

Inversion of the Debye-Wolf diffraction integral using an eigenfunction representation of the electric fields in the focal region

Matthew R. Foreman,¹ Sherif S. Sherif,² Peter R. T. Munro,¹
and Peter Török^{1*}

¹Blackett Laboratory, Department of Physics, Imperial College London, Prince Consort Road, London SW7 2BZ, UK

²Optics Group, Institute for Microstructural Sciences, Canadian National Research Council, 1200 Montreal Rd, Ottawa ON K1A 0R6 Canada

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

Abstract: The forward problem of focusing light using a high numerical aperture lens can be described using the Debye-Wolf integral, however a solution to the inverse problem does not currently exist. In this work an inversion formula based on an eigenfunction representation is derived and presented which allows a field distribution in a plane in the focal region to be specified and the appropriate pupil plane distribution to be calculated. Various additional considerations constrain the inversion to ensure physicality and practicality of the results and these are also discussed. A number of inversion examples are given.

© 2008 Optical Society of America

OCIS codes: (000.4430) Numerical approximation and analysis; (050.1960) Diffraction theory; (100.3190) Inverse problems; (180.0180) Microscopy

References and links

1. M. Endo, "Pattern formation method and exposure system" Patent No. 7094521 (2006)
2. U. Brand, G. Hester, J. Grochmalicki, and R. Pike "Super-resolution in optical data storage" *J. Opt. A: Pure Appl. Opt.* **1**, 794–800 (1999).
3. A. Rohrbach, J. Huisken, and E. H. K. Stelzer, "Optical trapping of small particles," in *Optical Imaging and Microscopy - Techniques and Advanced Systems*, P. Török and F.-J. Kao eds., (Springer, New York 2007).
4. S. Inoué, "Exploring living cells and molecular dynamics with polarized light microscopy," in *Optical Imaging and Microscopy - Techniques and Advanced Systems*, P. Török and F.-J. Kao eds., (Springer, New York 2007).
5. T. di Francia, "Super-gain antennas and optical resolving power," *Nuovo Cimento Suppl* **9**, 426–435 (1952).
6. D. R. Chowdhury, K. Bhattacharya, S. Sanyal, and A. K. Chakraborty, "Performance of a polarization-masked lens aperture in the presence of spherical aberration," *J. Opt. A: Pure Appl. Opt.* **4**, 98–104 (2002).
7. D. Gabor, "A new microscopic principle," *Nature (London)* **161**, 777–778 (1948).
8. W. H. Lee, "Computer-generated holograms: techniques and applications," *Prog. Opt.* **16**, 119232 (1978).
9. S.-S. Yu, B.J. Lin, A. Yen, C.-M. Ke, J. Huang, B.-C. Ho, C.-K. Chen, T.-S. Gau, H.-C. Hsieh, and Y.-C. Ku, "Thin-film optimization strategy in high numerical aperture optical lithography I - Principles," *J. Microlith. Microfab. Microsyst.* **4**, 043003 (2005).
10. S. S. Sherif, M. R. Foreman, and P. Török, "Eigenfunction expansion of the electric fields in the focal region of a high numerical aperture focusing system," *Opt. Express* (to be published).
11. M. A. A. Neil, T. Wilson and R. Juškaitis, "A wavefront generator for complex pupil function synthesis and point spread function engineering," *J. Microsc.* **197**, 219–223 (2000).

12. D. Slepian "Prolate spheroidal wave functions, Fourier analysis and uncertainty IV Extensions to many dimensions; Generalised prolate spheroidal functions," *Bell Syst. Tech. J.* **43**, 3009–3057 (1964).
13. D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty I," *Bell Syst. Tech. J.* **40**, 43–64 (1961).
14. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty II," *Bell Syst. Tech. J.* **40**, 65–84 (1961).
15. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty III The dimension of the space of essentially time- and band-limited signals," *Bell Syst. Tech. J.* **41**, 1295–1336 (1962).
16. A. W. Lohmann, R. G. Dorsch, D. Mendlovic, Z. Zalevsky and C. Ferreira, "Space bandwidth product of optical signals and systems," *J. Opt. Soc. Am. A* **13**, 470–473 (1996).
17. M. A. Neifeld, "Information, resolution, and space bandwidth product," *Opt. Lett.* **18**, 1477–1479 (1998).
18. J. C. Heurtley, "Hyperspheroidal functions - optical resonators with circular mirrors," *Proc. Symp. on Quasi Optics*, New York p.367 (1964)
19. B. R. Frieden "Evaluation, design and extrapolation methods for optical signals, based on use of the prolate functions," *Prog. Opt.* **9**, 311–407 (1971).
20. D. Slepian, "Some comments on Fourier analysis, uncertainty and modeling," *SIAM Review* **25**, 379–393 (1983).
21. E. Wolf "Electromagnetic diffraction in optical systems I. An integral representation of the image field," *Proc. Roy. Soc. A-Math Phys* **253**, 349–357 (1959).
22. B. Richards and E. Wolf "Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system," *Proc. Roy. Soc. A-Math Phys* **253**, 358–379 (1959).
23. P. Török, P. D. Hidgon and T. Wilson, "On the general properties of polarised light conventional and confocal microscopes," *Opt. Commun.* **148**, 300–315 (1998).
24. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed., (McGraw-Hill 1996).
25. B. Karczewski and E. Wolf, "Comparison of three theorems of electromagnetic diffraction at an aperture Part I: coherence matrices, Part II: The far field," *J. Opt. Soc. Am.* **56**, 1207–19 (1966).
26. S. S. Sherif and P. Török, "Pupil plane masks for super-resolution in high numerical aperture focussing," *J. Mod. Opt.* **51** 2007–2019 (2004).
27. R. Pike, D. Chana, P. Neocleous, and S. Jiang, "Superresolution in scanning optical systems," in *Optical Imaging and Microscopy - Techniques and Advanced Systems*, P. Török and F.-J. Kao eds., (Springer, New York 2007).
28. T. Zolezzi "Well-posedness criteria in optimization with application to the calculus of variations," in *Nonlinear Analysis: Theory, Methods and Applications*, **25**, 437–453 (1995).
29. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, (SIAM, Philadelphia, PA. 1997).
30. T. Ha, T. Enderle, D. S. Chemla, P. R. Selvin, and S. Weiss, "Single molecule dynamics studied by polarization modulation," *Phys. Rev. Lett.* **77**, 3979–3982 (1996).
31. B. Sick, B. Hecht, and L. Novotny "Orientational imaging of single molecules by annular illumination," *Phys. Rev. Lett.* **85** 4482–4485 (2000).
32. Y. Mushiaki, K. Matsumura, and N. Nakajima, "Generation of radially polarized optical beam mode by laser oscillation," *Proc. IEEE.* **60**, 1107–1109 (1972).
33. S. C. Tidwell, D. H. Ford, and W. D. Kimura, "Generating radially polarized beams interferometrically," *Appl. Opt.* **29**, 2234–2239 (1990).
34. Z. Bomzon, G. Biener, V. Kleiner, and E. Hasman, "Radially and azimuthally polarized beams generated by space-variant dielectric subwavelength gratings," *Opt. Lett.* **27**, 285–287 (2002).
35. K. C. Toussaint Jr., S. Park, J. E. Jureller, and N. F. Scherer, "Generation of optical vector beams with a diffractive optical element interferometer," *Opt. Lett.* **30**, 2846–2848 (2005).
36. M. A. A. Neil, F. Massoumian, R. Juškaitis, and T. Wilson "Method for the generation of arbitrary complex vector wave fronts," *Opt. Lett.* **27**, 1929–1931 (2002).
37. K. S. Youngworth and T. G. Brown, "Focusing of high numerical aperture cylindrical vector beams," *Opt. Express* **7** 77–87 (2000).
38. W. T. Welford "Use of Annular Apertures to Increase Focal Depth," *J. Opt. Soc. Am.* **50**, 749–753 (1960).
39. E. R. Dowski, Jr., and W. T. Cathey "Extended depth of field through wave-front coding," *Appl. Opt.* **34**, 1859–1866 (1995).
40. J. Ojeda-Castañeda, L. R. Berriel-Valdos, and E. Montes, "Spatial filter for increasing the depth of focus," *Opt. Lett.* **10**, 520–523 (1985).
41. T. C. Poon and M. Motamedi, "Optical/digital incoherent image processing for extended depth of field," *Appl. Opt.* **26**, 4612–4615 (1987).
42. S. S. Sherif and W. T. Cathey, "Depth of field control in incoherent hybrid imaging systems," in *Optical Imaging and Microscopy - Techniques and Advanced Systems*, P. Török and F.-J. Kao eds., (Springer, New York 2007).
43. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science* **220**, 671–680 (1983).
44. C. W. McCutcheon, "Generalised Aperture and the Three-Dimensional Diffraction Image," *J. Opt. Soc. Am.* **54**, 240–244 (1964).

