, Nugent KA. Noninterferometric phase imaging with partially coherent light. *Phys Rev Lett* 1998;80(12):2586.
- [160] Snigirev A, Snigireva I, Kohn V, Kuznetsov S, Schelokov I. On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Rev Sci Instrum* 1995;66(12):5486–92.
- [161] Nugent KA, Gureyev TE, Cookson DF, Paganin D, Barnea Z. Quantitative phase imaging using hard x rays. *Phys Rev Lett* 1996;77:2961.
- [162] Wilkins SW, Gureyev TE, Gao D, Pogany A, Stevenson AW. Phase-contrast imaging using polychromatic hard x-rays. *Nature* 1996;384(6607):335.

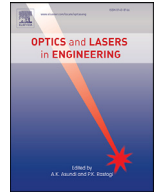
- [163] Cloetens P, Ludwig W, Baruchel J, Van Dyck D, Van Landuyt J, Guigay JP, Schlenker M. Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays. *Appl Phys Lett* 1999;75(19):2912–14.
- [164] Allman BE, McMahon PJ, Nugent KA, Paganin D, Jacobson DL, Arif M, Werner SA. Phase radiography with neutrons. *Nature* 2000;408:158.
- [165] McMahon PJ, Allman BE, Jacobson DL, Arif M, Werner SA, Nugent KA. Quantitative phase radiography with polychromatic neutrons. *Phys Rev Lett* 2003;91(14):145502.
- [166] Bajt S, Barty A, Nugent KA, McCartney M, Wall M, Paganin D. Quantitative phase-sensitive imaging in a transmission electron microscope. *Ultramicroscopy* 2000;83(1-2):67–73.
- [167] Allen LJ, Faulkner HML, Nugent KA, Oxley MP, Paganin D. Phase retrieval from the presence of first-order vortices. *Phys Rev E* 2001a;63(3):037602.
- [168] Allen LJ, Faulkner HML, Oxley MP, Paganin D. Phase retrieval and aberration correction in the presence of vortices in high-resolution transmission electron microscopy. *Ultramicroscopy* 2001b;88(2):85–97.
- [169] McMahon PJ, Barone-Nugent ED, Allman BE, Nugent KA. Quantitative phase-amplitude microscopy II: differential interference contrast imaging for biological TEM. *J Microsc* 2002;206(3):204–8.
- [170] Volkov VV, Zhu Y. Phase imaging and nanoscale currents in phase objects imaged with fast electrons. *Phys Rev Lett* 2003;91(4):043904.
- [171] Beleggia M, Schofield MA, Volkov VV, Zhu Y. On the transport of intensity technique for phase retrieval. *Ultramicroscopy* 2004;102(1):37–49.
- [172] Volkov VV, Zhu Y. Lorentz phase microscopy of magnetic materials. *Ultramicroscopy* 2004;98(2):271–81.
- [173] McVitie S, Cushley M. Quantitative fresnel lorentz microscopy and the transport of intensity equation. *Ultramicroscopy* 2006;106(4-5):423–31.
- [174] Petersen TC, Keast VJ, Paganin DM. Quantitative TEM-based phase retrieval of mgo nano-cubes using the transport of intensity equation. *Ultramicroscopy* 2008;108(9):805–15.
- [175] Barty A, Nugent KA, Paganin D, Roberts A. Quantitative optical phase microscopy. *Opt Lett* 1998;23(11):817–19.
- [176] Barty A, Nugent KA, Roberts A, Paganin D. Quantitative phase tomography. *Opt Commun* 2000;175(4-6):329–36.
- [177] Barone-Nugent ED, Barty A, Nugent KA. Quantitative phase-amplitude microscopy I: optical microscopy. *J Microsc* 2002;206(3):194–203.
- [178] Streibl N. Three-dimensional imaging by a microscope. *J Opt Soc Am A, JOSAA* 1985;2(2):121–7.
- [179] Sheppard CJR. Three-dimensional phase imaging with the intensity transport equation. *Appl Optics* 2002;41:5951–5.
- [180] Curl CL, Bellair CJ, Harris PJ, Allman BE, Roberts A, Nugent KA, Delbridge LMD. Quantitative phase microscopy: a new tool for investigating the structure and function of unstained live cells. *Clin Exp Pharmacol Physiol* 2004;31(12):896–901.
- [181] Curl CL, Bellair CJ, Harris PJ, Allman BE, Roberts A, Nugent KA, Delbridge LMD. Single cell volume measurement by quantitative phase microscopy (QPM): a case study of erythrocyte morphology. *Cell Physiol Biochem* 2006;17(5-6):193–200.
- [182] Ross GJ, Bigelow AW, Randers-Pehrson G, Peng CC, Brenner DJ. Phase-based cell imaging techniques for microbeam irradiations. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater At* 2005;241(1):387–91.
- [183] Dragomir NM, Goh XM, Curl CL, Delbridge LMD, Roberts A. Quantitative polarized phase microscopy for birefringence imaging. *Opt Express* 2007;15(26):17690–8.
- [184] Roberts A, Ampem-Lassen E, Barty A, Nugent KA, Baxter GW, Dragomir NM, Huntington ST. Refractive-index profiling of optical fibers with axial symmetry by use of quantitative phase microscopy. *Opt Lett* 2002;27(23):2061–3.
- [185] Ampem-Lassen E, Huntington ST, Dragomir NM, Nugent KA, Roberts A. Refractive index profiling of axially symmetric optical fibers: a new technique. *Opt Express* 2005;13(9):3277–82.
- [186] Dorrer C, Zuegel JD. Optical testing using the transport-of-intensity equation. *Opt Express* 2007;15(12):7165–75.
- [187] Darudi A, Shomali R, Tavassoly MT. Determination of the refractive index profile of a symmetric fiber preform by the transport of intensity equation. *Opt Laser Technol* 2008;40(6):850–3.
- [188] Kou SS, Waller L, Barbastathis G, Sheppard CJR. Transport-of-intensity approach to differential interference contrast (TI-DIC) microscopy for quantitative phase imaging. *Opt Lett* 2010;35(3):447–9.
- [189] Waller L, Tian L, Barbastathis G. Transport of intensity phase-amplitude imaging with higher order intensity derivatives. *Opt Express* 2010a;18(12):12552–61.
- [190] Waller L, Luo Y, Yang SY, Barbastathis G. Transport of intensity phase imaging in a volume holographic microscope. *Opt Lett* 2010b;35(17):2961–3.
- [191] Waller L, Kou SS, Sheppard CJR, Barbastathis G. Phase from chromatic aberrations. *Opt Express* 2010c;18(22):22817–25.
- [192] Kou SS, Waller L, Barbastathis G, Marquet P, Depierreux C, Sheppard CJR. Quantitative phase restoration by direct inversion using the optical transfer function. *Opt Lett* 2011;36(14):2671–3.
- [193] Almoró PF, Waller L, Agour M, Falldorf C, Pedrini G, Osten W, Hanson SG. Enhanced deterministic phase retrieval using a partially developed speckle field. *Opt Lett* 2012;37(11):2088–90.
- [194] Gorthi SS, Schonbrun E. Phase imaging flow cytometry using a focus-stack collecting microscope. *Opt Lett* 2012;37:707–9.
