Demonstration of an optimised focal field with long focal depth and high transmission obtained with the Extended Nijboer-Zernike theory

A. P. Konijnenberg, L. Wei,* **N. Kumar, L. Couto Correa Pinto Filho,**
L. Cisotto, S. F. Pereira, and H. P. Urbach

Optics Research Group, Delft University of Technology, Delft 2628 CH, Netherlands
*l.wei-11@tudelft.nl

Abstract: In several optical systems, a specific Point Spread Function (PSF) needs to be generated. This can be achieved by shaping the complex field at the pupil. The Extended Nijboer-Zernike (ENZ) theory relates complex Zernike modes on the pupil directly to functions in the focal region. In this paper, we introduce a method to engineer a PSF using the ENZ theory. In particular, we present an optimization algorithm to design an extended depth of focus with high lateral resolution, while keeping the transmission of light high (over 60%). We also have demonstrated three outcomes of the algorithm using a Spatial Light Modulator (SLM).

References and links

1. G. B. Airy, "On the diffraction of an annular aperture," Philos. Mag. **18**, 1–10 (1841).
2. J. Durnin and J. J. Miceli, Jr., "Diffraction-free beams," Phys. Rev. Lett. **58**, 1499–1501 (1987).
3. Y. Xu, J. Singh, C. Sheppard, and N. Chen, "Ultra long high resolution beam by multi-zone rotationally symmetrical complex pupil filter," Opt. Express **15**, 6409–6413 (2007).
4. H. Wang, C. J. R. Sheppard, K. Ravi, S. T. Ho, and G. Vienne, "Fighting against diffraction: apodization and near field diffraction structures," Laser Photonics Rev. **6**, 354–392 (2012).
5. B. J. Thompson, "Diffraction by semitransparent and phase annuli," J. Opt. Soc. Am. **55**, 145–148 (1965).
6. G. Toraldo di Francia, "Super-gain antennas and optical resolving power," Nuovo Cimento **9**, 426–438 (1952).
7. H. Wang and F. Gan, "High focal depth with a pure-phase apodizer," Appl. Opt. **40**, 5658–5662 (2001).
8. C. J. R. Sheppard, "Synthesis of filters for specified axial properties," J. Mod. Opt. **43**, 525–536 (1996).
9. J. Ojeda-Castañeda and L. R. Berriel-Valdos, "Arbitrarily high focal depth with finite apertures," Opt. Lett. **13**, 183–185 (1988).
10. H. Wang, G. Yuan, W. Tan, L. Shi, and T. Chong, "Spot size and depth of focus in optical data storage system," Opt. Eng. **46**, 065201 (2007).
11. L. Wei, "Evaluation of Extended Nijboer-Zernike theory as a tool for complex point spread function calculation," Master thesis (Delft University of Technology, 2012).
12. J. J. M. Braat, A. J. E. M. Janssen, P. Dirksen, and S. van Haver, "Assessment of optical systems by means of point spread functions," Prog. Opt. **51**, 349–468 (2008).
13. C. Maurer, A. Jesacher, S. Bernet, and M. Ritsch-Marte, "What spatial light modulators can do for optical microscopy," Laser Photonics Rev. **5**, 81–101 (2011).
14. A. J. E. M. Janssen, "ENZ approach for the computation of optical point spread function," J. Opt. Soc. Am. A **19**, 849–857 (2002).
15. A. J. E. M. Janssen, J. J. M. Braat, and P. Dirksen, "On the computation of the Nijboer-Zernike aberration integrals at arbitrary defocus," J. Mod. Opt. **51**, 687–703 (2004).

16. S. van Haver and A. J. E. M. Janssen, "Advanced analytic treatment and efficient computation of the diffraction integrals in the Extended Nijboer-Zernike theory," J. Eur. Opt. Soc. Rap. Publ. **8**, 13044 (2013).
17. Holoeye, "PLUTO Phase Only Spatial Light Modulator (Reflective)," http://holoeye.com/spatial-light-modulators/slm-pluto-phase-only/, June 24, 2013.
18. O. El Gawhary, A. Wiegmann, N. Kumar, S. F. Pereira, and H. P. Urbach, "Through-focus phase retrieval and its connection to the spatial correlation for propagating fields," Opt. Express **21**, 5550–5560 (2013).
19. H. Zhang, J. Xie, J. Liu, and Y. Wang, "Elimination of a zero-order beam induced by a pixelated spatial light modulator for holographic projection," Appl. Opt. **48**, 5834–5841 (2009).

1. Introduction

Engineering an elongated focal spot by manipulating the incident wavefront has been done for a long time. It has been known that by putting an annulus in the lens pupil the depth of focus is increased [1]. In fact, it turned out that if the annulus is made extremely narrow, a diffraction-free Bessel beam is approximated [2]. Another method which has been proposed to extend the depth of focus is by dividing the lens pupil in rings and modulate the phase and amplitude of each ring [3]. These and other methods are discussed in more detail in [4–10].

In this article, a new method of designing pupil function is proposed. It relies on a result of the Extended Nijboer-Zernike theory, as has been suggested by [11]. By exploiting a one-to-one correspondence between the Zernike polynomials which compose the pupil function and functions which compose the focal field [12], the number of degrees of freedom reduces to the number of polynomials with which the pupil is constructed, while the computation of the focal field is done by a quick matrix multiplication, rather than a time consuming diffraction integral. In particular, we present in this work an algorithm to find a pupil function which gives rise to an extended depth of focus for the case of a 0.4 numerical aperture, with low loss (less than 40%) of light intensity.

