

TUTORIAL REVIEW

Computational methods in vectorial imaging

Matthew R. Foreman and Peter Török*

*Blackett Laboratory, Department of Physics, Imperial College London,
Prince Consort Road, London SW7 2BZ, UK*

(Received 5 August 2010; final version received 14 September 2010)

In the search for higher resolution, modern day imaging systems frequently employ objective lenses with a high numerical aperture. Propagation of light through such lenses introduces a spatial variation in the polarisation across the beam profile, whilst the inherently large propagation angles also necessitates inclusion of additional transverse and axial electric field components in modelling. A full treatment of polarisation effects including such considerations has implications at all stages in the image formation process, namely; illumination, scattering from the sample, imaging and detection. This tutorial review considers each stage in turn and details the theories required for rigorous modelling and analysis. In particular a generalisation of the well known Jones calculus and ray tracing methods are shown to conveniently and accurately allow rigorous studies of high numerical aperture confocal and conventional polarised light microscopes, imaging samples of arbitrary complexity. Generalisation of the illumination to partially coherent, partially polarised systems is also briefly given. Whilst rigorous modelling techniques can prove time consuming a number of simplifications and approximations can be adopted, allowing computational gains to be achieved. Discussion in this vein is hence also presented.

Keywords: polarisation; vector imaging; microscopy; electromagnetic modelling; high numerical aperture

1. Introduction

The advent of the optical microscope by Janssen *c.* 1590 spawned a new and powerful research field, which has grown to encompass an extensive range of imaging modalities, such as dark field, fluorescence, spectroscopic studies and phase contrast for which Frits Zernike was awarded the Nobel Prize in 1953 [1–4]. Whilst scalar theory has proven to allow significant advances in the field to be made, it is easily recognised that any complete treatment of optical microscopes must consider the polarisation state of light. For example, many common models consider the state of polarisation to be unchanged by a lens, a fact that can not hold true in general. Refraction of light by a high numerical aperture (NA) lens, for instance, introduces a large longitudinal field component [5,6]. Fresnel transmission through the various surfaces of optical elements can also modify the polarisation state of light. Furthermore, biological samples often exhibit polarisation sensitive transmission such as optical birefringence, hence resulting in polarisation changes. Suitable frameworks within which to describe the vectorial properties of light are thus necessary for accurate modelling and design of real world setups. Given these considerations this article aims to present the concepts required for the modelling and computation of the structure of the electromagnetic field in the image

space of high NA, polarised light, coherent optical microscopes. Development of the theory in this paper will principally be in the context of microscopy; however, the theories discussed are more widely applicable to other vectorial imaging systems. Whilst much of the presented theory is detailed separately in the literature, it is hoped that the following unified presentation serves as a tutorial in the quasi-rigorous modelling of imaging by optical systems of almost arbitrary complexity.

Practically all optical systems are composed of multiple elements and surfaces through which light propagates. Analysis of such systems is a significant problem in optical design. Traditionally geometrical, or ray, techniques would be used in which arbitrary ray paths through an optical system are traced so as to calculate the aberrations present (or any other desired system analysis), for example by means of ABCD matrices [7]. With the increasing popularity of surface coatings on many components, thin film calculations became an equally important part of the design process to incorporate the polarisation transmission properties of these coatings [7]. Although this combined approach proves adequate for many systems in which propagation of light can be considered from a scalar viewpoint, a more advanced strategy is needed to accurately describe high NA imaging. Vectorial ray tracing

*Corresponding author. Email: peter.torok@imperial.ac.uk

provides a solution to this problem [8,9], which, by allowing a ray to represent a vectorial plane wave, can simultaneously account for both geometrical and polarisation effects. Section 2 thus aims to set out the foundations of this technique, built upon a generalisation of Jones calculus, from which results will be drawn throughout the remainder of the article. Ray tracing is generally considered to be an approximate method, however, Section 2 will also discuss conditions under which vectorial ray tracing is a fully rigorous and accurate calculation tool.

Typical microscope configurations incorporate four distinct stages in the image formation process: namely illumination of the sample; interaction of light with the sample; imaging of the scattered light; and detection. Sections 3 to 6 will consider each of these in turn. Specifically, Section 3 considers the Debye–Wolf diffraction integral, commonly used in focusing problems to calculate the illumination field structure incident upon an object, or sample. The Debye–Wolf integral represents a rigorous solution to Maxwell’s equations; however, semi-analytic and approximate methods can prove computationally quicker. A number of approximations and series representations are thus also presented in this vein. An extension for partially coherent illumination is also introduced.

Section 4 continues by considering how light interacts with the sample structure, and how the scattered field can be determined. In special cases analytic results can be found, for example point-like scatterers can be accurately modelled as dipole emitters. Short discussion is hence given to point scatterers; however, these results also prove useful in so-called discrete dipole approximations, which will thus also be briefly introduced. Rigorous methods fundamentally based on Maxwell’s equations, such as finite difference time domain methods, represent the most accurate methods and thus Section 4 concludes by covering some aspects pertinent to rigorous calculations.

Having calculated the field scattered from a sample, subsequent imaging to a detector is required. Section 5 hence applies the vectorial ray tracing technique to imaging systems. Considerations in this vein will highlight the necessity for vectorial modelling techniques by first considering the imaging properties of a high numerical aperture microscope in the absence of a scattering object. Development of imaging dipole fields will be presented due to the growing interest in single molecule imaging, but also due to the pivotal role they can play in the imaging of arbitrary scattered field structures. Finally, Section 6 considers a number of alternative detection configurations commonly employed in imaging systems, chiefly confocal, conventional and polarimetric detection geometries.

2. Vectorial ray tracing

Jones vectors (and the associated calculus), introduced in a series of eight papers by R. Clark Jones during the 1940s and 1950s [10–17], define the polarisation of a monochromatic transverse electromagnetic plane wave by specifying the complex amplitudes of the components of the electric field perpendicular to the direction of propagation. An optical ray can be used to describe the direction of propagation of a plane wave, and thus it is legitimate to also associate a Jones vector to a ray. Implicit in the definition of Jones vectors is, however, a restriction to describing collimated beams only (i.e. those with a zero component in the direction of propagation). Assuming rotational symmetry in the optical system and neglecting skew rays, it is still possible to reduce the description of obliquely propagating beams to a two-dimensional (2D) formalism, by working in a frame of reference in which one of the coordinate axes coincides with the direction of propagation. A restriction to collimated beams is not of consequence for the ‘crystal optics’ systems for which Jones vectors were originally designed; however, since modern optical systems combine polarisation changing elements, such as retardation plates and polarisers, with elements that change the direction of propagation, such as lenses and prisms, full three-dimensional (3D) generalisation is necessary.

Extension of Jones vectors to three dimensions can fortunately be simply achieved by specifying the complex amplitude of all three field components as opposed to only the transverse components. It is also commonly necessary to provide the propagation direction of a plane wave (for which $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)]$) as specified by the wave vector \mathbf{k} . A generalised Jones vector is thus defined as the complex field amplitudes of all three components of the electric field with reference to a fixed set of coordinate axes.

The generalisation of Jones matrices to describe 3D fields was first done heuristically by Török and co-workers [18,19] in 1995. This work was later further extended [20] to include the polarisation altering characteristics of high aperture lenses first described by Inoue and Kubota [21,22] and to describe conventional and confocal polarised light microscopes [23,24], which will be discussed in later sections.

Following [25], it is possible to define three types of generalised Jones matrices describing field transformations, interactions with surfaces, and the action of polarisation elements, such as retarders and lenses. In many situations, such as focusing or transmission through a prism, it is necessary to describe a change in the direction of propagation of a plane wave or equivalently a ray of light. Generally, rotation of a

ray is about a direction perpendicular to the meridional plane (i.e. the plane containing the ray and the optical axis of the system) in which the ray lies. It is thus the field components lying in the meridional plane that are rotated, whilst the perpendicular component is left unaffected. This can be simply described by the rotation matrix

$$\mathbb{L} = \begin{pmatrix} \cos \Delta\theta & 0 & \sin \Delta\theta \\ 0 & 1 & 0 \\ -\sin \Delta\theta & 0 & \cos \Delta\theta \end{pmatrix}, \quad (1)$$

where $\Delta\theta$ is the angle by which the ray is rotated. This matrix can account for a direction change due to refraction and reflection for example, however, if the change in direction occurs at a surface there will also be a change in amplitude of the perpendicular, s , and parallel, p , components as given by the Fresnel transmission and reflection coefficients [26]. This behaviour can be described using the matrices

$$\mathbb{t}_t = \begin{pmatrix} t_p & 0 & 0 \\ 0 & t_s & 0 \\ 0 & 0 & t_p \end{pmatrix}, \mathbb{r}_r = \begin{pmatrix} r_p & 0 & 0 \\ 0 & r_s & 0 \\ 0 & 0 & r_p \end{pmatrix}. \quad (2)$$

For more complex interfaces, such as stratified media, it is also possible to find similar matrices (for an example see Section 3.2). The matrices \mathbb{L} , \mathbb{t}_t and \mathbb{r}_r can only be applied when the electric field is decomposed into its s and p components. This field transformation requires a simple rotation of the Cartesian coordinate axes about the optic axis as expressed by the rotation matrix

$$\mathbb{R} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (3)$$

where ϕ is the angle of the meridional plane to the positive x -axis.

Coming finally to the transformations induced by polarisation elements, if the entrance and exit surfaces of the component are parallel (although possibly oblique to the incident ray vector) it is assumed, due to experimental evidence, that the direction of propagation of the ray is unaltered except for a potential lateral shift, which is irrelevant for a plane wave. Consequently, the longitudinal field component must be unchanged upon propagation through such a parallel surface component. Hence the generalised Jones matrix for an ideal linear retarder (Babinet-Soleil compensator) is

$$\mathbb{BS} = \begin{pmatrix} \cos \frac{\delta}{2} + i \cos 2\gamma \sin \frac{\delta}{2} & i \sin 2\gamma \sin \frac{\delta}{2} & 0 \\ i \sin 2\gamma \sin \frac{\delta}{2} & \cos \frac{\delta}{2} - i \cos 2\gamma \sin \frac{\delta}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (4)$$

where γ is the azimuth of the fast axis of the retarder and δ is the relative retardation between the E_x and E_y component, whilst that for a linear polariser is

$$\mathbb{LP} = \begin{pmatrix} \cos^2 \gamma & \sin \gamma \cos \gamma & 0 \\ \sin \gamma \cos \gamma & \sin^2 \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5)$$

Complex optical systems, formed from multiple optical components and in which diffraction effects are negligible can be modelled by successive application of generalised Jones matrices, i.e. for a system composed of N sequential elements with associated Jones matrices \mathbb{T}_j the whole system can be described by the composite Jones matrix

$$\mathbb{T} = \mathbb{T}_N \mathbb{T}_{N-1} \cdots \mathbb{T}_2 \mathbb{T}_1. \quad (6)$$

Vectorial ray tracing through optical components using Jones matrices in this instance differs from using Fresnel's equations at each individual surface of an optical system (a method used in [27] for example) since it determines an effective single interface, which represents the entire action of all optical elements present. It is, however, necessary to (a) determine the shape of the equivalent surface and (b) transform the effect of all individual surfaces to the single surface. Whilst in some cases this can be achieved theoretically, in general this analysis can only be done based on experimental studies [23].

Vectorial ray tracing in many circumstances is unfortunately only an approximate method of modelling optical systems, in much the same way as geometrical optics cannot accurately model many scalar systems. Nevertheless, it has proven of particular commercial importance having spawned a number of software packages, e.g. [28]. Furthermore, under a number of conditions, which shall be discussed more fully in the context of focusing of light in the next section, vectorial ray tracing can be a rigorous, and hence accurate, calculation tool.

3. Focused illumination

Many modern imaging systems, such as microscopes, often use pairs of lenses arranged in a $4f$ configuration, that is to say they are placed such that their foci coincide, for both illumination and image formation (see Figure 1). Since any well-corrected single lens used as an imaging device may be represented as a double lens system, it is evident that more complex optical setups can also be modelled as either a combination of one or more lens pairs [25]. A light source (or object) is placed in the front focal plane of the first lens, such that a corresponding illumination (or image) field results in the back focal plane of the second lens

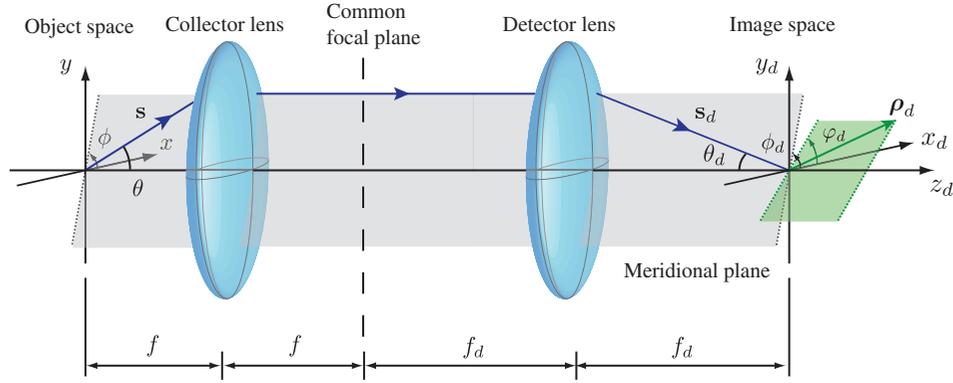


Figure 1. Simple $4f$ telecentric imaging setup. Nomenclature and notation shown is for a $4f$ system used in image formation, however, the setup can equally be used as the illumination component of a microscope. Positions in the object and image plane are defined by the position vectors ρ and ρ_d , respectively, whilst positions on the reference spheres associated with the collector and detector lens (assumed aplanatic and with $\text{NA} = \sin \alpha$ and $\text{NA}_d = \sin \alpha_d$) are defined by the coordinates (θ, ϕ) and (θ_d, ϕ_d) . Ray directions in the respective spaces are described by the normalised wavevectors \mathbf{s} and \mathbf{s}_d . (The colour version of this figure is included in the online version of the journal.)

in a phase correct manner. Critically, the aperture stop is placed in the common focal plane of the lens pair, yielding an afocal and telecentric system from both object and image spaces. By virtue of the telecentricity, the aperture stop appears at infinity from both the object and image side. Consequently computation of the illumination (or image) field structure requires the specification of a field distribution, i.e. a boundary condition, at infinity. Specification of the boundary value at infinity is equivalent to specifying the field in the back focal plane of the lens for a wave with infinitely small wavelength [25], that is to say it is only necessary to specify the field at this plane as predicted by geometrical optics as can be found using vectorial ray tracing.

A rigorous integral representation for the field in the focal region of a telecentric lens, known as the Debye–Wolf integral, can be found by applying Kirchoff’s boundary condition at the aperture stop. Explicitly (see for example [5,6,25,29,30] for a derivation)

$$\mathbf{E}(\boldsymbol{\rho}) = -\frac{ikf}{2\pi} \iint_{s_x^2 + s_y^2 \leq 1} \mathbf{e}(s_x, s_y) \exp(iks \cdot \boldsymbol{\rho}) \frac{ds_x ds_y}{s_z}, \quad (7)$$

where, with reference to Figure 2, $\boldsymbol{\rho} = (\rho \cos \varphi, \rho \sin \varphi, z)$ is the position vector of a point of observation in the focal region of the lens, $\mathbf{s} = (s_x, s_y, s_z) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ is a unit vector describing the direction of a geometrical ray, f is the focal length of the lens, $k = 2\pi/\lambda = \omega/c$ is the wavenumber of light of wavelength λ and frequency ω , and $\mathbf{e}(s_x, s_y)$ describes the field distribution on the Gaussian reference sphere located in the exit pupil of the system centered on the geometrical focus of the lens

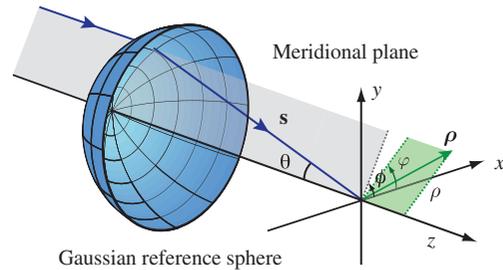


Figure 2. Coordinate system and geometry of the scaled Debye–Wolf diffraction integral. (The colour version of this figure is included in the online version of the journal.)

as seen from image space. Implicit in Equation (7) is the assumption that the point of observation is far from the pupil plane, such that evanescent waves can be neglected, hence yielding the stated domain of integration. The assumption $f \sin \alpha \gg \lambda$ has also been made.