45. J. Ojeda-Castañeda, L. R. Berriel-Valdos, and E. Montes, "Spatial filter for increasing the depth of focus," *Opt. Lett* **10**, 520–522 (1985).
 46. T. di Francia, "Super-gain antennas and, optical resolving power," *Nuovo Cimento* **9**, 426–438 (1952).
 47. Z. S. Hegedus and V. Sarafis, "Superresolving filters in confocally scanned imaging systems," *J. Opt. Soc. Am. A* **3**, 1892–1896 (1986).
 48. D. J. Innes and A. L. Bloom, "Design of optical systems for use with laser beams," *Spectra-Physics Laser Technical Bulletin* **5**, 1–10 (1966).
-

1. Introduction

Synthesis of arbitrary field distributions in optical systems is useful for a wide variety of applications including lithography [1], optical data storage [2], atomic manipulation [3] and polarisation microscopy [4]. A significant number of alternative methods by which to produce a desired field distribution exist such as apodisation or phase masks [5], polarisation structuring [6] and computer generated holograms [7, 8]. Numerical optimisation is however normally used to determine the appropriate mask or field distribution to use [9]. To the best of the authors' knowledge there does not currently exist an analytic method to invert the Debye-Wolf integral in the literature. This omission is addressed in this work whereby the Debye-Wolf integral is inverted using a series expansion also developed by some of the current authors [10].

In principle the new method allows an arbitrary field distribution to be specified in the focal region of a high numerical aperture (NA) lens and the appropriate weighting function, or equivalently the pupil plane distribution, to be calculated. Furthermore due to the simple form of the inversion it is amenable to numerical optimisation should extra constraints need to be introduced to the system. Such a constraint may include the pixelation of masking optics, a feature often encountered when using spatial light modulators (SLMs) [11] for example, which limits the level of fine structure producible in any physical mask and thus any pupil plane field distribution.

Since the expansion of the Debye-Wolf formula and its inversion are based on generalised prolate spheroidal functions they are introduced in the next section and a number of their basic properties are explored. A derivation of the inversion formula is then given in Section 3, whilst Section 4 contains some notes and caveats on the use of the formula. Some examples are given in Section 5 before concluding remarks are finally given in Section 6.

2. Theory of generalised prolate spheroidal functions

In this section various properties of generalised prolate spheroidal functions, which are the eigenfunctions of a finite two dimensional Fourier transform over a circular domain, are discussed. Although a number of different mathematical properties are considered derivations are omitted for brevity. Reference is however made to the works of Slepian, Landau and Pollock [12–15] where a full analysis can be found.

2.1. Space-bandwidth product

All optical devices are incapable of transmitting signals with arbitrarily high frequency content perfectly, but instead possess transfer functions which extend over a finite range. The resulting transmitted signal thus has a finite bandwidth denoted Ω . A bandlimited function cannot in itself also be space limited due to the uncertainty principle, however it is possible to define a region of spatial extent r_0 outside of which the function is negligible or of little interest. The product $c = r_0\Omega$ is then called the space-bandwidth product and is often used as a measure of the optical performance of a system [16, 17].

The space-bandwidth product is important for our discussion of prolate spheroidal functions since they are bandlimited functions whose form and behaviour is dependent upon the param-

ter c . This explicit dependence is however occasionally dropped in this work for clarity with the understanding the dependence still remains.

2.2. Eigenfunctions of the two dimensional finite Fourier integral

It can be shown [12] that the eigenfunctions of the *finite* two dimensional Fourier transform over a circular domain can be written in the form

$$\psi_{N,n}(c, r, \theta) = \Phi_{N,n}(c, r) \begin{cases} \cos N\theta \\ \sin N\theta \end{cases} \quad N = 0, 1, 2, \dots, \quad n = 0, 1, 2, \dots \quad (1)$$

where the $\Phi_{N,n}(c, r)$, known as the circular prolate spheroidal functions, are the eigenfunctions of the N^{th} order finite Hankel transform. The defining relation for these functions can thus be expressed

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

where J_N is the N^{th} order Bessel function of the first kind, ω and r are conjugate coordinates and $\lambda_{N,n}$ are the circular prolate spheroidal eigenvalues. Figures. 1 and 2 show the behaviour of the eigenvalues and eigenfunctions respectively which are further discussed in Section 2.4.

It should be noted that the circular prolate functions $\Phi_{N,n}$ used here are scaled versions of those developed by Slepian $\varphi_{N,n}$ such that

$$\Phi_{N,n}(c, r) = \left(\frac{\lambda_{N,n}}{r r_0}\right)^{1/2} \varphi_{N,n}(c, r/r_0) \quad (3)$$

$\varphi_{N,n}$ are also the solutions to the wave equation when expressed in a prolate spheroidal coordinate system.