The relevance of such a focal field with an extended depth of focus is apparent from the many applications of Bessel beams, which range from optical tweezers to certain forms of microscopy to barcode scanners. The solution which is presented here however, has the advantage that light passes through the pupil in more areas than just the outer ring, which is a standard method to create a Bessel beam, and thereby achieving higher transmittance.

Finally, three outcomes of the algrotihm have been obtained experimentally using a phase-only Spatial Light Modulator (SLM) to shape the pupil field [13].

2. Theory

In this section we briefly discuss theoretical results required for the optimization algorithm that is later presented in this paper. First we give the definition of the complex Zernike polynomials, in which any pupil function can be decomposed. Then, we explain a result from ENZ-theory [12,14], which relates the pupil Zernike modes Z_n^m to basic functions V_n^m in the field in the focal region.

2.1. Complex Zernike polynomials

Any complex function defined on the unit disk can be expanded in terms of complex Zernike polynomials

$$f = \sum_{n \geq |m|} \beta_n^m Z_n^m, \tag{1}$$

where the coefficients β_n^m can be complex and where the complex Zernike polynomials are defined as

$$Z_n^m(\rho, \phi) = R_n^{|m|}(\rho) e^{im\phi}, \tag{2}$$

where

$$R_n^m(\rho) = \sum_{s=0}^{p} \frac{(-1)^s(n-s)!}{s!(q-s)!(p-s)!}\rho^{n-2s},$$

$$p = \frac{1}{2}(n-|m|),$$

$$q = \frac{1}{2}(n+|m|).$$

(3)

This expansion determines both the phase and amplitude of the pupil function. The Zernike polynomials Z_n^m are a complete set of orthogonal functions on the unit disk [12].

2.2. Result from ENZ-theory

In [12] it is shown that there exists a one-to-one correspondence between the complex Zernike polynomials $Z_n^m(\rho,\theta)$ which compose the pupil field $E_0(\rho,\theta)$ and functions $V_n^{|m|}(r,f)$ which compose the electric field in the focal region $E(r,\phi,f)$:

$$E_0(\rho,\theta) = \sum_{n\geq|m|} \beta_n^m Z_n^m(\rho,\theta)$$

$$\longleftrightarrow$$

$$E(r,\phi,f) \propto \sum_{n\geq|m|} \beta_n^m V_n^{|m|}(r,f) 2\mathrm{i}^{|m|} \mathrm{e}^{\mathrm{i}m\phi},$$

(4)

where (ρ,θ) are polar coordinates, (r,ϕ,f) cylindrical coordinates, and β_n^m complex coefficients. An expression for the V_n^m-functions is given in [15] and is included in the Appendix.

The importance of this result lies in the fact that now the intensity distribution can be computed directly using the Zernike coefficients β_n^m and the precalculated V_n^m rather than using the input field $E_0(\rho,\theta)$ and the diffraction integral [16].

2.3. The optimization algorithm

In order to find good solutions efficiently, we rely on Eq. (4). This equation states that once one has found β such that

$$I = |E|^2 \propto \left| \sum_{k=0}^{N} \beta_{2k}^0 V_{2k}^0 \right|^2,$$

is close to a desired pattern, the corresponding pupil function follows immediately from

$$E_0(\rho,\phi) = \sum_{k=0}^{N} \beta_{2k}^0 Z_{2k}^0.$$

(5)

Note that choosing $m = 0$ implies that we only consider circularly symmetric pupil functions. Since the V_n^m's only need to be precalculated once, the intensity I can computed this way significantly faster than using the diffraction integral.

In the algorithm, only the intensity distributions along the lateral axis (x-axis) and along the optical axis (z-axis) are considered. More specifically, let $(0,0)$ be the Gaussian focal point. Then $I_x = I(x,z=0)$ is the intensity along the lateral axis in the focal plane and $I_z = I(x=0,z)$ is the intensity along the optical axis. Using a least squares method, the β_n^m coefficients are found such that I_x and I_z approximate two target functions $I_{x,\text{target}}$ and $I_{z,\text{target}}$ as shown in Fig. 1. After that, the transmittance of the pupil function is optimized using the Nelder-Mead algorithm. The algorithm is explained in more detail in the Appendix.

In Fig. 2 the cross sections of the phase and amplitude of the obtained pupil function are shown, accompanied by the intensity profiles in the focal region along the z-axis and x-axis. The focal spot for this pupil is longer in the axial direction than in the aberration-free case, meaning that the depth of focus is increased. Since this pupil involves a quite structured amplitude illumination, we have made some approximations in such a way that the pupil could more easily be implemented in an experimental setup. This is treated in detail in the following section.
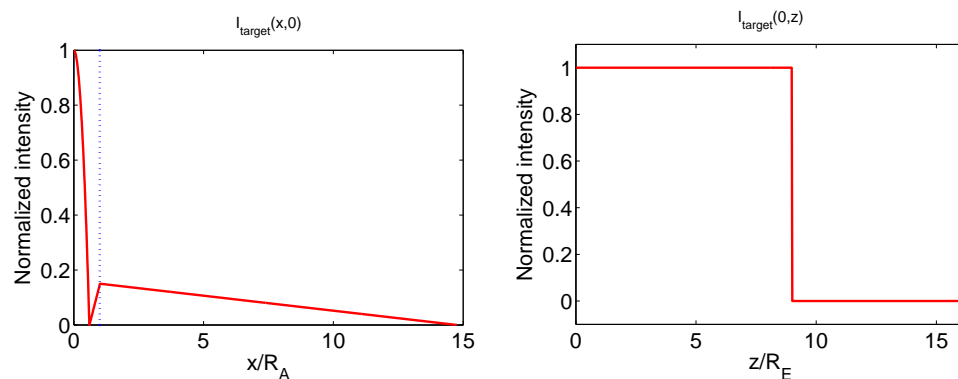