On a final note, non-telecentric systems are briefly mentioned. Such systems can be modelled by $4f$ arrangements in which the aperture stop is not placed at the common focal plane of lens pairs. In this situation, the Fresnel diffraction from the aperture stop to the exit pupil (or vice versa) must be incorporated into the solution of Maxwell’s equations. Such a scenario was considered in a scalar regime by Li and Wolf [31], and later in a vectorial regime by Török [32], giving rise to what is known as the scaled Debye–Wolf integral. This integral bears striking similarities with Equation (7) albeit position vectors are described in a transformed coordinate system. This article will however not consider this scenario any further.

3.1. Homogeneous media

Consider first applying Equation (7) to the description of focusing of light in a homogeneous medium. Noting that an element of solid angle over the reference sphere is given by $\frac{ds_x ds_y}{s_z} = \sin \theta d\theta d\phi$, the Debye–Wolf integral can be rewritten

$$\mathbf{E}(\rho, \varphi, z) = -\frac{ifk}{2\pi} \int_0^{2\pi} \int_0^\alpha \mathbf{e}(\theta, \phi) \times \exp[ik\rho \sin \theta \cos(\phi - \varphi)] e^{ikz \cos \theta} \sin \theta d\theta d\phi \quad (8)$$

where α is the semi-angle of convergence of the lens, such that the NA (assuming the lens is in air) is given by $\text{NA} = \sin \alpha$.

Consider then the specification of the geometric field distribution $\mathbf{e}(\theta, \phi)$ on the Gaussian reference sphere as calculated from the pupil plane distribution of an ideal focusing system as represented by the spatially dependent Jones vector $\tilde{\mathbf{E}}(\theta, \phi)$. An ideal lens acts to rotate a ray by an angle $\Delta\theta = \theta$ (see Figure 2) about an axis perpendicular to the meridional plane, as can be described by the generalised Jones matrix $\mathbb{L}(\theta)$. It is, however, necessary to first decompose the field into its constituent s and p components using the generalised Jones matrix $\mathbb{R}(\phi)$. Accordingly, the field on the reference sphere, as seen from the image side of the lens, is given by (neglecting skew rays)

$$\begin{aligned} \mathbf{e}(\theta, \phi) &= a(\theta) \mathbb{R}^{-1}(\phi) \cdot \mathbb{L}(\theta) \cdot \mathbb{R}(\phi) \cdot \tilde{\mathbf{E}}(\theta, \phi) \\ &= \mathbb{Q}(\theta, \phi) \cdot \tilde{\mathbf{E}}(\theta, \phi), \end{aligned} \quad (9)$$

where $\mathbb{Q}(\theta, \phi) = a(\theta) \mathbb{R}^{-1}(\phi) \cdot \mathbb{L}(\theta) \cdot \mathbb{R}(\phi)$ and the final rotation \mathbb{R}^{-1} is to transform from the s and p frame of reference back to the initial Cartesian frame. The scalar factor $a(\theta)$ is an apodisation factor that ensures energy is conserved when projecting from a plane to a sphere. For example $a(\theta) = \sqrt{\cos \theta}$ or $a(\theta) = 1$ if the lens satisfies the sine or Herschel condition, respectively [33].

To illustrate the preceding theory, a simple example is given, whereby a homogeneously, albeit arbitrarily, polarised beam, described by the Jones vector $\tilde{\mathbf{E}} = (\tilde{E}_x, \tilde{E}_y, 0)$, is focused by an aplanatic lens (one satisfying the sine condition). Expanding $\mathbb{Q}(\theta, \phi)$ gives

$$\mathbb{Q}(\theta, \phi) = \frac{a(\theta)}{2} \begin{pmatrix} q_1 + q_2 \cos 2\phi & q_2 \sin 2\phi & q_3 \cos \phi \\ q_2 \sin 2\phi & q_1 - q_2 \cos 2\phi & q_3 \sin \phi \\ -q_3 \cos \phi & -q_3 \sin \phi & q_4 \end{pmatrix}, \quad (10)$$

where

$$q_1 = \cos \theta + 1, \quad (11a)$$

$$q_2 = \cos \theta - 1, \quad (11b)$$

$$q_3 = 2 \sin \theta, \quad (11c)$$

$$q_4 = 2 \cos \theta, \quad (11d)$$

which in turn yields

$$\mathbf{e}(\theta, \phi) = \frac{a(\theta)}{2} \begin{pmatrix} \tilde{E}_x(q_1 + q_2 \cos 2\phi) + q_2 \tilde{E}_y \sin 2\phi \\ \tilde{E}_y(q_1 - q_2 \cos 2\phi) + q_2 \tilde{E}_x \sin 2\phi \\ -q_3(\tilde{E}_x \cos \phi + \tilde{E}_y \sin \phi) \end{pmatrix}. \quad (12)$$

Using the well known identity [34]

$$\begin{aligned} \int_0^{2\pi} \left\{ \begin{matrix} \sin m\alpha \\ \cos m\alpha \end{matrix} \right\} \exp[ia \cos(\alpha - \beta)] d\alpha \\ = 2\pi i^m \left\{ \begin{matrix} \sin m\beta \\ \cos m\beta \end{matrix} \right\} J_m(a), \end{aligned} \quad (13)$$

allows the azimuthal integration to be performed analytically yielding the final result

$$\mathbf{E}(\rho, \varphi, z) = -ikf \begin{pmatrix} \tilde{E}_x(I_0 + I_2 \cos 2\varphi) + \tilde{E}_y I_2 \sin 2\varphi \\ \tilde{E}_y(I_0 - I_2 \cos 2\varphi) + \tilde{E}_x I_2 \sin 2\varphi \\ -2iI_1(\tilde{E}_x \cos \varphi + \tilde{E}_y \sin \varphi) \end{pmatrix}, \quad (14)$$

where

$$I_0 = \int_0^\alpha \sqrt{\cos \theta} \sin \theta (\cos \theta + 1) J_0(k\rho \sin \theta) \exp(ikz \cos \theta) d\theta, \quad (15a)$$

$$I_1 = \int_0^\alpha \sqrt{\cos \theta} \sin^2 \theta J_1(k\rho \sin \theta) \exp(ikz \cos \theta) d\theta, \quad (15b)$$

$$I_2 = \int_0^\alpha \sqrt{\cos \theta} \sin \theta (\cos \theta - 1) J_2(k\rho \sin \theta) \exp(ikz \cos \theta) d\theta. \quad (15c)$$

Evaluating Equation (14) for an x -polarised beam ($\tilde{E}_x = 1, \tilde{E}_y = 0$), it is seen that a non-zero E_y and E_z component of polarisation is generated, except for the on-axis point (see Figure 3). The high NA of 0.95 assumed also gives rise to a significant asymmetry in the focused intensity spot (Figure 3(d)). As the NA of the focusing lens is increased, so both the component mixing and asymmetry become more significant.

3.2. Stratified media

Equation (14) is valid when considering focusing in a single homogeneous medium; however, frequently in

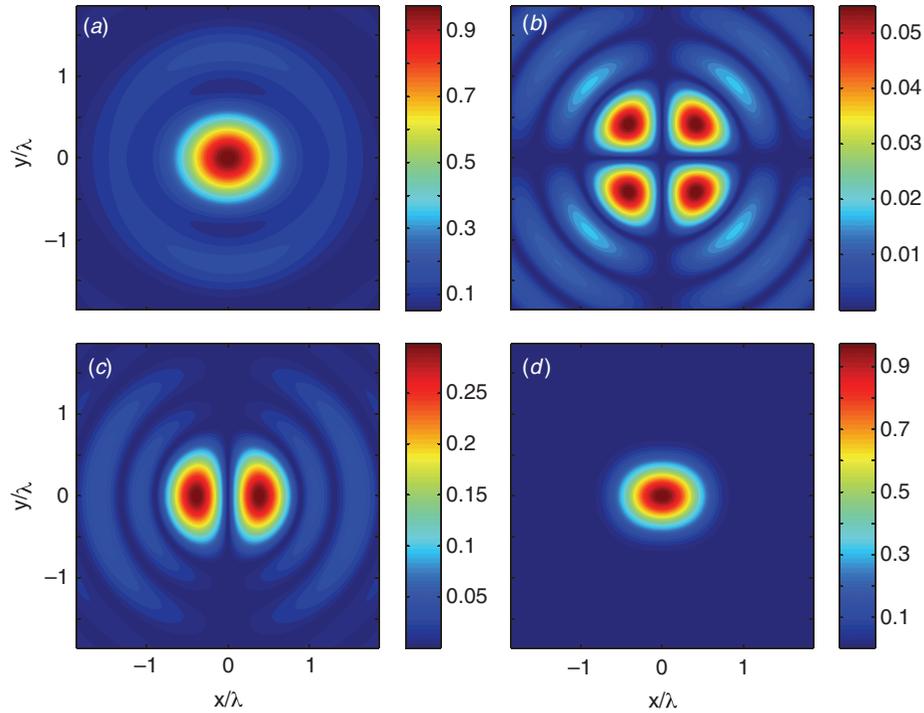


Figure 3. (a)–(c) Absolute magnitude of the Cartesian field components (E_x , E_y , E_z) and (d) optical intensity in the focal plane of a lens of $\text{NA} = 0.95$ for x -polarised illumination. (The colour version of this figure is included in the online version of the journal.)

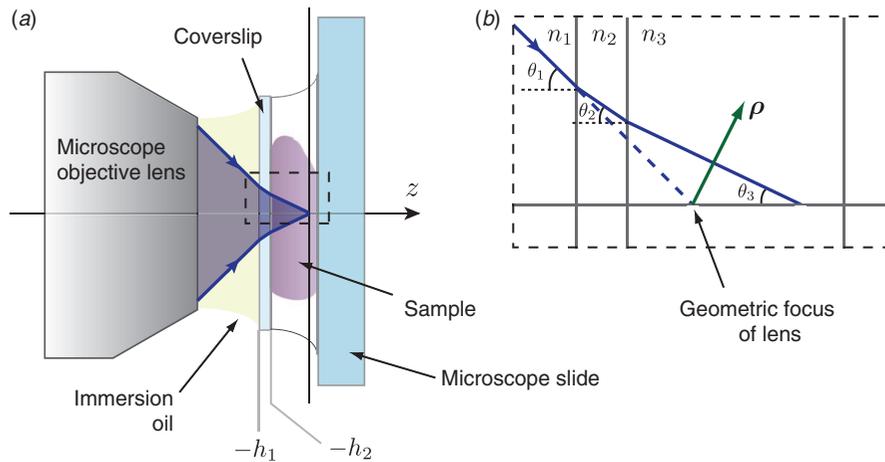


Figure 4. (a) Stratified sample structure that can easily arise when using immersion lenses in biological studies. (b) Zoomed view of dashed rectangle in (a) showing the geometry of the ray tracing problem. (The colour version of this figure is included in the online version of the journal.)

realistic scenarios it is necessary to focus light through a number of materials of differing refractive indices. In biological microscopy for example, samples are commonly mounted on a glass microscope slide with a protective cover slip. High spatial resolution is also frequently sought by use of oil immersion lenses, yielding an illumination arrangement similar to that shown in Figure 4(a). The need to describe focusing

through stratified media hence regularly arises. Vectorial ray tracing again can be applied such that the ray direction in the j th medium is described by the vector $\mathbf{s}_j = (s_{jx}, s_{jy}, s_{jz}) = (\sin \theta_j \cos \phi, \sin \theta_j \sin \phi, \cos \theta_j)$. No index is required for the azimuthal angle ϕ since it is assumed that refraction at each interface does not give rise to skew rays. Similarly, k_j denotes the wavenumber in each medium of refractive index n_j .

The interface between the j th and $(j+1)$ th media is assumed to lie at $z=h_j$, with $h_j < h_{j+1}$. The origin is taken as the geometric focus of the lens (see Figure 4(b)).

As a single ray propagates through a stratified medium it experiences partial transmission and reflection at each interface depending on its polarisation and angle of incidence, as is conveniently parameterised using Fresnel's transmission and reflection coefficients, explicitly given by [26]

$$t_s^{(j)} = \frac{2n_j \cos \theta_j}{n_j \cos \theta_j + n_{j+1} \cos \theta_{j+1}}, \quad (16)$$

$$t_p^{(j)} = \frac{2n_j \cos \theta_j}{n_{j+1} \cos \theta_j + n_j \cos \theta_{j+1}}, \quad (17)$$

$$r_s^{(j)} = \frac{n_j \cos \theta_j - n_{j+1} \cos \theta_{j+1}}{n_j \cos \theta_j + n_{j+1} \cos \theta_{j+1}}, \quad (18)$$

$$r_p^{(j)} = \frac{n_{j+1} \cos \theta_j - n_j \cos \theta_{j+1}}{n_{j+1} \cos \theta_{j+1} + n_j \cos \theta_{j+1}}, \quad (19)$$

respectively, for s and p polarised light passing from the j th medium to the $(j+1)$ th. Consider then converting to a frame of reference that propagates and rotates with a given transmitted ray. For a ray in the j th medium this can be achieved using the transformation matrix

$$\mathbb{P}^{(j)} = \begin{pmatrix} \cos \theta_j & 0 & -\sin \theta_j \\ 0 & 1 & 0 \\ \sin \theta_j & 0 & \cos \theta_j \end{pmatrix}. \quad (20)$$

At each interface the amplitude of the m ($=s, p$) polarised component reduces by a factor of $t_m^{(j)}$, whilst an optical phase of $\beta_j = k_j(h_{j-1} - h_j) \cos \theta_j$ is acquired via propagation through the j th medium. As such, if multiple reflections are neglected, the net transmission coefficient for light after propagating from the 1st interface to just after the $(N-1)$ th, is given by $T_m^{(N-1)} = t_m^{(N-1)} \prod_{j=1}^{N-2} t_m^{(j)} \exp(i\beta_{j+1})$. Realistically, however, a correction factor must also be introduced to account for multiple reflections and the resulting contribution to the total transmitted wave. A complete derivation is omitted here; however, it can be shown (see [24,26,35]) that

$$T_m^{(N-1)} = \frac{t_m^{(N-1)} \prod_{j=1}^{N-2} t_m^{(j)} \exp(i\beta_{j+1})}{D_m^{(N-1)}} \quad (21)$$

where the explicit form for $D_m^{(N-1)}$ is given in [24]. Whilst the form for $D_m^{(N-1)}$ is in general complicated it is possible, in certain cases, to consider this term to be unity, principally when the reflection coefficients $r_m^{(j)}$

are small, i.e. when the mismatch in refractive index of neighbouring media is small.

Following the above analysis for each individual ray in a focusing system, and noting that from Snell's law $k_{js_{jx}} = k_1 s_{1x}$, $k_{js_{jy}} = k_1 s_{1y}$ for all j , gives the field distribution in the N th medium as

$$\begin{aligned} \mathbf{E}_N(\boldsymbol{\rho}) &= -\frac{ik_1 f}{2\pi} \iint_{s_{1x}^2 + s_{1y}^2 \leq 1} \mathbf{e}_N(s_{1x}, s_{1y}) \\ &\quad \times \exp[ik_0(n_N h_{N-1} s_{Nz} - n_1 h_1 s_{1z})] \\ &\quad \times \exp(ik_1 \mathbf{s}_{1t} \cdot \boldsymbol{\rho}_t) \frac{ds_{1x} ds_{1y}}{s_{1z}}, \quad (22) \\ &= -\frac{ik_1 f}{2\pi} \int_0^{2\pi} \int_0^\alpha \mathbf{e}_N(\theta_1, \phi) \exp[ik_0 \Psi] \\ &\quad \times \exp[ik_1 \rho \sin \theta_1 \cos(\phi - \varphi)] e^{ik_N z \cos \theta_N} \\ &\quad \times \sin \theta_1 d\theta_1 d\phi \quad (23) \end{aligned}$$

where the subscript t denotes the transverse components of a vector, the phase term $\Psi = n_N h_{N-1} \cos \theta_N - n_1 h_1 \cos \theta_1$ accounts for propagation in the 1st and N th media and

$$\mathbf{e}_N(\theta_1, \phi) = a(\theta_1) \mathbb{R}^{-1} \cdot \mathbb{P}^{(N-1)-1} \cdot \mathbb{P}^{(N-1)} \cdot \mathbb{P}^{(1)} \cdot \mathbb{L} \cdot \mathbb{R} \cdot \tilde{\mathbf{E}}(\theta_1, \phi) \quad (24)$$

and

$$\mathbb{P}^{(N-1)} = \begin{pmatrix} T_p^{(N-1)} & 0 & 0 \\ 0 & T_s^{(N-1)} & 0 \\ 0 & 0 & T_p^{(N-1)} \end{pmatrix}. \quad (25)$$

Physically focusing through a stratified medium can give rise to an improvement in the resolution of a system, as for example with a solid immersion lens; however, the focused field distribution will in general suffer from spherical aberration.

