



HAL
open science

Adaptive optics for structured illumination microscopy

Delphine Débarre, Edward J. Botcherby, Martin J. Booth, Tony Wilson

► **To cite this version:**

Delphine Débarre, Edward J. Botcherby, Martin J. Booth, Tony Wilson. Adaptive optics for structured illumination microscopy. *Optics Express*, 2008, 16, pp.9290. 10.1364/OE.16.009290 . hal-00681925

HAL Id: hal-00681925

<https://hal.science/hal-00681925>

Submitted on 28 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptive optics for structured illumination microscopy

Delphine Débarre, Edward J. Botcherby, Martin J. Booth*
and Tony Wilson

Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ,
United Kingdom.

martin.booth@eng.ox.ac.uk

Abstract: We implement wave front sensor-less adaptive optics in a structured illumination microscope. We investigate how the image formation process in this type of microscope is affected by aberrations. It is found that aberrations can be classified into two groups, those that affect imaging of the illumination pattern and those that have no influence on this pattern. We derive a set of aberration modes ideally suited to this application and use these modes as the basis for an efficient aberration correction scheme. Each mode is corrected independently through the sequential optimisation of an image quality metric. Aberration corrected imaging is demonstrated using fixed fluorescent specimens. Images are further improved using differential aberration imaging for reduction of background fluorescence.

© 2008 Optical Society of America

OCIS codes: (010.1080) Adaptive Optics; (010.7350) Wave-front sensing; (180.2520) Fluorescence microscopy;(180.6900) Three-dimensional microscopy.

References and links

1. J. B. Pawley, ed., *Handbook of Biological Confocal Microscopy*, 3rd Edition (Springer, New York, 2006).
2. J. A. Conchello and J. W. Lichtman, "Optical sectioning microscopy," *Nature Methods* **2**(12), 920–931 (2005).
3. M. A. A. Neil, R. Juškaitis, and T. Wilson, "Method of obtaining optical sectioning by using structured light in a conventional microscope," *Opt. Lett.* **22**, 1905–1907 (1997).
4. M. J. Booth, "Adaptive optics in microscopy," *Philos. Transact. A Math. Phys. Eng. Sci.* **365**, 2829–2843 (2007).
5. M. J. Booth, M. A. A. Neil, R. Juškaitis, and T. Wilson, "Adaptive aberration correction in a confocal microscope," *Proc. Nat. Acad. Sci.* **99**, 5788–5792 (2002).
6. D. Débarre, M. J. Booth, and T. Wilson, "Image based adaptive optics through optimisation of low spatial frequencies," *Opt. Express* **15**, 8176–8190 (2007).
7. H. Hopkins, "The use of diffraction-based criteria of image quality in automatic optical design," *Opt. Acta* **13**, 343–69 (1966).
8. D. Karadaglić and T. Wilson, "Image formation in structured illumination wide-field fluorescence microscopy," *Micron* (2008, in press).
9. E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics* (Chapman and Hall/CRC, 2003).
10. M. A. A. Neil, M. J. Booth, and T. Wilson, "New modal wavefront sensor: a theoretical analysis," *J. Opt. Soc. Am. A* **17**, 1098–1107, (2000).
11. M. J. Booth, T. Wilson, H.-B. Sun, T. Ota, and S. Kawata, "Methods for the characterisation of deformable membrane mirrors," *Appl. Opt.* **44**(24), 5131–5139 (2005).
12. A. Leray and J. Mertz, "Rejection of two-photon fluorescence background in thick tissue by differential aberration imaging," *Opt. Express* **14**, 10,565–10,573 (2006).

1. Introduction

Optical sectioning microscopy is widely used to provide three-dimensional fluorescence images of biological specimens. A common way of obtaining this sectioning ability is through point

scanning methods such as confocal or multiphoton microscopy [1, 2]. An alternative is to use a wide-field technique such as structured illumination (SI) microscopy, which retains the sectioning ability of confocal microscopy, but can be implemented in a conventional microscope using an incoherent light source, and without the need for scanning [3]. In this technique, the image of a grid is projected on the specimen so as to produce a one-dimensional sinusoidal excitation pattern in the focal plane of the objective lens. The resulting fluorescence image, consisting of both in-focus and out-of-focus fluorescence emission, is acquired by a camera. Several images are taken, each corresponding to a different grid position. As the grid pattern appears only in the focal plane, it is possible to extract an optical section from the spatially modulated component of the images via a simple calculation.

As with all microscopes, aberrations also detrimentally affect imaging in SI microscopy, leading to lower image intensity and reduced resolution. Image quality can be restored using the techniques of adaptive optics, where an adaptive element, such as a deformable mirror (DM), is used to correct the aberrations [4]. Traditional adaptive optics systems use a wavefront sensor to measure aberrations. An alternative approach is to use model-based, wavefront sensor-less schemes, in which the aberration correction is indirectly optimised through the application of a short sequence of trial aberrations. By using an appropriate combination of optimisation metric, modal aberration expansion and aberration estimator algorithm, correction can be achieved with a minimal number of measurements. The use of these efficient correction schemes is of particular interest in biological imaging, where the reduced number of measurements minimises photobleaching and damage on the sample. Such schemes have been demonstrated in confocal microscopy [5] and more recently in incoherent transmission imaging [6].

In this paper we describe wavefront sensorless adaptive optics implemented in a SI microscope. In Section 2, we study the image formation process in the SI microscope and investigate the effects of aberrations on imaging performance. It is shown that the final image quality depends predominantly on the imaging efficiency of the illumination pattern's spatial frequency. This imaging efficiency is affected much more by some aberration modes than by others. Consequently, different aberration modes can have significantly different effects on the final sectioned image. We therefore present in Section 3 a general method that provides an optimum modal expansion of the aberration and suggests an efficient aberration correction scheme. In Sections 4 to 6, we apply this method to the SI microscope and show how an optimum set of aberration modes can be derived theoretically and confirmed experimentally. In the subsequent sections, we describe how the adaptive optics scheme was implemented and applied to the correction of aberrations in biological specimens. We discuss how the general scheme presented in this paper can be applied to other adaptive optics systems.

2. Image formation in a structured illumination microscope

A structured illumination fluorescence microscope can be modelled as a two stage incoherent imaging system. In such a system, the image $I(\mathbf{n})$ is formed as the convolution of the object function $f(\mathbf{n})$ and the intensity point spread function in the focal plane $h(\mathbf{n})$:

$$I(\mathbf{n}) = h(\mathbf{n}) * f(\mathbf{n}) , \quad (1)$$

where \mathbf{n} is the coordinate vector in the plane perpendicular to the optical axis, and $*$ is the convolution operation. Alternatively, we can consider the imaging process in the frequency domain and write:

$$I(\mathbf{n}) = \text{FT}^{-1} [H(\mathbf{m})F(\mathbf{m})] , \quad (2)$$

where FT is the Fourier transform operation, \mathbf{m} is the two-dimensional spatial frequency coordinate vector, $H(\mathbf{m})$ is the optical transfer function (OTF), which is equivalent to the two-dimensional FT of $h(\mathbf{n})$, and $F(\mathbf{m})$ is the FT of $f(\mathbf{n})$. For simplicity, we present here expressions

that describe the imaging of two-dimensional objects, but this approach is readily extendable to three-dimensional imaging. The OTF can be calculated as the autocorrelation of the effective pupil function, $P(\mathbf{r})$:

$$H(\mathbf{m}) = P(\mathbf{r}) \otimes P^*(\mathbf{r}) = \frac{1}{A} \iint_{D(\mathbf{m})} P(\mathbf{r} - \mathbf{m}) P^*(\mathbf{r}) dA, \quad (3)$$

where \otimes is the correlation operation, A is the pupil area, \mathbf{r} is the position vector, P^* is the complex conjugate of P and $D(\mathbf{m})$ is the region of overlap of the offset pupils. We assume a circular pupil with unity radius, so $D(\mathbf{m})$ is defined by $|\mathbf{r}| \leq 1$ and $|\mathbf{r} - \mathbf{m}| \leq 1$ (see Fig. 1, part b1). If there is no amplitude variation across the pupil, then we can express $P(\mathbf{r}) = \exp[j\Phi(\mathbf{r})]$, where $\Phi(\mathbf{r})$ is the phase aberration and $j = \sqrt{-1}$. The OTF becomes

$$H(\mathbf{m}) = \frac{1}{\pi} \iint_{D(\mathbf{m})} \exp[j\Delta\Phi(\mathbf{m}, \mathbf{r})] d\mathbf{r} \quad (4)$$

with

$$\Delta\Phi(\mathbf{m}, \mathbf{r}) = \Phi(\mathbf{r} - \mathbf{m}) - \Phi(\mathbf{r}). \quad (5)$$

The aberration difference function $\Delta\Phi(\mathbf{m}, \mathbf{r})$, rather than the aberration $\Phi(\mathbf{r})$, is therefore the important quantity when considering the effect of aberrations in the incoherent imaging of a particular spatial frequency \mathbf{m} [7].

The SI microscope relies upon the projection of a physical grid pattern into the focal plane of the specimen. We assume that the grid object is a sinusoidal transmission mask with unity modulation depth and spatial frequency vector \mathbf{g} , of the form $1 + \cos(\psi + \mathbf{g} \cdot \mathbf{n})$. The variable ψ is the spatial phase shift of the pattern that depends on the grid displacement along the direction of \mathbf{g} . Using equation 2, the excitation pattern formed by its image in the specimen can thus be written as:

$$I_{exc}(\mathbf{n}) = H_{exc}(\mathbf{0}) + H_{exc}(\mathbf{g}) \frac{e^{(j\psi + j\mathbf{g} \cdot \mathbf{n})}}{2} + H_{exc}(-\mathbf{g}) \frac{e^{-(j\psi + j\mathbf{g} \cdot \mathbf{n})}}{2}, \quad (6)$$

where $H_{exc}(\mathbf{m})$ is the FT of $h_{exc}(\mathbf{r})$, the intensity PSF of the excitation path imaging system. This illumination pattern excites fluorescence in the specimen, whose fluorophore distribution is described by the object function $f(\mathbf{n})$. The generated fluorescence is therefore given by the product $f(\mathbf{n})I_{exc}(\mathbf{n})$. If we now consider the image $I_i(\mathbf{n})$ of this fluorescence obtained on the camera for a position i of the grid corresponding to $\psi = \psi_i$, we can write

$$I_i(\mathbf{n}) = [f(\mathbf{n})I_{exc,i}(\mathbf{n})] * h_{em}(\mathbf{n}), \quad (7)$$

where $h_{em}(\mathbf{n})$ is the intensity PSF of the emission path of the microscope. For simplicity, we assume from here on that the excitation and emission wavelengths are close and that the aberrations are equal for the excitation and the emission pathways. Hence $h_{exc}(\mathbf{n}) = h_{em}(\mathbf{n}) = h(\mathbf{n})$ and equivalently $H_{exc}(\mathbf{m}) = H_{em}(\mathbf{m}) = H(\mathbf{m})$.

The sectioned image is retrieved from the three images $I_i(\mathbf{n})$ corresponding to grid positions $\psi_i = 0, \frac{2\pi}{3}$ and $\frac{4\pi}{3}$ respectively, using the formula [3]:

$$I_{sect}(\mathbf{n}) = \sqrt{\sum_{i \neq j} (I_i - I_j)^2}. \quad (8)$$

It can be shown that this image, when expressed in terms of spatial frequencies, is given by [8]:

$$I_{sect}(\mathbf{n}) = \frac{3}{2} |H(\mathbf{g}) \text{FT}^{-1} [F(\mathbf{m}) H(\mathbf{g} + \mathbf{m})]|. \quad (9)$$

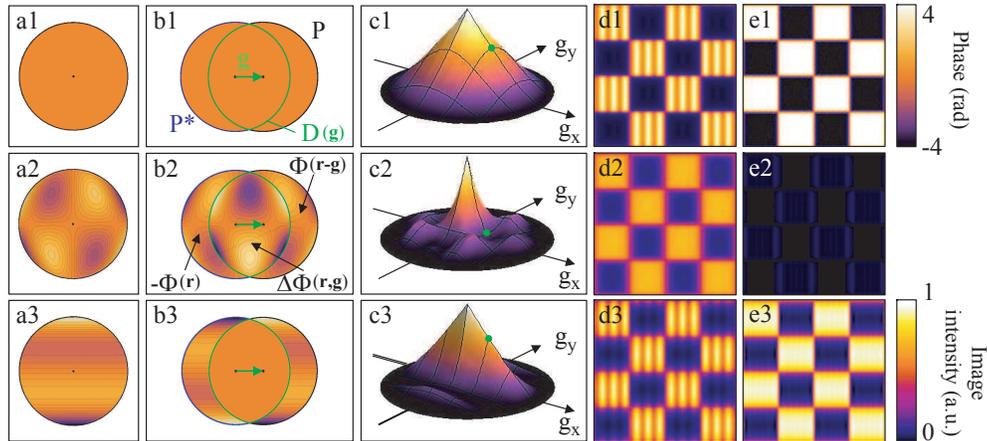


Fig. 1. Illustration of the effects of grid modes and non-grid modes: (a1-3) the pupil function $P(\mathbf{r})$ showing zero phase aberration, an example grid mode, and an example non-grid mode, respectively. The latter two aberrations have a rms phase of 1 rad. (b1) the geometry used for calculating the OTF, showing the two pupil functions offset by the grid frequency \mathbf{g} ; in (b2) the overlap region shows $\Delta\Phi \neq 0$ for the grid mode; in (b3) the overlap region shows $\Delta\Phi = 0$ for the non-grid mode. (c1-3) show the corresponding OTFs; \mathbf{g} is indicated by the green dots. (d1-3) simulated images of a checkerboard pattern for one grid position. (e1-3) corresponding sectioned images derived using Eq. (8).

It is clear from this expression that the imaging efficiency of the grid's spatial frequency, represented by $H(\mathbf{g})$, dominates the overall intensity of the sectioned image. The first $H(\mathbf{g})$ term is due to the imaging of the grid pattern onto the specimen on the illumination side. The term $H(\mathbf{g} + \mathbf{m})$ appears because the object spectrum is modulated by the grid frequency when imaged onto the camera. Hence, those aberrations that affect the grid frequency have the most significant effect on the SI microscope.