2.3. Orthogonality and completeness of the generalised prolate spheroidal functions

Heurtley has shown [18] that the functions satisfying the integral equation (2) are orthogonal and complete over the finite region $0 \leq r \leq r_0$ i.e.

$$\int_0^{r_0} \Phi_{N,n}(c, r) \Phi_{N,m}(c, r) r dr = \lambda_{N,n} \delta_{nm} \quad (4)$$

and

$$\sum_{n=0}^{\infty} \lambda_{N,n}^{-1} \Phi_{N,n}(c, r) \Phi_{N,n}(c, r') = \delta(r - r')/r \quad \text{for } 0 \leq r, r' \leq r_0 \quad (5)$$

where δ_{nm} is the Kronecker delta and $\delta(r - r')$ is the Dirac delta function centered on $r = r'$. Furthermore the prolate functions possess the unique property that they are also orthogonal and complete on the infinite interval $0 \leq r \leq \infty$.

Noting that sinusoidal functions are also complete and orthogonal it is possible to expand any two dimensional bandlimited function in terms of generalised prolate functions

$$f(r, \phi) = \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} A_{N,n} \Phi_{|N|,n}(c, r) \exp(iN\phi) \quad (6)$$

where it has been elected to write $\psi_{N,n}$ in terms of exponentials as opposed to the trigonometric functions of Eq. (1). The coefficients $A_{N,n}$ can be calculated using the orthogonality property

$$A_{N,n} = \frac{1}{2\pi \lambda_{|N|,n}} \int_0^{2\pi} \int_0^{r_0} f(r, \phi) \Phi_{|N|,n}(c, r) \exp(-iN\phi) r dr d\phi. \quad (7)$$

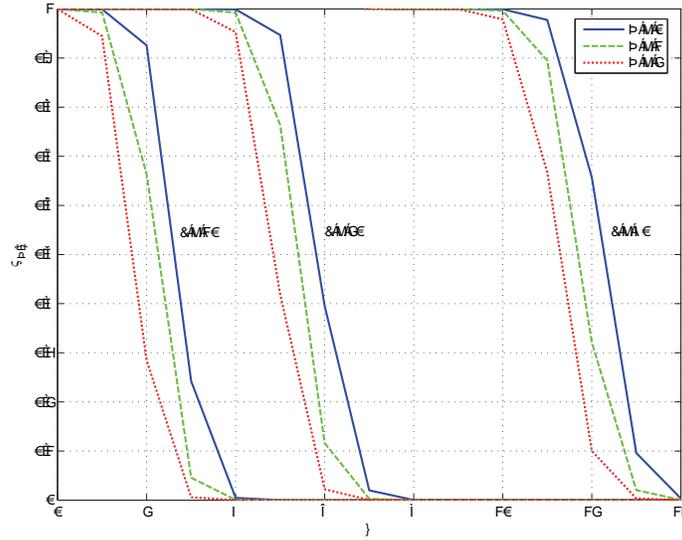


Fig. 1. Circular prolate spheroidal eigenvalues for different orders (N and n) and space bandwidth products c .

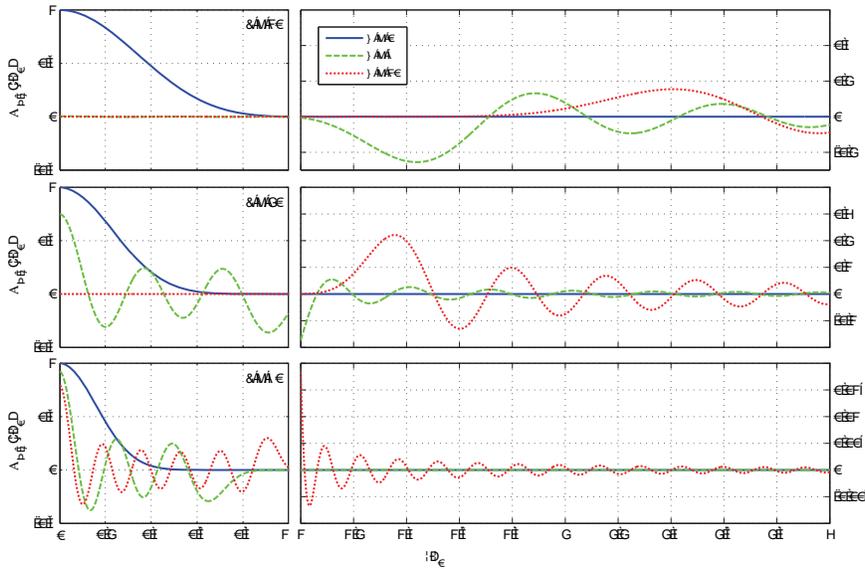


Fig. 2. Circular prolate spheroidal functions for different orders n for $N = 0$ and for different space bandwidth products c . Prolate functions plotted have been normalised so that $\Phi_{0,0}(0) = 1$. Note the different vertical scales between plots.

2.4. Energy concentration property of generalised prolate spheroidal functions

Given a bandlimited function the question may be asked as to how concentrated the function can be in the spatial domain in terms of its energy distribution. This is of particular interest in optics, for example when trying to improve the resolution in an imaging system or to extend the depth of field where large sidelobe structures are undesirable. Traditionally the encircled energy of a function $f(r, \phi)$ within a circular region of radius r_0 is defined as

$$I_{enc} = \int_0^{2\pi} \int_0^{r_0} |f(r, \phi)|^2 r dr d\phi \bigg/ \int_0^{2\pi} \int_0^{\infty} |f(r, \phi)|^2 r dr d\phi \quad (8)$$

Expansion of $f(r, \phi)$ by using Eq. (6) and subsequent substitution into Eq. (8) gives

$$I_{enc} = \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} |A_{N,n}|^2 \lambda_{|N|,n} \bigg/ \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} |A_{N,n}|^2 \quad (9)$$

where the orthogonality condition (4) and the analogous equation for the infinite interval (see for example [19]) have also been used. From Fig. 1 it can be seen that the eigenvalues lie in the range $0 \leq \lambda_{N,n} \leq 1$ and monotonically decrease with N and n and as such the encircled energy takes its maximum value of $I_{enc}^{max} = \lambda_{0,0}$ when

$$f(r, \phi) = A_{0,0} \Phi_{0,0}(c, r) \quad (10)$$

More generally the eigenvalue $\lambda_{N,n}$ is a measure of the fraction of energy contained within the circular region defined by $0 \leq r \leq r_0$ and $0 \leq \phi < 2\pi$ [20]. This feature can be seen in Figs. 1 and 2. Considering first the $c = 10$ case it is noted that the eigenvalues drop off rapidly at $n \sim 3$. As such when the $n = 0$ order is plotted in Fig. 2 it is non-zero when $r \leq 0$ and essentially (although not precisely) zero outside. Higher order modes, $n = 5$ and 10 , however display the converse behaviour.

For the $c = 20$ case the eigenvalues remain close to unity up to higher orders and instead decrease at $n \sim 6$. When plotted the $n = 0$ mode displays the same properties as before, but now the $n = 5$ mode shows contributions for all values of r considered. With an eigenvalue of 8.46×10^{-9} the $n = 10$ order again contains negligible energy within the central region.

Finally considering the $c = 40$ case the eigenvalues do not fall off until $n \sim 13$ meaning the plotted orders have only a small contribution for $r \geq r_0$.