Fig. 1. Example of $I_{\text{target}}(x,0)$ and $I_{\text{target}}(0,z)$ which may be used to find an initial solution $\vec{\beta}_1$. Here $R_A = \frac{0.61\lambda}{\text{NA}}$ and $R_E = \frac{n\lambda}{\text{NA}^2}$. The vertical line in the figure of $I_{\text{target}}(x,0)$ indicates where $x = R_A$. These target functions are approximated using a least-squares method.
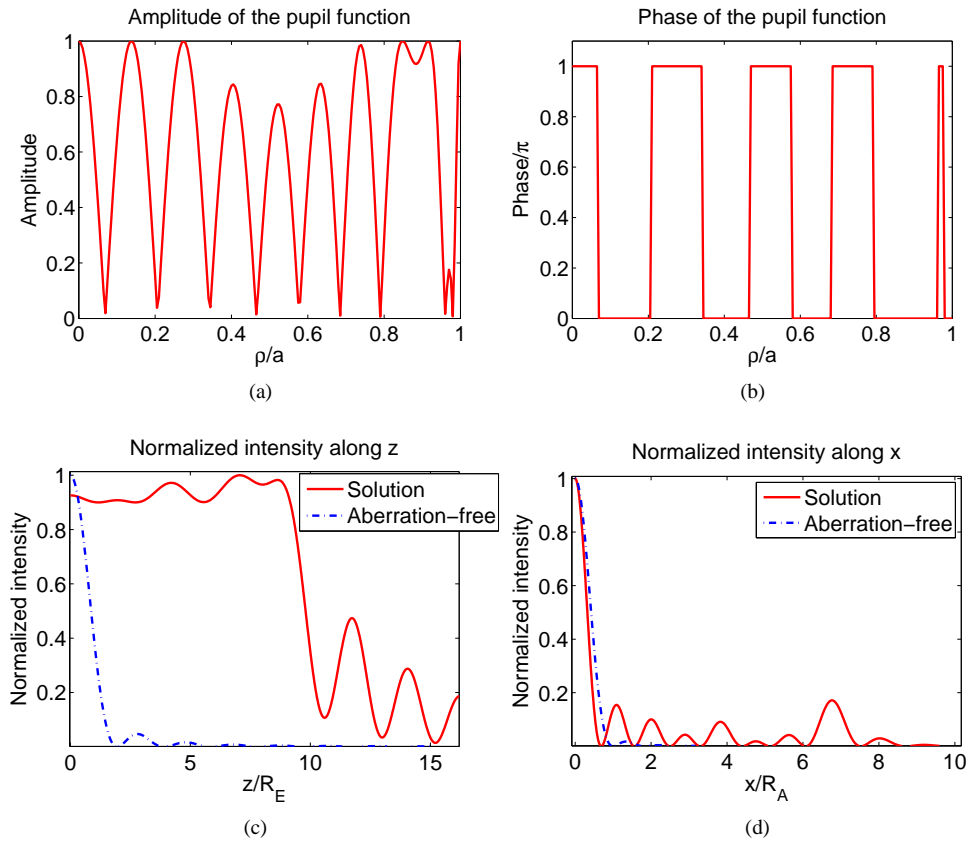
Fig. 2. Figures a and b show the cross sections of the amplitude and phase respectively of one of the pupil functions found with the algorithm. The pupil function is circularly symmetric, so only the cross sections along the radius (which is in this case normalized to the radius *a* of the aperture) are needed. This pupil function produces an elongated focal spot along the optical axis and a high lateral resolution in the focal plane, as is shown in Figures c and d respectively. The Zernike coefficients are given in Table 1 in the row C2.

2.4. *Creating binary pupil functions*

In order to simplify the experimental implementation, the pupil functions found by the algorithm (for example the one shown in Fig. 2) are made binary. Consider the β_n^m shown in Table 1. According to theory these coefficients should give pupil functions which give rise to an extended depth of focus and a lateral resolution below the diffraction limit. To construct the binary pupil functions we first define

$$\tilde{E}_0(\rho, \theta) = \sum_{k=0}^{N} \beta_{2k}^0 Z_{2k}^0(\rho, \theta), \qquad (6)$$

Table 1. The complex Zernike coefficients used to create the pupil functions '*Phase*' (P), '*Complex1*' (C1) and '*Complex*' (C2) which are tested experimentally.

	β_0^0	β_2^0	β_4^0	β_6^0	β_8^0	β_{10}^0	β_{12}^0	β_{14}^0	β_{16}^0	β_{18}^0	β_{20}^0	β_{22}^0	β_{24}^0	β_{26}^0
P	0.342	0.688	0.183	-0.555	-1.009	-0.041	1.147	-1.619	-0.392	0.840	-0.592	0.035	-0.114	0.145
C1	-0.232	-0.503	-0.352	0.208	0.533	0.325	-0.521	-1.172	1.377	0.013	-0.835	0.531	-0.003	0.317
C2	0.183	0.370	0.192	-0.270	-0.463	-0.257	0.747	0.674	0.442	-0.890	0.215	1.429	-1.316	-0.122

where Z_n^m are the complex Zernike polynomials as defined in Eq. (2). The pupil function for '*Phase*' (named so because only the phase is modulated) is obtained using

$$E_{0,\text{phase}}(\boldsymbol{\rho}) = \text{sgn}(\tilde{E}_0(\boldsymbol{\rho})). \tag{7}$$