3.3. Approximations to the Debye–Wolf integral

Whilst the Debye–Wolf integral is a powerful and popular tool in high NA focusing problems, it may be acceptable for numerical accuracy and rigour to be sacrificed in favour of computation times and physical insight. Accordingly, this section discusses a number of approximate and semi-analytic representations.

3.3.1. Series representations

With a view to simplifying the Debye–Wolf integral a number of series expansions have historically been proposed. For example, the exponential terms can be expanded using the formulae of Watson [34], Gradshteyn and Ryzhik [36] and Agrawal and

Pattanayak [37]. Kant has furthermore reported a series expansion based on the Gegenbauer polynomials [38]. More physical representations have also been proposed in which the expansion terms carry a greater significance in real world systems. For example, Braat and colleagues represented the field in the focal region of a high NA lens as a series using Nijboer-Zernike functions [39] hence providing an insight in terms of optical aberration theory (it should however be noted that these functions are not the best for high NA focusing where it is more suitable to establish aberration functions based on spherical harmonics (and orthogonal on a cap of a unit sphere) that, for the low NA case, revert to the circular polynomials of Zernike [40]). Sheppard and Török obtained the field components as a multipole expansion [41], whilst Sherif and co-workers determined an eigenfunction representation of the focused field [42,43], providing insight into the physical restrictions imposed by the finite bandwidth of the lens. A comparison of the computational speed of a number of these expansions was considered in [44]. Two of these expansions will be discussed in greater detail in what follows, namely the eigenfunction and the multipole expansions. The two expansions are illustrative examples of expansions using scalar and vectorial modes. Depending on the specific application either may be more appropriate.

3.3.1.1. Eigenfunction expansion. The Debye–Wolf integral represents the field in the focal region of a lens as the superposition of plane waves of varying strength and polarisation propagating in different directions. The strength and polarisation of each plane wave is determined by the field distribution at infinity over the Gaussian reference sphere as can be determined by vectorial ray tracing as discussed above. As such, if computing the field structure on the focal plane the Debye–Wolf integral can be considered as a component-wise two-dimensional finite Fourier transform from the reference sphere to the focal plane. An additional phase term can be introduced into the kernel of the Fourier transform if a defocus plane is of greater interest or if aberrations are present.

Given the structure of the Debye–Wolf integral, the question can then be raised as to which illumination field patterns are unchanged by focusing of the lens, i.e. what distributions represent the eigenfunctions of the lens operator. Interestingly, this question is synonymous with the question of maximal energy concentration in a given circular domain in the focal plane. Such considerations are of importance in, for example, super resolution studies. Eigenfunctions to the 1D finite Fourier transform were first determined by

Slepian and colleagues [45–48], who also gave an extension to a 2D Fourier transform over a circular domain. The result is separable in the radial and azimuthal directions whereby the radial eigenfunctions $\Phi_{N,n}(c, r)$ are defined by

$$\int_0^{r_0} J_N(\omega r) \Phi_{N,n}(c, r) r dr = (-1)^n \left(\frac{r_0}{\Omega}\right) \lambda_{N,n}^{1/2} \Phi_{N,n}\left(c, \frac{\omega r_0}{\Omega}\right) \quad (26)$$

where $J_m(\dots)$ is the Bessel function of the first kind of order m , $\Phi_{N,n}(c, r)$ are known as the circular prolate spheroidal functions, ω and r are conjugate coordinates and $\lambda_{N,n}$ are the circular prolate spheroidal eigenvalues. The azimuthal eigenfunctions are sinusoidal, or complex exponential, functions. A detailed discussion of the assorted properties of the prolate spheroidal functions can be found in [49]; however, it is important to note their dependence on c , a parameter known as the space bandwidth product. The space bandwidth product is frequently used as a measure of system performance [50,51].

Using Slepian's work, it is possible to show that the field in the focal region of a high NA lens is given by [42]

$$\begin{aligned} \mathbf{E}(\rho, \varphi, z) = & -ikf \left(\frac{u_\alpha}{k\rho_{\max}} \right) \\ & \times \sum_{m=-\infty}^{\infty} \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} i^{|N|} \mathbf{A}_{m,N,n} (-1)^n \\ & \times \sqrt{\lambda_{|N|,n}} \Phi_{|N|,n} \left(c, \frac{u_\alpha \rho}{\rho_{\max}} \right) J_m(kz) e^{iN\varphi}, \end{aligned} \quad (27)$$

where $u_\alpha = \sin \alpha$, ρ_{\max} is the field of view in the focal space and $c = ku_\alpha \rho_{\max}$. $\mathbf{A}_{m,N,n}$ are vector expansion coefficients of the space-limited function

$$\mathbf{a}(u, \phi) = \frac{\mathbf{e}(u, \phi)}{\sqrt{1-u^2}} \exp \left[im \left(\frac{\pi}{2} - \sin^{-1} u \right) \right], \quad (28)$$

such that

$$\mathbf{a}(u, \phi) = \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} \mathbf{A}_{m,N,n} \Phi_{|N|,n}(u, c) \exp(iN\phi)$$

and

$$\begin{aligned} \mathbf{A}_{m,N,n} = & \frac{1}{2\pi \lambda_{|N|,n}} \int_0^{2\pi} \int_0^{u_\alpha} \mathbf{a}(u, \phi) \Phi_{|N|,n}(u, c) \\ & \times \exp(-iN\phi) u du d\phi. \end{aligned} \quad (29)$$

Closer investigation of Equation (27) reveals a number of useful computational properties. In particular the series is rapidly convergent in the radial, azimuthal and axial directions and has been shown to

give good quantitative agreement with direct integration methods [42]. A particularly important result to come from the eigenfunction representation stems from the considerations of the energy contained within a given circular domain in the focal plane. Specifically, the circular prolate spheroidal eigenvalues give a measure of the integrated electric energy density, with larger eigenvalues corresponding to greater concentration of energy and with the lowest order eigenfunction $(N, n) = (0, 0)$ giving the greatest energy concentration of any focused field distribution. This dependence means that higher orders are energetically less significant in the focused distribution and hence the dominant orders are those with lower spatial frequencies or equivalently are less tightly packed. The dominant modes in the focal region hence provide a further means to determine the resolution of the optical system.

Particular benefit can also be drawn from the inherent structure of an eigen representation due to its suitability for the solution of inverse problems in which arbitrary focused field distributions are sought. Synthesis of arbitrary field distributions in optical systems is useful for a wide variety of applications including lithography [52], optical data storage [53], atomic manipulation [54] and polarisation microscopy [55]. Numerical optimisation is frequently used to determine the appropriate illumination field structure required; however, the eigen expansion allows semi-analytic inversion. Specifically, given Equation (27) for the field in the focal region of a high NA lens and the orthogonality of the generalised prolate spheroidal functions [49] it is possible to derive the relation [43]

$$\sum_{m=-\infty}^{\infty} J_m(kz) \mathbf{A}_{m,N,n} = \frac{i}{kf} \left(\frac{u_\alpha}{k\rho_{\max}} \right) \frac{(-1)^n}{i^{|N|}} \lambda_{|N|,n}^{-1/2} \mathbf{B}_{N,n}, \quad (30)$$

where

$$\mathbf{B}_{N,n} = \frac{1}{2\pi\lambda_{|N|,n}} \int_0^{2\pi} \int_0^{\rho_{\max}} \mathbf{E}(\rho, \varphi, z) \Phi_{|N|,n} \left(\frac{u_\alpha \rho}{\rho_{\max}} \right) \times \exp(-iN\varphi) \rho d\rho d\varphi. \quad (31)$$

Unfortunately, Equation (30) cannot be solved uniquely to determine the coefficients $\mathbf{A}_{m,N,n}$, which define the illumination field required to produce a specified field $\mathbf{E}(\rho, \varphi, z)$, but can, however, form the basis for numerical optimisation techniques due to the relatively few number of terms required in the summation of Equation (27). A unique solution can, however, be achieved on the focal plane i.e. when $z = 0$

whereby $J_m(0) = 0$ for $m \neq 0$, yielding the simple scaling relation

$$\mathbf{A}_{N,n} = \frac{i}{kf} \left(\frac{u_\alpha}{k\rho_{\max}} \right) \frac{(-1)^n}{i^{|N|}} \lambda_{|N|,n}^{-1/2} \mathbf{B}_{N,n} \quad (32)$$

where the subscript m has now been dropped. The simple scaling relation encapsulated in Equation (32) constitutes the basic inversion formula for the Debye–Wolf integral.

Unfortunately the constraints imposed by Maxwell's equations are not automatically fulfilled by the developed series since each field component is expanded separately, thus presenting complications when attempting to solve inverse problems. In particular, in the specification of an inverse focusing problem one (or more) field components must be left unconstrained. Consider, for example attempting to reduce the width of the intensity profile by specifying the E_x field component as a Dirac delta function centred on the origin. Only one component of the focused field is specified since this introduces two degrees of freedom into the inversion problem as required for polarisation structuring. E_x is hence written in the form

$$E_x(\rho, \varphi, 0) = \frac{1}{\rho} \delta \left(\frac{u_\alpha \rho}{\rho_{\max}} \right) = \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} \lambda_{|N|,n}^{-1} \Phi_{|N|,n} \times \left(c, \frac{u_\alpha \rho}{\rho_{\max}} \right) \Phi_{|N|,n}(c, 0) \exp(iN\varphi), \quad (33)$$

where the second step has used the completeness property of the generalised prolate spheroidal functions [49]. Applying the inversion formula (Equation (32)) and noting $\Phi_{|N|,n}(0) = 0$ for $N \neq 0$, immediately gives

$$A_{N,n}^x = \begin{cases} \frac{i}{kf} \left(\frac{u_\alpha}{k\rho_{\max}} \right) (-1)^n \lambda_{0,n}^{-1/2} \Phi_{0,n}(c, 0) & \text{for } N = 0 \\ 0 & \text{for } N \neq 0 \end{cases}, \quad (34)$$

where the superscript x denotes the x -component of $\mathbf{A}_{N,n}$.

In practice, however, the illumination specified by Equation (34) does not achieve super resolution since insufficient control is exerted on the E_y and E_z components of the focused field. As such, when a delta function is specified for the E_x component, energy is pushed into the E_y component. The resultant focused distribution is then essentially that of a uniformly y -polarised beam for which there is no resolution improvement.

A further attempt to achieve super-resolution can be made by specifying both the E_x and E_y focused field components as Dirac delta functions. Following the same logic as above this means $A_{N,n}^x = A_{N,n}^y$ as given by

Equation (34). Since $a^j(u, \phi) = \sum_{n=0}^{\infty} A_{0,n}^j \Phi_{0,n}(c, u)$ (for $j = x, y, z$) the required incident field distributions can be found using the inverse of Equation (34) and are given by

$$\begin{aligned}\tilde{E}_x(u, \phi) &= \frac{1}{\sqrt{(1-u^2)} Q_{11} Q_{22} - Q_{12} Q_{21}} \sum_{n=0}^{\infty} A_{0,n}^x \Phi_{0,n}(c, u), \\ \tilde{E}_y(u, \phi) &= \frac{1}{\sqrt{(1-u^2)} Q_{11} Q_{22} - Q_{12} Q_{21}} \sum_{n=0}^{\infty} A_{0,n}^y \Phi_{0,n}(c, u),\end{aligned}\quad (35)$$

where Q_{pq} denotes the (p, q) th element of \mathbb{Q} .

Having specified two field components on the focal plane means there are four degrees of freedom within the system (phase and amplitude for each field component specified). Such a situation would correspond to the combination of polarisation structuring, apodisation and phase modulation in the pupil plane. However, since $A_{N,n}^x$ and $A_{N,n}^y$ are both real the weighting functions $a^x(u, \phi)$ and $a^y(u, \phi)$ are real (the factor of i in Equation (34) represents a global phase and can safely be ignored). Projecting back to the pupil plane is not a complex operation and hence the field in the pupil plane is also real. It is thus apparent that the field in the pupil plane is linearly polarised and only binary phase modulation is necessary. In [43] this super resolution mask is considered in greater detail. Whilst it is found that the size of the focal spot can be reduced, the relative intensity contained in the central focal spot compared to the sidelobe structure rapidly worsens as more terms in the summation of Equation (35) are considered.

This super-resolution example highlights some caveats that can arise with component-wise scalar expansions; however, they can still prove important in applications in which the response of a system to each field component differs or needs to be tailored. For example optimal coupling of light into single molecules has also been considered in [43]. Such field sensitivity may also be exhibited by nonlinear materials.

3.3.2. Multipole expansion

Any solution to Maxwell's equation can be written as the superposition of electric and magnetic multipole fields with strengths $p_{l'm'}^E$ and $p_{l'm'}^M$. Multipole representations are often adopted in electromagnetic calculations which possess spherical symmetry. Accordingly, within a focusing context they give a natural alternative means to represent the geometric field distributions over the Gaussian reference sphere of a lens [41] and as such are considered here. Multipole representations have been useful in showing, for example that a homogeneously x -polarised illumination contains only

$m = -1$ orders and has equal proportions of electric and magnetic multipoles (albeit oscillating $\pi/2$ out of phase) [41]. Studies on the peak intensities in focused fields can also be simplified using a multipole representation [56].

The components for the electric field components for the (l', m') th order, expressed in spherical polar coordinates are given by [57]

$$E_{r'l'm'}^E = l'(l'+1) \frac{j_{l'}(kr)}{r} Y_{l'}^{m'}(\theta, \phi) \quad (36a)$$

$$E_{\theta l'm'}^E = \frac{1}{r} \frac{d}{dr} (r j_{l'}(kr)) \frac{\partial}{\partial \theta} Y_{l'}^{m'}(\theta, \phi) \quad (36b)$$

$$E_{\phi l'm'}^E = \frac{im'}{\sin \theta} \frac{1}{r} \frac{d}{dr} (r j_{l'}(kr)) Y_{l'}^{m'}(\theta, \phi) \quad (36c)$$

$$E_{r'l'm'}^M = 0 \quad (37a)$$

$$E_{\theta l'm'}^M = -\frac{km'}{\sin \theta} j_{l'}(kr) Y_{l'}^{m'}(\theta, \phi) \quad (37b)$$

$$E_{\phi l'm'}^M = -ik j_{l'}(kr) \frac{\partial}{\partial \theta} Y_{l'}^{m'}(\theta, \phi) \quad (37c)$$

where $j_{l'}(kr)$ is a spherical Bessel function of the first kind. $Y_{l'}^{m'}(\theta, \phi)$ are the spherical harmonics of order $l' = 1, 2, 3, \dots, m' = -l, -l+1, \dots, l$ defined by

$$Y_{l'}^{m'}(\theta, \phi) = \left(\frac{(2l'+1)(l'-m')!}{4\pi(l'+m')!} \right)^{1/2} P_{l'}^{m'}(\cos \theta) \exp(im'\phi) \quad (38)$$

$$= a_{l'm'} P_{l'}^{m'}(\cos \theta) \exp(im'\phi) \quad (39)$$

where $P_{l'}^{m'}$ are the associated Legendre polynomials. The associated magnetic field components for each multipole can easily be found by using the curl equations, although the explicit forms shall not be given here.

Vectorial multipoles are solutions to Maxwell's equation over all space, such that a general electric field can be expressed in the form

$$\mathbf{E}(\boldsymbol{\rho}) = \sum_{\mu} \sum_{l', m'} p_{l'm'}^{\mu} \mathbf{E}_{l'm'}^{\mu}(\boldsymbol{\rho}) \quad (40)$$

where μ denotes either electric (E) or magnetic (M) multipoles. Computation of the field structure in the focal region of a lens therefore reduces to a problem of determining the multipole strengths on the reference sphere, which then automatically yields the field $\mathbf{E}(\boldsymbol{\rho})$. Since the reference sphere is located at infinity, the far field forms of the multipole fields are required. Asymptotic expressions for the spherical Bessel

functions [58], which describe the radial behaviour, then give the far field multipoles as

$$e_{\theta l' m'}^{E(j)} = -(-i)^{l'+1} \frac{\partial}{\partial \theta} Y_{l'}^{m'}(\theta, \phi) \quad (41a)$$

$$e_{\phi l' m'}^{E(j)} = -(-i)^{l'} \frac{m'}{\sin \theta} Y_{l'}^{m'}(\theta, \phi) \quad (41b)$$

$$e_{\theta l' m'}^{M(j)} = -(-i)^{l'} \frac{m'}{\sin \theta} Y_{l'}^{m'}(\theta, \phi) \quad (42a)$$

$$e_{\phi l' m'}^{M(j)} = (-i)^{l'+1} \frac{\partial}{\partial \theta} Y_{l'}^{m'}(\theta, \phi) \quad (42b)$$

where the radial component is identically zero for both electric and magnetic dipoles as is required for transverse waves.