3. The Debye-Wolf diffraction integral

Having discussed the basic theory of the generalised prolate spheroidal functions their use in the inversion of the Debye-Wolf integral is now discussed.

The Debye-Wolf integral describes the electric field distribution at a point p with Cartesian coordinates $\mathbf{r}_p = (x_p, y_p, z_p)$ in the vicinity of the focal region of a telecentric, high NA lens as shown in Fig. 3 and can be written [21, 22]

$$\mathbf{E}(\mathbf{r}_p) = -\frac{i}{\lambda} \iint_{\Theta} \frac{\mathbf{a}(s_x, s_y)}{s_z} \exp(iks \cdot \mathbf{r}_p) ds_x ds_y \quad (11)$$

where λ is the wavelength of the illuminating light, $\mathbf{a}(s_x, s_y)$ is the strength vector of a geometric ray at the Gaussian reference sphere centered on the focal point, $\mathbf{s} = (s_x, s_y, s_z)$ is a unit ray vector and Θ is the domain of the exit pupil. The strength vector is easily modified by introduction of suitable optics in the pupil plane which can be written in terms of the pupil coordinates in the general form

$$\mathbf{a}(u, \phi) = g(u, \phi) e^{i\Psi(u, \phi)} \mathcal{L}(u, \phi) \mathbf{e}(u, \phi) \quad (12)$$

where $u = \sin \theta$, $g(u, \phi)$ and $\Psi(u, \phi)$ describe the amplitude and phase variation introduced to the incident field distribution

$$\mathbf{e} = \begin{pmatrix} e_x(u, \phi) \\ e_y(u, \phi) \end{pmatrix} \quad (13)$$

by an apodisation and phase mask respectively and \mathcal{L} is the sub generalised Jones matrix given by

$$\mathcal{L} = \begin{pmatrix} (1 + \sqrt{1 - u^2}) - (1 - \sqrt{1 - u^2}) \cos 2\phi & -(1 - \sqrt{1 - u^2}) \sin 2\phi \\ -(1 - \sqrt{1 - u^2}) \sin 2\phi & (1 + \sqrt{1 - u^2}) + (1 - \sqrt{1 - u^2}) \cos 2\phi \\ -2u \cos \phi & -2u \sin \phi \end{pmatrix} \quad (14)$$

which describes the action of the lens and maps the field to the Gaussian reference sphere [23].

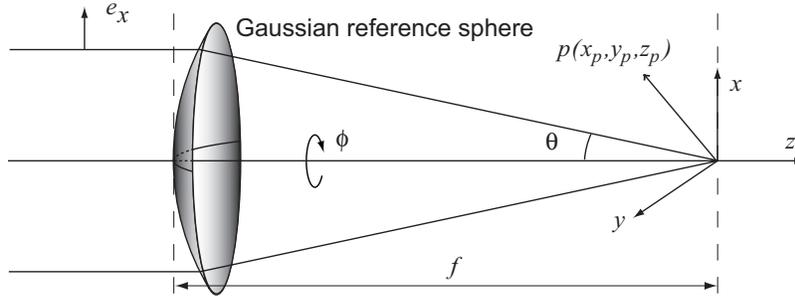


Fig. 3. Coordinate system and geometry of Debye-Wolf diffraction integral.

3.1. Eigenfunction expansion of the Debye-Wolf integral

Earlier work by some of the current authors [10] has shown that for a circular aperture the Debye-Wolf integral can be represented using a series expansion based on Bessel functions and generalised prolate spheroidal functions

$$E_j(\rho_p, \phi_p, z_p) = 2iA \left(\frac{\alpha'}{k\rho_p^{max}} \right) \sum_{m=-\infty}^{\infty} \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} i^{|N|} A_{m,N,n}^j (-1)^n \sqrt{\lambda_{|N|,n}} J_m(kz_p) \exp(iN\phi_p) \Phi_{|N|,n} \left(\frac{\alpha' \rho_p}{\rho_p^{max}} \right) \quad (15)$$

where A is a constant as per [22], $k = 2\pi/\lambda$ is the wavenumber of the illuminating light, $\alpha' = \sin \alpha$ is the NA of the lens assumed to be in air, ρ_p^{max} is the field of view in the focal space, $A_{m,N,n}^j$ are expansion coefficients and a switch to cylindrical polar coordinates has been made such that

$$x_p = \rho_p \cos \phi_p \quad y_p = \rho_p \sin \phi_p \quad z_p = z_p. \quad (16)$$

The space bandwidth product in this case is given by $c = k\alpha' \rho_p^{max}$ since the spatial cutoff frequency of a lens is $\omega_0 = k\alpha'$ [24].

It is noted that Eq. (15) omits a minus sign from the E_x component when compared to the original formulation [10] since here it is assumed the sign difference is absorbed into the weighting function that is expanded to give the coefficients $A_{m,N,n}^j$. This is done for convenience reasons only.

3.2. Inversion of the Debye-Wolf integral

Given Eq. (15) for the field in the focal region of a high NA lens and the orthogonality of the generalised prolate spheroidal functions as described by Eq. (4) it is possible to invert the Debye-Wolf integral as follows.

Consider multiplying both sides of Eq. (15) by the generalised prolate spheroidal function of order Q, q and integrating over a plane in the focal region. This yields the result

$$\begin{aligned} \int_0^{2\pi} \int_0^{\rho_p^{\max}} E_j(\rho_p, \phi_p, z_p) \Phi_{|Q|,q} \left(\frac{\alpha' \rho_p}{\rho_p^{\max}} \right) \exp(-iQ\phi_p) \rho_p d\rho_p d\phi_p \\ = 2iA \left(\frac{\alpha'}{k\rho_p} \right) \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \sum_{n=0}^{\infty} i^{|N|} A_{m,N,n}^j (-1)^n \lambda_{|N|,n}^{3/2} J_m(kz_p) \delta_{QN} \delta_{qn} \end{aligned} \quad (17)$$

The Kronecker deltas eliminate all but a single term within the double summation over N and n , namely the term for which $Q = N$ and $q = n$. Thus

$$E_{N,n}^j = \frac{iA}{\pi} \left(\frac{\alpha'}{k\rho_p^{\max}} \right) (-1)^n i^{|N|} \lambda_{|N|,n}^{1/2} \sum_{m=-\infty}^{\infty} J_m(kz_p) A_{m,N,n}^j \quad (18)$$

where

$$E_{N,n}^j = \frac{1}{2\pi \lambda_{|N|,n}} \int_0^{2\pi} \int_0^{\rho_p^{\max}} E_j(\rho_p, \phi_p, z_p) \Phi_{|N|,n} \left(\frac{\alpha' \rho_p}{\rho_p^{\max}} \right) \exp(-iN\phi_p) \rho_p d\rho_p d\phi_p \quad (19)$$