Note that this is well-defined, because using only Zernike modes with $m = 0$ means that $\tilde{E}_0(\boldsymbol{\rho})$ is real. The pupil functions for '*Complex 1*' and '*Complex 2*' (named so because both amplitude and phase are modulated) are obtained using

$$E_{0,\text{complex}}(\boldsymbol{\rho}) = \begin{cases} 0 & \text{if } |\tilde{E}_0(\boldsymbol{\rho})| \leq t \\ \text{sgn}(\tilde{E}_0(\boldsymbol{\rho})) & \text{if } |\tilde{E}_0(\boldsymbol{\rho})| > t, \end{cases} \tag{8}$$

where it turns out by trial-and-error that $t = 0.57$ is a good choice for '*Complex 1*' and $t = 0.55$ for '*Complex 2*'. The resulting pupil functions which will be assigned to the SLM are shown in Figs. 3(a)–3(c). Note that in the regions where the amplitude should be 0, a phase ramp is added so that light hitting that region of the SLM is tilted away. The corresponding theoretical intensity distributions in the focal region produced by these pupil functions are shown in Figs. 3(d)–3(f). For '*Complex 2*', the pupil function that gives the longest focal depth, 61% of the light is transmitted.

3. Experiment

In the experiment we used a Holoeye PLUTO liquid crystal Spatial Light Modulator (SLM) to assign the pattern to the laser beam. The specifications of the SLM are given in [17]. A schematic of the experimental setup is shown in Fig. 4. A He-Ne laser with a wavelength of 633nm is used as the source. Its light is coupled to a single-mode optical fiber after which it emerges as a point source (by approximation). The light is then collimated using a lens with a focal length of 80cm and cut off by an aperture. The collimated light passes through the beamsplitter, hits the SLM perpendicularly, and is reflected by the same beamsplitter. The plane of the SLM is conjugated with the plane of the entrance pupil of a 0.4 NA microscope objective using two lenses with a focal length of 30cm. The light in the focal region of the 0.4 NA microscope objective is then imaged onto a CCD camera using a 0.9 NA microscope objective and a tube lens. The 0.4 NA microscope objective can be translated back and forth using a piezo-stage, which allows us to scan different planes of the focal field. An interferometer is used to determine the exact position of the piezo-stage [18]. Several important details of the experimental setup are summarized in Table 2.

3.1. Measurement method

First, we need to realize that not all the light that hits the SLM gets modulated by it due to interpixel space and a 'wrong' polarization. Thus in order to separate the modulated from the unmodulated beam an additional phase ramp is added on top of the images shown in Fig. 3 [19].

To view different focal planes, a lens phase is added to the image assigned to the SLM. The strength of this lens phase can be expressed as the Zernike defocus coefficient α_2^0 (note that

(a) Phase   (b) Complex 1   (c) Complex 2
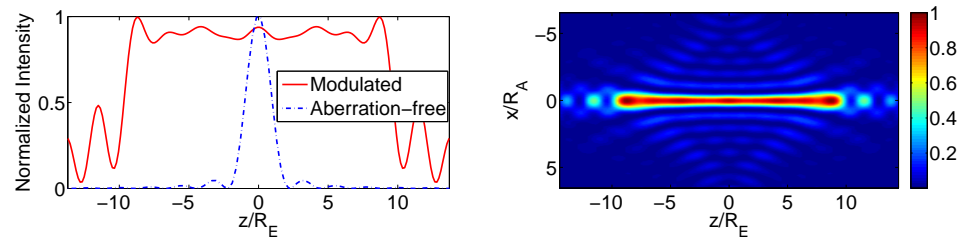

(d) Phase

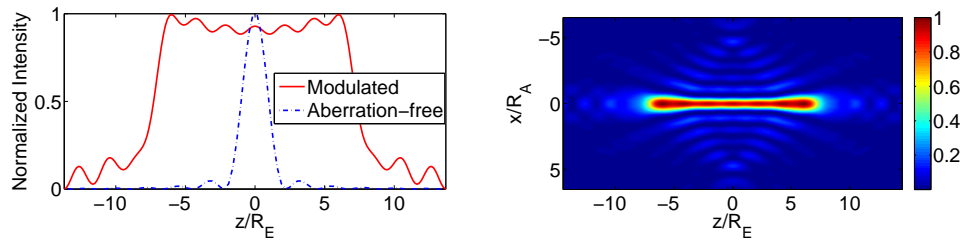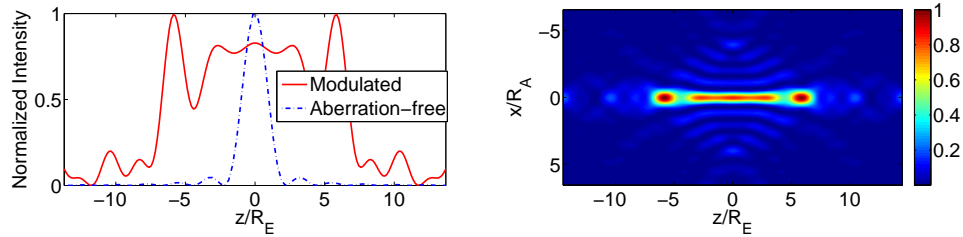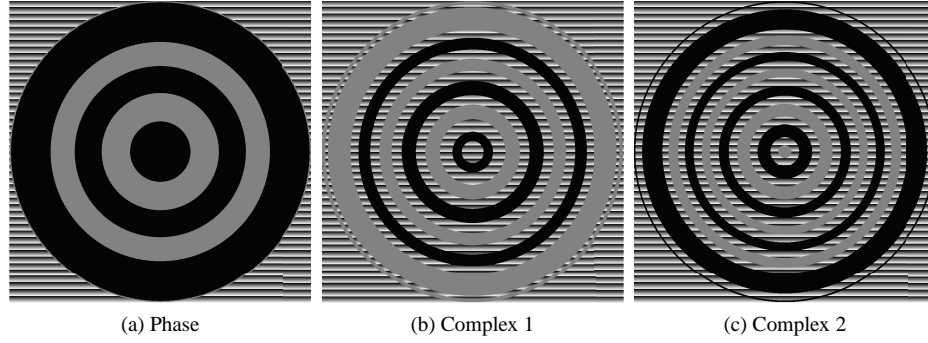