Matching of the fields is most easily achieved if the illumination field is considered in terms of spherical field components (e_r, e_θ, e_ϕ) where $e_r=0$, $e_\theta=e_x(\theta, \phi) \cos \phi + e_y(\theta, \phi) \sin \phi$ and $e_\phi=e_y(\theta, \phi) \cos \phi - e_x(\theta, \phi) \sin \phi$, whereby

$$e_\theta(\theta, \phi) = \sum_{\mu} \sum_{l', m'} p_{l' m'}^{\mu} E_{\theta l' m'}^{\mu}(\theta, \phi) \quad (43a)$$

$$e_\phi(\theta, \phi) = \sum_{\mu} \sum_{l', m'} p_{l' m'}^{\mu} E_{\phi l' m'}^{\mu}(\theta, \phi) \quad (43b)$$

where the transmission coefficients of the lens have been assumed to be equal to unity for simplicity. Equations (43) are coupled due to the two sets of coefficients $p_{l' m'}^E$ and $p_{l' m'}^M$, however, they can be decoupled by considering the linear combinations

$$e_\theta(\theta, \phi) \pm i e_\phi(\theta, \phi) = \sum_{l', m'} q_{l' m'}^{\pm} e^{\mp i \frac{\pi}{2}} f_{l' m'}^{\pm} \quad (44)$$

where

$$f_{l' m'}^{\pm} = \left(\frac{\partial Y_{l'}^{m'}(\theta, \phi)}{\partial \theta} \mp \frac{m'}{\sin \theta} Y_{l'}^{m'}(\theta, \phi) \right) \quad (45)$$

and $q_{l' m'}^{\pm} = p_{l' m'}^E \pm i p_{l' m'}^M$. Assuming a lexicographic ordering of the mode indices (l', m') , the expansion coefficients $p_{l' m'}^{\pm}$ can be determined using the method of moments. For example using the moment functions $f_{l' m'}^{\pm}$ gives

$$\mathbf{q}^{\pm} = \mathbb{F}_{\pm}^{-1} \mathbf{f}_{\pm} \quad (46)$$

where $\mathbf{q}^{\pm} = [q_{l' m'}^{\pm}]$ is a vector formed from stacking the coefficients $q_{l' m'}^{\pm}$,

$$\mathbf{F}_{\pm} = \left[\left[\int \int_S f_{l' m'}^{\pm} f_{l' m'}^{\pm*} dS \right] \right] \quad (47)$$

$$\mathbf{f}_{\pm} = \left[\left[\int \int_S (e_\theta(\theta, \phi) \pm i e_\phi(\theta, \phi)) f_{l' m'}^{\pm*} dS \right] \right] \quad (48)$$

and integration is performed over the reference sphere. In such calculations the matrix equation can be poorly conditioned if a large number of multipole orders are used and the multipole modes and moment functions are not normalised to unity over a sphere and hence it is better suited to slower spatially varying fields distributions.

3.3.3. Approximate forms of focusing integrals

In scenarios where the focused field distribution must be calculated many times, for example in lithographic mask design, rapid calculation times are desired. As such it may be acceptable to compromise high numerical accuracy associated with exact evaluation of the focusing integrals in favour of approximate, yet quicker methods. Numerical evaluation of an integral function entails the computation of the kernel at all points over the integration domain, followed by a weighted summation. Approximation of an integral by an analytic function immediately affords computational gains since fewer function evaluations are required. In this section, it is shown how suitable approximations can be made to the diffraction integrals as was developed in [59]. Whilst the method is illustrated by further considering the example given in Section 3.1, the technique can be applied more generally, as will be illustrated in Section 5.4.

Section 3.1 considered the focusing of an arbitrarily, although homogeneously polarised incident beam. It was seen that analytic results could be found for the field in the focal region of a high NA aperture lens; however, numerical evaluation of the integral functions $I_0(\rho, z)$, $I_1(\rho, z)$ and $I_2(\rho, z)$ is required. It is these integrals that shall be approximated in this section.

The first step in approximating the I integrals is to perform the change of variable $u = \sin \theta / \sin \alpha$, such that Equations (15) assume the typical form

$$I_n(\rho, z) = \int_0^1 g_n(u) J_n(ku\rho \sin \alpha) \exp\left[ikz\sqrt{1-u^2 \sin^2 \alpha}\right] u du. \quad (49)$$

The phase term represents a spherical wavefront, which can be approximated as a paraboloidal wavefront. By the principle of stationary phase, the major contributions to the integral are from the stationary points $u=0$ and $u=1$. As such, the paraboloid is chosen so as to match the spherical wavefront at the edge of the pupil such that [41]

$$I_n(\rho, z) = e^{ikz} \int_0^1 g_n(u) J_n(ku\rho \sin \alpha) \exp\left[-i\frac{su^2}{2}\right] u du \quad (50)$$

where $s = 4kz \sin^2(\alpha/2)$. Taylor expanding the kernel of Equation (50) around the point $u = 0$, and integrating term by term allows the integral to be evaluated. That said, however, only the leading orders contribute significantly, such that higher order terms can be neglected. If restriction is made to the focal plane ($s = 0$) then Equation (50) can be evaluated analytically to give

$$I_{0a}(\rho) \approx A_0 \left(\frac{2J_1(k\rho \sin \alpha)}{k\rho \sin \alpha} \right) \quad (51)$$

$$I_{1a}(\rho) \approx A_1 \left(\frac{J_2(k\rho \sin \alpha)}{k\rho \sin \alpha} \right) \quad (52)$$

$$I_{2a}(\rho) \approx A_2 \left(\frac{J_3(k\rho \sin \alpha)}{2k\rho \sin \alpha} \right) \quad (53)$$

where the approximate forms have also been normalised so as to give the same values as the exact formulae near the focus, such that

$$A_0 = C_{3/2} + C_{5/2} \quad (54)$$

$$A_1 = \frac{4}{\sin \alpha} (C_{3/2} - C_{7/2}) \quad (55)$$

$$A_2 = \frac{12}{\sin^2 \alpha} (C_{3/2} - C_{5/2} - C_{7/2} + C_{9/2}) \quad (56)$$

where $C_n = (1 - \cos^n \alpha)/n$.

Figure 5 compares the approximate and exact integrals, for different NAs. As would be expected, greater quantitative agreement is seen at lower NAs, whilst discrepancies are principally in the sidelobe structure, however, these errors are small even for large NAs. Computationally, it is also interesting to note

that up to a normalisation factor the approximate forms of the integrals are independent of the NA of the focusing system. As such, the integrals need not be re-evaluated for investigation of different apertures, again presenting a speed advantage in optimisation studies.

3.4. Focusing of partially coherent, partially polarised light

Theory detailed thus far has been restricted to the focusing of fully coherent, quasi-monochromatic light. What follows in this section is a more general treatment in which the focusing of quasi-monochromatic, but spatially partially coherent and partially polarised light, is considered. Such considerations are, for example, of growing importance due to development of X-ray lasers [60]. Whilst the discussion below is presented for focusing in homogeneous media, it can be simply extended to stratified media.

Correlation and similar second-order metrics are particularly attractive in coherence theory due to their relation to physically observable quantities, such as intensity or fringe visibility, and as such the cross-spectral density matrix (CSDM), which quantifies the correlation of two quasi-monochromatic field vectors at different positions [61] will be used in this section. In the focal region the CSDM is denoted $\mathbb{W}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \langle \mathbf{E}(\boldsymbol{\rho}_1) \mathbf{E}^\dagger(\boldsymbol{\rho}_2) \rangle$, where \dagger denotes the Hermitian transpose operation. Substituting Equation (8) into the definition of the CSDM yields

$$\begin{aligned} \mathbb{W}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = & |K|^2 \int_0^{2\pi} \int_0^{2\pi} \int_0^\alpha \int_0^\alpha \mathbb{w}(\theta_1, \phi_1, \theta_2, \phi_2) \exp[ik\Delta_{12}] \\ & \times \exp[ikz(\cos \theta_1 - \cos \theta_2)] \\ & \times \sin \theta_1 \sin \theta_2 d\theta_1 d\theta_2 d\phi_1 d\phi_2 \end{aligned} \quad (57)$$

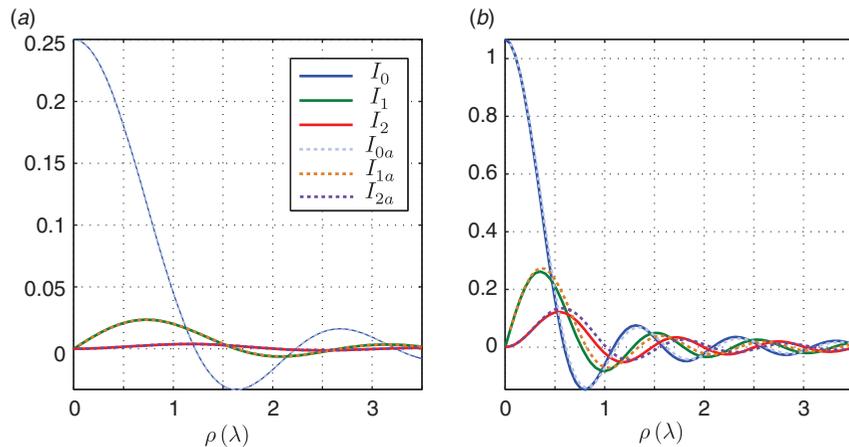


Figure 5. Numerical comparison of the focusing I integrals when calculated using exact (solid plots) and approximate (dashed plots) expressions for a NA of (a) 0.5 and (b) 1. (The colour version of this figure is included in the online version of the journal.)

where $K = ik/(2\pi)$, $\Delta_{12} = \rho_1 \sin \theta_1 \cos(\phi_1 - \varphi_1) - \rho_2 \sin \theta_2 \cos(\phi_2 - \varphi_2)$. Only correlations over a single transverse plane in the focal region are considered (i.e. $z_1 = z_2 = z$ is assumed), since knowledge of the CSDM in a single plane is sufficient to calculate the CSDM on any transverse focal plane, for example, by using the Wolf equations [61]. It should also be noted that the CSDM on the Gaussian reference sphere has been defined as $\mathbb{w}(\theta_1, \phi_1, \theta_2, \phi_2) = \langle \mathbf{e}(\theta_1, \phi_1) \mathbf{e}^\dagger(\theta_2, \phi_2) \rangle$. This can be regarded as an angular correlation function.

In many applications it may be more useful to define the focused CSDM in terms of the CSDM in the back-focal plane of the lens denoted $\tilde{\mathbb{W}}(\theta_1, \phi_1, \theta_2, \phi_2) = \langle \tilde{\mathbf{E}}(\theta_1, \phi_1) \tilde{\mathbf{E}}^\dagger(\theta_2, \phi_2) \rangle$. Using this definition and Equation (9) gives

$$\mathbb{w}(\theta_1, \phi_1, \theta_2, \phi_2) = \mathbb{Q}(\theta_1, \phi_1) \cdot \tilde{\mathbb{W}}(\theta_1, \phi_1, \theta_2, \phi_2) \cdot \mathbb{Q}^\dagger(\theta_2, \phi_2). \quad (58)$$

Substituting Equation (58) into Equation (57) yields

$$\begin{aligned} \mathbb{W}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) &= |K|^2 \int_0^{2\pi} \int_0^{2\pi} \int_0^\alpha \int_0^\alpha \mathbb{Q}(\theta_1, \phi_1) \cdot \tilde{\mathbb{W}}(\theta_1, \phi_1, \theta_2, \phi_2) \\ &\quad \times \cdot \mathbb{Q}^\dagger(\theta_2, \phi_2) \exp[ik\Delta_{12}] \\ &\quad \times \exp[ikz(\cos \theta_1 - \cos \theta_2)] \\ &\quad \times \sin \theta_1 \sin \theta_2 d\theta_1 d\theta_2 d\phi_1 d\phi_2. \end{aligned} \quad (59)$$

Computation of four-fold integrals as embodied in Equations (57) and (59) can prove computationally burdensome and hence often simplifications are sought. Again modal representations can provide these simplifications; however, for partially coherent light these representations take a different form than for fully coherent light. In particular, the CSDM is expanded, using the so-called coherent mode expansion. Coherent mode expansion may be scalar based [61] or vector based [62]. Consideration here will be given only to vector-based expansion since it has been demonstrated that these are more suitable for considering focusing problems, since they easily accommodate the inter-component mixing [63]. A vector-based coherent mode expansion, expresses the CSDM on the reference sphere in the form

$$\mathbb{w}(\theta_1, \phi_1, \theta_2, \phi_2) = \sum_{n=0}^{\infty} \lambda_n \boldsymbol{\Psi}_n(\theta_1, \phi_1) \boldsymbol{\Psi}_n^\dagger(\theta_2, \phi_2) \quad (60)$$

where $\boldsymbol{\Psi}_n(\theta, \phi)$ represents a coherent field distribution and λ_n are the expansion coefficients. It should be noted that Equation (60) requires only a single summation over the index n . This requires the coherent modes to satisfy the Fredholm equation

$$\begin{aligned} \int_0^{2\pi} \int_0^\alpha \mathbb{w}(\theta_1, \phi_1, \theta_2, \phi_2) \boldsymbol{\Psi}_n(\theta_1, \phi_1, \theta_2, \phi_2) \sin \theta_1 d\theta_1 d\phi_1 \\ = \lambda_n \boldsymbol{\Psi}_n(\theta_2, \phi_2). \end{aligned} \quad (61)$$

Accordingly, the focused CSDM is given by

$$\mathbb{W}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \sum_{n=0}^{\infty} \lambda_n \mathbf{C}_n(\boldsymbol{\rho}_1) \mathbf{C}_n^\dagger(\boldsymbol{\rho}_2) \quad (62)$$

where

$$\begin{aligned} \mathbf{C}_n(\boldsymbol{\rho}_l) &= K \int_0^{2\pi} \int_0^\alpha \boldsymbol{\Psi}_n(\theta_l, \phi_l) \\ &\quad \times \exp[ik\rho_l \sin \theta_l \cos(\phi_l - \varphi_l)] e^{ikz \cos \theta_l} \sin \theta_l d\theta_l d\phi_l. \end{aligned} \quad (63)$$

Equation (63) shows that given the coherent modes on the reference sphere, the individual coherent modes can be calculated using conventional coherent theories, i.e. \mathbf{C}_n is found by inputting $\boldsymbol{\Psi}_n$ into the Debye–Wolf diffraction integral. As such, the fourfold integral of Equations (57) and (59) are reduced to more manageable two-fold integrals. Similar relations can be derived if the CSDM in the back focal plane is expanded as opposed to that over the reference sphere. Details will not be given here since they follow the same logic as already presented; however, the interested reader is referred to [63]. Further simplifications can also be shown to be achievable under certain symmetry conditions as also detailed in [63].

4. Scattering model

Having calculated the structure of the illumination field in Section 3, the next stage in determining the image field structure is to calculate how light is scattered from a sample. Unfortunately, the scattered field cannot in general be calculated analytically except for a few specific structures possessing a high degree of symmetry, such as spheroids [64], thus necessitating numerical methods. Whilst rigorous methods, in which the scattered field satisfies Maxwell's equations, provide the most accurate answer, approximate formulations can be adopted in a number of specific scenarios. Two such approximations will be first discussed here, before a brief discussion is given to more rigorous numerical techniques.

4.1. Small dielectric scatterers

The first example that will be considered in this section is scatterers of dimensions much smaller than the wavelength, i.e. $ka \ll 1$ where a denotes the largest dimension of the object. For small dielectric scatterers it is reasonable to use a dipole approximation [65] in which the scattered far field is given by

$$\mathbf{e}(\theta, \phi) = -\mathbf{s} \times (\mathbf{s} \times \mathbf{p}) \frac{\exp(iks)}{s^3}, \quad (64)$$

where \mathbf{s} describes a position vector in the far field ($s = |\mathbf{s}|$), however, the problem remains to find the appropriate dipole moment \mathbf{p} . To do so the scatterer will be assumed to lie on axis at the focus of the illumination lens, such that the illumination field is given by $\mathbf{E}(\mathbf{0}) = \mathbf{E}_0$.