By inspection of Eqs. (4) and (5), it can be seen that $H(\mathbf{g})$ is unaffected by aberrations for which $\Delta\Phi(\mathbf{g}, \mathbf{r})$ is a constant. Conversely, $H(\mathbf{g})$ is strongly affected by aberrations for which $\Delta\Phi(\mathbf{g}, \mathbf{r})$ has a large variance. It is therefore useful to separate aberration modes into two groups: those that affect the grid frequency (referred to hereon as "grid modes") and those that have no influence on this frequency ("non-grid modes"). The effects of these modes are illustrated in Fig. 1. It is clear that grid modes have a significant influence on the intensity of the sectioned image, whereas non-grid modes have comparatively little effect. The non-grid modes do however affect the resolution.

3. Derivation of a general optimisation scheme

The specification of a modal aberration correction scheme requires the choice of three components: the aberration representation (the mathematical functions used to describe the aberrations), the optimisation metric (a quantity representing the image quality) and the estimator (the algorithm for estimating the correction aberration). In order to obtain an optimum scheme in which the number of required measurements is minimised, it is useful to consider these three components as linked. We would like to choose a metric function M whose maximum corresponds to the highest quality image. For convenience, we will assume that this maximum occurs for zero aberration, $\Phi = 0$. One way to achieve efficient optimisation is to find an aber-

ration expansion for which the modes act independently on the metric. This would allow the independent optimisation of each mode. This can be achieved if the metric is of the form

$$M = M_0 - \sum_i x_i^2, \quad (10)$$

where M_0 is the value of the metric for zero aberration and the coefficients x_i represent aberration mode amplitudes. When M is expressed in this form, it is clear that independent maximisation with respect to each x_i is possible. Furthermore, as the function is quadratic, the maximum can be found directly from three measurements of M corresponding to three different values of x_i [6]. In practice, these measurements would be taken with three different trial aberrations introduced by the correction element. However, M only takes the form shown in Eq. (10) if the aberration representation is appropriately chosen. In this section we explain a generally applicable process that facilitates this choice.

Let us assume that the aberrations are represented by an expansion over a complete (but as yet undefined) set of modes $\{X_i\}$ with a set of coefficients $\{a_i\}$ so that $\Phi = \sum_i a_i X_i$. It is also convenient to represent this aberration as a vector \mathbf{a} , whose elements are the coefficients $\{a_i\}$. In general, for sufficiently small aberrations, M would be approximated by a quadratic polynomial:

$$M \approx M_0 - \sum_i \sum_j \alpha_{ij} a_i a_j = M_0 - \mathbf{a}^T \mathbf{A} \mathbf{a}, \quad (11)$$

where the constants α_{ij} are the elements of the matrix \mathbf{A} . In this form, the combination of the metric and the aberration expansion is not ideal for efficient optimisation as the coefficients do not act independently. As this expression represents a maximum of M , the matrix \mathbf{A} must be positive semi-definite, i.e. $\mathbf{a}^T \mathbf{A} \mathbf{a} \geq 0$ for all $\mathbf{a} \neq \mathbf{0}$. Furthermore, any positive semi-definite matrix can be converted into a diagonal matrix by:

$$\mathbf{A} = \mathbf{V} \mathbf{B} \mathbf{V}^T, \quad (12)$$

where \mathbf{B} is a diagonal matrix with elements $B_{ii} = \beta_i$ and where the columns of the orthogonal matrix \mathbf{V} are the eigenvectors of \mathbf{A} . Naturally, the diagonal elements of \mathbf{B} are also the eigenvalues of \mathbf{A} . The optimisation metric then becomes

$$M \approx M_0 - \mathbf{a}^T \mathbf{V} \mathbf{B} \mathbf{V}^T \mathbf{a} = M_0 - \mathbf{b}^T \mathbf{B} \mathbf{b} = M_0 - \sum_i \beta_i b_i^2, \quad (13)$$

with $\mathbf{b} = \mathbf{V}^T \mathbf{a}$. Eq. (13) has the desired form shown in Eq. (10). The process of converting Eq. (11) into the form of Eq. (13) is equivalent to obtaining an alternative expansion of the aberration function in terms of a new set of modes $\{Y_i\}$ so that $\Phi = \sum_i b_i Y_i$, where the new modes can be calculated as $Y_i = \sum_j V_{ij} X_j$.

We now outline the procedure for obtaining the matrix \mathbf{A} . As this is a positive semi-definite matrix, it must also be equivalent to a Gram matrix (the matrix of inner products of a particular set of basis functions) [9]. It follows that \mathbf{A} can be calculated if one has an appropriate definition for the inner product, $\alpha_{ij} = \langle X_i, X_j \rangle$. This can sometimes be obtained from an expansion of the metric M in terms of the aberration coefficients a_i . This process is illustrated in the following sections. Alternatively, \mathbf{A} can be determined empirically, by measuring the behaviour of M in the vicinity of the maximum. This process is demonstrated in Section 6.

We note that the metric M may be insensitive to certain aberration modes and therefore the maximum value may also occur for other values of $\Phi \neq 0$ (as an example, measurements in most microscopes would be insensitive to a constant phase offset). Equivalently, the inner product would be degenerate, meaning that $\langle Y_i, Y_i \rangle = 0$ for certain i . This property is useful as it enables the separation of the set of aberration modes into those that affect the chosen metric and those

that have no influence. In the latter case, the expression in Eq. (13) would contain zero values for each β_i corresponding to a mode Y_i that did not influence the metric. Clearly, it would not be desirable to include these modes in any correction scheme.

4. Optimisation metric for structured illumination microscopy

A commonly used function that satisfies the required properties of an optimisation metric is the image sharpness, defined as:

$$M = \iint_{-\infty}^{+\infty} I_{sect}^2(\mathbf{n}) d\mathbf{n} . \quad (14)$$

With the help of Eq. (9) and using Parseval's theorem, it becomes

$$M = \iint_S |F(\mathbf{m})|^2 |H(\mathbf{g})H(\mathbf{g} + \mathbf{m})|^2 d\mathbf{m} , \quad (15)$$

where we have neglected the premultiplying constant. The region of integration S is the circular region of support of the offset OTF $H(\mathbf{g} + \mathbf{m})$. If we assume that the sample frequency spectrum takes significant values only for small values of $|\mathbf{m}|$ (an approximation valid for many practical samples), we obtain using a Taylor expansion (see Appendix B):

$$M = \left(\iint_S |F(\mathbf{m})|^2 d\mathbf{m} \right) |H(\mathbf{g})|^4 + \frac{1}{2} \left(\iint_S |F(\mathbf{m})|^2 m^2 d\mathbf{m} \right) |H(\mathbf{g})|^2 \nabla^2 \left(|H(\mathbf{g})|^2 \right) , \quad (16)$$

where ∇^2 is the Laplacian operator and where we take $\nabla^2 t(\mathbf{g})$ to mean $[\nabla_{\mathbf{m}'}^2 t(\mathbf{m}')]_{\mathbf{m}'=\mathbf{g}}$. In each term of this equation, the effects of aberrations are separated from the object properties. This enables us to design a correction scheme that is mostly independent of the object structure. For a given object, the integral terms in parentheses are constants and Eq. (16) can be written as

$$M = F_0 |H(\mathbf{g})|^4 + F_1 |H(\mathbf{g})|^2 \nabla^2 \left(|H(\mathbf{g})|^2 \right) , \quad (17)$$

where F_0 and F_1 are constants. This equation suggests a two stage scheme for aberration correction in the SI microscope, firstly correcting grid modes and thereafter non-grid modes.