- [195] Waller L, Tsang M, Ponda S, Yang SY, Barbastathis G. Phase and amplitude imaging from noisy images by Kalman filtering. *Opt Express* 2011;19(3):2805–15.
- [196] Xue B, Zheng S, Cui L, Bai X, Zhou F. Transport of intensity phase imaging from multiple intensities measured in unequally-spaced planes. *Opt Express* 2011;19(21):20244–50.
- [197] Bie R, Yuan X-H, Zhao M, Zhang L. Method for estimating the axial intensity derivative in the TIE with higher order intensity derivatives and noise suppression. *Opt Express* 2012;20(7):8186–91.
- [198] Zheng S, Xue B, Xue W, Bai X, Zhou F. Transport of intensity phase imaging from multiple noisy intensities measured in unequally-spaced planes. *Opt Express* 2012;20(2):972–85.
- [199] Martínez-Carranza J, Falaggis K, Kozacki T. Optimum measurement criteria for the axial derivative intensity used in transport of intensity-equation-based solvers. *Opt Lett* 2014a;39(2):182–5.
- [200] Falaggis K, Kozacki T, Kujawinska M. Optimum plane selection criteria for single-beam phase retrieval techniques based on the contrast transfer function. *Opt Lett* 2014;39(1):30–3.
- [201] Zuo C, Chen Q, Asundi A. Boundary-artifact-free phase retrieval with the transport of intensity equation: fast solution with use of discrete cosine transform. *Opt Express* 2014a;22(8):9220.
- [202] Zuo C, Chen Q, Li H, Qu W, Asundi A. Boundary-artifact-free phase retrieval with the transport of intensity equation II: applications to microlens characterization. *Opt Express* 2014b;22(15):18310.
- [203] Huang L, Zuo C, Idir M, Qu W, Asundi A. Phase retrieval with the transport-of-intensity equation in an arbitrarily shaped aperture by iterative discrete cosine transforms. *Opt Lett* 2015;40(9):1976.
- [204] Zuo C, Chen Q, Huang L, Asundi A. Phase discrepancy analysis and compensation for fast Fourier transform based solution of the transport of intensity equation. *Opt Express* 2014c;22(14):17172.
- [205] Zuo C, Chen Q, Yu Y, Asundi A. Transport-of-intensity phase imaging using Savitzky-Golay differentiation filter - theory and applications. *Opt Express* 2013a;21(5):5346–62.
- [206] Sun J, Zuo C, Chen Q. Iterative optimum frequency combination method for high efficiency phase imaging of absorptive objects based on phase transfer function. *Opt Express* 2015;23(21):28031.
- [207] Zuo C, Chen Q, Asundi A. Light field moment imaging: comment. *Opt Lett* 2014d;39(3):654.
- [208] Zuo C, Chen Q, Tian L, Waller L, Asundi A. Transport of intensity phase retrieval and computational imaging for partially coherent fields: The phase space perspective. *Opt Lasers Eng* 2015a;71:20–32.
- [209] Zuo C, Sun J, Li J, Zhang J, Asundi A, Chen Q. High-resolution transport-of-intensity quantitative phase microscopy with annular illumination. *Sci Rep* 2017;7(1):7654.
- [210] Li J, Chen Q, Zhang J, Zhang Y, Lu L, Zuo C. Efficient quantitative phase microscopy using programmable annular LED illumination. *Biomed Opt Express* 2017a;8(10):4687–705.
- [211] Li J, Chen Q, Sun J, Zhang J, Pan X, Zuo C. Optimal illumination pattern for transport-of-intensity quantitative phase microscopy. *Opt Express* 2018a;26(21):27599–614.
- [212] Zuo C, Sun J, Zhang J, Hu Y, Chen Q. Lensless phase microscopy and diffraction tomography with multi-angle and multi-wavelength illuminations using a LED matrix. *Opt Express* 2015b;23(11):14314–28.
- [213] Li J, Chen Q, Zhang J, Zhang Z, Zhang Y, Zuo C. Optical diffraction tomography microscopy with transport of intensity equation using a light-emitting diode array. *Opt Lasers Eng* 2017b;95:26–34.
- [214] Li J, Chen Q, Sun J, Zhang J, Ding J, Zuo C. Three-dimensional tomographic microscopy technique with multi-frequency combination with partially coherent illuminations. *Biomed Opt Express* 2018b;9(6):2526–42.
- [215] Li J, Chen Q, Sun J, Zhang J, Zuo C. Multimodal computational microscopy based on transport of intensity equation. *J Biomed Opt* 2016;21(12):126003.
- [216] Zuo C, Chen Q, Qu W, Asundi A. Noninterferometric single-shot quantitative phase microscopy. *Opt Lett* 2013b;38(18):3538.
- [217] Zuo C, Chen Q, Qu W, Asundi A. High-speed transport-of-intensity phase microscopy with an electrically tunable lens. *Opt Express* 2013c;21(20):24060.
- [218] Kirchhoff G. Zur theorie der lichtstrahlen. *Ann Phys* 1883;254(4):663–95.
- [219] Rayleigh L. On the passage of waves through apertures in Plane screens, and allied problems. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 1897;43(263):259–72. Publisher: Taylor & Francis
- [220] Sommerfeld A. Mathematische theorie der diffraction. *Math Ann* 1896;47(2):317–74.
- [221] Guigay JP. Fourier transform analysis of fresnel diffraction patterns and in-line holograms. *Optik (Stuttgart)* 1977;49:121–5.
- [222] Gilbarg D, Trudinger NS. Elliptic partial differential equations of second order. Springer; 2015.
- [223] Nugent KA. Partially coherent diffraction patterns and coherence measurement. *J Opt Soc Am A* 1991;8(10):1574–9.
- [224] Bhamra KS, et al. Partial differential equations. PHI Learning Pvt Ltd; 2010.
- [225] Courant R, Hilbert D. Methods of Mathematical Physics: Partial Differential Equations. John Wiley & Sons; 2008.
- [226] Gureyev TE, Raven C, Snigirev A, Snigireva I, Wilkins SW. Hard x-ray quantitative non-interferometric phase-contrast microscopy. *J Phys Appl Phys* 1999;32(5):563.
- [227] Pinhasi SV, Alimi R, Perelmutter L, Eliezer S. Topography retrieval using different solutions of the transport intensity equation. *J Opt Soc Am A* 2010;27(10):2285–92.
- [228] Xue B, Zheng S. Phase retrieval using the transport of intensity equation solved by the FMG-CG method. *Opt - Int J Light Electron Opt* 2011;122(23):2101–6.
- [229] Voitikhovich VV. Phase-retrieval problem and orthogonal expansions: curvature sensing. *J Opt Soc Am A* 1995;12(10):2194–202.
- [230] Ros S, Acosta E, Bar S. Modal phase estimation from wavefront curvature sensing. *Opt Commun* 1996;123(4):453–6.
- [231] Volkov VV, Zhu Y, De Graef M. A new symmetrized solution for phase retrieval using the transport of intensity equation. *Micron* 2002;33(5):411–16.