Consider first a small dielectric scatterer. The presence of the scatterer gives rise to a displacement field given by $\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}_0$, where $\epsilon = 1 + \chi$ is the dielectric tensor and χ is the polarisability tensor. It should be noted that the scatterer has been assumed to possess no strong optical non-linearity, such that the polarisation density induced in the sample is given by $\mathbf{P} = \epsilon_0 \chi \mathbf{E}_0$. Since $ka \ll 1$ the scatterer then radiates as an electric dipole with dipole moment \mathbf{p} given by [65,66]

$$\mathbf{p} = k^2 a^3 \epsilon_r \mathbf{E}_0. \quad (65)$$

Although in general the crystal axes of the scatterer need not coincide with the Cartesian coordinate system, such a case does give rise to algebraic simplifications. In particular the dielectric tensor is diagonal and of the form $\epsilon_r = \text{diag}[\epsilon_{xx}^r, \epsilon_{yy}^r, \epsilon_{zz}^r]$, with

$$\epsilon_{kk}^r = \frac{n_{kk}^2 - n_m^2}{n_{kk}^2 + 2n_m^2} \quad (66)$$

where a non-magnetic dielectric has been assumed, n_{kk} denotes the refractive index of the dielectric crystal along a principal axis and n_m denotes the refractive index of the medium in which the scatterer is embedded.

Scattered dipole fields may, however, also arise in alternative scenarios, such as in the imaging of fluorescent molecules [67]. In these cases the radiation dipole moment is typically set by the orientation of the molecule. If the molecule is free to rotate then it will do so, such that its potential energy, $\mathbf{p} \cdot \mathbf{E}_0$, is minimised, i.e. the resultant dipole moment will lie parallel to the illumination field. If, however, the molecule is fixed, for example fluorescent tags rigidly attached to structures in a sample, the effective dipole moment is proportional to $|\mathbf{p}_{\text{mol}} \cdot \mathbf{E}| \mathbf{p}_{\text{mol}}$, where \mathbf{p}_{mol} denotes the inherent dipole moment of the molecule. Magnetic materials are also likely to give rise to magnetic dipoles for which the scattered far field is of the form

$$\mathbf{e}(\theta, \phi) = -(\mathbf{s} \times \mathbf{p}) \frac{\exp(iks)}{s^2}. \quad (67)$$

More consideration will be given to magnetic dipoles later in Section 5 since they will be seen to constitute a key element in rigorous imaging calculations.

4.2. Discrete dipole approximation

The next level of approximation beyond a single dipole is to model a continuous, extended object structure as a 3D array of electric dipoles of appropriate strength, as was first proposed by DeVoe in 1964 [68,69], with later extension by Purcell and Pennypacker [70]. Each dipole in the array oscillates in response to the incident field and the electric field arising from all the other dipoles. The discrete-dipole treatment is hence also known as the coupled-dipole method. The net scattered field is then the coherent superposition of the radiated field from each individual dipole.

Key to discrete dipole calculations is determining the relationship between the permittivity of the original object (and permeability for magnetic media) and the polarisability (and magnetisation) and hence the effective dipole moment, for each individual dipole. Suitable positioning of dipoles is also a core consideration. The Clausius–Mossotti relation is perhaps the most well known method, having been originally used by Purcell and Pennypacker; however, this only yields accurate scattering results in the static limit (i.e. when the frequency of the light $\omega \rightarrow 0$) and for infinite cubic lattices and are thus of limited potential. Corrections for non-static, radiating systems have, however, also been proposed by a number of authors in the field. A fuller discussion of these can be found in [71,72]. Schimura and Milster [73] have also developed an improved algorithm within the context of vectorial diffraction problems, based on lattice dispersion relations.

A full discussion of the discrete dipole method is beyond the scope of this work, however, computationally it is important to note that modelling of object features is restricted by the dipole spacing. Furthermore, scattering calculations from large objects can require significant memory resources, due to the large number of discrete dipoles needed for accurate results and the associated coupling between them. Utilisation of, for example, complex conjugate gradient and fast Fourier transform algorithms, can however help address such issues [74]. The discrete dipole method furthermore, cannot simulate conducting materials well, which generally have large refractive indices.

4.3. Rigorous numerical methods

A number of rigorous calculation tools exist, with the finite element method (FEM) [75,76], the method of moments (MOM) [77] and the finite difference time domain (FDTD) method [78] among the most popular. More recently, new techniques have also been developed including a Green's tensor formalism [79] and rigorous coupled wave theory [80,81], which all derive

from Maxwell's equations. It is unfortunately beyond the scope of this article to discuss each of these methods in detail; however, it should be noted that in terms of determining the image field structure in a given microscope, any method can be used. A brief explanation and discussion of the FDTD method will, however, be given as an example of relevant considerations due to its popularity.

The FDTD is a popular technique, partly due to the ease with which it can be implemented in comparison to alternative methods. This simplicity stems, not only from the algorithm itself, but also the natural manner in which it can be mapped into a practical code. FDTD methods are also particularly memory efficient and consequently capable of modelling scattering from larger aperiodic structures than the MOM and Green's tensor method, or of comparable size to the FEM method. That said, the FEM can suffer from stability problems arising from matrix inversion, an issue that does not arise when using the FDTD method. Furthermore, the FDTD method calculates both the electric and magnetic field vectors simultaneously, hence circumventing the need for further calculations should both field quantities be deemed of interest.

Naturally the FDTD method does however suffer from a number of drawbacks. For example, the FDTD method is inherently a time domain method and thus problems and complications can arise when modelling dispersive media. Additionally, the FDTD method employs a regular orthogonal sampling grid, which is less suited to the the modelling of complex objects, since the typical resolution of object features is limited by the grid spacing. Curved and sloping surfaces must also be approximated by a staircase function, a feature that can only be avoided by using the FEM or MOM, since these methods allow for use of an irregular grid.

Computationally, the FDTD method iteratively solves the curl equations

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \quad (68a)$$

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t}, \quad (68b)$$

where σ is the conductivity and ϵ is the permittivity tensor of the scattering region. It is, however, necessary to discretise the equations, which for the FDTD method is achieved using a Yee discretisation [82] in which an indexing system (i, j, k) denotes the position $(i\Delta x, j\Delta y, k\Delta z)$, where Δx , Δy and Δz are the separation of sample points in the directions of the coordinate axes. A typical separation of sample points is $\lambda/20$. A discretisation in the time domain is also required in which the time steps are Δt and are indexed by n .

Electric and magnetic fields are, however, calculated one half time step ($\Delta t/2$) out of sync. Using this discretisation scheme Yee has shown that the partial differential Equations (68) can be approximated to second order by means of difference equations of the form

$$\begin{aligned} E_x^{n+1/2}(i, j+1/2, k+1/2) &= \alpha(i, j+1/2, k+1/2) E_x^{n-1/2} \\ &\times (i, j+1/2, k+1/2) + \beta(i, j+1/2, k+1/2) \\ &\times \frac{H_z^n(i, j+1, k+1/2) - H_z^n(i, j, k+1/2)}{\Delta y} \\ &- \beta(i, j+1/2, k+1/2) \\ &\times \frac{H_y^n(i, j+1/2, k+1) - H_y^n(i, j+1/2, k)}{\Delta z} \end{aligned} \quad (69)$$

and similarly for other field electric and magnetic field components. α and β are functions of Δt and the material properties σ and ϵ . Before numerical calculations are possible it is necessary to terminate the computational domain. Commonly this is achieved by specification of a boundary condition over the outer domain of the region of interest [83–85]. For open region scattering problems, a perfectly matched layer (PML) [86], which is a layer of absorbing material with a very low reflection coefficient, can be used. Upon specification of an incident field it is possible then to calculate the scattered field at all positions and time by iteration. Whilst it is possible to use an analytic specification of the incident field to determine the incident field at all points and time in the computation domain [87], this can frequently prove computationally expensive and is often eschewed in favour of more approximate methods, such as specifying the field over a single plane near the sample [87]. Frequently, pulsed illuminations are also used to limit computation in the time domain. The scattered field at the central wavelength of the pulse can then be calculated by means of a Fourier analysis.

By way of example, optical data storage (ODS) is briefly considered. To date, advances in ODS technology have been realised by successive increases in the NA of the illumination optics (0.45, 0.6, and 0.85 for CD, DVD and Blu-ray, respectively) and decreases in the wavelength of light employed (780, 650 and 405 nm for CD, DVD and Blu-ray, respectively), resulting in a decrease in the focal spot size and hence data densities achievable. Potential improvements in ODS via NA increase and wavelength reduction are, however, now reaching their limits. For improvements beyond those achievable using immersion lenses, recourse must be made to alternative strategies, such as multiplexing in which a single pit can store more than a single bit of information. For example, the form birefringence

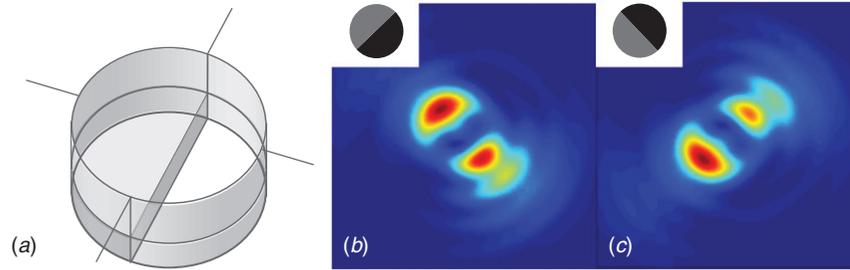


Figure 6. (a) Asymmetric pit structure, exhibiting form birefringence, suitable for polarisation multiplexing in ODS applications. (b)–(c) Scattered intensity distributions as determined using a rigorous FDTD code for differently oriented pit structures (see inset) (from [88]). Reproduced with permission. (The colour version of this figure is included in the online version of the journal.)

inherent in asymmetric scatterers allows information to be encoded in the orientation of a data pit and read out using the polarisation state of light. One such asymmetric data pit is shown in Figure 6(a) (reproduced from [88]). Rigorous scattering calculations were performed for two such pit structures oriented at different angles, when illuminated with a tightly focused linearly polarised beam (a Blu-ray specification for the illumination optics was assumed). Scattered intensity distributions, shown in Figure 6, exhibit a significant dependence upon the angle of the edge within the pit, thus demonstrating the viability of such pits for data multiplexing. Furthermore, scattered distributions are seen to differ significantly from that of a single dipole, despite the subwavelength dimensions assumed for the pit, hence demonstrating the need for more accurate scattering tools in this domain. Theoretical, calculations of this nature prove vital in the development of new ODS technology, for instance.

To conclude this section, brief consideration is given to scattering calculations when an extended sample is illuminated with partially coherent and partially polarised light, as was discussed earlier in Section 3.4 (it should be noted that a point scatterer will always give rise to a spatially coherent scattered field). Whilst a partially coherent, partially polarised illumination implies a statistical treatment of scattering is needed, the coherent methods discussed above can fortunately be employed. In particular, a coherent scattering calculation can be performed for each coherent mode present in the illumination field. The resulting CSDM of the scattered field can then be determined by summing the ‘scattered’ coherent modes in a similar fashion to Equation (62). Unfortunately scattering calculations of this nature require significant computational time, particularly for highly incoherent light, and hence no numerical examples will be presented here or in the following discussion on imaging.

5. Imaging

It is in this section that attention is now given to the calculation of the field structure in the image space of an imaging system. Principally, the input into such a calculation is the scattered field calculated in the preceding section. Whilst imaging of arbitrary electromagnetic fields is possible, as will be covered in Section 5.3, it is necessary to initially consider simpler source fields. Simplest of the cases to be considered will be to image the illumination field, i.e. assume no object structure is present in the imaging system, however, calculations will then be extended to consider the imaging of a dipole field. In turn, imaging of a single dipole allows imaging of arbitrary fields in a similar, albeit rigorous, fashion to the discrete dipole approximation.

5.1. No specimen

To highlight the importance of using a vectorial analysis in high NA systems, the imaging properties of a polarising microscope is studied when no sample is present. The vectorial ray tracing developed thus far is sufficient to consider such a scenario. Consider for example the optical system shown in Figure 7(a) in which a uniformly x -polarised beam $\tilde{\mathbf{E}}_0 = (1, 0, 0)$ is focused and collected by identical high NA lenses. An analyser oriented at an angle γ to the x -axis, is then placed in the back focal plane of the objective lens, the effect of which can be described by the generalised Jones matrix $\mathbb{L}\mathbb{P}(\gamma)$. The field structure directly after the analyser is then given by [23]

$$\tilde{\mathbf{E}}_1 = \mathbb{L}\mathbb{P} \cdot \mathbb{R}^{-1} \cdot \mathbb{L}^{-1} \cdot \mathbb{I}_2 \cdot \mathbb{R} \cdot \mathbb{R}^{-1} \cdot \mathbb{L} \cdot \mathbb{I}_1 \cdot \mathbb{R} \cdot \tilde{\mathbf{E}}_0 \quad (70)$$

$$= \mathbb{L}\mathbb{P} \cdot \begin{pmatrix} T^+ + T^- \cos 2\phi \\ T^- \sin 2\phi \\ 0 \end{pmatrix} \quad (71)$$

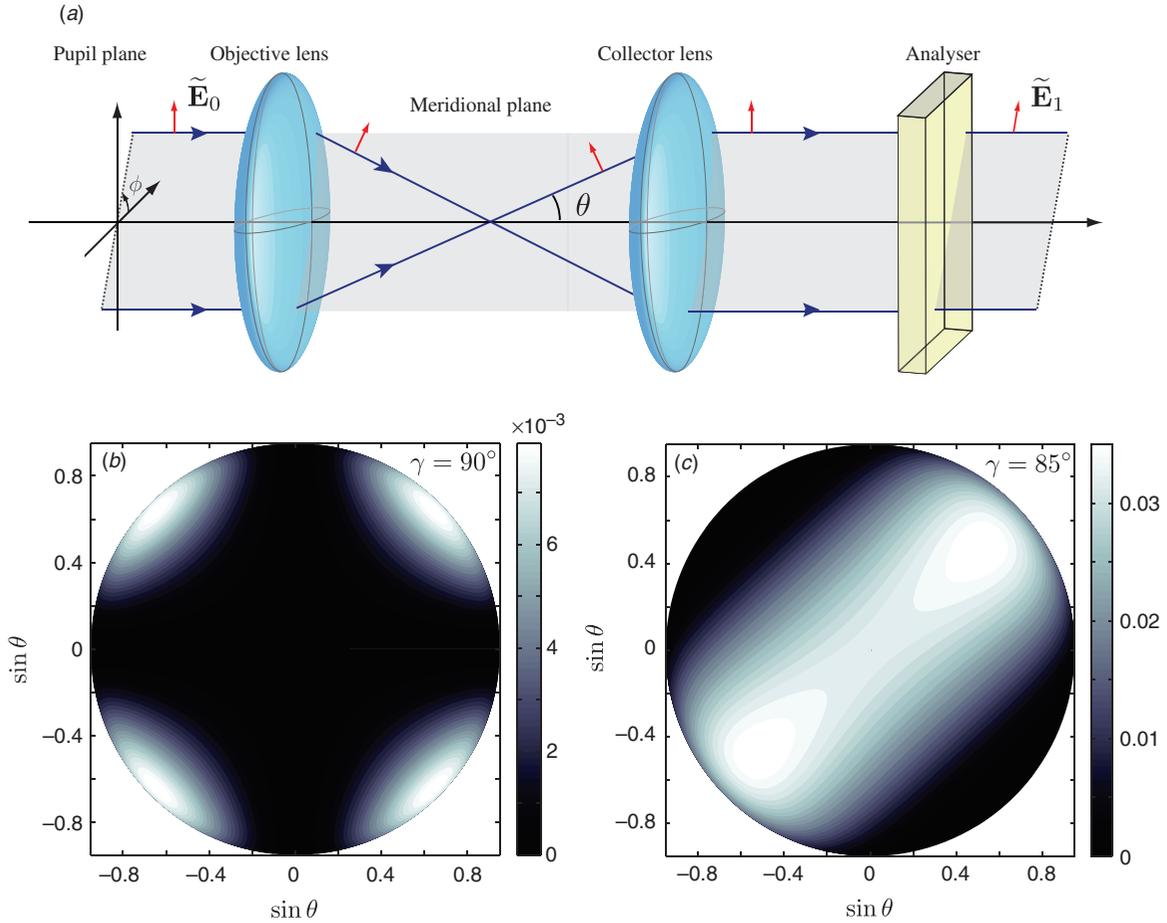


Figure 7. (a) Polarising microscope in which an analyser set at an angle γ is placed in the back focal plane of the collector lens. Red arrows show polarisation direction for some typical rays, assuming an x polarised illumination. Lenses are assumed to be identical. (b) Maltese cross intensity distribution seen just after the analyser for $\gamma = 90^\circ$ and (c) $\gamma = 85^\circ$. (The colour version of this figure is included in the online version of the journal.)

where $T^\pm = (t_{p1}t_{p2} \pm t_{s1}t_{s2})$ such that the intensity distribution is given by [23]

$$I_1 = (\sin \gamma + \cos \gamma)^2 [T^+ + T^- \cos 2\phi] \times \cos \gamma + T^- \sin 2\phi \sin \gamma^2. \quad (72)$$

Importantly, it is noted that the intensity distribution in this case is non-uniform (despite the uniform illumination) due to the polarisation dependent transmission properties of the lens. The resulting characteristic Maltese cross pattern [21] for an analyser oriented perpendicular to the angle of illumination polarisation is shown in Figure 7(b), where total extinction may be naively expected. Figure 7(c) shows the resulting intensity distribution for $\gamma = 85^\circ$ for comparison. Experimental verification of this result can be found in, for example [23]. The consequence of a non-zero extinction will be further considered in Section 6.