Trivial algebraic rearrangements yield an infinite set of linear equations,

$$\sum_{m=-\infty}^{\infty} J_m(kz_p) A_{m,N,n}^j = -\frac{i\pi}{A} \left(\frac{k\rho_p^{\max}}{\alpha'} \right) \frac{(-1)^n}{i^{|N|}} \lambda_{|N|,n}^{-1/2} E_{N,n}^j \quad (20)$$

which unfortunately cannot be solved uniquely to determine the desired coefficients $A_{m,N,n}^j$, but can however form the basis for numerical optimisation techniques, an example of which is given in Section 5. Unique solution can however be achieved on the focal plane i.e. when $z_p = 0$ whereby

$$J_m(kz_p)|_{z_p=0} = \begin{cases} 1 & \text{for } m = 0 \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

yielding the simple relation

$$A_{N,n}^j = -\frac{i\pi}{A} \left(\frac{k\rho_p^{\max}}{\alpha'} \right) \frac{(-1)^n}{i^{|N|}} \lambda_{|N|,n}^{-1/2} E_{N,n}^j \quad (22)$$

where the subscript m has now been dropped. This equation shows that the coefficients of the expansion of the weighting function are merely a scaled version of the coefficients of the expansion of the field in the focal plane as would be expected for an eigenfunction expansion. This is the basic inversion formula for the Debye-Wolf integral.

It is noted here that since the generalised prolate spheroidal functions are scalar functions they cannot be true eigenfunctions of the vectorial problem, however as just shown they are eigenfunctions on a component-wise basis. Furthermore for low NA systems the polarisation properties of light become less important often allowing a scalar treatment to be used and hence the prolate functions can then be interpreted as strict eigenfunctions of focusing by a lens.

4. Some notes on inversion and electric field specification

Although a formula to invert the Debye-Wolf integral has now been derived numerous problems may be encountered if it is used incorrectly. In this section some principles and caveats to use of Eq. (22) are thus presented.

4.1. Degrees of freedom

The underlying purpose of inversion of the Debye-Wolf integral is to provide a means by which to generate a desired field distribution. As it stands Eq. (22) describes how to find one component of the required strength vector $\mathbf{a}(s_x, s_y)$ to produce a *single* desired field component in the focal plane. Conceivably it would be possible to use Eq. (22) to calculate all three components of the strength vector, however such a naïve approach would not guarantee physicality or realisability. Maxwell's equations mean that at best only two field components can be specified and used for inversion, however there is no restriction on which components are chosen.

Furthermore, since some form of additional optics, e.g. a pupil plane mask, must be introduced into the system so as to modify the weighting function there are additional constraints on the specification of the electric field on the focal plane. These constraints arise from the degrees of freedom of the introduced optics. To illustrate this point consider use of an apodisation mask in the exit pupil of the system. This introduces only a single degree of freedom to the system, that is to say only the amplitude of the field in the pupil plane can be modified and not its phase. In turn this translates to the requirement that the field component specified in the focal region must be complex Hermitian. Combination of a phase and apodisation mask would however provide two degrees of freedom allowing an arbitrary phase and amplitude profile to be specified for one field component in the focal plane. Assuming more degrees of freedom than are present in a particular optical setup will lead to inconsistent inversion results that will not reproduce the desired field distribution and should hence be avoided.

4.2. Field specification away from the focal plane

Inversion was previously restricted to the focal plane since it is not possible to solve a set of $N \times n$ equations for $m \times N \times n$ unknowns uniquely. This restriction can however be circumvented under certain circumstances.

If only a single field component is specified on a plane in the focal region, but not necessarily the focal plane it is then possible to propagate this field to the focal plane by means of scalar techniques such as the angular spectrum method. Once the field on the focal plane has been obtained in this manner Eq. (22) can be used as prescribed in this paper.

Alternatively if two field components are specified then it is again possible to propagate the field to the focal plane however vector formulations, such as *e*- and *m*- theory must instead be used [25].

4.3. Extrapolation and encircled energy

Specification of a desired field distribution in the focal plane over an infinite region is not only impractical, but also superfluous to physical requirements. As such the inversion formula assumes the field is specified over a finite region of maximum extent ρ_p^{max} . So as to ensure the completeness of the prolate functions over the specification area it is necessary to use the appropriate space-bandwidth product $c = k\alpha'\rho_p^{max}$ when calculating the coefficients from Eq. (19).

Perhaps the most important issue arising from only specifying a finite area is the resulting behaviour of the field outside of this region. Superresolution is a concept in which the synthesis of a focal spot smaller than the Rayleigh diffraction limit is attempted [26, 27], and provides a

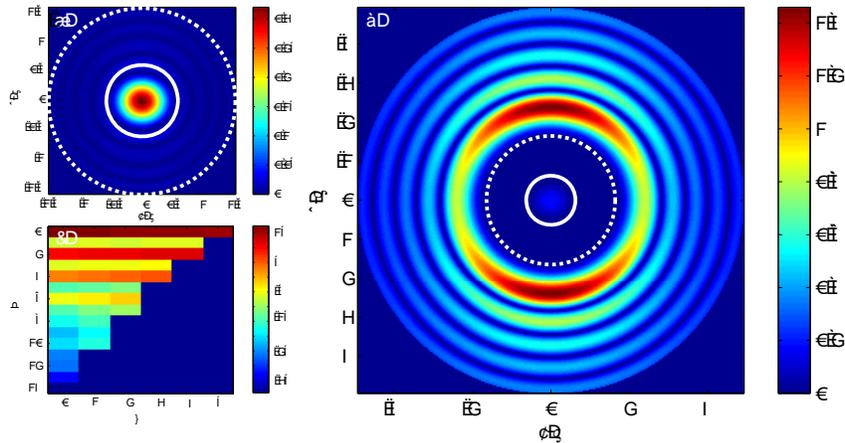


Fig. 4. When trying to produce a structure smaller than the diffraction limit as shown in (a) significant energy is pushed outside the specification area (b) as bounded by the dashed line. The solid line shows the size of the Airy disc. This can be understood from the high order contributions to the specified field as shown in (c) which plots $\ln(|A_{N,n}|^2 \lambda_{N,n})$ to allow comparison between all modes

good example to highlight how this can be of relevance.

Consider specifying a sub-diffraction focal spot in E_x as shown in Fig. 4(a) over a circle of radius ~ 1.6 times that of the Airy disc. Expansion of the specified field in terms of generalised prolate spheroidal functions as per Eqs. (6) and (19) allows extrapolation of the field beyond this region since [19]

$$f(r) = \sum_{n=0}^{\infty} \lambda_{N,n}^{-1} \Phi_{N,n}(r) \int_0^{r_0} f(r') \Phi_{N,n}(r') r' dr' \quad \text{for} \quad \begin{array}{l} r > 0 \text{ if } N > 0 \\ r \geq 0 \text{ if } N = 0 \end{array} \quad (23)$$

This equation states that with knowledge of the function $f(r)$ over a finite region $0 \leq r \leq r_0$ it is possible to extrapolate to all values of $r > 0$ and is a consequence of the duality of completeness and orthogonality of the circular prolate spheroidal functions.