- [232] Frank J, Altmeyer S, Wernicke G. Non-interferometric, non-iterative phase retrieval by green's functions. *J Opt Soc Am A* 2010;27(10):2244–51.
- [233] Parvizi A, Müller J, Funken SA, Koch CT. A practical way to resolve ambiguities in wavefront reconstructions by the transport of intensity equation. *Ultramicroscopy* 2015;154:1–6.
- [234] Martínez-Carranza J, Falaggis K, Kozacki T, Kujawinska M. Effect of imposed boundary conditions on the accuracy of transport of intensity equation based solvers. *Modeling Aspects in Optical Metrology IV*, vol 8789. International Society for Optics and Photonics; 2013. 87890N.
- [235] Acosta E, Ros S, Soto M, Voitsekovich VV. Role of boundary measurements in curvature sensing. *Opt Commun* 1999;169(1):59–62.
- [236] Campbell C. Wave-front sensing by use of a green's function solution to the intensity transport equation: comment. *J Opt Soc Am A* 2007;24(8):2480–1.
- [237] Woods SC, Campbell HI, Greenaway AH. Wave-front sensing by use of a Green's function solution to the intensity transport equation: reply to comment. *J Opt Soc Am A* 2007;24(8):2482–4.
- [238] Woods SC, Greenaway AH. Wave-front sensing by use of a Green's function solution to the intensity transport equation. *J Opt Soc Am A* 2003;20(3):508–12.
- [239] Han IW. New method for estimating wavefront from curvature signal by curve fitting. *Opt Eng* 1995;34(4):1232–7.
- [240] Ishizuka A, Mitsuishi K, Ishizuka K. Direct observation of curvature of the wave surface in transmission electron microscope using transport intensity equation. *Ultramicroscopy* 2018a;194:7–14.
- [241] Ishizuka A, Ishizuka K, Mitsuishi K. Boundary-artifact-free observation of magnetic materials using the transport of intensity equation. *Microsc Microanal* 2018b;24(S1):924–5.
- [242] Mehrabkhani S, Wefelnberg L, Schneider T. Fourier-based solving approach for the transport-of-intensity equation with reduced restrictions. *Opt Express* 2018;26(9):11458.
- [243] Zhang J, Chen Q, Sun J, Tian L, Zuo C. On a universal solution to the transport-of-intensity equation. 2019. 1912.07371.
- [244] Schmalz JA, Gureyev TE, Paganin DM, Pavlov KM. Phase retrieval using radiation and matter-wave fields: Validity of teague's method for solution of the transport-of-intensity equation. *Phys Rev A* 2011;84(2).
- [245] Shanker A, Sczyrba M, Connolly B, Kalk F, Neureuther A, Waller L. Critical assessment of the transport of intensity equation as a phase recovery technique in optical lithography. *Optical Microlithography XXVII*, vol 9052. International Society for Optics and Photonics; 2014. 90521D.
- [246] Sheppard CJR, Mao XQ. Three-dimensional imaging in a microscope. *J Opt Soc Am A* 1989;6(9):1260–9.
- [247] Hopkins HH, Barham PM. The influence of the condenser on microscopic resolution. *Proc Phys Soc B* 1950;63(10):737.
- [248] Goodman JW. *Statistical optics*. Wiley classics library ed. New York: Wiley; 2000. ISBN 978-0-471-39916-2 978-0-471-01502-4
- [249] Sheppard CJR. Defocused transfer function for a partially coherent microscope and application to phase retrieval. *J Opt Soc Am A* 2004;21(5):828–31.
- [250] McCutchen CW. Generalized Aperture and the Three-Dimensional Diffraction Image. *J Opt Soc Am* 1964;54(2):240–4.
- [251] Wilson T. *Theory and Practice of Scanning Optical Microscopy*. 1 edition. London ; Orlando: Academic Pr; 1984. ISBN 978-0-12-757760-9.
- [252] Gu M. *Advanced optical imaging theory*, vol 75. Springer Science & Business Media; 2000.
- [253] Pogany A, Gao D, Wilkins SW. Contrast and resolution in imaging with a microfocus x-ray source. *Rev Sci Instrum* 1997;68(7):2774–82.
- [254] Cloetens P, Ludwig W, Boller E, Helfen L, Salvo L, Mache R, Schlenker M. Quantitative phase contrast tomography using coherent synchrotron radiation. In: *Developments in X-ray tomography III*, vol. 4503. International Society for Optics and Photonics; 2002. p. 82–91.
- [255] Mayo SC, Miller PR, Wilkins SW, Davis TJ, Gao D, Gureyev TE, Paganin D, Parry DJ, Pogany A, Stevenson AW. Quantitative x-ray projection microscopy: phase-contrast and multi-spectral imaging. *J Microsc* 2002;207(2):79–96.
- [256] Gureyev TE, Pogany A, Paganin DM, Wilkins SW. Linear algorithms for phase retrieval in the Fresnel region. *Opt Commun* 2004a;231(1-6):53–70.
- [257] Gureyev TE, Mayo S, Wilkins SW, Paganin D, Stevenson AW. Quantitative in-line phase-contrast imaging with multienergy x rays. *Phys Rev Lett* 2001;86(25):5827–30.
- [258] Wu X, Liu H. A general theoretical formalism for x-ray phase contrast imaging. *J X-ray Sci Technol* 2003;11(1):33–42.
- [259] Guigay JP, Langer M, Boistel R, Cloetens P. Mixed transfer function and transport of intensity approach for phase retrieval in the fresnel region. *Opt Lett* 2007;32(12):1617–19.
- [260] Langer M, Cloetens P, Guigay J-P, Peyrin F. Quantitative comparison of direct phase retrieval algorithms in in-line phase tomography. *Med Phys* 2008;35(10):4556–66.
- [261] Cloetens P, Barrett R, Baruchel J, Guigay J-P, Schlenker M. Phase objects in synchrotron radiation hard x-ray imaging. *J Phys D: Appl Phys* 1996;29(1):133.
- [262] Zabler S, Cloetens P, Guigay J-P, Baruchel J, Schlenker M. Optimization of phase contrast imaging using hard x rays. *Rev Sci Instrum* 2005;76(7):073705.
- [263] Roddier F. Curvature sensing: a diffraction theory. *NAOARD Note* 1987;87-3.
- [264] Paganin D, Barty A, McMahon PJ, Nugent KA. Quantitative phase-amplitude microscopy. III. the effects of noise. *J Microsc* 2004a;214(1):51–61.
- [265] Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* 1964;36(8):1627–39.
- [266] Gorry PA. General least-squares smoothing and differentiation of nonuniformly spaced data by the convolution method. *Anal Chem* 1991;63(5):534–6.
- [267] Orfanidis SJ. *Introduction to signal processing*. Prentice-Hall, Inc; 1995.
- [268] Luo J, Ying K, He P, Bai J. Properties of savitzky-golay digital differentiators. *Digital Signal Process* 2005;15(2):122–36.
- [269] Gureyev TE. Composite techniques for phase retrieval in the fresnel region. *Opt Commun* 2003;220(1):49–58.
- [270] Donnadieu P, Verdier M, Berthom G, Mur P. Imaging a dense nanodot assembly by phase retrieval from TEM images. *Ultramicroscopy* 2004;100(1-2):79–90.
- [271] Martínez-Carranza J, Falaggis K, Kozacki T. Optimum plane selection for transport-of-intensity-equation-based solvers. *Appl Optics* 2014b;53(30):7050–8.
- [272] Soto M, Acosta E. Improved phase imaging from intensity measurements in multiple planes. *Appl Optics* 2007;46(33):7978–81.
- [273] Jingshan Z, Claus RA, Dauwels J, Tian L, Waller L. Transport of intensity phase imaging by intensity spectrum fitting of exponentially spaced defocus planes. *Opt Express* 2014;22(9):10661–74.
- [274] Martínez-Carranza J, Falaggis K, Kozacki T. Multi-filter transport of intensity equation solver with equalized noise sensitivity. *Opt Express* 2015;23(18):23092–107.
- [275] Bostan E, Froustey E, Nilchian M, Sage D, Unser M. Variational phase imaging using the transport-of-intensity equation. *IEEE Trans Image Process* 2016;25(2):807–17.
- [276] Langer M, Cloetens P, Peyrin F. Regularization of phase retrieval with phase-attenuation duality prior for 3-d holotomography. *IEEE Trans Image Process* 2010;19(9):2428–36.
- [277] Zuo C, Chen Q, Qu W, Asundi A. Direct continuous phase demodulation in digital holography with use of the transport-of-intensity equation. *Opt Commun* 2013d;309:221–6.
- [278] Tian L, Petruccielli JC, Barbastathis G. Nonlinear diffusion regularization for transport of intensity phase imaging. *Opt Lett* 2012a;37(19):4131–3.
- [279] Lee P-C. Phase retrieval method for in-line phase contrast x-ray imaging and denoising by regularization. *Opt Express* 2015;23(8):10668.
- [280] Wu X, Liu H. X-ray cone-beam phase tomography formulas based on phase-attenuation duality. *Opt Express* 2005;13(16):6000–14.
- [281] Wolf E, et al. *Introduction to the Theory of Coherence and Polarization of Light*. Cambridge University Press; 2007.
- [282] Mandel L, Wolf E. *Optical coherence and quantum optics*. Cambridge university press; 1995.
- [283] Zernike F. The concept of degree of coherence and its application to optical problems. *Physica* 1938;5(8):785–95.
- [284] Mandel L, Wolf E. Spectral coherence and the concept of cross-spectral purity*. *J Opt Soc Am* 1976;66(6):529–35.
- [285] Chatfield C. *The analysis of time series: theory and practice*. Springer; 2013.
- [286] Wolf E. New theory of partial coherence in the space-frequency domain. Part I: spectra and cross spectra of steady-state sources. *J Opt Soc Am* 1982;72(3):343–51.
- [287] van Cittert PH. Die wahrscheinliche schwingungsverteilung in einer von einer lichtquelle direkt oder mittels einer linse beleuchteten ebene. *Physica* 1934;1(1):201–10.
- [288] Zernike F. Diffraction and optical image formation. *Proc Phys Soc* 1948;61(2):158.
- [289] Gori F. Directionality and spatial coherence. *Opt Acta Int J Opt* 1980;27(8):1025–34.
- [290] Wolf E. New spectral representation of random sources and of the partially coherent fields that they generate. *Opt Commun* 1981;38(1):3–6.
- [291] Starikov A, Wolf E. Coherent-mode representation of Gaussian Schell-model sources and of their radiation fields. *J Opt Soc Am* 1982;72(7):923–8.
- [292] Gori F, Santarsiero M, Borghi R, Guattari G. Intensity-based modal analysis of partially coherent beams with Hermite-Gaussian modes. *Opt Lett* 1998;23(13):989–91.
- [293] Allen L, Beijersbergen MW, Spreeuw RJC, Woerdman JP. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys Rev A* 1992;45(11):8185–9.
- [294] Schell A. A technique for the determination of the radiation pattern of a partially coherent aperture. *IEEE Trans Antennas Propag* 1967;15(1):187–8.
- [295] Nugent KA. A generalization of Schell's theorem. *Opt Commun* 1990;79(5):267–9.
- [296] Nugent KA. Coherent methods in the x-ray sciences. *Adv Phys* 2010;59(1):1–99.
- [297] Siegman AE. New developments in laser resonators. In: *Optical resonators*, vol. 1224. International Society for Optics and Photonics; 1990. p. 2–14.
- [298] Wigner EP. On the quantum correction for thermodynamic equilibrium. In: *Part I: Physical Chemistry. Part II: Solid State Physics*. Springer; 1997. p. 110–20.
- [299] Dolin L. Beam description of weakly-inhomogeneous wave fields. *Izv Vyssh Uchebn Zaved Radiofiz* 1964;7:559–63.
- [300] Walther A. Radiometry and coherence. *J Opt Soc Am* 1968;58(9):1256–9.
- [301] Walther A. Radiometry and coherence. *J Opt Soc Am* 1973;58(9):1256–9.
- [302] Bastiaans MJ. The wigner distribution function applied to optical signals and systems. *Opt Commun* 1978;25(1):26–30.
- [303] Bastiaans MJ. Application of the wigner distribution function to partially coherent light. *J Opt Soc Am A* 1986;3(8):1227–38.
- [304] Bastiaans MJ. Wigner distribution function and its application to first-order optics. *J Opt Soc Am* 1979a;69(12):1710–16.
- [305] Woodward PM. *Probability and information theory, with applications to radar*. Pergamon science series. Electronics and waves, a series of monographs. New York: McGraw-Hill; 1953.
- [306] Papoulis A. Ambiguity function in fourier optics. *J Opt Soc Am* 1974;64(6):779–88.
- [307] Testorf M, Hennelly B, Ojeda-Castaneda J. *Phase-Space Optics: Fundamentals and Applications*. 1 edition. New York: McGraw-Hill Education; 2009.
- [308] Boashash B. Estimating and interpreting the instantaneous frequency of a signal. i. fundamentals. *Proc IEEE* 1992;80(4):520–38.
- [309] McBRIDE AC, Kerr FH. On namias's fractional fourier transforms. *IMA J Appl Math* 1987;39(2):159–75.
- [310] Lohmann AW. Image rotation, wigner rotation, and the fractional fourier transform. *J Opt Soc Am A* 1993;10(10):2181–6.