It should be noted that a transmission geometry is assumed throughout this work, yet reflection

geometries are frequently employed in optical microscopes. Whilst fundamentally the ray tracing concept is valid and can be adopted in modelling of reflection type imaging systems, some subtleties arise with regards to the coordinate systems used to describe ray directions. Accordingly, slight variations in the forms of the following (and preceding) equations occur. Exact details will not be given here, for the sake of clarity and consistency, however, it is important for any modeller to be aware of the geometry of their system. The assumption of a transmission geometry will thus be maintained. The reader is referred to [89,90], for examples of modelling reflection geometries.

5.2. Single dipole/point scatterer

Electric and magnetic dipoles play a pivotal role in vectorial imaging [89,91]. For example the vectorial

Green's tensors are related to dipole sources [79]. Furthermore, electric dipole emitters represent a good model for single molecules, which are currently receiving much attention in the literature, e.g. [92–95]. Due to their importance, the preceding theory will now be applied to the problem of imaging a single electric dipole. Imaging of a single magnetic dipole will also be considered since this will prove to be critical in the imaging of arbitrary field distributions as discussed in the next section. To do so, the simple transmission geometry of Figure 1 is assumed, where the collector and detector lens have numerical apertures $\text{NA} = \sin \alpha$ and $\text{NA}_d = \sin \alpha_d$, respectively. Both lenses will also be assumed ideal such that Fresnel transmission is neglected for simplicity. The dipole, with moment $\mathbf{p} = (p_x, p_y, p_z)$ is assumed to lie on the optical axis in the front focal plane of the first lens, whilst the detector is placed in the focal plane of the second lens.

The field distribution arising from the radiating dipole source on a sphere located in the pupil of the first lens at infinity, can be found by considering the far-field distribution of an electric and magnetic dipole, as given by Equations (64) and (67), respectively, with $\mathbf{s} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T$. The exponential terms are, however, constant over a sphere and can hence be dropped. Applying the inverse of Equation (9), whereby $\tilde{\mathbf{E}}(\theta, \phi) = \mathbb{Q}^{-1}(\theta, \phi) \cdot \mathbf{e}(\theta, \phi)$, gives the collimated field originating from an electric dipole after the collector lens as

$$\tilde{\mathbf{E}}(\theta, \phi) = \frac{1}{2\sqrt{\cos \theta}} \times \begin{pmatrix} (q_1 + q_2 \cos 2\phi)p_x + q_2 \sin 2\phi p_y - q_3 \cos \phi p_z \\ q_2 \sin 2\phi p_x + (q_1 - q_2 \cos 2\phi)p_y - q_3 \sin \phi p_z \\ 0 \end{pmatrix}, \quad (73)$$

where $q_i(\theta)$ are given by Equation (11).

Refocusing of the collimated field gives rise to a field, $\mathbf{E}_d(\boldsymbol{\rho}_d = (\rho_d, \varphi_d, z_d))$, in the image plane as can be found by evaluation of the Debye–Wolf diffraction integral. The calculations are very similar to those presented in Section 3 and are therefore omitted here for brevity. Full exposition can, however, be found in [96,91] where it is shown that

$$\mathbf{E}_d(\boldsymbol{\rho}_d) = \begin{pmatrix} p_x(K_0^A + K_2^A \cos 2\varphi_d) + p_y K_2^A \sin 2\varphi_d + 2ip_z K_1^A \cos \varphi_d \\ p_x K_2^A \sin 2\varphi_d + p_y(K_0^A - K_2^A \cos 2\varphi_d) + 2ip_z K_1^A \sin \varphi_d \\ -2i(p_x \cos \varphi_d + p_y \sin \varphi_d)K_1^B - 2p_z K_0^B \end{pmatrix}, \quad (74)$$

where

$$K_0^A = \int_0^{\alpha_d} \sqrt{\frac{\cos \theta_d}{\cos \theta}} \sin \theta_d (1 + \cos \theta \cos \theta_d) J_0(k\rho_d \sin \theta_d) \times \exp[ikz_d \cos \theta_d] d\theta_d \quad (75a)$$

$$K_0^B = \int_0^{\alpha_d} \sqrt{\frac{\cos \theta_d}{\cos \theta}} \sin^2 \theta_d \sin \theta J_0(k\rho_d \sin \theta_d) \times \exp[ikz_d \cos \theta_d] d\theta_d \quad (75b)$$

$$K_1^A = \int_0^{\alpha_d} \sqrt{\frac{\cos \theta_d}{\cos \theta}} \sin \theta_d \sin \theta \cos \theta J_1(k\rho_d \sin \theta_d) \times \exp[ikz_d \cos \theta_d] d\theta_d \quad (75c)$$

$$K_1^B = \int_0^{\alpha_d} \sqrt{\frac{\cos \theta_d}{\cos \theta}} \sin^2 \theta_d \cos \theta J_1(k\rho_d \sin \theta_d) \times \exp[ikz_d \cos \theta_d] d\theta_d \quad (75d)$$

$$K_2^A = \int_0^{\alpha_d} \sqrt{\frac{\cos \theta_d}{\cos \theta}} \sin \theta_d (1 - \cos \theta \cos \theta_d) J_2(k\rho_d \sin \theta_d) \times \exp[ikz_d \cos \theta_d] d\theta_d. \quad (75e)$$

These K integrals can be evaluated numerically when complemented with the aplanatic condition $\sin \theta = \beta \sin \theta_d$, where $\beta = f_d/f$ is the magnification of the imaging system as determined by the ratio of the focal lengths f and f_d of the two lenses. If the second focusing lens is of a low NA, as is often the case, then it is possible to approximate Equation (74) since $K_1^A \approx K_0^B \approx 0$. This will be assumed henceforth.

As an example the image field arising from an electric dipole with moment $\mathbf{p} = (1, 0, 0)$ has been calculated assuming the setup and parameters outlined for the system of Figure 1. Figure 8 shows the magnitude of the Cartesian field components \mathbf{E} in the focal plane. Again the effect of the high NA lenses can be seen by the generation of non-zero E_y and E_z field components, although the E_z component is small by virtue of the assumption of a low NA detector lens.

For a magnetic dipole a similar analysis (assuming a low NA detector lens) can be followed [89] which yields

$$\mathbf{E}_d(\boldsymbol{\rho}_d) = \begin{pmatrix} p_y(K_0^A + K_2^A \cos 2\varphi_d) - p_x K_2^A \sin 2\varphi_d \\ p_y K_2^A \sin 2\varphi_d - p_x(K_0^A - K_2^A \cos 2\varphi_d) \\ 2i(p_x \sin \varphi_d - p_y \cos \varphi_d)K_1^B \end{pmatrix}, \quad (76)$$

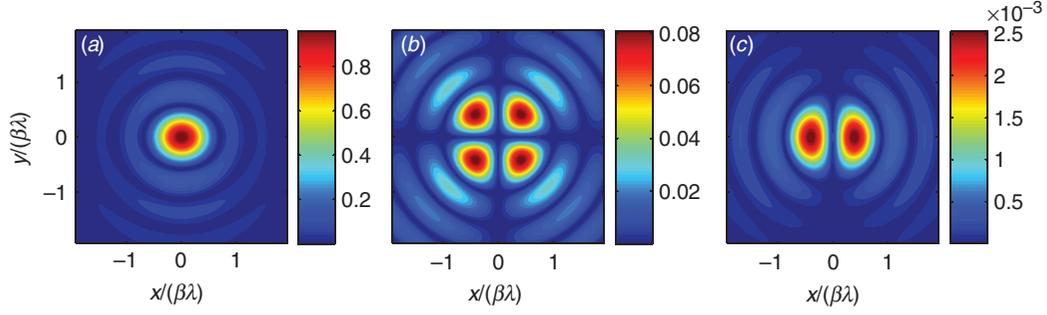


Figure 8. (a)–(c) Absolute magnitude of the Cartesian field components (E_x, E_y, E_z) in the focal plane of the imaging setup of Figure 1, as seen when imaging an electric dipole with moment $\mathbf{p} = (1, 0, 0)$. Fields have been normalised such that the peak intensity is unity. (The colour version of this figure is included in the online version of the journal.)

where

$$K_0^A = \int_0^{\alpha_d} \sqrt{\frac{\cos \theta_d}{\cos \theta}} \sin \theta_d (\cos \theta + \cos \theta_d) J_0(k \rho_d \sin \theta_d) \times \exp[ikz_d \cos \theta_d] d\theta_d \quad (77a)$$

$$K_1^B = \int_0^{\alpha_d} \sqrt{\frac{\cos \theta_d}{\cos \theta}} \sin^2 \theta_d J_1(k \rho_d \sin \theta_d) \exp[ikz_d \cos \theta_d] d\theta_d \quad (77b)$$

$$K_2^A = \int_0^{\alpha_d} \sqrt{\frac{\cos \theta_d}{\cos \theta}} \sin \theta_d (\cos \theta - \cos \theta_d) J_2(k \rho_d \sin \theta_d) \times \exp[ikz_d \cos \theta_d] d\theta_d. \quad (77c)$$

In a similar fashion to Section 3.2, samples are frequently embedded into stratified media. As such it is necessary to incorporate out-propagation through such a structure. As with the treatment in Section 3.2, this problem can be approached by tracing individual rays through the optical system (see [25] and [32] for a full discussion). The media are labelled such that the dipole is embedded in the N th medium in which the wave number is k_N , and such that the medium index decreases with further distance from the dipole, such that the last medium has refractive index n_1 . The interface between the j th and $(j-1)$ th interface is located at $z = h_{j-1}$, whilst the geometric focus of the collector lens is again taken to lie at the origin.

Calculation of the image of an embedded dipole reduces to calculating the field in the front focal plane of the detector lens $\tilde{\mathbf{E}}_d(\theta_d, \phi)$. Whilst the theory developed in Section 3 could then be used to calculate the field structure in the focal region of the detector lens, by first determining the geometrical

approximation to the field on the Gaussian reference sphere, which is then input into the Debye–Wolf integral, if the detector lens is a low NA lens conventional Fourier theory can be used. As such, the field in the focal plane of the detector lens (or detector plane) is given simply by the component-wise Fourier transform of $\tilde{\mathbf{E}}_d(\theta_d, \phi)$, i.e.

$$\mathbf{E}_d(\boldsymbol{\rho}_d) = -\frac{ik_d f}{2\pi} \iint_{\Omega_d} \tilde{\mathbf{E}}_d(\mathbf{s}_d) \exp(ik_d \mathbf{s}_d \cdot \boldsymbol{\rho}_d) \times \exp(ik_0 \Psi) ds_{dx} ds_{dy} \quad (78)$$

where $\Psi = n_N s_N z h_{N-1} - n_1 s_1 z h_1$ again accounts for propagation from the dipole to the N th interface and in the 1st medium. Immediately it is thus seen that the longitudinal field component will be identically zero in the detector plane, since $\tilde{E}_{dz} = 0$, as is a good approximation to reality under a paraxial approximation.

Accounting for Fresnel transmission at each interface of the stratified media yields

$$\tilde{\mathbf{E}}_d(\mathbf{s}_d) = \frac{1}{\sqrt{\cos \theta}} \mathbb{R}^{-1}(\phi) \cdot \mathbb{L}(\theta_1)^{-1} \cdot \mathbb{I}'(N-1) \cdot \mathbb{P}^{(N)}(\theta_N) \cdot \mathbb{R}(\phi) \cdot \mathbf{e}(\theta_N, \phi) \quad (79)$$

where $\mathbf{e}(\theta, \phi)$ is defined by Equations (64) and (67). The dashed notation on the matrix \mathbb{I}' denotes propagation of a wave from the dipole to the lens as opposed to transmission of light from the lens to the dipole as denoted earlier by undashed quantities. Caution must be taken here to ensure that the correct Fresnel transmission coefficients are used.

Upon substituting Equation (79) into Equation (78) and performing the analytic integration over the azimuthal angle ϕ in an analogous manner to Section 3 it is found that

$$\mathbf{E}_d^E(\boldsymbol{\rho}_d) = \begin{pmatrix} p_x(K_0^E + K_2^E \cos 2\varphi_d) + p_y K_2^E \sin 2\varphi_d - 2ip_z K_1^E \cos \varphi_d \\ p_x K_2^E \sin 2\varphi_d + p_y(K_0^E - K_2^E \cos 2\varphi_d) - 2ip_z K_1^E \sin \varphi_d \\ 0 \end{pmatrix} \quad (80)$$

and

$$\mathbf{E}_d^M(\rho_d) = \begin{pmatrix} p_y(K_0^M + -K_2^M \cos 2\varphi_d) + p_x K_2^M \sin 2\varphi_d - 2ip_z K_1^M \cos \varphi_d \\ p_y K_2^M \sin 2\varphi_d - p_x(K_0^M + K_2^M \cos 2\varphi_d) + 2ip_z K_1^M \sin \varphi_d \\ 0 \end{pmatrix} \quad (81)$$

where the superscripts E and M denote the electric and magnetic dipole case, respectively, and

$$K_0^E = \int_0^\alpha \frac{1}{\sqrt{\cos \theta_1}} (T'_s + T'_p \cos \theta_N) \sin 2\theta_1 J_0(k_1 \rho_d \sin \theta_1) \times \exp(-ik_1 z_d \cos \theta_1) \exp(ik_0 \Psi) d\theta_1 \quad (82a)$$

$$K_1^E = \int_0^\alpha \frac{1}{\sqrt{\cos \theta_1}} T'_p \sin \theta_N \sin 2\theta_1 J_1(k_1 \rho_d \sin \theta_1) \times \exp(-ik_1 z_d \cos \theta_1) \exp(ik_0 \Psi) d\theta_1 \quad (82b)$$

$$K_2^E = \int_0^\alpha \frac{1}{\sqrt{\cos \theta_1}} (T'_s - T'_p \cos \theta_N) \sin 2\theta_1 J_2(k_1 \rho_d \sin \theta_1) \times \exp(-ik_1 z_d \cos \theta_1) \exp(ik_0 \Psi) d\theta_1 \quad (82c)$$

and

$$K_0^M = \int_0^\alpha \frac{1}{\sqrt{\cos \theta_1}} (T'_p + T'_s \cos \theta_N) \sin 2\theta_1 J_0(k_1 \rho_d \sin \theta_1) \times \exp(-ik_1 z_d \cos \theta_1) \exp(ik_0 \Psi) d\theta_1 \quad (83a)$$

$$K_1^M = \int_0^\alpha \frac{1}{\sqrt{\cos \theta_1}} T'_s \sin \theta_N \sin 2\theta_1 J_1(k_1 \rho_d \sin \theta_1) \times \exp(-ik_1 z_d \cos \theta_1) \exp(ik_0 \Psi) d\theta_1 \quad (83b)$$

$$K_2^M = \int_0^\alpha \frac{1}{\sqrt{\cos \theta_1}} (T'_p - T'_s \cos \theta_N) \sin 2\theta_1 J_2(k_1 \rho_d \sin \theta_1) \times \exp(-ik_1 z_d \cos \theta_1) \exp(ik_0 \Psi) d\theta_1. \quad (83c)$$

Here z_d denotes the distance from the stratified medium to the lens.