For grid modes, the variation of M is dominated by the first term of Eq. (17); non-grid modes by definition have no effect on this term. Hence, following the scheme presented in Section 3, we can determine the appropriate inner product using $M \sim F_0 |H(\mathbf{g})|^4$ and thus derive the set of grid modes. The first stage of optimisation is performed using this subset of aberrations.

The remaining aberration then consists solely of non-grid modes, for which we can consider $H(\mathbf{g})$ to be constant. For this second stage of optimisation, the metric therefore varies as $M \sim \nabla^2 \left(|H(\mathbf{g})|^2 \right)$. This expression can be used to obtain a different inner product from which the non-grid modes can be derived. We then perform optimisation based upon these modes. In the next section we derive these two inner products explicitly.

5. Explicit expression of the metric as a function of modal coefficients

For small amplitudes of $\Delta\Phi$, the exponential term in Eq. (4) can be expanded as a Taylor series, giving

$$H(\mathbf{m}) = \frac{1}{\pi} \iint_{D(\mathbf{m})} d\mathbf{r} + \frac{j}{\pi} \iint_{D(\mathbf{m})} \Delta\Phi(\mathbf{m}, \mathbf{r}) d\mathbf{r} - \frac{1}{2\pi} \iint_{D(\mathbf{m})} [\Delta\Phi(\mathbf{m}, \mathbf{r})]^2 d\mathbf{r} . \quad (18)$$

If M is dominated by the first term in Eq. (17), we can write

$$M = F_0 \left| H_0(\mathbf{g}) + \frac{j}{\pi} \iint_{D(\mathbf{g})} \Delta\Phi(\mathbf{g}, \mathbf{r}) d\mathbf{r} - \frac{1}{2\pi} \iint_{D(\mathbf{g})} \Delta\Phi^2(\mathbf{g}, \mathbf{r}) d\mathbf{r} \right|^4, \quad (19)$$

where $H_0(\mathbf{g})$ is the OTF in the absence of aberrations. For a fixed grid frequency, this is a constant, so will be referred to simply as H_0 . Considering only terms up to the second order, Eq. (19) can be written:

$$M = F_0 \left\{ H_0^4 - 2H_0^2 \left[\frac{H_0}{\pi} \iint_{D(\mathbf{g})} [\Delta\Phi(\mathbf{g}, \mathbf{r})]^2 d\mathbf{r} - \left(\frac{1}{\pi} \iint_{D(\mathbf{g})} \Delta\Phi(\mathbf{g}, \mathbf{r}) d\mathbf{r} \right)^2 \right] \right\} \quad (20)$$

Let the pupil phase be represented by the modal expansion $\Phi = \sum_i a_i X_i$. We define corresponding aberration difference modes as $\Delta X_i(\mathbf{g}, \mathbf{r}) = X_i(\mathbf{r} - \mathbf{g}) - X_i(\mathbf{r})$ so that $\Delta\Phi = \sum_i a_i \Delta X_i$. Eq. (20) then becomes

$$M = F_0 \left\{ H_0^4 - 2H_0^2 \sum_i \sum_j a_i a_j \left[\frac{H_0}{\pi} \iint_{D(\mathbf{g})} \Delta X_i \Delta X_j d\mathbf{r} - \frac{1}{\pi^2} \iint_{D(\mathbf{g})} \Delta X_i d\mathbf{r} \iint_{D(\mathbf{g})} \Delta X_j d\mathbf{r} \right] \right\}, \quad (21)$$

where for brevity we have omitted the arguments of the $\Delta X_i(\mathbf{g}, \mathbf{r})$ terms. Comparison of Eqs. (11) and (21) shows that the expression in square brackets represents the inner product required to determine the coefficients α_{ij} . The metric is therefore

$$M = M_0 - \sum_i \sum_j a_i a_j \langle X_i, X_j \rangle, \quad (22)$$

where $M_0 = F_0 H_0^4$ and where the (degenerate) inner product is defined by

$$\langle X_i, X_j \rangle_{\text{grid}} = 2F_0 H_0^2 \left[\frac{H_0}{\pi} \iint_{D(\mathbf{g})} \Delta X_i \Delta X_j d\mathbf{r} - \frac{1}{\pi^2} \iint_{D(\mathbf{g})} \Delta X_i d\mathbf{r} \iint_{D(\mathbf{g})} \Delta X_j d\mathbf{r} \right]. \quad (23)$$

This inner product is used to derive the grid modes when the specimen is a thin object located in the focal plane. However in the case of a more realistic sample with finite thickness, the axial extent of the specimen should also be taken into account. Following the same principles, an expression for such an inner product is derived in Appendix A.

As the inner product $\langle \cdot, \cdot \rangle_{\text{grid}}$ is degenerate for non-grid modes, for the second stage of correction we need another inner product that is derived using the second term of Eq. (17). The grid pattern in the specimen is unaffected by non-grid modes. It follows that the axial sectioning strength of the microscope does not vary with the amplitude of these aberrations and, hence, we only need consider the plane $z = 0$ in the following analysis. It is shown in Appendix B that the new inner product is defined, to the lowest order in \mathbf{m} , as

$$\langle X_i, X_j \rangle_{\text{non-grid}} = F_1 \left[\frac{H_0}{\pi} \iint_{D(\mathbf{g})} \nabla X_i' \nabla X_j' d\mathbf{r} - \frac{1}{\pi^2} \iint_{D(\mathbf{g})} \nabla X_i' d\mathbf{r} \iint_{D(\mathbf{g})} \nabla X_j' d\mathbf{r} \right], \quad (24)$$

where ∇ is the gradient operator and X_i' represents $X_i(\mathbf{r} - \mathbf{g})$.

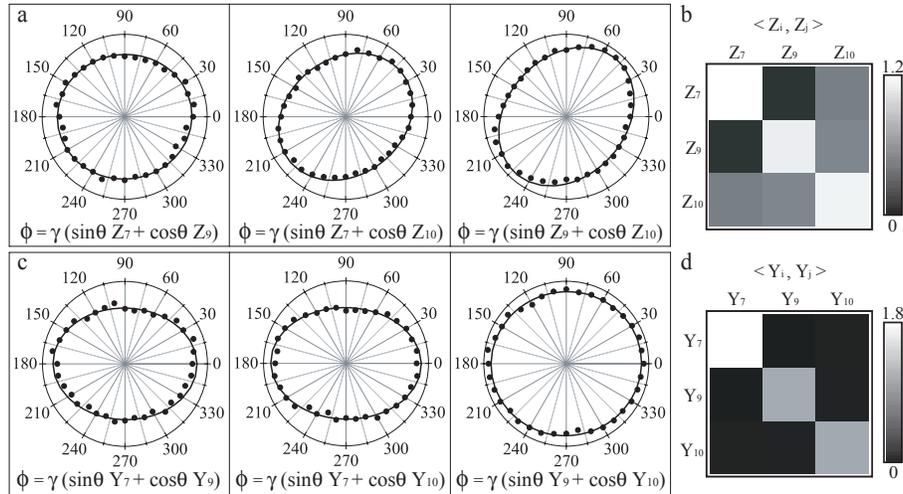


Fig. 2. Experimental determination of aberration modes. (a), the metric M is plotted as a function of the aberrated phase $\Phi(\mathbf{r})$ for a number of basis modes (here 3 Zernike modes: coma ($z=7$) and trefoil ($z=9$ and $z=10$)), and the resulting curves are fitted to a multidimensional ellipsoid. (b), the fitting parameters are used to construct the experimental \mathbf{A} matrix and determine the new set of modes $\{Y_i(\mathbf{r})\}$. (c), a similar plot of the metric M , derived using the new set of modes. The main axes of the fitted ellipsoid now correspond to pure $\{Y_i(\mathbf{r})\}$ modes. (d), the experimental \mathbf{A} matrix for the new modes is diagonal, confirming that the cross-talk between the modes has been cancelled. Although only three modes are shown here, this principle can be extended to an arbitrary number.