- [311] Mendlovic D, Ozaktas HM. Fractional Fourier transforms and their optical implementation: I. *J Opt Soc Am A* 1993;10(9):1875–81.
- [312] Ozaktas HM, Kutay MA. The fractional fourier transform. In: 2001 European Control Conferenc (ECC); 2001. p. 1477–83.
- [313] Lohmann AW, Soffer BH. Relationships between the Radon-Wigner and fractional Fourier transforms. *J Opt Soc Am A* 1994;11(6):1798–801.
- [314] Raymer MG, Beck M, McAlister D. Complex wave-field reconstruction using phase-space tomography. *Phys Rev Lett* 1994;72(8):1137–40.
- [315] McAlister DF, Beck M, Clarke L, Mayer A, Raymer MG. Optical phase retrieval by phase-space tomography and fractional-order Fourier transforms. *Opt Lett* 1995;20(10):1181–3.
- [316] McKee CB, O'Shea PG, Madey JMJ. Phase space tomography of relativistic electron beams. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 1995;358(1):264–7.
- [317] Tian L, Lee J, Oh SB, Barbastathis G. Experimental compressive phase space tomography. *Opt Express* 2012b;20(8):8296–308.
- [318] Nazarathy M, Shamir J. First-order optics—a canonical operator representation: lossless systems. *J Opt Soc Am* 1982;72(3):356–64.
- [319] Gerrard A, Burch JM. Introduction to Matrix Methods in Optics. Courier Corporation; 1975. ISBN 978-0-486-68044-6.
- [320] Collins SA. Lens-system diffraction integral written in terms of matrix optics. *J Opt Soc Am* 1970;60(9):1168–77.
- [321] Brenner K-H, Lohmann AW. Wigner distribution function display of complex 1d signals. *Opt Commun* 1982;42(5):310–14.
- [322] Bastiaans MJ. Transport equations for the wigner distribution function in an inhomogeneous and dispersive medium. *Opt Acta Int J Opt* 1979b;26(11):1333–44.
- [323] Winston R, Welford WT. Geometrical vector flux and some new nonimaging concentrators. *J Opt Soc Am* 1979;69(4):532–6.
- [324] Naik DN, Pedrini G, Takeda M, Osten W. Spectrally resolved incoherent holography: 3d spatial and spectral imaging using a mach-zehnder radial-shearing interferometer. *Opt Lett* 2014;39(7):1857.
- [325] Naik DN, Pedrini G, Osten W. Recording of incoherent-object hologram as complex spatial coherence function using Sagnac radial shearing interferometer and a Pockels cell. *Opt Express* 2013;21(4):3990–5.
- [326] Iaconis C, Walmsley IA. Direct measurement of the two-point field correlation function. *Opt Lett* 1996;21(21):1783–5.
- [327] Marks DL, Stack RA, Brady DJ. Three-dimensional coherence imaging in the fresnel domain. *Appl Optics* 1999;38(8):1332–42.
- [328] Kak AC, Slaney M, Wang G. Principles of computerized tomographic imaging. *Med Phys* 2002;29(1). 107–107
- [329] Radon J. On the determination of functions from their integral values along certain manifolds. *IEEE T Med Imaging* 1986;5(4):170–6.
- [330] Tu J, Tamura S. Wave field determination using tomography of the ambiguity function. *Phys Rev E* 1997;55(2):1946–9.
- [331] Dragoman D, Dragoman M, Brenner KH. Tomographic amplitude and phase recovery of vertical-cavity surface-emitting lasers by use of the ambiguity function. *Opt Lett* 2002;27(17):1519–21.
- [332] Liu X, Brenner KH. Reconstruction of two-dimensional complex amplitudes from intensity measurements. *Opt Commun* 2003;225(1-3):19–30.
- [333] Testorf ME, Semichaevsky A. Phase retrieval and phase-space tomography from incomplete data sets. In: *Image Reconstruction from Incomplete Data III*, vol. 5562. International Society for Optics and Photonics; 2004. p. 38–49.
- [334] Bartelt HO, Brenner K-H, Lohmann AW. The wigner distribution function and its optical production. *Opt Commun* 1980;32(1):32–8.
- [335] Zhang Z, Levoy M. Wigner distributions and how they relate to the light field. In: 2009 IEEE International Conference on Computational Photography (ICCP); 2009. p. 1–10. doi:10.1109/ICCPHOT.2009.5559007.
- [336] Waller L, Situ G, Fleischer JW. Phase-space measurement and coherence synthesis of optical beams. *Nat Photonics* 2012;6(7):474–9.
- [337] Ng R, Levoy M, Brdif M, Duval G, Horowitz M, Hanrahan P. Light field photography with a hand-held plenoptic camera. *Comput Sci Tech Rep* 2005;2:1–11.
- [338] Tian L, Zhang Z, Petrucci JC, Barbastathis G. Wigner function measurement using a lenslet array. *Opt Express* 2013;21(9):10511.
- [339] Stoklasa B, Motka L, Rehacek J, Hradil Z, Sanchez-Soto LL. Wavefront sensing reveals optical coherence. *Nat Commun* 2014;5:ncmms4275.
- [340] Banaszek K, Wdkiewicz K. Direct probing of quantum phase space by photon counting. *Phys Rev Lett* 1996;76(23):4344–7.
- [341] Chapman HN. Phase-retrieval x-ray microscopy by wigner-distribution deconvolution. *Ultramicroscopy* 1996;66(3):153–72.
- [342] Chakraborty T, Petrucci JC. Source diversity for transport of intensity phase imaging. *Opt Express* 2017;25(8):9122–37.
- [343] Hamilton DK, Sheppard C. Differential phase contrast in scanning optical microscopy. *J Microsc-oxford* 1984;133(1):27–39.
- [344] Tian L, Waller L. Quantitative differential phase contrast imaging in an LED array microscope. *Opt Express* 2015b;23(9):11394–403.
- [345] Fan Y, Sun J, Chen Q, Pan X, Tian L, Zuo C. Optimal illumination scheme for isotropic quantitative differential phase contrast microscopy. *Photon Res* 2019;7(8):890–904. doi:10.1364/PRJ.7.000890.
- [346] Iglesias I. Pyramid phase microscopy. *Opt Lett* 2011;36(18):3636–8.
- [347] Parthasarathy AB, Chu KK, Ford TN, Mertz J. Quantitative phase imaging using a partitioned detection aperture. *Opt Lett* 2012;37(19):4062–4.
- [348] Zuo C, Sun J, Feng S, Zhang M, Chen Q. Programmable aperture microscopy: A computational method for multi-modal phase contrast and light field imaging. *Opt Laser Eng* 2016b;80:24–31.
- [349] Gureyev TE, Nesterets YI, Paganin DM, Pogany A, Wilkins SW. Linear algorithms for phase retrieval in the fresnel region. 2. partially coherent illumination. *Opt Commun* 2006;259(2):569–80.
- [350] Zysk AM, Schoonover RW, Carney PS, Anastasio MA. Transport of intensity and spectrum for partially coherent fields. *Opt Lett* 2010;35(13):2239–41.
- [351] Petrucci JC, Tian L, Barbastathis G. The transport of intensity equation for optical path length recovery using partially coherent illumination. *Opt Express* 2013;21(12):14430–41.
- [352] Gureyev TE, Paganin DM, Stevenson AW, Mayo SC, Wilkins SW. Generalized eikonal of partially coherent beams and its use in quantitative imaging. *Phys Rev Lett* 2004b;93(6):068103.
- [353] Ng R. Fourier slice photography. In: *ACM Transactions on Graphics (TOG)*, vol. 24. ACM; 2005. p. 735–44.
- [354] Sheppard CJR. Partially coherent microscope imaging system in phase space: effect of defocus and phase reconstruction. *J Opt Soc Am A* 2018;35(11):1846.
- [355] Paganin DM, Labriet H, Brun E, Berujon S. Single-image geometric-flow x-ray speckle tracking. *Phys Rev A* 2018;98(5):053813.
- [356] Lu L, Sun J, Zhang J, Fan Y, Chen Q, Zuo C. Quantitative phase imaging camera with a weak diffuser. *Front Phys* 2019a;7:777.
- [357] Zdora M-C. State of the art of x-ray speckle-based phase-contrast and dark-field imaging. *J Imaging* 2018;4(5):60.
- [358] Chellappan KV, Erden E, Urey H. Laser-based displays: a review. *Appl Optics* 2010;49(25):F79–98.
- [359] Liba O, Lew MD, SoRelle ED, Dutta R, Debasish S, Moshfeghi DM, Chu S, de La Zerda A. Speckle-modulating optical coherence tomography in living mice and humans. *Nat Commun* 2017;8:15845.
- [360] Bates R. Astronomical speckle imaging. *Phys Rep* 1982;90(4):203–97.
- [361] Høgmoen K, Pedersen HM. Measurement of small vibrations using electronic speckle pattern interferometry: theory. *J Opt Soc Am* 1977;67(11):1578–83.
- [362] Boas DA, Dunn AK. Laser speckle contrast imaging in biomedical optics. *J Biomed Opt* 2010;15(1):011109.
- [363] Lu L, Sun J, Zhang J, Fan Y, Chen Q, Zuo C. Speckle quantitative phase imaging based on coherence effect compensation. Seventh International Conference on Optical and Photonic Engineering (icOPEN 2019), vol 11205. International Society for Optics and Photonics; 2019b. 112050I.
- [364] Orth A, Crozier KB. Light field moment imaging. *Opt Lett* 2013;38(15):2666–8.
- [365] Liu J, Xu T, Yue W, Sun J, Situ G. Light-field moment microscopy with noise reduction. *Opt Express* 2015;23(22):29154–62.
- [366] Jenkins MH, Long JM, Gaylord TK. Multifilter phase imaging with partially coherent light. *Appl Optics* 2014;53(16):D29–39.
- [367] Jenkins MH, Gaylord TK. Quantitative phase microscopy via optimized inversion of the phase optical transfer function. *Appl Optics* 2015a;54(28):8566–79.
- [368] Bao Y, Gaylord TK. Two improved defocus quantitative phase imaging methods: discussion. *J Opt Soc Am A* 2019;36(12):2104–14.
- [369] Martínez-Carranza J, Falaggis K, Kozacki T. Enhanced lateral resolution for phase retrieval based on the transport of intensity equation with tilted illumination. In: Popescu G, Park Y, editors. *Quantitative Phase Imaging II*, vol. 9718. SPIE; 2016. p. 65–74. doi:10.1117/12.2212876. International Society for Optics and Photonics
- [370] Zhu Y, Shanker A, Tian L, Waller L, Barbastathis G. Low-noise phase imaging by hybrid uniform and structured illumination transport of intensity equation. *Opt Express* 2014a;22(22):26696–711. doi:10.1364/OE.22.026696.
- [371] Zhu Y, Zhang Z, Barbastathis G. Phase imaging for absorptive phase objects using hybrid uniform and structured illumination transport of intensity equation. *Opt Express* 2014b;22(23):28966–76. doi:10.1364/OE.22.028966.
- [372] Chakraborty T, Petrucci JC. Optical convolution for quantitative phase retrieval using the transport of intensity equation. *Appl Optics* 2018;57(1):A134–41.
- [373] Bao Y, Dong GC, Gaylord TK. Weighted-least-squares multi-filter phase imaging with partially coherent light: characteristics of annular illumination. *Appl Optics* 2019;58:137.
- [374] Kak AC, Slaney M. Principles of Computerized Tomographic Imaging. SIAM; 1988. ISBN 978-0-89871-927-7.
- [375] Wolf E. Three-dimensional structure determination of semi-transparent objects from holographic data. *Opt Commun* 1969;1(4):153–6.
- [376] Haeblerl O, Belkebir K, Giovannini H, Sentenac A. Tomographic diffractive microscopy: basics, techniques and perspectives. *J Mod Opt* 2010;57(9):686–99.
- [377] Rappaz B, Marquet P, Cuche E, Emery Y, Depeursinge C, Magistretti PJ. Measurement of the integral refractive index and dynamic cell morphology of living cells with digital holographic microscopy. *Opt Express* 2005;13(23):9361–73.
- [378] Lauer V. New approach to optical diffraction tomography yielding a vector equation of diffraction tomography and a novel tomographic microscope. *J Microsc* 2002;205(Pt 2):165–76.
- [379] Choi W. Tomographic phase microscopy and its biological applications. *3D Res* 2012;3(4).
- [380] Jenkins MH, Gaylord TK. Three-dimensional quantitative phase imaging via tomographic deconvolution phase microscopy. *Appl Optics* 2015b;54(31):9213–27.
- [381] Soto JM, Rodrigo JA, Alieva T. Label-free quantitative 3d tomographic imaging for partially coherent light microscopy. *Opt Express* 2017;25(14):15699–712.
- [382] Devaney AJ. Inverse-scattering theory within the Rytov approximation. *Opt Lett* 1981;6(8):374–6.
- [383] Devaney AJ. A filtered backpropagation algorithm for diffraction tomography. *Ultrason Imaging* 1982;4(4):336–50.
- [384] Sung Y, Choi W, Fang-Yen C, Badizadegan K, Dasari RR, Feld MS. Optical diffraction tomography for high resolution live cell imaging. *Opt Express* 2009;17(1):266–77.
- [385] Fiddy MA. Inversion of optical scattered field data. *J Phys D: Appl Phys* 1986;19(3):301.