(e) Complex 1

(f) Complex 2

Fig. 3. Figures a, b and c show the pupil masks according to Table 1 which are assigned to the SLM. In the black regions the phase is 0. In the grey regions the phase is π. In the regions with a phase ramp the light is tilted away, which corresponds to the amplitude being modulated to 0. Figures d, e and f show the theoretical intensity distributions in the focal region and the corresponding profiles along the optical axis for the three pupil functions. For comparison, the aberration-free case (uniform illumination) is included.

Table 2. Details of the setup used to measure the focal field of a wavefront modulated by the PLUTO SLM and focused by a microscope objective. The schematic of the setup is shown in Fig. 4.

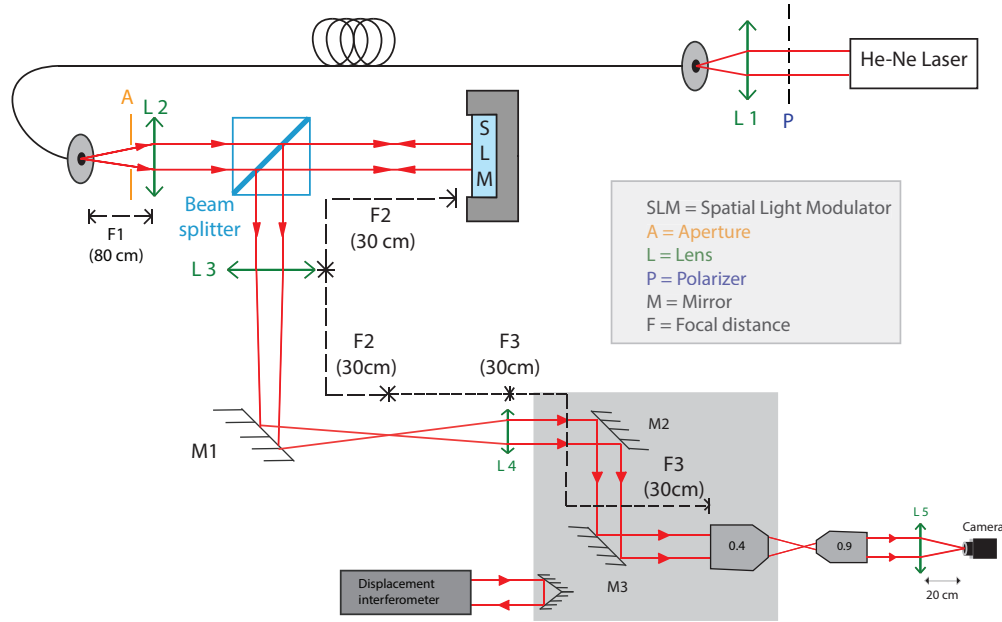| Wavelength	633nm
NA Microscope objective (focusing)	0.4
NA Microscope objective (imaging)	0.9
CCD resolution	1280×960
CCD pixel size	$4.7\mu m \times 4.7\mu m$
Magnification	50.4

Fig. 4. Schematic of the experimental setup used to modulate the phase of the laser beam and scan through the focal field of the 0.4 NA microscope objective.

the corresponding Zernike polynomial is real and describes the phase only, as opposed to the coefficients and Zernike polynomials used in the optimization which are complex).

However, if we want to compare measurement results to the theoretical predictions, we first need to find out how the Zernike defocus coefficient α_2^0 relates to the Rayleigh unit R_E, i.e. which value for α_2^0 corresponds to a shift of 1 R_E. This has been done by scanning through focus without a phase mask, and observing for what value for α_2^0 the intensity has dropped to $1/e$ times the maximum intensity. This value of α_2^0 corresponds to 1 R_E. In Fig. 5 the result of this measurement is shown.