The resultant field from extrapolation of the field distribution of Fig. 4(a) is shown in Fig. 4(b). It can be seen that a significant fraction of energy is pushed out into the sidelobe/peripheral structures. When calculated the encircled energy is 6.86×10^{-4} . This behaviour arises since the high order modes contribute significantly as shown in 4(c) and thus energy is pushed out of the specification area as discussed in Section 2.4.

4.4. Noise amplification

Hadamard defined a number of criteria which a mathematical problem must meet to be well-posed [28], namely that a unique solution exists that depends continuously on the data i.e. is stable. Inverse problems, such as that considered in the current article, are however in general ill-posed i.e. violate one or more of these conditions. Predominantly such a situation arises due to sensitivity to the initial data which in the problem under consideration is a specified field distribution. In terms of inversion of the Debye-Wolf integral errors in the specified field arise since the series expansion must be truncated for computational purposes.

The condition number κ is a commonly used quantity which measures the amplification \mathcal{A} of noise and errors in the initial data to the final inversion [29] such that $\mathcal{A} \propto \kappa$. When using an

eigenfunction inversion method the condition number can be defined as the ratio of the largest and smallest non-zero eigenvalue, that is

$$\kappa = \frac{\lambda_{0,0}}{\lambda_{N,n}^{\min}} \approx \frac{1}{\lambda_{N,n}^{\min}} \quad (24)$$

where $\lambda_{N,n}^{\min}$ is the value of the smallest eigenvalue used in the truncated series expansion. It is thus advisable to use orders that lie within or close to the plateau of eigenvalues of Fig. 1 to reduce noise amplification. Since small eigenvalues (high orders) correspond to high frequency components better inversion will be obtained for smoother, slower varying fields.

4.5. Pixelation

A final consideration that may arise in many practical systems is that of pixelation. Exact reproduction of the required pupil plane field distribution is generally not possible in practise due to the pixelated nature of the liquid crystal SLMs often used to implement complex masks [11] and as such an error on the focused field distribution is introduced. Choosing individual pixel values so as to minimise this error is then a further problem. Fortunately since focusing is a unitary transformation i.e. one in which the inner product is conserved, minimisation of the root mean square (RMS) error in the focal plane is equivalent to minimising the RMS error in the exit pupil between the ideal and the pixelated mask. Doing so requires that the jk^{th} pixel of the SLM be set such that the output field is the average of the ideal profile over the domain Π_{jk} of the pixel i.e.

$$\mathbf{e}_{jk} = \frac{1}{S_{jk}} \iint_{\Pi_{jk}} \mathbf{e}(u, \phi) u du d\phi \quad (25)$$

where S_{jk} denotes the area of the jk^{th} pixel. Minimisation can also be performed on the Gaussian reference sphere, although this would require projection back to pupil plane to determine the appropriate SLM configuration.

5. Examples

5.1. Polarisation structuring

In this section a few examples are given so as to illustrate the inversion procedure, the first of which considers trying to determine the orientation of a single fluorescent molecule. This often entails the use of a high NA optical system which provides the better resolution needed to select individual fluorophores. Many existing methods are limited to determination of the transverse angle [30, 31] and as such it would be desirable to couple light into the transverse orientation efficiently so as to improve the signal to noise ratio. A fluorophore, modelled as a fixed electric dipole of moment \mathbf{p} , illuminated by a field \mathbf{E} re-radiates light as if it had an effective dipole moment proportional to $\mathbf{p} \cdot \mathbf{E}$. Efficient coupling thus entails minimising the longitudinal component of the focused field. Inverting a field specification of $E_z = 0$ gives a zero strength vector, meaning that such a specification cannot be achieved via apodisation or phase masks. However a beam with a non-uniform polarisation distribution can be used. There exist numerous methods to produce these so-called vector beams including: modification of laser cavities, via introduction of polarisation sensitive components such that only modes with the desired polarisation structure can lase [32]; interferometric methods, which superpose orthogonal polarisation states with appropriate phase and intensity profiles [33]; and subwavelength gratings, which act as uniaxial crystals whose structure determines the birefringence [34].

Although these methods are typically used to generate radially and azimuthally polarised vector beams it is possible to make more arbitrary vector beams by similar methods [35] or

alternatively by using computer generated holograms and SLMs [36]. Of these [36] is perhaps the most versatile being capable of dynamic modulation. Due to the pixelated nature of the SLMs it can however introduce undesired diffraction effects and often requires complex algorithms. Toussaint *et al.* [35] avoided these issues albeit at the cost of reduced light throughput and complexity of the required optical setup.

In the ideal non-pixelated case the weighting function appropriate to a vector beam input is given by Eq. (12) with $g(u, \phi) = 1$ and $\Psi(u, \phi) = 0$. It has been observed that azimuthally polarised light when focused has a very weak longitudinal component [37]. Results from inversion agree with this observation as shown in Fig. 5, however a true azimuthal pattern is not seen due to an angular ambiguity in the inversion meaning one half of the pattern is rotated by 180° . If the calculated polarisation structure is re-input into the forward focusing problem the maximum value of the longitudinal component is of order 10^{-17} ; a number most likely attributable to numerical noise and inversion hence gives a suitable solution.

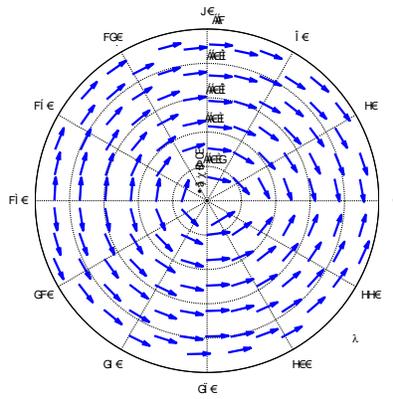


Fig. 5. Polarisation structure of illuminating beam required to give zero longitudinal field component in the focus of a high aperture lens as found by inversion ($NA = 0.966$, $c = 20$). Arrows indicate plane of polarisation of field at each point.