- [386] Chen B, Stammes JJ. Validity of diffraction tomography based on the first born and the first rytov approximations. *Appl Optics* 1998;37(14):2996–3006.
- [387] Dändliker R, Weiss K. Reconstruction of the three-dimensional refractive index from scattered waves. *Opt Commun* 1970;1(7):323–8.
- [388] Bao Y, Gaylord TK. Quantitative phase imaging method based on an analytical nonparaxial partially coherent phase optical transfer function. *J Opt Soc Am* 2016;33(11):2125.
- [389] Bao Y, Gaylord TK. Clarification and unification of the obliquity factor in diffraction and scattering theories: discussion. *J Opt Soc Am A* 2017;34:1738–45.
- [390] Sheppard CJR, Matthews HJ. Imaging in high-aperture optical systems. *J Opt Soc Am A* 1987;4(8):1354–60.
- [391] Sheppard CJR, Kawata Y, Kawata S, Gu M. Three-dimensional transfer functions for high-aperture systems. *J Opt Soc Am A* 1994;11(2):593–8.
- [392] Choi W, Fang-Yen C, Badizadegan K, Oh S, Lue N, Dasari RR, Feld MS. Tomographic phase microscopy. *Nat Methods* 2007;4(9):717.
- [393] Kim K, Yoon H, Diez-Silva M, Dao M, Dasari RR, Park Y. High-resolution three-dimensional imaging of red blood cells parasitized by *Plasmodium falciparum* and in situ hemozoin crystals using optical diffraction tomography. *J Biomed Opt* 2013;19(1):011005.
- [394] Cotte Y, Toy F, Jourdain P, Pavillon N, Boss D, Magistretti P, Marquet P, Depeursing C. Marker-free phase nanoscopy. *Nat Photonics* 2013;7(2):113–17.
- [395] Charrière F, Marian A, Montfort F, Kuehn J, Colomb T, Cuche E, Marquet P, Depeursing C. Cell refractive index tomography by digital holographic microscopy. *Opt Lett* 2006;31(2):178.
- [396] Charrère F, Pavillon N, Colomb T, Depeursing C, Heger TJ, Mitchell EAD, Marquet P, Rappaz B. Living specimen tomography by digital holographic microscopy: morphometry of testate amoeba. *Opt Express* 2006;14(16):7005.
- [397] Soto JM, Rodrigo JA, Alieva T. Partially coherent illumination engineering for enhanced refractive index tomography. *Opt Lett* 2018;43(19):4699.
- [398] Blanchard PM, Fisher DJ, Woods SC, Greenaway AH. Phase-diversity wave-front sensing with a distorted diffraction grating. *Appl Optics* 2000;39(35):6649–55.
- [399] Di Martino JM, Ayubi GA, Dalchiele EA, Alonso JR, Fernández A, Flores JL, Perciante CD, Ferrari JA. Single-shot phase recovery using two laterally separated defocused images. *Opt Commun* 2013;293:1–3.
- [400] Lee JW, Ku J, Waller L, Barbastathis G. Transport of intensity imaging applied to quantitative optical phase tomography. In: *Digital Holography and Three-Dimensional Imaging*. Optical Society of America; 2011. p. DTuD2.
- [401] Nguyen T, Nehmetallah G, Tran D, Darudi A, Soltani P. Fully automated, high speed, tomographic phase object reconstruction using the transport of intensity equation in transmission and reflection configurations. *Appl Optics* 2015;54(35):10443–53.
- [402] Nguyen T, Nehmetallah G. Non-interferometric tomography of phase objects using spatial light modulators. *J Imaging* 2016;2(4):30.
- [403] Phillips KG, Velasco CR, Li J, Kolatkar A, Luttgen M, Bethel K, Duggan B, Kuhn P, McCarty OJT. Optical quantification of cellular mass, volume, and density of circulating tumor cells identified in an ovarian cancer patient. *Front Oncol* 2012a;2:72.
- [404] Phillips KG, Jacques SL, McCarty OJT. Measurement of single cell refractive index, dry mass, volume, and density using a transillumination microscope. *Phys Rev Lett* 2012b;109(11):118105.
- [405] Paganin D, Gureyev TE, Mayo SC, Stevenson AW, Nesterets YI, Wilkins SW. X-ray omni microscopy. *J Microsc* 2004b;214(3):315–27.
- [406] Li Y, Di J, Ma C, Zhang J, Zhong J, Wang K, Xi T, Zhao J. Quantitative phase microscopy for cellular dynamics based on transport of intensity equation. *Opt Express* 2018c;26(1):586.
- [407] Zheng J, Zuo C, Gao P, Nienhaus GU. Dual-mode phase and fluorescence imaging with a confocal laser scanning microscope. *Opt Lett* 2018;43(22):5689.
- [408] Zheng J-J, Yao B-L, Shao X-P. Nonresonant background suppression in wide-field coherent anti-stokes raman scattering microscopy with transport of intensity equation based phase imaging. *Acta Phys Sin* 2017;66(11):114206.
- [409] Zheng J, Akimov D, Heuke S, Schmitt M, Yao B, Ye T, Lei M, Gao P, Popp J. Vibrational phase imaging in wide-field CARS for nonresonant background suppression. *Opt Express* 2015;23(8):10756–63.
- [410] Liao J, Wang Z, Zhang Z, Bian Z, Guo K, Nambiar A, Jiang Y, Jiang S, Zhong J, Choma M, Zheng G. Dual light-emitting diode-based multichannel microscopy for whole-slide multiplane, multispectral and phase imaging. *J Biophotonics* 2018;11(2):e201700075.
- [411] Li Y, Di J, Wu W, Shang P, Zhao J. Quantitative investigation on morphology and intracellular transport dynamics of migrating cells. *Appl Optics* 2019a;58(34):G162.
- [412] Ma L, Rajshekhkar G, Wang R, Bhaduri B, Sridharan S, Mir M, Chakraborty A, Iyer R, Prasanth S, Millet L, Gillette MU, Popescu G. Phase correlation imaging of unlabeled cell dynamics. *Sci Rep* 2016;6(1):32702.
- [413] Bian Y, Yao Y, Liu Q, Xu Y, Kuang C, Li H, Liu X. Assessment of tissues- inhomogeneous optical properties based on a portable microscope under partially coherent illumination. *Opt Commun* 2019;434:145–51.
- [414] Rajput SK, Kumar M, Quan X, Morita M, Furuyashiki T, Awatsuji Y, Tajahuerce E, Matoba O. Three-dimensional fluorescence imaging using the transport of intensity equation. *J Biomed Opt* 2019;25(3):032004.
- [415] Mandula O, Kleman J-P, Lacroix F, Allier C, Fiole D, Herv L, Blandin P, Kraemer DC, Morales S. Phase and fluorescence imaging with a surprisingly simple microscope based on chromatic aberration. *Opt Express* 2020;28(2):2079–90.
- [416] Wittkopp JM, Khoo TC, Carney S, Pisala K, Bahreini SJ, Tubbesing K, Tubbesing K, Mahajan S, Sharikova A, Petruccielli JC, Khmaladze A. Comparative phase imaging of live cells by digital holographic microscopy and transport of intensity equation methods. *Opt Express* 2020;28(5):6123–33.