6. Experimental determination of aberration modes

In order to verify the theoretical derivation, we developed a complementary empirical method for ascertaining the optimal set of modes for aberration correction. The method relies on the principle outlined in Fig. 2: for any pair (i, j) of initial modes with $i \neq j$, the value of M is measured for a number of aberrated pupil phases $\Phi(\mathbf{r}, \theta)$ obtained by combining the basis modes $X_i(\mathbf{r})$ and $X_j(\mathbf{r})$ with constant amplitude γ :

$$\Phi(\mathbf{r}, \theta) = \gamma \cos \theta X_i(\mathbf{r}) + \gamma \sin \theta X_j(\mathbf{r}), \quad (25)$$

where θ varies between 0 and 2π . In the vicinity of its maximum, the contours of M as a function of the coefficients $\{a_{ij}\}$ are ellipsoidal. Therefore, the resulting curves were simultaneously fitted to a multidimensional ellipsoid defined as

$$\sum_{i,j} \alpha_{ij} a_i a_j = c \quad (26)$$

where c is a constant. This fitting provides us with the coefficients α_{ij} and hence the matrix \mathbf{A} . This matrix is identical to that derived in Section 5, but is obtained without recourse to the inner product calculation. The same approach can be used as before to derive the new set of basis modes $\{Y_i\}$. An example of this orthogonalisation process is presented in Fig. 2. These results were obtained using the experimental set-up described in the next section and the value of the metric M was calculated from images of a thin inhomogeneous fluorescent sheet located in the focal plane of the focussing objective, for various values of $\Phi(\mathbf{r}, \theta)$. In order to test the

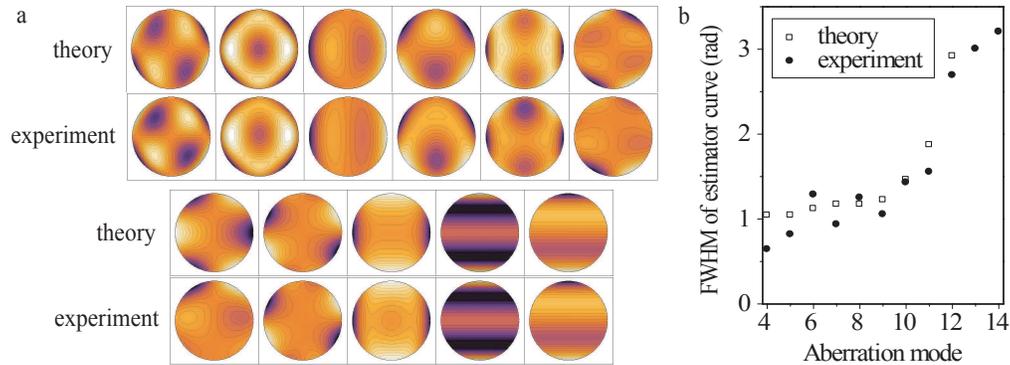


Fig. 3. Set of modes used for aberration correction. (a), set of modes determined experimentally and theoretically using an initial set of 11 Zernike modes ($z=5$ to 15). The modes are ordered by decreasing eigenvalue. The last two modes have no influence on the imaging of the grid. All the modes have a root mean square (rms) phase amplitude of 1 rad. (b), Influence of the modes on the metric M , determined theoretically using the mode eigenvalues, and experimentally as the FWHM of the curve of M as a function of aberration amplitude in a single mode. The last two modes influence only the sample frequency spectrum, and hence their eigenvalues depend on the sample and cannot be determined theoretically.

validity of this method, the same process was repeated with the newly determined modes. We obtained a diagonal matrix, confirming that the new set of modes does not exhibit cross-talk and can be used for sequential correction of the aberrations.

The comparison between the experimentally determined modes and their theoretically determined counterparts is presented in Fig. 3. The two sets were obtained using eleven Zernike modes ($z = 5$ to $z = 15$ using the indexing scheme of Reference [10]) as the basis functions $\{X_i\}$. We note that piston, tip, tilt and defocus were not used as basis functions as they should not be included in correction schemes for three dimensional imaging to avoid changing the imaged region in the sample. Good agreement was found between the two approaches. The observed discrepancies were attributed to the residual error in the control of the aberrated phase in the pupil plane. It should be noted here that since the derivation was independent of the object structure, this characterization was only performed once, and the same set of modes was used for all the following experiments.

We measured the variation of M as a function of aberration amplitude for each of the $\{Y_i\}$ modes. These curves were accurately described by Gaussian functions; the full width at half maximum (FWHM) for each of the curves is displayed in Fig. 3b. In the case of grid modes, the obtained values can be related to the eigenvalues β_i and show a good agreement with the calculated values. It should be noted that the width of each curve also determines the domain of validity of the quadratic approximation used in section 3, which is approximately the same as the FWHM. For greater aberration amplitudes, we found experimentally that correction of each mode could still be performed independently with good accuracy by taking into account the Gaussian shape of the curves in the correction algorithm. The range of the correction was thus only limited by the aberration amplitude accessible to our DM.

7. Results

We demonstrated the aberration correction scheme in a structured illumination microscope based around a modified IX70 inverted microscope (Olympus), incorporating a deformable

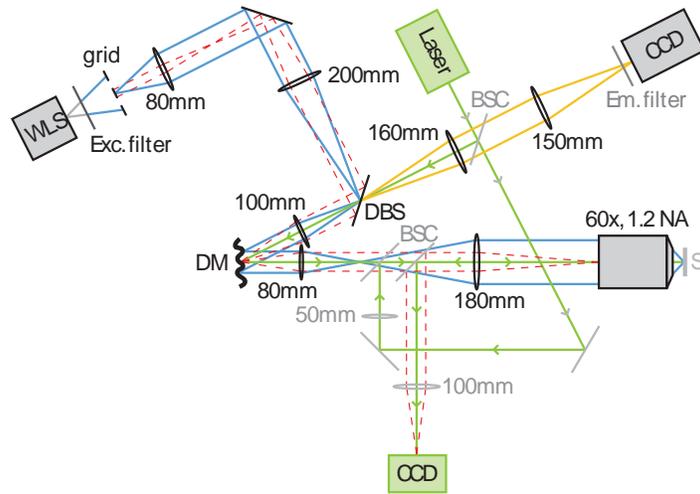


Fig. 4. Schematic of the structured illumination microscope with aberration correction. WLS, white light source. DM, deformable mirror. DBS, dichroic beamsplitter. BSC, beam-splitter cube. S, sample. The blue rays mark the illumination path; the detection path is shown in yellow. The green path represents the Mach Zehnder interferometer used to characterise the deformable mirror – this is not used during the imaging experiments.