5.2. Extended depth of field

As a second example extension of the depth of field in imaging systems is treated. In its most basic form extension of the depth of field can be considered as a problem of reducing the intolerance to defocus as judged by some pre-agreed figure of merit. Extended depth of field (EDF) in imaging systems has been considered by a number of researchers and engineers since it can become an important issue when imaging three dimensional objects and for design tolerances in optical systems. By far the most commonplace technique of extending the depth of field in an imaging system is by means of pupil plane engineering [38–40]. Other techniques also exist, including axial scanning and hybrid systems employing post-detection signal processing [41, 42], however a discussion of such methods is beyond the scope of this article. Here a numerical example is given in which the incident beam is assumed to be uniformly x polarised. Consequently only the E_x field component contributes to the axial behaviour which is thus specified as

$$E_x(0, 0, z_p) = E_0 \text{rect}\left(\frac{z_p}{w}\right) \quad (26)$$

where E_0 is a constant and w denotes the half width of the rect function. On axis Eq. (20) reduces to

$$\sum_{m=-\infty}^{\infty} J_m(kz_p) A_{m,0,n}^j = -\frac{i\pi}{A} \left(\frac{k\rho_p^{max}}{\alpha'} \right) \frac{(-1)^n}{\lambda_{0,n}^{-1/2}} E_{0,n}^j \quad (27)$$

since $\Phi_{|N|,n}(0) = 0$ for $N \neq 0$ i.e. only $N = 0$ orders contribute on axis. Using Eqs. (19), (26) and (27) it is possible to numerically optimise the coefficients to find a good solution to the problem. A popular method of doing this is that of simulated annealing [43] in which random steps are taken with a probability that depends on a control parameter T which is slowly reduced. In simulated annealing a loss function is defined which is analogous to the energy in an annealing process. For the current example this was taken as the Hilbert angle ψ_H as defined by

$$\cos \psi_H = \frac{\langle |E_x(0,0,z_p)|^2, |E_x^{opt}(0,0,z_p)|^2 \rangle}{\| |E_x(0,0,z_p)|^2 \|^{1/2} \| |E_x^{opt}(0,0,z_p)|^2 \|^{1/2}} \quad (28)$$

where

$$\langle |E_x(0,0,z_p)|^2, |E_x^{opt}(0,0,z_p)|^2 \rangle = \int_{-\infty}^{\infty} |E_x(0,0,z_p)|^2 |E_x^{opt}(0,0,z_p)|^2 dz_p \quad (29)$$

and

$$\| |E_x(0,0,z_p)|^2 \|^2 = \int_{-\infty}^{\infty} |E_x(0,0,z_p)|^4 dz_p \quad (30)$$

which is a measure of the similarity between the shape of the desired and optimised distributions $E_x(0,0,z_p)$ and $E_x^{opt}(0,0,z_p)$ respectively [42] ranging from 0 if they are identical, to $\pi/2$ if they are orthogonal. Suitable truncation points for termination of the series in Eq. (27) can be determined as discussed in [10] so as to ensure convergence of the field expansion and was found to be $m_0 = 42$. Rejecting eigenvalues smaller than 10^{-5} , so as to limit noise amplification gave $n_0 = 9$. The resulting intensity profile as found from the 378 optimised coefficients is shown in Fig. 6(a) as compared to the desired rect function. The minimum Hilbert angle found was approximately $\frac{7\pi}{200}$.

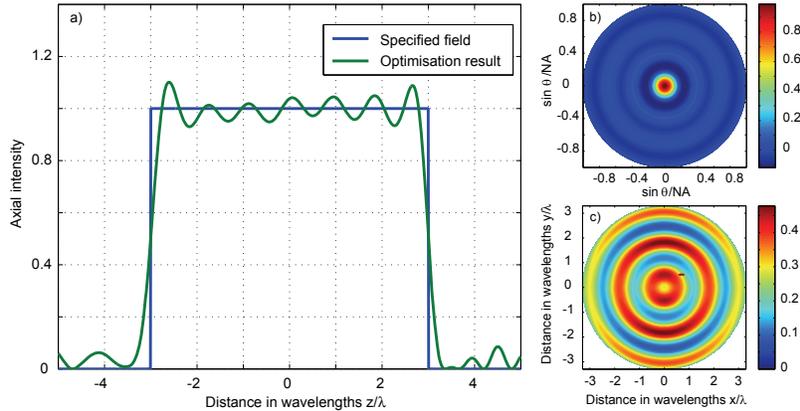


Fig. 6. (a) Comparison between the desired axial field profile and that found from a simulated annealing optimisation algorithm. (b) Apodisation mask required to produce the optimised distribution (NA = 0.966). (c) Resulting intensity distribution on the focal plane showing sidelobe pattern.

The corresponding apodisation mask required to produce the optimised axial behaviour is shown in Fig. 6(b) and is very similar in form to a sinc mask as would be expected from McCutchen's theorem [44, 45]. Significant energy is however contained in the sidelobe structure and a fuller treatment may hence also incorporate a constraint of the sidelobe height by including additional terms in the loss function.

5.3. Superresolution

As a final example the topic of superresolution is again considered. The use of apodising pupil plane masks for such a purpose has previously been considered [46, 47], however use of a polarisation structured beam to obtain superresolution is attempted here.

In an attempt to reduce the width of the intensity profile consider specifying the E_x field component as a Dirac delta function centered on the origin. Only one component of the focused field is specified since this introduces 2 degrees of freedom into the inversion problem as required for polarisation structuring. E_x is hence written in the form

$$E_x(\rho_p, \phi_p, 0) = \frac{1}{\rho_p} \delta\left(\frac{\alpha' \rho_p}{\rho_p^{max}}\right) \delta(\phi_p) \quad (31)$$

$$= \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} \lambda_{|N|,n}^{-1} \Phi_{|N|,n}\left(c, \frac{\alpha' \rho_p}{\rho_p^{max}}\right) \Phi_{|N|,n}(c, 0) \exp(iN\phi) \quad (32)$$

where the second step has used the completeness property of the generalised prolate spheroidal functions (c.f. Eq. (4)). Applying the inversion formula Eq. (22) and noting $\Phi_{|N|,n}(0) = 0$ for $N \neq 0$ immediately gives

$$A_{N,n}^x = \begin{cases} -\frac{i}{2A} \left(\frac{k\rho_p^{max}}{\alpha'}\right) (-1)^n \lambda_{0,n}^{-3/2} \Phi_{0,n}(c, 0) & \text{for } N = 0 \\ 0 & \text{for } N \neq 0 \end{cases} \quad (33)$$

Using Eq. (12) and noting that for a purely polarised structured beam $e_x^2 = 1 - e_y^2$ a quadratic equation in terms of e_x can be found, from which the required incident field distributions can be found. In practice however this method does not achieve superresolution for the simple reason that insufficient control is exerted on the y and z components of the focused field. As such when a delta function is specified for the x component energy is pushed into the y component. The resultant focused distribution is then essentially identical to that of a uniformly y polarised beam for which there is no resolution improvement.

Consider then specifying both the E_x and E_y focused field components to be Dirac delta functions. By the same logic this means $A_{N,n}^x = A_{N,n}^y$ as given by Eq. (33). Since $a_j(u, \phi) = \sum_{n=0}^{\infty} A_{0,n}^j \Phi_{0,n}(c, u)$ the required incident field distributions can be found using Eqs. (12), (33) and are given by

$$\begin{aligned} e_x(u, \phi) &= \frac{1}{\sqrt{(1-u^2)}} \frac{\mathcal{L}_{21} - \mathcal{L}_{22}}{\mathcal{L}_{11}\mathcal{L}_{22} - \mathcal{L}_{12}\mathcal{L}_{21}} \sum_{n=0}^{\infty} A_{0,n}^x \Phi_{0,n}(c, u) \\ e_y(u, \phi) &= \frac{1}{\sqrt{(1-u^2)}} \frac{\mathcal{L}_{12} - \mathcal{L}_{11}}{\mathcal{L}_{11}\mathcal{L}_{22} - \mathcal{L}_{12}\mathcal{L}_{21}} \sum_{n=0}^{\infty} A_{0,n}^x \Phi_{0,n}(c, u) \end{aligned} \quad (34)$$

where \mathcal{L}_{pq} denotes the pq^{th} element of \mathcal{L} as given in Eq. (14). The $(1-u^2)^{-1/2}$ factor is required to conserve energy when projecting from the surface of the Gaussian reference sphere to the pupil plane [48].