- [417] Rodrigo JA, Soto JM, Alieva T. Fast label-free microscopy technique for 3d dynamic quantitative imaging of living cells. *Biomed Opt Express* 2017;8(12):5507–17.
- [418] Ma X, Zhou B, Su Z, Zhang Z, Peng J, Zhong J. Label-free 3d imaging of weakly absorbing samples using spatially-incoherent annular illumination microscopy. *Ultramicroscopy* 2019;200:97–104.
- [419] Li J, Matlock A, Li Y, Chen Q, Zuo C, Tian L. High-speed in vitro intensity diffraction tomography. *Adv Photon* 2019b;1(6):066004.
- [420] Soto JM, Mas A, Rodrigo JA, Alieva T, Domínguez-Bernal G. Label-free bioanalysis of leishmania infantum using refractive index tomography with partially coherent illumination. *J Biophotonics* 2019;12(9):e201900030.
- [421] Frank J, Matrisch J, Horstmann J, Altmeyer S, Wernicke G. Refractive index determination of transparent samples by noniterative phase retrieval. *Appl Optics* 2011;50(4):427–33.
- [422] Shomali R, Darudi A, Nasiri S. Application of irradiance transport equation in aspheric surface testing. *Opt - Int J Light Electron Opt* 2012;123(14):1282–6.
- [423] Pan S, Ma J, Zhu R, Ba T, Zuo C, Chen F, Dou J, Wei C, Zhou W. Real-time complex amplitude reconstruction method for beam quality m 2 factor measurement. *Opt Express* 2017;25(17):20142–55.
- [424] Bunsen M, Tateyama S. Detection method for the complex amplitude of a signal beam with intensity and phase modulation using the transport of intensity equation for holographic data storage. *Opt Express* 2019;27(17):24029–42.
- [425] Seo S, Su T-W, Derek KT, Erlinger A, Ozcan A. Lensfree holographic imaging for on-chip cytometry and diagnostics. *Lab Chip* 2009;9(6):777–87.
- [426] Greenbaum A, Ozcan A. Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy. *Opt Express* 2012;20(3):3129.
- [427] Luo W, Zhang Y, Grcs Z, Feizi A, Ozcan A. Propagation phasor approach for holographic image reconstruction. *Sci Rep* 2016;6:srep22738.
- [428] Bishara W, Su T-W, Coskun AF, Ozcan A. Lensfree on-chip microscopy over a wide field-of-view using pixel super-resolution. *Opt Express* 2010;18(11):11181.
- [429] Greenbaum A, Zhang Y, Feizi A, Chung P-L, Luo W, Kandukuri SR, Ozcan A. Wide-field computational imaging of pathology slides using lens-free on-chip microscopy. *Sci Transl Med* 2014;6(267). 267ra175–267ra175
- [430] Bishara W, Sikora U, Mudanyali O, Su T-W, Yaglidere O, Luckhart S, Ozcan A. Holographic pixel super-resolution in portable lensless on-chip microscopy using a fiber-optic array. *Lab Chip* 2011;11(7):1276–9.
- [431] Luo W, Zhang Y, Feizi A, Grcs Z, Ozcan A. Pixel super-resolution using wavelength scanning. *Light Sci Appl* 2015;5(4):e16060.
- [432] Zhang J, Chen Q, Li J, Sun J, Zuo C. Lensfree dynamic super-resolved phase imaging based on active micro-scanning. *Opt Lett* 2018;43(15):3714.
- [433] Zhang Y, Wu Y, Zhang Y, Ozcan A. Color calibration and fusion of lens-free and mobile-phone microscopy images for high-resolution and accurate color reproduction. *Sci Rep* 2016;6(1).
- [434] Zhang J, Sun J, Chen Q, Zuo C. Resolution analysis in a lens-free on-chip digital holographic microscope. *IEEE T Comput Imag* 2020;6:697–710.
- [435] Wang H, Chen N, Zheng S, Liu J, Situ G. Fast and high-resolution light field acquisition using defocus modulation. *Appl Optics* 2018;57(1):A250–6.
- [436] Chen N, Zuo C, Lam EY, Lee B. 3d imaging based on depth measurement technologies. *Sensors* 2018;18(11):3711.
- [437] Pandey N, Ghosh A, Khare K. Two-dimensional phase unwrapping using the transport of intensity equation. *Appl Optics* 2016;55(9):2418.
- [438] Zuo C. Connections between transport of intensity equation and two-dimensional phase unwrapping. *arXiv preprint arXiv:170403950* 2017.
- [439] Martínez-Carranza J, Falaggis K, Kozacki T. Fast and accurate phase-unwrapping algorithm based on the transport of intensity equation. *Appl Optics* 2017;56(25):7079–88.
- [440] Zhou W-J, Guan X, Liu F, Yu Y, Zhang H, Poon T-C, Banerjee PP. Phase retrieval based on transport of intensity and digital holography. *Appl Optics* 2018;57(1):A229–34.
- [441] Cheng H, Wang J, Gao Y, Zhang Q, Wei S. Phase unwrapping based on transport-of-intensity equation with two wavelengths. *Opt Eng* 2019;58(5):1–7.
- [442] Zhao Z, Zhang H, Xiao Z, Du H, Zhuang Y, Fan C, Zhao H. Robust 2d phase unwrapping algorithm based on the transport of intensity equation. *Meas Sci Technol* 2019;30(1):015201.
- [443] Zhang C, He W, Wu J, Peng X. Optical cryptosystem based on phase-truncated fresnel diffraction and transport of intensity equation. *Opt Express* 2015;23(7):8845–54.
- [444] Yoneda N, Saita Y, Komuro K, Nobukawa T, Nomura T. Transport-of-intensity holographic data storage based on a computer-generated hologram. *Appl Optics* 2018;57(30):8836–40.
- [445] Rajput SK, Matoba O. Security-enhanced optical voice encryption in various domains and comparative analysis. *Appl Optics* 2019;58(11):3013–22.
- [446] Sui L, Zhao X, Huang C, Tian A, Anand A. An optical multiple-image authentication based on transport of intensity equation. *Opt Lasers Eng* 2019;116:116–24.
- [447] Yang Z, Zhan Q. Single-shot smartphone-based quantitative phase imaging using a distorted grating. *PLoS one* 2016;11(7):e0159596.
- [448] Meng X, Huang H, Yan K, Tian X, Yu W, Cui H, Kong Y, Xue L, Liu C, Wang S. Smartphone based hand-held quantitative phase microscope using the transport of intensity equation method. *Lab Chip* 2017;17(1):104–9.
- [449] Zuo C, Chen Q, Asundi A. Comparison of digital holography and transport of intensity for quantitative phase contrast imaging. In: *Fringe 2013*. Springer; 2014. p. 137–42.
- [450] Lubk A, Guzzinati G, Börmert F, Verbeeck J. Transport of intensity phase retrieval of arbitrary wave fields including vortices. *Phys Rev Lett* 2013;111(17):173902.
- [451] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.