mirror (DM – Boston Micromachines Corp., Multi-DM), a 60x, 1.2NA water objective lens with coverslip correction (Olympus) and a 30- μm period grid (Optigrd, Thalès) (see Fig. 4). With the magnification of our system, the spatial frequency of the grid in the pupil plane was about 4 times smaller than the incoherent frequency cutoff, corresponding to a theoretical axial resolution of 0.46 μm [8]. The grid was illuminated by a white-light source with a narrowband excitation filter centered on 488 nm (Chroma) and imaged in the focal plane of the objective after being reflected off the DM, which was conjugated to the pupil plane of the objective. The fluorescence was then detected on a CCD camera through a broadband emission filter and a dichroic beamsplitter (both Chroma), after reflection off the DM. In order to precisely control the aberrations induced by the DM, an interferometer was added to the set-up, incorporating a 532nm laser and a CCD camera conjugated with both the pupil plane of the objective and the DM. By placing a mirror in the focal plane of the objective, this permitted us to measure the aberrated phase in the pupil plane and hence to determine the combinations of control signals sent to the mirror that would produce a given aberration mode [11]. The control signals were restricted to ensure that the DM operated in a linear range; this range was found to cover approximately 85% of the full deflection of the DM. As the relative position of the mirror and the objective were fixed, the characterisation was performed only once, and the same control matrix was then used in the rest of the experiments. As a final step, we used an appropriate control signal offset to ensure the initial flatness of the mirror lay within the measurement uncertainty of our set-up (rms phase ≤ 0.02 rad).

Using the experimentally determined modes shown in Fig. 3, we performed aberration correction on a fixed mouse intestine sample (FluoCells prepared slide 4, Molecular Probes, USA) using a multi-dimensional quadratic maximization algorithm: for each aberration mode, the metric M was measured when adding a given amount of the considered mode, then again when subtracting the same amount. Along with the measured value of M when no aberration was added, this allowed us to estimate the initial aberration present in each of the assessed modes

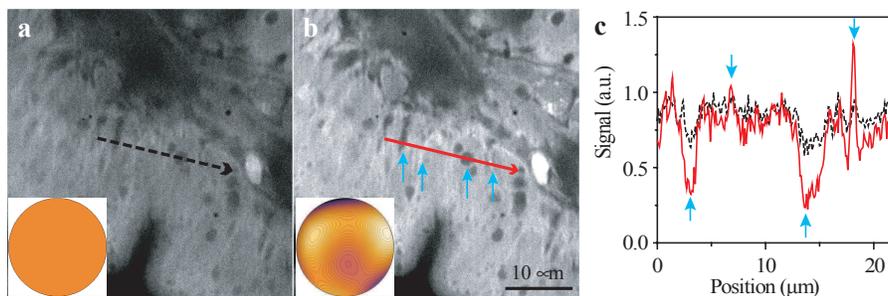


Fig. 5. Aberration correction in SI microscopy. A fluorescent mouse intestine sample was imaged before (a) and after (b) aberration correction. The two images are displayed with the same color table. Insert, phase induced by the mirror in the pupil plane of the objective. The rms phase after correction is 0.61 rad. (c), profile along the lines drawn on the images. Both profiles have been normalized so that their mean value is identical. As a result of the resolution improvement, the contrast of small sample features (blue arrows) are better defined after (red solid line) rather than before (black dotted line) correction. The imaging depth was approximately $10\mu\text{m}$. The coverslip thickness was $170\mu\text{m}$.

and hence to correct these aberrations [6]. After each mode was optimised, images of the sample were acquired with the full aberration correction applied. These exhibit significant intensity and resolution improvement compared to the uncorrected image (Fig. 5). Here we want to emphasise that the use of the appropriate metric and aberration modes means that the correction of N modes was performed using only $2N + 1$ measurements, thereby minimizing the increase in illumination time of the sample required for the correction. The acquisition time per sectioned image was 300-500ms and the total time required for correction of 11 modes was 7-12s, depending on the sample.

An additional benefit of aberration correction for structured illumination microscopy is that the adaptive element can also be used to improve the rejection of the out-of-focus fluorescence, as recently demonstrated by Leray and Mertz in two-photon microscopy [12]. Indeed, when imaging a thick specimen, noise fluctuations in the fluorescence signal between the three successive widefield images result in a large out-of-focus background in the calculated sectioned images. As it arises from fluorescence generated outside the focal plane, this background is weakly sensitive to the presence of aberrations, whereas the in-focus calculated signal depends strongly on the visibility of the grid pattern on the camera, and hence on the amount of aberration in the grid modes. Therefore, by applying large aberrations in a combination of grid modes, the grid pattern is suppressed and only the out-of-focus noise can be measured when reconstructing the sectioned image. By subtracting this aberrated image from the original sectioned image, the fluorescent background can be efficiently removed, leading to greatly improved contrast of the in-focus structures (Fig. 6). This improvement comes however at the expense of additional exposure of the specimen and slight increase of the noise in the resulting images. When combined with adaptive aberration correction, this process leads to significant improvement in image quality.

As the aberrations have been optimised over an axial thickness equal to the one of the images, the correction can be adapted for different imaging depths in the sample. This permits improvement of the image quality throughout an axially extended sample. Indeed, as the aberrations can change significantly with depth, using the same correction phase for different depths can result in a degradation of the image quality compared to the case where no correction is applied, as

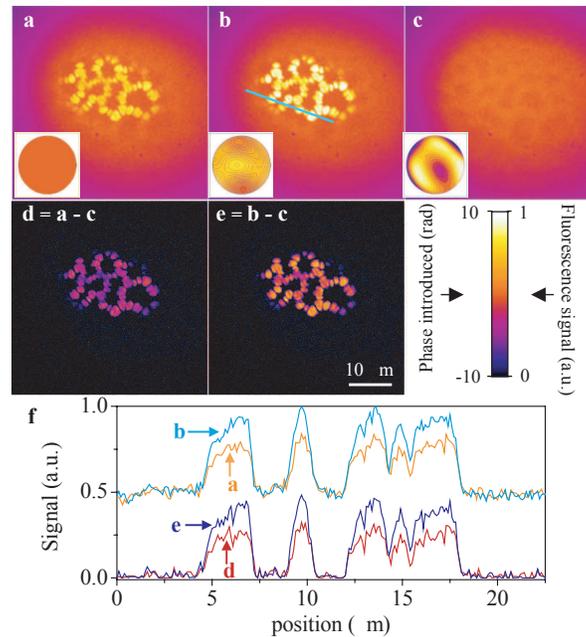


Fig. 6. Aberration correction and out-of-focus fluorescence rejection in SI microscopy. Axially sectioned images of a pollen grain without (a) and with (b) aberration correction, and with large induced aberration (c). The phase induced by the mirror is shown as an insert. In (c), the phase was a combination of the first 4 aberration modes in Fig. 3 with total rms amplitude of 3 radians. Background-free images were obtained by subtracting the highly aberrated image (c) from the images obtained before and after aberration correction, giving (d) and (e) respectively. (f), profiles obtained along the line drawn in (b) for images (a),(b),(d) and (e): aberration correction increases the intensity of the structures while background subtraction improves the contrast. The imaging depth was approximately $30\mu\text{m}$. The coverslip thickness was $170\mu\text{m}$.

demonstrated in Fig. 7.

8. Discussion and conclusion

We have shown that the image formation in SI microscopy is strongly dependent upon aberration modes that affect the imaging of the grid frequency. These grid modes cause a significant reduction in the intensity of the sectioned image and lower resolution. Conversely, non-grid modes have little effect on the final image intensity. It is important to note that both of these sets of modes would strongly affect the imaging quality in confocal and other sectioning microscopes. We are therefore led to the conclusion that the SI microscope, in comparison to other sectioning microscopes, is more susceptible to certain aberrations (grid modes) and more resilient to others (non-grid modes).