3.2. Results and discussion

Once the calibrations are done, we obtained the focal field intensity distributions for the three pupil functions (as in Table 1) and compared them with the theoretical predictions as shown in Fig. 6. In Fig. 7 the lateral width (along the x-axis) of the focal spot for the aberration-free case and for the pupil function '*Complex 2*' are compared. Although the measurement results clearly indicate an extended depth of focus of a length similar to the prediction, there are still some
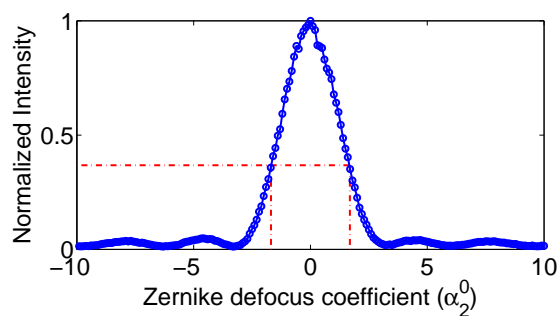
Fig. 5. Result of the measurement performed to determine how the Zernike defocus coefficient α_2^0 relates to the Rayleigh unit R_E. By seeing for what value of α_2^0 the intensity has dropped by a factor $1/e \approx 0.37$, we know which α_2^0 corresponds to $1\ R_E$ (as indicated by the red dashed lines). In this case it turns out to be $\alpha_2^0 \approx 1.69$.
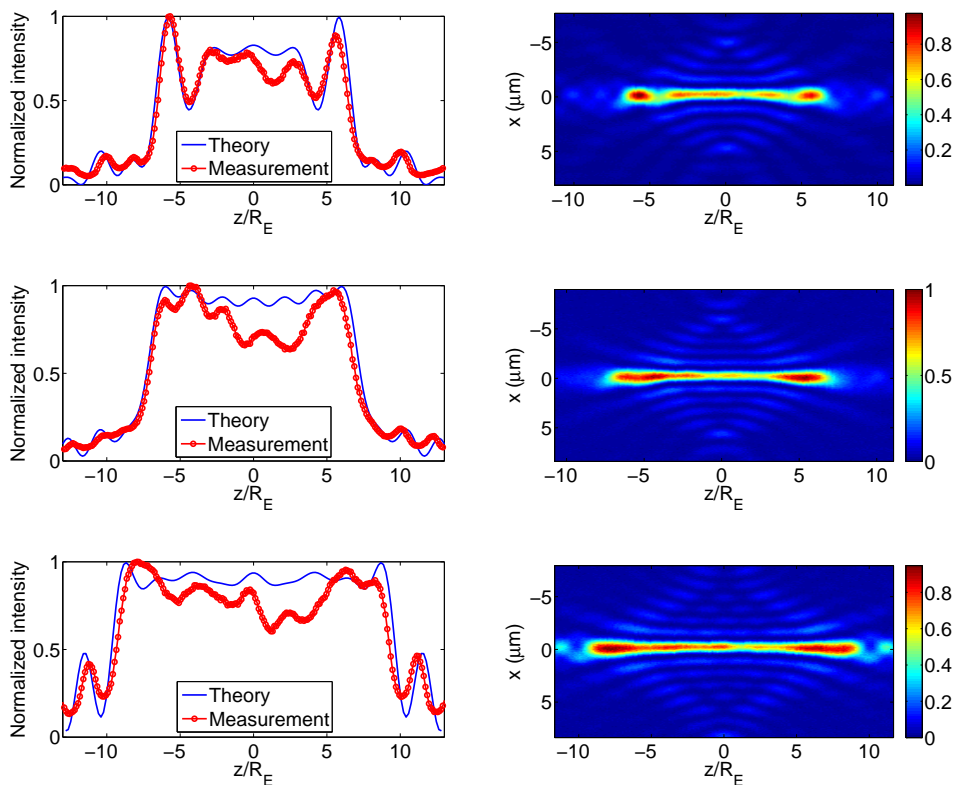

Fig. 6. The experimental results of intensity distributions of the focal fields for the three pupil functions (from top to bottom) 'Phase', 'Complex 1' and 'Complex 2'. The left column shows the on-axis intensity distributions. The right column shows the through-focus intensity distributions in the xz-plane.
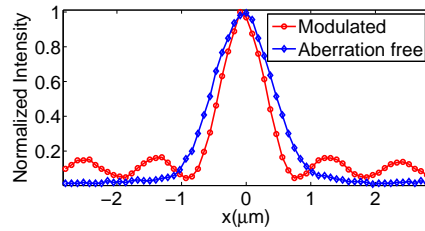
Fig. 7. Comparison of the widths of the focal spots for the aberration free case and for the pupil function '*Complex 2*'.
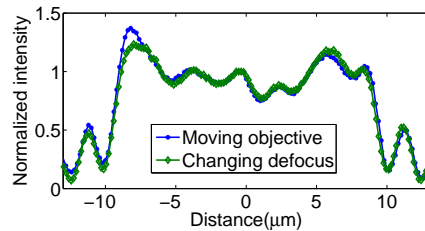

Fig. 8. Comparison of the two scanning methods (adding a lens phase and moving the microscope objective). The graphs have been normalized to the intensity at $z = 0$, since that is the point where no lens phase is added and thus the measurements are the same.

discrepancies. To find out the possible causes, we compared the cases where the through focus scans were made by adding a lens phase, or by moving the 0.4 NA microscope objective. The comparison is shown in Fig. 8. It turns out that these measurements give somewhat different results, but it still does not account for the mismatch with the theory. A possible explanation may be that because the 0.9 NA objective used for imaging is infinity corrected, so whenever a plane other than the focal plane is imaged, aberrations may come into play. This is particularly relevant in our case since the focal depth is around 18 R_E. Nevertheless, the experimental results have demonstrated that pupil masks indeed create elongated focal spots and that the lateral resolution at the focal plane is below the diffraction limit.

4. Conclusion

In conclusion, we used the Extended Nijboer-Zernike theory to obtain a focused field with elongated focal length (up to 18 Rayleigh distances) with diffraction-limited spot size, while keeping the transmittance over 60%. Since this theory is based on functions that can be pre-calculated, the optimisation parameters are reduced to a limited number of Zernike coefficients that compose the pupil, allowing fast calculation of the focal field. We have found not only a way to create elongated focal spots, but also we have shown more generally that ENZ theory has the potential to be used for pupil engineering, as had been suggested by [11]. We also demonstrated experimentally three pupil functions generated by an SLM that produce elongated focal spots.