Having specified two field components on the focal region plane means there are four degrees of freedom within the system. Such a situation could correspond to the combination of

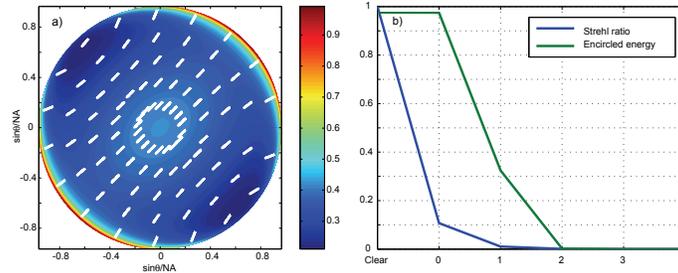


Fig. 7. (a) Colour plot showing the transmittance of the apodising mask in the pupil plane, whilst white lines represent the plane of oscillation of the electric field vector. (b) Variation of the Strehl intensity ratio and the encircled energy for the E_x and E_y components as the mask order n_0 is increased.

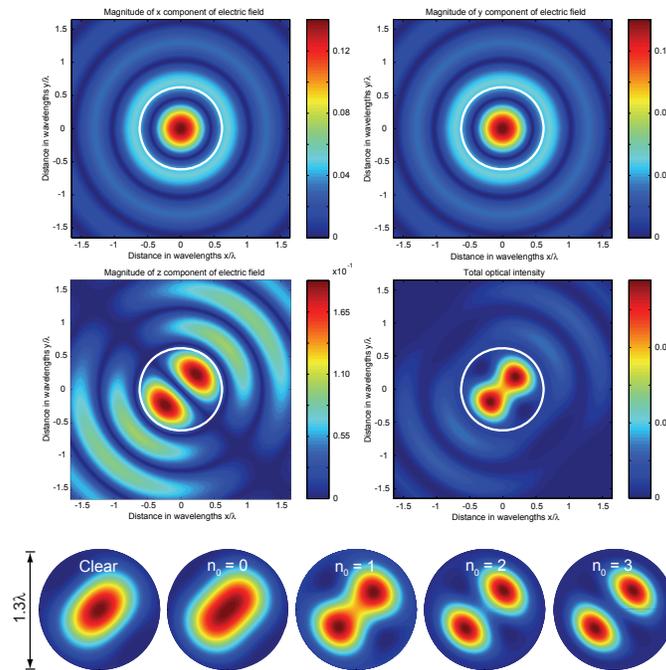


Fig. 8. Optical distribution in the focal plane for an apodised and polarised structured beam with truncation point $n_0 = 1$ (top) for $NA = 0.966$ and $c = 4$. White circles again denote the extent of the Airy disc. Variation of the central intensity focal spot over the Airy disc as mask order n_0 is increased (bottom). Note the intensity scales differ with each plot, but have been equalised for easy comparison.

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. 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 no phase

modulation is necessary. Apodisation is however necessary as can be seen by considering

$$g(u, \phi) = (|e_x|^2 + |e_y|^2)^{1/2} = \frac{\sqrt{2 - u^2(1 - \sin 2\phi)}}{2(1 - u^2)} \sum_{n=0}^{\infty} A_{0,n}^x \Phi_{0,n}(c, u) \quad (35)$$

Renormalisation of the incident field expressions (Eq. (34)) would be necessary to ensure the mask is passive. Practically the series in Eqs. (34) must be truncated at say $n = n_0$, meaning the pupil and focal plane field distributions will differ from the ideal case in a way that is dependent on the truncation point. Figure 7(a) represents the required pupil plane distribution for $n_0 = 1$ whilst Fig. 8 shows the corresponding optical distribution in the focal plane. The shown distributions were calculated assuming $NA = 0.966$ and a value of $c = 4$ corresponding to a field of view in the focal plane approximately the size of the Airy disc.

From Fig. 8 it can be seen that there has been a resolution gain in the E_x and E_y distributions as compared to a clear aperture with uniform illumination, however there is little gain in the intensity focal spot. This again arises from a redistribution of energy to the unconstrained field component E_z which is then dominant in the final intensity profile. Furthermore due to the presence of the apodising mask this arrangement also has a low optical efficiency as can be seen in the plot of the Strehl intensity ratio as shown in Fig. 7(b) as a function of the truncation order n_0 . At high n_0 this quantity loses its meaning however since the central peak essentially vanishes with respect to the large sidelobes as would be expected from the discussion in Section 4.3. The performance of this particular superresolution setup consequently worsens as n_0 is increased.

6. Conclusions

In this work the inversion of the Debye-Wolf integral has been considered to the extent that an inversion formula has been derived and given. This new result provides a means by which to calculate the required weighting function and hence pupil plane distribution to generate a desired field distribution in the focal region. Although it is not impossible to solve for a general plane in the focal region a restriction to the focal plane was made since this greatly simplifies the calculations.

A number of caveats to the use of the inversion formula have also been given. These include consideration of the degrees of freedom present in the system which restricts how fully the focal distribution can be specified to maintain physicality.

Perhaps the most important result to come from this work though stems from the considerations of the energy contained within the field specification area. By considering the meaning of the prolate functions a means to construct “reasonable” focal distributions also becomes apparent. In the sense of energy distribution an optimum field can be constructed by using only the prolate functions with large eigenvalues as a basis in the focal plane. This also produces better inversion results due to reduced noise amplification.

In conclusion it should be said that although much focus has been placed on “exact” inversion the developed formula is also highly suitable for numerical optimisation, for which there already exists a vast range of tools and knowledge. Such suitability arises since only relatively few orders are required for reasonably accurate results and hence the number of optimisation parameters i.e. expansion coefficients, is also small. This is especially true for synthesis of axial or circularly symmetric transverse distributions since in this case only the $N = 0$ modes have non-zero coefficients.

Further extension of this work would be to investigate the inversion of a specified intensity distribution in the focal region, which poses further problems due to the lose of phase information. The reported method is unfortunately unsuitable for this purpose since although it is theoretically possible to generate an arbitrary distribution in one or two field components,

Maxwell's equations dictate this is at the cost of dominant features arising in the unconstrained components.

Acknowledgments

The authors would like to thank Paul Abbott and Peter Falloon from the Physics Department of the University of Western Australia for supplying the Mathematica code to compute Slepian's generalised spheroidal functions. Thanks are also extended to Arthur van de Nes and Carl Paterson for many productive discussions on this topic. The authors further acknowledge the financial support of the EU via NANOPRIM Contract No. NMP3-CT-2007-033310