Whilst the SI microscope relies upon a relatively simple optical principle, the image formation process has a complex mathematical description. Similarly, the derivation of a model-based, sensorless, adaptive optical scheme is a complex process. However, our results show that the scheme is effective in correcting specimen-induced and system aberrations and restoring image quality. For the samples presented here, we found that the aberration mainly consisted of

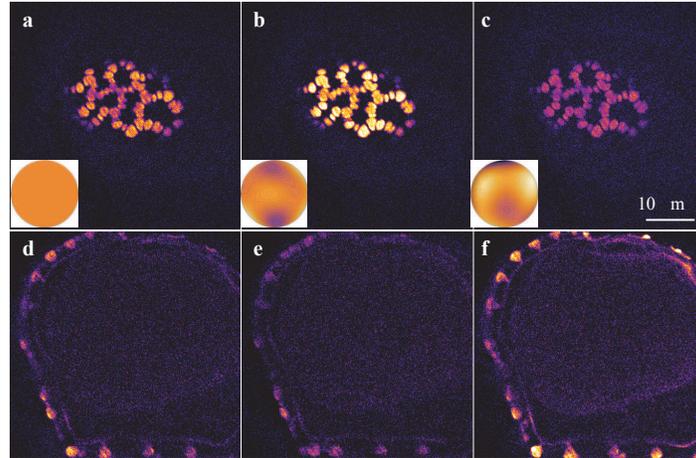


Fig. 7. Correction variation with imaging depth. Pollen grain images after background subtraction, at the top of the grain (a,b and c) and around the equator $20\mu\text{m}$ below (d,e and f). The images were acquired without aberration correction (a and d), with the correction optimised for the top of the grain (b and e), and for the equator (c and f). Images in the same row are displayed with the same color code and the phase induced by the DM is shown as an insert. The appropriate correction settings for one plane (b and f) clearly deteriorate the image quality in another plane (c and e). The rms phase after correction is 0.37 rad in (b) and 0.63 rad in (c). The imaging depth was approximately $30 - 50\mu\text{m}$. The coverslip thickness was $170\mu\text{m}$.

astigmatism, coma and spherical aberration modes. The magnitude of the astigmatism component was similar across various samples, suggesting that it arises from our optical system. The amplitude of coma and spherical aberration, however, varied significantly across different samples, or at various depth inside the same sample (see Fig. 7), indicating that these aberrations are induced by the specimens.

The adaptive scheme described here has significant advantages over model-free algorithms in that the aberration correction can be estimated using a small number of measurements ($2N + 1$ for N aberration modes). Moreover, as the scheme is mostly independent of the object structure, the appropriate modes have only to be determined once and the same scheme can be used for any specimen. We have also shown that aberration correction can be effectively combined with background subtraction to further improve SI microscope images. In the results presented here, aberration correction was performed as an average over an image frame and therefore would not correct for any local variations in aberrations. If these variations were found to be significant, the image could be formed from several sub-images for which independent aberration correction would be performed.

We have presented a general method that provides an optimal aberration expansion for a chosen optimisation metric. This relied upon the derivation of an inner product from a mathematical model of the imaging process, followed by an orthogonalisation process applied to a set of basis functions, such as the Zernike functions. This process reveals a wealth of information about the effects of different aberration modes on an imaging system – for the SI microscope, it enabled us to derive the sets of grid modes and non-grid modes. This method could equally be applied to any sectioning microscope to derive aberration expansions that are best suited to that application.

Acknowledgments

D. Débarre was supported by the Délégation Générale pour l'Armement and the Human Frontier Science Program. M. J. Booth was a Royal Academy of Engineering/EPSRC Research Fellow. This work was supported by a grant from the Royal Society.

Appendix A: Inner product for grid modes with a three-dimensional specimen

In order to take into account the axial extension of the sample, we introduce the parameter z as the axial distance between the considered plane and the focal plane. With this notation, the incoherent OTF can be generalized as:

$$H(\mathbf{m}, z) = \frac{1}{\pi} \iint_{D(\mathbf{g})} \exp[j\Delta\Phi(\mathbf{g}, \mathbf{r}) + j\Delta\Phi_d(\mathbf{g}, \mathbf{r}, z)] d\mathbf{r}, \quad (27)$$

with $\Delta\Phi_d(\mathbf{g}, \mathbf{r}, z) = az(|\mathbf{r} - \mathbf{g}|^2 - r^2)/2$ and $a = (8\pi n/\lambda) \sin^2(\alpha/2)$ and where $n \sin(\alpha)$ is the numerical aperture of the objective lens and λ is the wavelength in vacuum. Using this expression, the excitation pattern on the sample for each spatial phase shift ψ_i is given by:

$$I_{exc,i}(\mathbf{n}, z) = H(\mathbf{0}, z) + H(\mathbf{g}, z) \frac{e^{j\psi_i + j\mathbf{g}\cdot\mathbf{n}}}{2} + H(-\mathbf{g}, z) \frac{e^{-j\psi_i + j\mathbf{g}\cdot\mathbf{n}}}{2} \quad (28)$$

The image obtained on the camera is then obtained by integrating the intensity corresponding to different planes in the sample:

$$I_i(\mathbf{n}) = \int_{-\infty}^{+\infty} \left\{ H(\mathbf{0}, z) \text{FT}^{-1} [F(\mathbf{m}, z) H(\mathbf{m}, z)] + \frac{e^{j\psi_i}}{2} H(\mathbf{g}, z) \text{FT}^{-1} [F(\mathbf{m} - \mathbf{g}, z) H(\mathbf{m}, z)] \right. \\ \left. + \frac{e^{-j\psi_i}}{2} H(-\mathbf{g}, z) \text{FT}^{-1} [F(\mathbf{m} + \mathbf{g}, z) H(\mathbf{m}, z)] \right\} dz, \quad (29)$$

and accordingly the axially sectioned image can be expressed as:

$$I_{sect}(\mathbf{n}) = \frac{3}{2} \left| \int_{-\infty}^{+\infty} H(\mathbf{g}, z) \text{FT}^{-1} [F(\mathbf{m}, z) H(\mathbf{g} + \mathbf{m}, z)] dz \right|. \quad (30)$$

This can be simplified by assuming that the two-dimensional spatial frequency spectrum of the sample, $F(\mathbf{m}, z)$, varies slowly with z at the scale of the axial resolution of the sectioned image, such that we can write $F(\mathbf{m}, z) \approx F(\mathbf{m})$. Under this assumption, Parseval's theorem yields:

$$M = \iint_{-\infty}^{+\infty} I_{sect}^2(\mathbf{n}) d\mathbf{n} \approx \iint_S |F(\mathbf{m})|^2 \left| \int_{-\infty}^{+\infty} H(\mathbf{g}, z) H(\mathbf{g} + \mathbf{m}, z) dz \right|^2 d\mathbf{m}. \quad (31)$$

Here we restrict ourselves to grid modes, and hence consider that the effect of aberrations on the metric is dominated by the changes in the grid pattern intensity. In this case we can write $H(\mathbf{g} + \mathbf{m}, z) \approx H(\mathbf{g}, z)$ and:

$$M \approx F_0 \left| \int_{-\infty}^{+\infty} H(\mathbf{g}, z)^2 dz \right|^2. \quad (32)$$