5.3. Arbitrary fields

Given the scattered field from an object, as may be calculated via the methods discussed in Section 4, it is possible to calculate the appropriate image in a similar fashion to that of a single dipole. To see this use is made of another vectorial diffraction integral:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{2\pi} \nabla \times \iint_S \mathbf{n} \times \mathbf{E}_S(\mathbf{r}') \frac{\exp(ikr)}{r} dS \quad (84)$$

where \mathbf{r}' denotes a position of a planar surface S with normal \mathbf{n} directed towards the half space containing the point described by position vector \mathbf{r} and $r = |\mathbf{r} - \mathbf{r}'|$. $\mathbf{E}_S(\mathbf{r}')$ denotes the field distribution on S . All sources

and sinks are assumed to lie in the half space not

containing \mathbf{r} . This diffraction integral can be derived directly from Maxwell's equations, assuming that \mathbf{E} satisfies the vectorial radiation condition [25]. Such a derivation was performed, for example, by Török [25]. Karczewski and Wolf also provided a similar derivation, from which the name 'm-theory' derives; however, they assumed that S represented a finite-sized aperture.

If the order of differentiation and integration is interchanged in Equation (84) it can be shown [89] that Equation (84) can be rewritten in the form

$$\mathbf{E}(\mathbf{r}) = \frac{ik}{2\pi} \iint_S \mathbf{r} \times (\mathbf{n} \times \mathbf{E}_S(\mathbf{r}')) \frac{\exp(ikr)}{r^2} dS \quad (85)$$

in the far field. Whilst this integral will not be used to directly calculate the image of an arbitrary field it does provide significant insight to allow the earlier methods to be used. Specifically, Equation (85) reveals that each point \mathbf{r}' on S , can be considered as the source of a secondary spherical wave with a polarisation of $\mathbf{r} \times (\mathbf{n} \times \mathbf{E}_S(\mathbf{r}'))$. Comparison of this field distribution with that of a magnetic dipole discussed in Section 5.2 shows that the secondary point sources are equivalent to magnetic dipoles with dipole moment given by $\mathbf{n} \times \mathbf{E}_S(\mathbf{r}')$.

Consider then the image of an off-axis dipole in a stratified medium. Displacement of a dipole to an off-axis position $\boldsymbol{\rho}_{dp} = (\rho_{dp}, \varphi_{dp}, 0)^T$ can be modelled by assuming shift invariance of the imaging system [89]. Shift invariant imaging can, for example, be achieved by use of highly corrected $4f$ imaging systems. Given the preceding arguments, final calculation of the image of an arbitrary field distribution can now be approached. To do so requires the image of each equivalent magnetic dipole in Equation (85) to be calculated and then the resulting image field structure from each superposed coherently, such that

$$\mathbf{E}(\boldsymbol{\rho}_d) = \frac{ik_0}{2\pi} \iint_S \mathbf{E}_d^{\text{on-axis}}(\boldsymbol{\rho}_d - \beta \boldsymbol{\rho}_{dp}, \mathbf{n} \times \mathbf{E}(\boldsymbol{\rho}_{dp})) d\boldsymbol{\rho}_{dp} \quad (86)$$

where the explicit dependence on the dipole moment ($\mathbf{n} \times \mathbf{E}(\boldsymbol{\rho}_{dp})$) and the position in object space has been given. Practical implementation of Equation (86) will involve a discretisation of the integral and thus is similar to a discrete-dipole treatment, however, the foundations of Equation (86) are built on a more

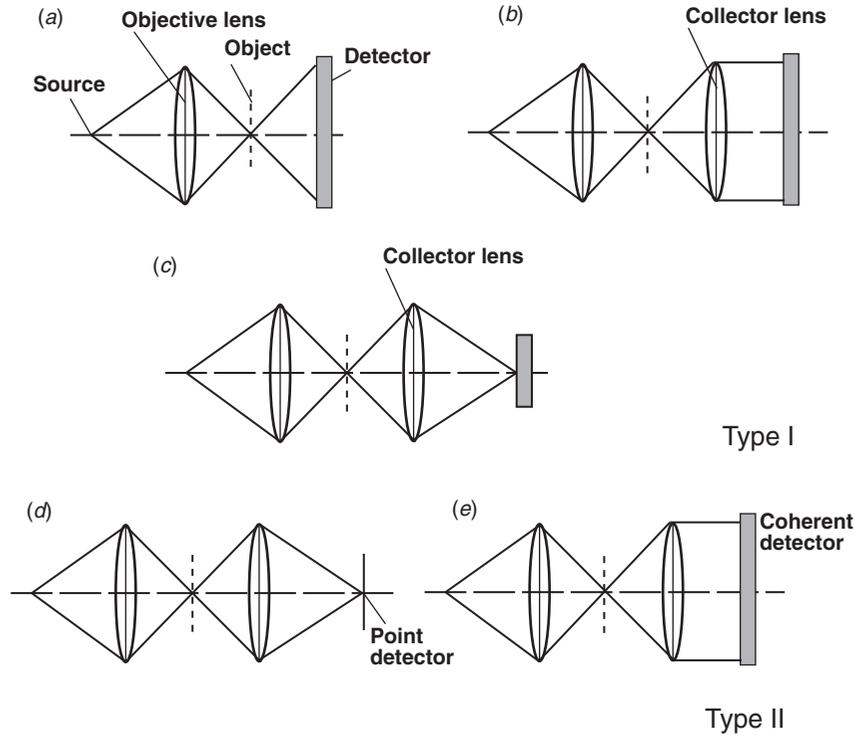


Figure 9. Different imaging system geometries following [97].

rigorous basis and differ since magnetic dipoles are considered as opposed to electric dipole emitters.

5.4. Approximate forms of imaging integrals

As with the focusing I integrals it may be desirable to approximate the imaging K integrals (Equations (75)), if they are to be calculated many times. This is again possible following the procedure set out in Section 3.5. Ultimately it is found

$$K_0^A = A_0^A \left(\frac{2J_1(k\rho \sin \alpha)}{k\rho \sin \alpha} \right), \quad (87a)$$

$$K_0^B = A_0^B \left(\frac{J_2(k\rho \sin \alpha)}{k\rho \sin \alpha} \right), \quad (87b)$$

$$K_1^A = A_1^A \left(\frac{J_3(k\rho \sin \alpha)}{2k\rho \sin \alpha} \right), \quad (87c)$$

$$K_1^B = A_1^B \left(\frac{2J_2(k\rho \sin \alpha)}{(k\rho \sin \alpha)^2} - \frac{J_3(k\rho \sin \alpha)}{k\rho \sin \alpha} \right), \quad (87d)$$

$$K_2^A = A_2^A \left(\frac{J_2(k\rho \sin \alpha)}{k\rho \sin \alpha} \right). \quad (87e)$$

Unfortunately, analytic determination of the normalisation coefficients A_j^k is in general not possible, and hence must be determined numerically. They will, in general, have a dependence on the NA of the objective and detector lens.

6. Detection

Detection constitutes the final stage in any imaging system and it is here that this stage is now considered. A number of alternative detection configurations exist, such as so-called conventional and confocal setups [97]. Polarimetric detection is also growing in popularity and is frequently employed in astronomy, biomedical diagnosis and material studies [98–100] for example.

6.1. Conventional and confocal detection

Conventional microscopy images a large area of a sample in a single measurement such that each point is imaged in parallel. Accordingly, if a CCD were placed in the image plane of the objective lens a whole image could be obtained in a single instant. Frequently, however, a sequential imaging paradigm, in which an image is built up pixel by pixel can prove more flexible and is thus commonly used. Consider then the scanning configurations depicted in Figure 9. In Figure 9(a)

a point source is imaged onto a sample, which transmits part of the incident light. A large area detector is also placed behind the sample that produces an output current proportional to the total intensity of light incident upon it. For each sample position, a single intensity reading is taken, so as to build up the total image by raster scanning the sample position. Equivalently, a second lens can be placed after the sample so as to either collimate, or image, the transmitted light onto a wide area detector as shown in Figure 9(b) and (c), respectively. Since the resulting image is proportional to the incident intensity, this detection configuration is frequently termed incoherent detection. It has also been shown for paraxial systems that a so-called Type 1 scanning microscope possesses the same imaging properties as a conventional microscope [101] and hence is frequently, albeit rather misleadingly, termed a conventional microscope.

If, instead of a large area detector a single point detector is used, or equivalently a pinhole placed (or imaged) in front of the large detector (Figure 9(d)) the image built up is sensitive to the amplitude of the transmitted light [102]. By analogy to Figure 9(b) if the wide area detector is a coherent detector, i.e. field sensitive, the same result is achieved as shown schematically in Figure 9(e). Coherent detectors can practically be implemented by focusing light into a single-mode fibre.

In general, confocal microscopes provide better optical resolution than conventional microscopes. This can heuristically be understood by noting that in a conventional microscope out-of-focus planes in a sample contribute to the in-focus image, giving rise to a blurred image. Confocal microscopes, however, reject these out-of-focus planes (i.e. give an optical section of the sample) to produce clearer, higher resolution images.

Following the preceding discussion it is possible to determine the readout signal for both a conventional and confocal imaging system (denoted with subscripts conv and conf, respectively) such that

$$I_{\text{conv}} = \int_0^{2\pi} \int_0^{\alpha_d} |\mathbf{E}_d(\boldsymbol{\rho}_d)|^2 \sin \theta_d d\theta_d d\phi \quad (88a)$$

$$I_{\text{conf}} = \left| \int_0^{2\pi} \int_0^{\alpha_d} \mathbf{E}_d(\boldsymbol{\rho}_d) \sin \theta_d d\theta_d d\phi \right|^2 \quad (88b)$$

where integration is performed over the exit pupil of the detector lens (it should be noted that integration could equivalently be performed over the exit pupil of the collimating lens or after the sample over the Gaussian reference sphere). The physical basis for the difference between Equations (88) is perhaps easiest to understand by considering the geometries of

Figure 9(b) and 9(e), however, each geometry is equivalent.

Returning to the earlier example of a simple polarising microscope, given above in Section 5.1, the different properties of confocal and conventional can be investigated in the absence of a sample. Respectively, these configurations can be modelled using the 4f system shown in Figure 7, where a large detector is now placed immediately after the analysers. Noting that the action of the analyser gives a fully polarised field, a scalar treatment can be used. Consequently the field incident on the detector is given by [23]

$$U_d(\theta, \phi) = (\sin \gamma + \cos \gamma)([T^+ + T^- \cos 2\phi] \times \cos \gamma + T^- \sin 2\phi \sin \gamma). \quad (89)$$

Substituting Equation (89) into Equations (88) yields (dropping a constant multiplier)

$$I_{\text{conv}} = 2 \cos^2 \gamma \left[\int_0^{\alpha_d} T^+ J_0(k\rho_d \sin \theta) \sin 2\theta d\theta \right]^2 + \left[\int_0^{\alpha_d} T^- J_2(k\rho_d \sin \theta) \sin 2\theta d\theta \right]^2 \quad (90a)$$

and

$$I_{\text{conf}} = \cos^2 \gamma \left| \int_0^{\alpha_d} T^+ J_0(k\rho_d \sin \theta) \sin 2\theta d\theta \right|^2. \quad (90b)$$

Equations (90) allow the detector signal to be estimated for both incoherent and coherent polarisation microscopes with a linearly polarised illumination. Both configurations are seen to exhibit the familiar $\cos^2 \gamma$ dependence on the analyser orientation, however the incoherent configuration gives rise to a background term, which increases as the NA of the collector lens increases. In practice, this background appears as a depolarisation of the light. In general it can hence prove difficult to distinguish depolarisation induced by a sample from that introduced by use of high NA lens. Confocal setups do not suffer this problem if ideal polarisers are used. A more general treatment of polarising microscopes in which arbitrary illumination states and polarisation state analysers are used can be found in [23]. A discussion of geometries of the form of Figure 9(d) with non-pointlike detectors is also considered.

6.2. Polarimetric detection

Polarimetric detection, is similar in some respects to the polarisation microscope discussed above, in the sense that polarised light is passed through an analyser and an image taken. Nevertheless, polarimetric

detection inherently contains greater information since the previous examples consider forming the image for a single given polarisation state (as set by the analyser orientation), whilst polarimetric detection instead gives an image in which the complete polarisation state of the light incident onto the detector is known. To measure the complete state of light, multiple measurements must be taken, so as to uniquely determine all the polarisation parameters. Consider, for example, the division of amplitude polarimeter (DOAP), as originally proposed by Azzam [103] and shown in Figure 10(a). Light incident into the DOAP is divided into four different detection arms in which different analysers are placed.

Each detector hence records an intensity that is proportional to the projection of the incident polarisation state onto a basis measurement state, given by

$$I_{\text{conv},j} = \iint_{S_d} |\mathbb{T}_j \mathbf{E}_d(\boldsymbol{\rho}_d)|^2 d\boldsymbol{\rho}_d \quad (91a)$$

$$I_{\text{conf},j} = \left| \iint_{S_d} \mathbb{T}_j \mathbf{E}_d(\boldsymbol{\rho}_d) d\boldsymbol{\rho}_d \right|^2 \quad (91b)$$

where \mathbb{T}_j represents the generalised Jones matrix for the analyser in the j th detection arm.

Although measurement of the state of polarisation of light is in itself informative, polarimetric techniques are also frequently used to obtain polarisation dependent properties of a sample, such as its birefringence. This domain is commonly known as Mueller matrix polarimetry. Therefore, to extend Equations (91) consider inputting a homogeneously polarised beam into a high NA microscope. Each position on the entrance pupil of the illuminating lens gives rise to a different plane wave incident upon the sample as discussed in Section 3. Each of these plane waves in turn scatters and gives rise to a spectrum of plane waves propagating from the sample. The resulting, total spectrum of scattered plane waves is, in essence, the same as the field over the reference sphere of the collector lens (see Sections 4 and 5). It is then possible to relate the field distribution $\tilde{\mathbf{E}}_0(\theta_0, \phi_0)$ in the entrance pupil of the microscope to that in the exit pupil plane of the collector lens, $\tilde{\mathbf{E}}_d(\theta_d, \phi_d)$ via the equation

$$\tilde{\mathbf{E}}_d(\theta_d, \phi_d) = \int_0^{2\pi} \int_0^\alpha \mathbb{J}(\theta_0, \phi_0, \theta_d, \phi_d) \tilde{\mathbf{E}}_0(\theta_0, \phi_0) \sin \theta_0 d\theta_0 d\phi_0 \quad (92)$$

where $\mathbb{J}(\theta_0, \phi_0, \theta_d, \phi_d)$ is known as the Jones pupil [104].

It is, however, unusual to work with field vectors in polarimetry, since they are not directly measurable. Instead Stokes vectors, denoted $\mathbf{S} = (S_0, S_1, S_2, S_3)^T$ where S_j are known as the Stokes parameters,

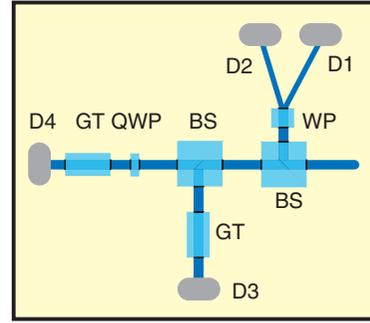


Figure 10. Schematic of a simple DOAP design proposed by Azzam [103]. Notation is as follows: BS - beam splitter, WP - Wollaston prism, QWP - quarter wave plate, GT - Glan Thompson polariser and D - detector. (The colour version of this figure is included in the online version of the journal.)

are often used in preference. Each Stokes parameter in turn can be considered as giving the difference in intensity of horizontal and vertical polarised components in a beam, the difference in intensity of components polarised at $\pm 45^\circ$ and finally the difference in intensity of right- and left-handed circularly polarised components of the beam. In analogy to Equation (92), it is possible to relate the illumination Stokes vector to the distribution in the exit pupil of the collector lens via

$$\tilde{\mathbf{S}}_d(\theta_d, \phi_d) = \int_0^{2\pi} \int_0^\alpha \mathbb{M}(\theta_0, \phi_0, \theta_d, \phi_d) \tilde{\mathbf{S}}_0(\theta_0, \phi_0) \sin \theta_0 d\theta_0 d\phi_0 \quad (93)$$

where $\mathbb{M}(\theta_0, \phi_0, \theta_d, \phi_d)$ is known as the Mueller pupil. Consideration of Equations (91)–(93) shows (see [105]) that the intensity detected by each detector is dependent on the average of the Jones (Mueller) pupil if a confocal (conventional) configuration is used and hence different imaging properties result. The practical implications will, in general, depend on the sample, however standard behaviour can be determined. For example, Jones matrices are capable of representing a phase change induced by a sample, whilst Mueller matrices are not. This effect implies that confocal Mueller matrix polarimeters will be more sensitive to samples with an optical retardation. Averaging performed by the setup furthermore means that neither setup is suitable for measuring highly symmetric polarisation distributions. Finally, it can be seen via the discussions in Sections 5.1 and 6.1 that a non-uniform distribution is generated by the use of high numerical aperture optics. Whilst for confocal setups this symmetry can fortunately negate detrimental effects and can even be utilised (see, for example [95]), in a conventional setup it gives rise to an apparent depolarisation of the light being measured, even if light is pointwise fully polarised.