Appendix A: An expression for the V_n^m functions

An expression of V_n^m is given in [15]:

$$V_n^m(r,f) = \varepsilon_m \exp\left(\frac{1}{2}\mathrm{i}f\right) \sum_{k=0}^{\infty} (2k+1)\mathrm{i}^k j_k\left(\frac{1}{2}f\right) \sum_{l_{\min}}^{k+p} (-1)^l w_{kl} \frac{J_{|m|+2l+1}(2\pi r)}{2\pi r}, \qquad (9)$$

in which $\varepsilon_m = \mathrm{e}^{\frac{m-|m|}{2}}$, $p = (n-|m|)/2$, $q = (n+|m|)/2$, and the minimum bound of l in the summation is given as:

$$l_{\min} = \left\{ \begin{array}{ll} p-k, & \text{if } 0 < k < p \\ 0, & \text{if } p \le k \le q \\ k-q, & \text{if } q < k \end{array} \right.$$

$$w_{kl} = \sum_{s=0}^{p} \sum_{t=0}^{\min(k,s)} f_{ps}^{|m|} b_{kst} g_{k+s-2t,l}^{|m|}, \qquad (10)$$

$l \in [\max(0, k-q, p-k), k+p]$. When $m = 0$, we have:

$$w_{k,k+p-2j} = b_{kpj}; \ j = 0, 1, \cdots, \min(k,p). \qquad (11)$$

where

$$f_{ps}^m = (-1)^{p-s}\left[\frac{2s+1}{p+s+1} \binom{m+p-s-1}{p-s}\binom{m+p+s}{s} \bigg/ \binom{p+s}{s} \right], s = 0, \cdots, p \qquad (12)$$

$$\begin{aligned} g_{ul}^m &= \frac{m+2l+1}{m+u+l+1}\left[\binom{m}{u-l}\binom{u+l}{u} \bigg/ \binom{m+l+u}{u} \right] \\ &= \frac{m+2l+1}{m+u+l+1}\binom{m}{u-l}\prod_{i=1}^{u} \frac{l+i}{l+m+i}, u = l, \cdots, l+m, \end{aligned} \qquad (13)$$

$$b_{s_1 s_2 t} = \frac{2s_1 + 2s_2 - 4t + 1}{2s_1 + 2s_2 - 2t + 1}\left(\frac{A_{s_1-t}A_t A_{s_2-t}}{A_{s_1+s_2-t}} \right), t = 0, \cdots, \min(s_1, s_2), \qquad (14)$$

where $A_t = \binom{2t}{t}$. For $m = 0$:

$$f_{ps}^0 = \delta_{ps}, \ g_{ul}^0 = \delta_{ul}, \qquad (15)$$

δ is Kronecker's delta.

Appendix B: The optimization algorithm

B.1. Formulating the problem

The goal is to construct a radially symmetric pupil function $E_0(\rho)$ defined in the exit pupil (which is the unit disk) such that the intensity distribution in the focal field $I(x,y,z) = |E(x,y,z)|^2$ has a sharp peak near $(x,y) = (0,0)$ for a long range along the optical z-axis. This means that the lateral resolution is high and the depth of focus long. In order to give a more precise formulation, we introduce the following simplifications and conventions:

★ We only consider pupil functions which can be expressed as a linear combination of N Zernike polynomials of the form Z_n^0, N being finite: $E_0(\rho) = \sum_{k=0}^{N} \beta_{2k}^0 Z_{2k}^0$. The pupil function then depends on ρ only and therefore has circular symmetry, and the electric field $E(r,\phi,z)$ in the focal region does not depend on ϕ.

★ We only consider real Zernike coefficients β. This implies that $I(x,y,z)$ is symmetric in z.

★ The cross-sections $I(x,z) = I(x,0)$ and $I(x,z) = I(0,z)$ are a sufficiently good representation of $I(x,z)$ in the entire positive xz-plane, hence we only consider the cross-sections.

★ The transverse coordinates in the focal region are normalized by the Airy radius $R_A = \frac{0.61\lambda}{\text{NA}}$, and z is normalized by the Rayleigh unit $R_E = \frac{n\lambda}{\text{NA}^2}$. R_A denotes the lateral width of the focal spot of a focused plane wave, while R_E denotes the length of the spot along the optical axis for this field.

We are now able to define our decision variables:

- Define $\vec{\beta} = [\beta_0^0, \beta_2^0 \cdots \beta_N^0]^T$.
- Define x_0 to be the location of the local minimum of $I(x,0)$ closest to the optical axis.
- Define z_0 to be the location of the point with smallest value where $I(0,z) < 0.9$.
- Define T to be the transmittance of the pupil:

$$T = 2 \int\limits_0^1 \rho |E_0(\rho)|^2 \mathrm{d}\rho.$$

Note that the maximum T is obtained when $|E_0(\rho)| = 1$, in which case $T = 1$.
- Define M to be $M = \max\limits_{x > x_0} I(x,0)$.

Using these definitions, the problem can be formulated as follows:
Find $\vec{\beta}$ such that:

◇ x_0 is minimized,
◇ z_0 is maximized,
◇ T is maximized,

subject to

◇ $M < 0.2$.

B.2. Finding solutions

In order to find good solutions efficiently, we rely on Eq. (4). This equation states that once one has found β such that

$$I = |E|^2 \propto \left(\left| \sum_{k=0}^N \beta_{2k}^0 V_{2k}^0 \right| \right)^2,$$

provides a good result, the correspondig pupil function follows immediately using

$$E_0(\rho, \phi) = \sum_{k=0}^N \beta_{2k}^0 Z_{2k}^0. \tag{16}$$

B.2.1. Precalculations

We start by defining grid points on x and z where we want to calculate the focal field intensity, and the points ρ where we want to find the pupil function

$$\mathbf{x} := \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_{\text{max}} \end{pmatrix}, \mathbf{z} := \begin{pmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_{\text{max}} \end{pmatrix}, \rho := \begin{pmatrix} \rho_1 \\ \cdot \\ \cdot \\ \cdot \\ \rho_{\text{max}} \end{pmatrix}.$$