- [452] Sinha A, Lee J, Li S, Barbastathis G. Lensless computational imaging through deep learning. *Optica* 2017;4(9):1117.
- [453] Rivenson Y, Zhang Y, Gnaaydn H, Teng D, Ozcan A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light-Sci Appl* 2018;7(2):17141.
- [454] Feng S, Chen Q, Gu G, Tao T, Zhang L, Hu Y, Yin W, Zuo C. Fringe pattern analysis using deep learning. *Adv Photon* 2019;1(2):025001.
- [455] Barbastathis G, Ozcan A, Situ G. On the use of deep learning for computational imaging. *Optica* 2019;6(8):921.



Corrigendum to “Transport of intensity equation: A tutorial” Optics and Lasers in Engineering, Volume 135 (2020) 106187[☆]

Chao Zuo^{a,b,*}, Jiaji Li^{a,b}, Jiasong Sun^{a,b}, Yao Fan^{a,b}, Jialin Zhang^{a,b}, Linpeng Lu^{a,b}, Runnan Zhang^{a,b}, Bowen Wang^{a,b}, Lei Huang^c, Qian Chen^b

^a Smart Computational Imaging Laboratory (SCILab), Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

^b Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

^c Brookhaven National Laboratory, NSLS II 50 Rutherford Drive, Upton, NY 11973-5000, United States

1. Text correction

The authors regret that the following typos in the original text and should be corrected in this corrigendum:

- In the last paragraph of Page 4 left column of original article, “A major branch of interferometric phase measurement techniques is...” should be “A major branch of non-interferometric phase measurement techniques is...”.
- In the upper line of Eq. (75) of in the original article “the paraxial coherent transfer function of Eq. (67) can be rewritten as”, the equation number here should be Eq. (71) instead of Eq. (67).
- In the upper paragraph of Eq. (160) of in the original article, the sentence “When $\sigma_s \gg \sigma_g$, the source is spatially coherent, and when $\sigma_s \ll \sigma_g$, the source is almost spatially incoherent” should be “When $\sigma_s \gg \sigma_g$, the source is spatially incoherent, and when $\sigma_s \ll \sigma_g$, the source is almost spatially coherent”.
- The title of subsection 6.3.1 should be “Ideal imaging model under partially coherent illumination”.
- In the last paragraph of Page 70 right column of original article, “Eq. (328) is quite similar to...” should be “Eq. (329) is quite similar to...”.
- In the next paragraph of Eq. (181) “(see Subsection 6.2.5 for details)”, the subsection number should be Subsection 6.2.6 instead of Subsection 6.2.5.
- In the upper line of Eq. (393) “the absorption is assumed to be proportional to RI $\hat{P}(\mathbf{u}) \approx \varepsilon \hat{A}(\mathbf{u})$ ” should be $\hat{A}(\mathbf{u}) \approx \varepsilon \hat{P}(\mathbf{u})$.

2. Equation correction

The authors regret the typos in the following equations, and the correct formulations are listed in this corrigendum:

$$U_i(\mathbf{x}) = U_0(\mathbf{x}) \otimes h(\mathbf{x}) \quad (59)$$

$$I_i(\mathbf{x}) = U_i(\mathbf{x})U_i^*(\mathbf{x}) = |U_0(\mathbf{x}) \otimes h(\mathbf{x})|^2 \quad (60)$$

$$\bar{z} = 2k \left(1 - \sqrt{1 - NA^2} \right) \Delta z \stackrel{\text{paraxial}}{\approx} kNA^2 \Delta z \quad (73)$$

$$\hat{I}_{\Delta z}(\mathbf{u}) = \hat{I}_{\Delta z}^{\phi=0}(\mathbf{u}) + 2[\sin(\pi\lambda\Delta z|\mathbf{u}|^2) - \pi\lambda\Delta z|\mathbf{u}|^2 \cos(\pi\lambda\Delta z|\mathbf{u}|^2)] \mathcal{F}\{I_0(\mathbf{x})\phi(\mathbf{x})\} - \cos(\pi\lambda\Delta z|\mathbf{u}|^2) \frac{\lambda\Delta z}{2\pi} \mathcal{F}\{\nabla \cdot [I_0(\mathbf{x})\nabla\phi(\mathbf{x})]\} \quad (90)$$

$$U(\mathbf{x}, t) = U(\mathbf{x}) \exp(-j\omega_0 t) = A(\mathbf{x}) \exp[j(\phi(\mathbf{x}) - \omega_0 t)] \quad (105)$$

$$\tan \theta = -\lambda \Delta z \quad (223)$$

$$\begin{aligned} \hat{T}(\mathbf{u}_1)\hat{T}^*(\mathbf{u}_2) &= a_0^2 \delta(\mathbf{u}_1)\delta(\mathbf{u}_2) \\ &+ a_0 \delta(\mathbf{u}_2) [\hat{\eta}(\mathbf{u}_1) + j\hat{\phi}(\mathbf{u}_1)] \\ &+ a_0 \delta(\mathbf{u}_1) [\hat{\eta}(\mathbf{u}_2) - j\hat{\phi}(\mathbf{u}_2)] \end{aligned} \quad (248)$$

$$\begin{aligned} \hat{U}_{3D}(u_x, u_y, u_z) &= \int \hat{U}_{2D}(u_x, u_y, 0) e^{j2\pi z \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}} e^{-j2\pi z u_z} dz \\ &= \int \hat{U}_{2D}(u_x, u_y, 0) e^{-j2\pi z \left(u_z - \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}\right)} dz \\ &= \hat{U}_{2D}(u_x, u_y, 0) \delta\left(u_z - \sqrt{\left(\frac{1}{\lambda}\right)^2 - u_x^2 - u_y^2}\right) \end{aligned} \quad (324)$$

$$\begin{aligned} H(\rho, l) &= \int P(\rho) e^{jkz\sqrt{1-\lambda^2\rho^2}} e^{-j2\pi z \eta} dz \\ &= P(\rho) \delta\left(\eta - \sqrt{\left(\frac{1}{\lambda}\right)^2 - \rho^2}\right) \end{aligned} \quad (329)$$

$$I(\mathbf{r}) = W_I(\mathbf{r}, \mathbf{r}) = \iint W_S(\mathbf{r}_1, \mathbf{r}_2) T(\mathbf{r}_1) T_S^*(\mathbf{r}_2) h(\mathbf{r} - \mathbf{r}_1) h^*(\mathbf{r} - \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (366)$$

[☆] The authors regret to request the corrigendum to the full manuscript to provide more accurate and complete content about Transport of intensity equation. The authors would like to apologize for any inconvenience caused.