When measuring the polarisation properties of samples, it will hence not be feasible to distinguish between system and sample induced depolarisation and hence careful consideration should be given to depolarising samples [94].

7. Conclusions

This article was intended to serve as a tutorial for the modelling of vectorial imaging systems. Accordingly it was first necessary to formalise the concept of vectorial ray tracing, which was argued to provide a rigorous and accurate method to determine the infinite boundary conditions in high numerical aperture focusing and imaging integrals, such as the Debye–Wolf diffraction integral. Application of vectorial ray tracing allowed not only fields of arbitrary polarisation and coherence properties to be focused through complex stratified media, but also the development of various modal representations of a focused field. Modal representations in turn allow physical insight to be gained into the focusing operation, improved computational efficiency and also simple implementation of setups employing structured illumination. Diffraction integrals can commonly prove burdensome to compute and, as such, approximate analytic formulae were presented as an alternative; a high degree of accuracy was demonstrated.

Interaction of the illumination field with a given sample structure also plays a vital role in modelling of imaging systems. Although full consideration of this topic is beyond the scope of this article, the principles of an assortment of scattering tools were discussed briefly, such as the discrete dipole approximation and the FDTD method. Having calculated the field scattered from a sample (by whatever method is selected), it is then necessary to propagate this field through the remainder of the imaging system. It was shown that this can be achieved by considering a continuous array of equivalent magnetic dipoles. Given the shift invariant imaging properties of many optical systems this can easily be determined once the image of a single dipole is known. Results were hence also given in this vein, with accompanying approximations that can be made. Finally, the importance of polarisation effects was highlighted by consideration of a simple polarising microscope in which non-uniform and non-zero intensities were seen for certain analyser configurations, in contrast to intuitive expectations.

Ultimately, in any experimental setup, measurements must be taken. The form of the output signal however depends on the imaging modality employed. For example, confocal microscopy was seen to produce a signal that is dependent on the spatial average of the

electric field, whilst conventional microscopy yields an output dependent on the average of the intensity. Extending such considerations to a polarimetric detection, it was also argued that depolarisation can be introduced for a conventional polarimetric detection, in addition to the loss of phase sensitivity.

References

- [1] Zernike, F. *Physica* **1942**, *9*, 686–698.
- [2] Zernike, F. *Physica* **1942**, *9*, 974–986.
- [3] Herman, S. *Fluorescence Microscopy*, 2nd ed.; Taylor & Francis: New York, 1998.
- [4] Boustany, N.N.; Boppart, S.A.; Backman, V. *Annu. Rev. Biomed. Eng.* **2010**, *12*, 285–314.
- [5] Wolf, E. *Proc. Roy. Soc. London A* **1959**, *253*, 349–357.
- [6] Richards, B.; Wolf, E. *Trans. Opt. Inst. Pet.* **1959**, *253*, 358–379.
- [7] Brooker, G. *Modern Classical Optics*; Oxford University Press: Oxford, 2003.
- [8] Waluschka, E. *Opt. Eng.* **1989**, *28*, 86–89.
- [9] McGuire, J.P. Jr; Chipman, R.A. *J. Opt. Soc. Am. A* **1990**, *7*, 1614–1626.
- [10] Jones, R.C. *J. Opt. Soc. Am.* **1941**, *31*, 488–493.
- [11] Jones, R.C. *J. Opt. Soc. Am.* **1941**, *31*, 493–499.
- [12] Jones, R.C. *J. Opt. Soc. Am.* **1941**, *31*, 500–503.
- [13] Jones, R.C. *J. Opt. Soc. Am.* **1942**, *32*, 486–493.
- [14] Jones, R.C. *J. Opt. Soc. Am.* **1947**, *37*, 107–110.
- [15] Jones, R.C. *J. Opt. Soc. Am.* **1947**, *37*, 110–112.
- [16] Jones, R.C. *J. Opt. Soc. Am.* **1948**, *38*, 110–112.
- [17] Jones, R.C. *J. Opt. Soc. Am.* **1956**, *46*, 126–131.
- [18] Török, P.; Varga, P.; Laczik, Z.; Booker, G.R. *J. Opt. Soc. Am. A* **1995**, *12*, 325–332.
- [19] Török, P.; Varga, P.; Booker, G.R. *J. Opt. Soc. Am. A* **1995**, *12*, 2136–2144.
- [20] Higdon, P.D.; Török, P.; Wilson, T. *J. Microsc.* **1998**, *193*, 127–141.
- [21] Inoué, S. *Exp. Cell. Bio.* **1952**, *3*, 199–208.
- [22] Kubota, H.; Inoué, S. *J. Opt. Soc. Am.* **1959**, *49*, 191–198.
- [23] Török, P.; Higdon, P.D.; Wilson, T. *Opt. Commun.* **1998**, *148*, 300–315.
- [24] Török, P.; Varga, P. *Appl. Opt.* **1997**, *36*, 2305–2312.
- [25] Török, P. An Imaging Theory for Advanced, High Numerical Aperture Optical Microscopes. DSc Thesis, 2003.
- [26] Born, M.; Wolf, E. *Principles of Optics*, 7th ed.; Cambridge University Press, Cambridge, 1980.
- [27] Hardy, A.; Treves, D. *J. Opt. Soc. Am.* **1973**, *63*, 85–90.
- [28] Chipman, R.A.; Stacy, J.E. *COSMIC Program NPO-17273*; University of Georgia: Athens, 1983.
- [29] Ignatowsky, V.S. *Trans. Opt. Inst. Pet.* **1919**, *1*, 1–36.
- [30] Luneburg, R.K. *Mathematical Theory of Optics*; University of California Press: USA, 1966.
- [31] Li, Y.; Wolf, E. *J. Opt. Soc. Am. A* **1984**, *1*, 801–808.
- [32] Török, P. *J. Opt. Soc. Am. A* **1998**, *15*, 3009–3015.
- [33] Innes, D.J.; Bloom, A.L. *Spec.-Phys. Laser Tech. Bull.* **1966**, *5*, 1–10.

- [34] Watson, G.N. *A Treatise on the Theory of Bessel Functions*; Cambridge University Press: Cambridge, 1995.
- [35] Flagello, D.G.; Milster, T.; Rosenbluth, A.E. *J. Opt. Soc. Am. A* **1996**, *13*, 53–64.
- [36] Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series and Products*; Elsevier Academic Press: USA, 1980.
- [37] Agrawal, G.P.; Pattanayak, D.N. *J. Opt. Soc. Am.* **1979**, *69*, 575–578.
- [38] Kant, R. *J. Mod. Opt.* **1993**, *40*, 2293–2310.
- [39] Braat, J.J.M.; Dirksen, P.; Janssen, A.J.E.M.; van des Nes, A.S. *J. Opt. Soc. Am. A* **2003**, *20*, 2281–2292.
- [40] Sheppard, C.J.R. *Appl. Opt.* **1988**, *27*, 4782–4786.
- [41] Sheppard, C.J.R.; Török, P. *J. Mod. Opt.* **1997**, *44*, 803–818.
- [42] Sherif, S.S.; Foreman, M.R.; Török, P. *Opt. Express* **2008**, *16*, 3397–3407.
- [43] Foreman, M.R.; Sherif, S.S.; Munro, P.R.T.; Török, P. *Opt. Express* **2008**, *16*, 4901–4917.
- [44] Török, P.; Hewlett, S.J.; Varga, P. *J. Mod. Opt.* **1997**, *44*, 493–503.
- [45] Slepian, D.; Pollak, H.O. *Bell System Tech. J.* **1961**, *40*, 43–64.
- [46] Landau, H.J.; Pollak, H.O. *Bell System Tech. J.* **1961**, *40*, 65–84.
- [47] Landau, H.J.; Pollak, H.O. *Bell System Tech. J.* **1962**, *41*, 1295–1336.
- [48] Slepian, D. *Bell System Tech. J.* **1964**, *43*, 3009–3057.
- [49] Frieden, B.R. *Progress in Optics IX*; North Holland: Amsterdam, 1971; pp 311–407.
- [50] Lohmann, A.W.; Dorsch, R.G.; Mendlovic, D.; Zalevsky, Z.; Ferreira, C. *J. Opt. Soc. Am. A* **1996**, *13*, 470–473.
- [51] Neifeld, M.A. *Opt. Lett.* **1998**, *18*, 1477–1479.
- [52] Endo, M., Pattern Formation Method and Exposure System. Patent No. 7094521, 2006.
- [53] Brand, U.; Hester, G.; Grochmalicki, J.; Pike, R. *J. Opt. A: Pure Appl. Opt.* **1999**, *1*, 794–800.
- [54] Rohrbach, A.; Huisken, J.; Stelzer, E.H.K. Optical Trapping of Small Particles. In *Optical Imaging and Microscopy – Techniques and Advanced Systems*, 2nd ed.; Springer-Verlag: New York, 2007.
- [55] Inoué, S. Exploring Living Cells and Molecular Dynamics with Polarized Light Microscopy. In *Optical Imaging and Microscopy – Techniques and Advanced Systems*, 2nd ed.; Springer-Verlag: New York, 2007.
- [56] Moore, N.J.; Alonso, M.A.; Sheppard, C.J.R. *J. Opt. Soc. Am. A* **2007**, *24*, 3115–3122.
- [57] Stratton, J.A. *Electromagnetic Theory*; McGraw-Hill: New York, 1941.
- [58] Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions*; Dover: USA, 1970.
- [59] Sheppard, C.J.R.; Török, P. *Optik* **1997**, *105*, 77–82.
- [60] Sommargren, G.E.; Seppala, L.G. *Appl. Opt.* **1993**, *32*, 6938–6944.
- [61] Wolf, E. *Introduction to the Theory of Coherence and Polarization of Light*; Cambridge University Press: Cambridge, 2007.
- [62] Gori, F.; Santarsiero, M.; Simon, R.; Piquero, G.; Borghi, R.; Guattari, G. *J. Opt. Soc. Am. A* **2003**, *20*, 78–84.
- [63] Foreman, M.R.; Török, P. *J. Opt. Soc. Am. A* **2009**, *26*, 2470–2479.
- [64] Török, P.; Higdon, P.D.; Juškaitis, R.; Wilson, T. *Opt. Commun.* **1998**, *155*, 335–341.
- [65] van de Hulst, H.C. *Light Scattering by Small Particles*; Dover: USA, 1981.
- [66] Török, P. *Opt. Lett.* **2000**, *181*, 7–18.
- [67] Higdon, P.D.; Török, P.; Wilson, T. *J. Microsc.* **1999**, *193*, 127–141.
- [68] DeVoe, H. *J. Chem. Phys.* **1964**, *43*, 393–400.
- [69] DeVoe, H. *J. Chem. Phys.* **1965**, *43*, 3199–3208.
- [70] Purcell, E.M.; Pennypacker, C.R. *Astrophys. J.* **1973**, *186*, 705–714.
- [71] Draine, B.T.; Goodman, J. *Astrophys. J.* **1993**, *405*, 685–697.
- [72] Lakhtakia, A.; Mulholland, G.W. *J. Res. Natl. Inst. Stand. Technol.* **1993**, *98*, 699–716.
- [73] Shimura, K.; Milster, T.D. *J. Opt. Soc. Am. A* **2001**, *18*, 2895–2900.
- [74] Draine, B.T.; Flatau, P.J. *J. Opt. Soc. Am. A* **1994**, *11*, 1491–1499.
- [75] Wei, X.; Wachters, A.J.; Urbach, H.P. *J. Opt. Soc. Am. A* **2007**, *24*, 866–881.
- [76] Şendur, K.; Challener, W.; Peng, C. *J. Appl. Phys.* **2004**, *96*, 2743–2752.
- [77] Pocklington, H. *Proc. Cam. Phil. Soc.* **1897**, *9*, 324–332.
- [78] Liu, J.; Xu, B.; Chong, T.C. *Jap. J. Appl. Phys.* **2000**, *39*, 687–692.
- [79] Martin, O.; Dereux, A.; Girard, C. *J. Opt. Soc. Am. A* **1994**, *11*, 1073–1080.
- [80] Moharam, M.G.; Gaylord, T.K. *J. Opt. Soc. Am.* **1981**, *71*, 811–818.
- [81] Moharam, M.G.; Gaylord, T.K. *J. Opt. Soc. Am.* **1983**, *73*, 451–455.
- [82] Yee, K. *IEEE Trans. Antennas Propag.* **1966**, *14*, 302–307.
- [83] Bayliss, A.; Turkel, E. *Commun. Pure App. Math.* **1980**, *23*, 707–725.
- [84] Enquist, B. *Math. Comput.* **1977**, *31*, 629–651.
- [85] Mur, G. *IEEE Trans. Electromag. Compat.* **1981**, *23*, 377–382.
- [86] Berenger, J.P. *J. Comput. Phys.* **1994**, *114*, 185–200.
- [87] Taflove, A.; Hagness, S. *Computational Electrodynamics: The Finite Difference Time Domain Method*; Artech House: Boston, MA, 2000.
- [88] Munro, P.R.T. Personal communication 2009.
- [89] Munro, P.R.T.; Török, P. *Opt. Express* **2007**, *15*, 9293–9307.
- [90] Török, P.; Munro, P.R.T.; Kriezis, E.E. *Opt. Express* **2008**, *16*, 507–523.
- [91] Török, P.; Higdon, P.D.; Wilson, T. *Opt. Commun.* **1998**, *45*, 1681–1698.
- [92] Ha, T.; Enderle, T.; Chemla, D.S.; Selvin, P.R.; Weiss, S. *Phys. Rev. Lett.* **1996**, *77*, 3979–3982.
- [93] Jovin, T.M.; Bartholdi, M.; Vaz, W.L.C.; Austin, R.H. *Ann. N.Y. Acad. Sci.* **1981**, *366*, 176–196.

- [94] Foreman, M.R.; Sherif, S.S.; Török, P. *Opt. Express* **2007**, *15*, 13597–13606.
- [95] Foreman, M.R.; Macías Romero, C.; Török, P. *Opt. Lett.* **2008**, *33*, 1020–1022.
- [96] Munro, P.R.T. Application of Numerical Methods to High Numerical Aperture Imaging. Ph.D. Thesis, 2006.
- [97] Wilson, T.; Sheppard, C. *Theory and Practice of Scanning Optical Microscopy*; Academic Press: London, 1984.
- [98] Readhead, A.C.S.; Myers, S.T.; Pearson, T.J.; Sievers, J.L.; Mason, B.S.; Contaldi, C.R.; Bond, J.R.; Bustos, R.; Altamirano, P.; Achermann, C.; Bronfman, L.; Carlstrom, J.E.; Cartwright, J.K.; Casassus, S.; Dickinson, C.; Holzappel, W.L.; Kovac, J.M.; Leitch, E.M.; May, J.; Padin, S.; Pogosyan, D.; Pospieszalski, M.; Pryke, C.; Reeves, R.; Shepherd, M.C.; Torres, S. *Science* **2004**, *306*, 836–844.
- [99] Oldenbourg, R. *Nature* **1996**, *381*, 811–812.
- [100] Ellis, J.; Dogariu, A. *Phys. Rev. Lett.* **2005**, *95*, 203905.
- [101] Welford, W.T. *J. Micros.* **1972**, *96*, 105–107.
- [102] Sheppard, C.J.R.; Choudhury, A. *J. Mod. Opt.* **1976**, *24*, 1051–1073.
- [103] Azzam, R.M.A.; Bashara, N.M. *J. Mod. Opt.* **1982**, *29*, 685–689.
- [104] Totzeck, M.; Gräupner, P.; Heil, T.; Göhnermeier, A.; Dittmann, O.; Krähmer, D.; Kamenov, V.; Ruoff, J.; Flagello, D. *J. Microlithogr. Microfabr., Microsyst.* **2005**, *4*, 031108.
- [105] Munro, P.R.T.; Török, P. *Opt. Lett.* **2008**, *33*, 2428–2430.
- [106] Török, P.; Kao, F.J., Eds. *Optical Imaging and Microscopy - Techniques and Advanced Systems*, 2nd ed.; Springer-Verlag: New York, 2007.