We then proceed by calculating the V_{2k} and Z_{2k}^0 for those points

$$\mathbf{V}_{x,k}(\mathbf{x}) = \begin{pmatrix} V_{2k}^0(x_1,0) \\ . \\ . \\ . \\ V_{2k}^0(x_{\max},0) \end{pmatrix}, \mathbf{V}_{z,k}(\mathbf{z}) = \begin{pmatrix} V_{2k}^0(0,z_1) \\ . \\ . \\ . \\ V_{2k}^0(0,z_{\max}) \end{pmatrix}, \mathbf{Z}_{\rho,k}(\rho) = \begin{pmatrix} Z_{2k}^0(\rho_1) \\ . \\ . \\ . \\ Z_{2k}^0(\rho_{\max}) \end{pmatrix}.$$

If we now define the matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{R}$ as

$$\mathbf{X} = \begin{bmatrix} V_{x,0}(\mathbf{x}) & V_{x,1}(\mathbf{x}) & ... & V_{x,N}(\mathbf{x}) \end{bmatrix},$$
$$\mathbf{Z} = \begin{bmatrix} V_{z,0}(\mathbf{z}) & V_{z,1}(\mathbf{z}) & ... & V_{z,N}(\mathbf{z}) \end{bmatrix},$$
$$\mathbf{R} = \begin{bmatrix} Z_{\rho,0}(\rho) & Z_{\rho,1}(\rho) & ... & Z_{\rho,N}(\rho) \end{bmatrix},$$

then $I(x,0)$, $I(0,z)$ and $|E_0|^2$ are calculated and normalized according to

$$I(x,0) = \frac{|\mathbf{X}\vec{\beta}|^2}{\max|\mathbf{X}\vec{\beta}|^2}, I(0,z) = \frac{|\mathbf{Z}\vec{\beta}|^2}{\max|\mathbf{Z}\vec{\beta}|^2}, |E_0(\rho)|^2 = \frac{|\mathbf{R}\vec{\beta}|^2}{\max|\mathbf{R}\vec{\beta}|^2}.$$

Note that these matrices need to be calculated only once, which makes this method of finding the focal field distribution much more efficient than calculating a diffraction integral repeatedly.

B.2.2. The algorithm

We start the algorithm by defining the target functions $I_{\text{target}}(x,0)$ and $I_{\text{target}}(0,z)$ which roughly resemble our desired intensity distribution as in Fig. 1. This is shown in the flowchart in Fig. 9. Then we find $\vec{\beta}$ which closely match these targets using the built-in Matlab algorithm `lsqnonlin`. The merit function f is defined as:

$$f = \left| \begin{bmatrix} I_{\text{target}}(x,0) & I_{\text{target}}(0,z) \end{bmatrix} - \begin{bmatrix} I(x,0) & I(0,z) \end{bmatrix} \right|^2 \tag{17}$$

If we feed `lsqnonlin` a random starting $\vec{\beta}_0$, it will apply the *trust-region-reflective algorithm* using f and returns a $\vec{\beta}_1$. We keep creating random starting $\vec{\beta}_0$ until the output is satisfactory.

Having found an initial solution we proceed by improving x_0, z_0, T and D using another Matlab function, `fminsearch`. Given a starting point, it applies the *Nelder-Mead algorithm*, a heuristic search method, to find minima. Of course, while improving T, we do not want x_0 or z_0 to worsen. Therefore we use the following scheme:

- Given a starting solution $\vec{\beta}_1$, we calculate x_0^* and z_0^*.
- We define a function $MT(\vec{\beta})$ which, given a certain $\vec{\beta}$, calculates x_0, z_0, and T, and returns

$$MT(\vec{\beta}) = \begin{cases} \frac{1}{T(\vec{\beta})} & \text{if } x_0(\vec{\beta}) \leq x_0^*, z_0(\vec{\beta}) \geq z_0^* \text{ and } M < 0.2 \\ \infty & \text{otherwise} \end{cases} \tag{18}$$

So basically x_0^* and z_0^* are constraints for x_0 and z_0.
- We apply the Nelder-Mead algorithm to the function $MT(\vec{\beta})$, using the starting solution $\vec{\beta}_1$. It then returns a new $\vec{\beta}$, say $\vec{\beta}_2$, with improved T.
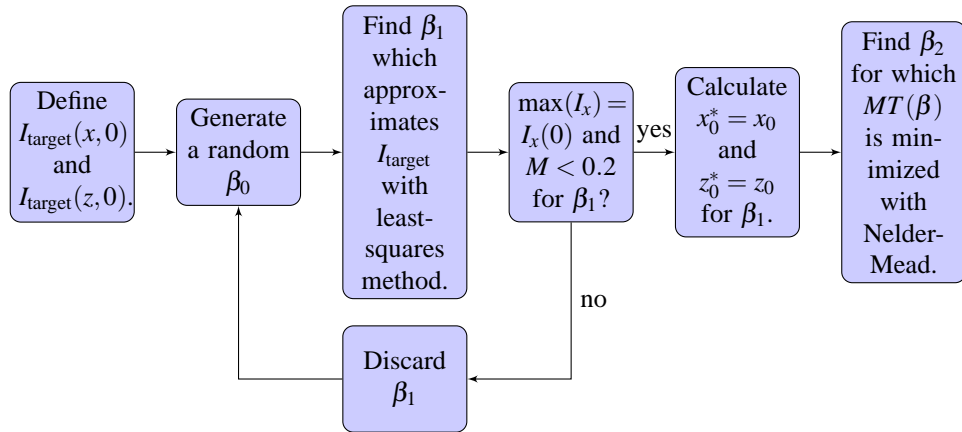
The algorithm is summarized in Fig. 9.

Fig. 9. Flowchart of the algorithm

Acknowledgments