DOI of original article: [10.1016/j.optlaseng.2020.106187](https://doi.org/10.1016/j.optlaseng.2020.106187)

* Corresponding author at: Smart Computational Imaging Laboratory (SCILab), Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China.

E-mail addresses: zuochao@njust.edu.cn (C. Zuo), chenqian@njust.edu.cn (Q. Chen).

<https://doi.org/10.1016/j.optlaseng.2021.106672>

0143-8166/© 2021 Published by Elsevier Ltd.

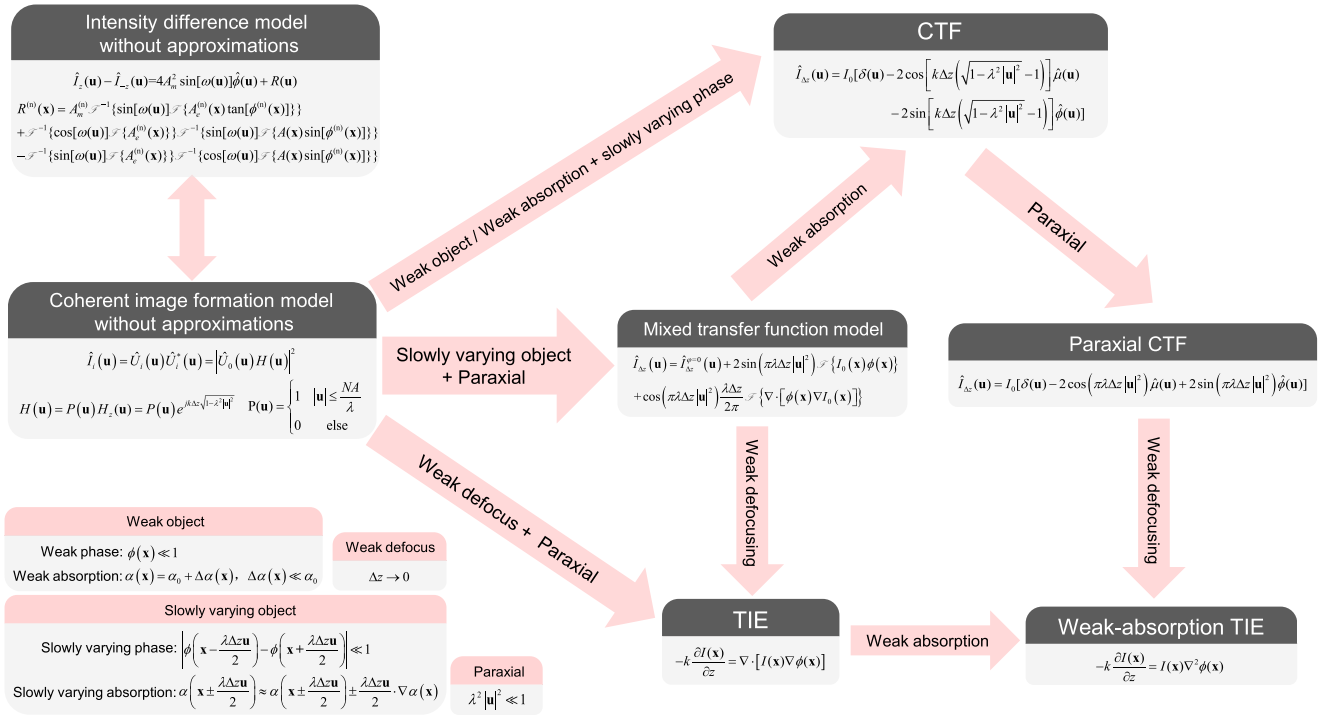


Fig. 26. Mathematical models, establishment conditions, and the intrinsic relationship of different phase retrieval algorithms

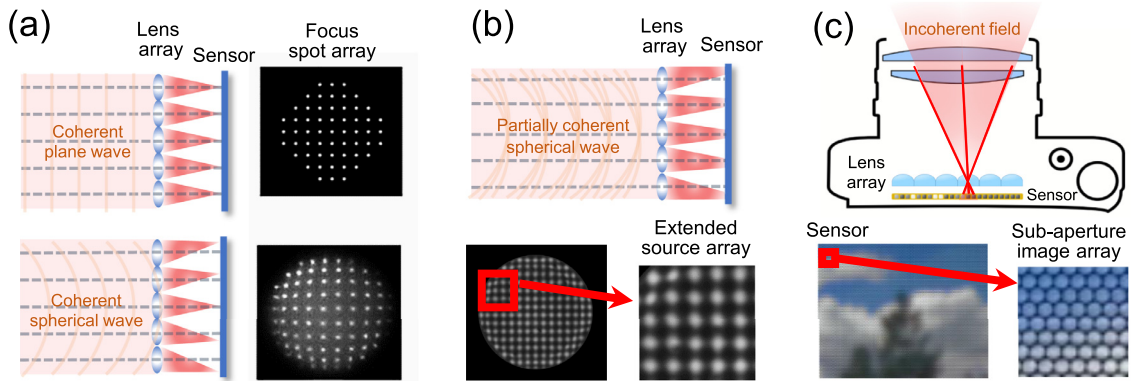


Fig. 63. Principle of the Shack-Hartmann sensor and light-field camera. (a) For coherent field, the Shack-Hartmann sensor forms a focus spot array sensor signal; (b) for a partially coherent field, the Shack-Hartmann sensor forms an extended source array sensor signal; (c) for incoherent imaging, the light-field camera produces a 2D sub-aperture image array.

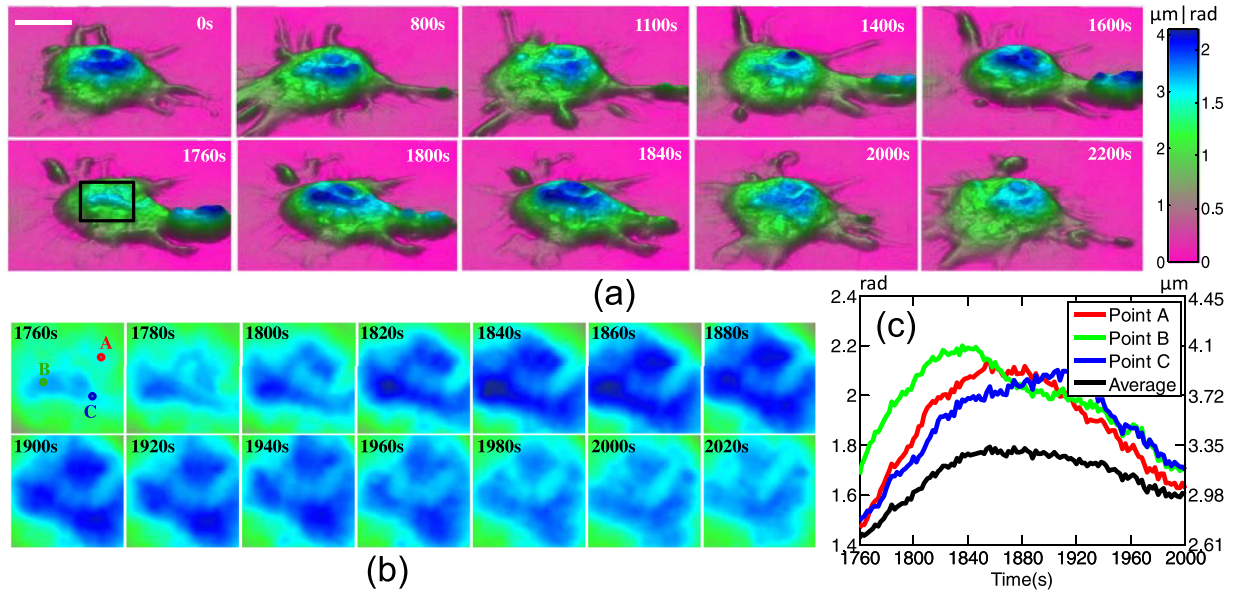


Fig. 102. Dynamic TIE phase imaging of macrophage phagocytosis. (a) Color-coded phase profiles at different stages of phagocytosis; (b) phase maps of the nuclear region of the macrophage (the black square in (a)) during the internalization stage of phagocytosis; (c) phase/thickness variation with time of three points, indicated by the dots in (b): (A) red, (B) green, and (C) blue, and the average of the whole square region (bottom, black curve). Scale bar: 10 μm .

$$I_z(\mathbf{r}) = \int I(\mathbf{u}) e^{j2\pi z\eta} d\eta \quad (389)$$

$$\hat{P}(\mathbf{u}) = \frac{[\hat{I}(\mathbf{u}) - B\delta(\mathbf{u})] H_p^*(\mathbf{u})}{|H_p(\mathbf{u})|^2 + \beta} \quad (394)$$

3. Figure correction

The authors regret that the following figures were processed incorrectly and the revised figures are presented in this corrigendum:

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Redistribution and use of MATLAB source code must retain the copyright notice, list of conditions, the disclaimer from the authors.