For small aberration amplitudes, the optical transfer function can be approximated by a Taylor expansion, giving:

$$H(\mathbf{g}, z) = \frac{1}{\pi} \iint_{D(\mathbf{g})} \zeta(\mathbf{g}, \mathbf{r}, z) d\mathbf{r} + \frac{j}{\pi} \iint_{D(\mathbf{g})} \Delta\Phi(\mathbf{g}, \mathbf{r}) \zeta(\mathbf{g}, \mathbf{r}, z) d\mathbf{r} - \frac{1}{2\pi} \iint_{D(\mathbf{g})} \Delta\Phi(\mathbf{g}, \mathbf{r})^2 \zeta(\mathbf{g}, \mathbf{r}, z) d\mathbf{r} \quad (33)$$

where the defocus term ζ is defined as

$$\zeta(\mathbf{g}, \mathbf{r}, z) = \exp[j\Delta\Phi_d(\mathbf{g}, \mathbf{r}, z)] \quad (34)$$

From here on, for notational brevity, we omit the explicit dependence of $\Delta\Phi(\mathbf{g}, \mathbf{r})$ and $\zeta(\mathbf{g}, \mathbf{r}, z)$ on their arguments. The first term of Eq. (33) corresponds to the defocused OTF in the absence of aberration; this is denoted as H_0 in the following expressions. Introducing $H(\mathbf{g}, z)$ into Eq. (32) yields:

$$\begin{aligned} M \approx F_0 \left| \int_{-\infty}^{+\infty} H_0^2 dz \right|^2 - F_0 \left\{ \frac{2}{\pi^2} \int_{-\infty}^{+\infty} H_0^2 dz \int_{-\infty}^{+\infty} \left[\iint_{D(\mathbf{g})} \Delta\Phi \zeta d\mathbf{r} \right]^2 dz \right. \\ \left. + \frac{2}{\pi} \int_{-\infty}^{+\infty} H_0^2 dz \int_{-\infty}^{+\infty} H_0 \iint_{D(\mathbf{g})} \Delta\Phi^2 \zeta d\mathbf{r} dz - \frac{4}{\pi^2} \left[\int_{-\infty}^{+\infty} H_0^2 \iint_{D(\mathbf{g})} \Delta\Phi \zeta d\mathbf{r} dz \right]^2 \right\}, \end{aligned} \quad (35)$$

As a result, the inner product derived from the metric is defined here as:

$$\begin{aligned} \langle X_i, X_j \rangle_{grid\ z} = F_0 \left\{ \frac{2}{\pi^2} \int_{-\infty}^{+\infty} H_0^2 dz \int_{-\infty}^{+\infty} \left[\iint_{D(\mathbf{g})} \Delta X_i \zeta d\mathbf{r} \right] \left[\iint_{D(\mathbf{g})} \Delta X_j \zeta d\mathbf{r} \right] dz \right. \\ \left. + \frac{2}{\pi} \int_{-\infty}^{+\infty} H_0^2 dz \int_{-\infty}^{+\infty} H_0 \iint_{D(\mathbf{g})} \Delta X_i \Delta X_j \zeta d\mathbf{r} dz \right. \\ \left. - \frac{4}{\pi^2} \left[\int_{-\infty}^{+\infty} H_0^2 \iint_{D(\mathbf{g})} \Delta X_i \zeta d\mathbf{r} dz \right] \left[\int_{-\infty}^{+\infty} H_0^2 \iint_{D(\mathbf{g})} \Delta X_j \zeta d\mathbf{r} dz \right] \right\} \end{aligned} \quad (36)$$

Appendix B: Inner product for non-grid modes

Here we consider non-grid modes, which do not affect the grid pattern projected on the sample. For those aberrations, the sectioning strength is independent of the aberration amplitude, and hence we only need consider the plane $z = 0$ in the following analysis. Starting with:

$$M = \iint_S |F(\mathbf{m})|^2 |H(\mathbf{g})H(\mathbf{g} + \mathbf{m})|^2 d\mathbf{m}, \quad (37)$$

we assume $|F(\mathbf{m})|^2$ takes significant values only for small \mathbf{m} and expand $H(\mathbf{g} + \mathbf{m})$ as a Taylor series up to the second order term [9]:

$$M = \iint_S |F(\mathbf{m})|^2 |H(\mathbf{g})|^2 \left(|H(\mathbf{g})|^2 + (\mathbf{m} \cdot \nabla) |H(\mathbf{g})|^2 + \frac{1}{2} [(\mathbf{m} \cdot \nabla)^2 |H(\mathbf{g})|^2] \right) d\mathbf{m} \quad (38)$$

where we take $(\mathbf{m} \cdot \nabla)t(\mathbf{g})$ to mean $[(\mathbf{m} \cdot \nabla_{\mathbf{m}'})t(\mathbf{m}')]_{\mathbf{m}'=\mathbf{g}}$. If we further assume that the object is real, then $F(-\mathbf{m}) = F^*(\mathbf{m})$ and hence:

$$\begin{aligned} \iint_S |F(\mathbf{m})|^2 \mathbf{m} d\mathbf{m} &= \mathbf{0} \\ \iint_S |F(\mathbf{m})|^2 m_x m_y d\mathbf{m} &= 0 \end{aligned} \quad (39)$$

We also assume that the object structure (and hence frequencies) are not predominantly aligned with one direction, so that $\iint_S |F(\mathbf{m})|^2 m_x^2 d\mathbf{m} \approx \iint_S |F(\mathbf{m})|^2 m_y^2 d\mathbf{m}$, and we obtain:

$$M = F_0 |H(\mathbf{g})|^4 + F_1 |H(\mathbf{g})|^2 \nabla^2 (|H(\mathbf{g})|^2), \quad (40)$$

where ∇^2 is the Laplacian operator and:

$$F_1 = \frac{1}{2} \iint_S |F(\mathbf{m})|^2 m^2 d\mathbf{m} . \quad (41)$$

Using Eq. (18), we find that for small aberration difference amplitudes

$$|H(\mathbf{g} + \mathbf{m})|^2 \approx |H_0(\mathbf{g} + \mathbf{m})|^2 - \frac{|H_0(\mathbf{g} + \mathbf{m})|}{\pi} \iint_{D(\mathbf{g} + \mathbf{m})} \Delta\Phi(\mathbf{g} + \mathbf{m}, \mathbf{r})^2 d\mathbf{r} + \left[\frac{1}{\pi} \iint_{D(\mathbf{g} + \mathbf{m})} \Delta\Phi(\mathbf{g} + \mathbf{m}, \mathbf{r}) d\mathbf{r} \right]^2 \quad (42)$$

Under the assumption that \mathbf{m} is small, we can write:

$$\begin{aligned} \Delta\Phi(\mathbf{g} + \mathbf{m}, \mathbf{r}) &= \Delta\Phi(\mathbf{g}, \mathbf{r}) + \Phi(\mathbf{r} - \mathbf{g}) - \Phi(\mathbf{r} - \mathbf{g} - \mathbf{m}) \\ &\approx \Delta\Phi(\mathbf{g}, \mathbf{r}) + (\mathbf{m} \cdot \nabla) \Phi(\mathbf{r} - \mathbf{g}) . \end{aligned} \quad (43)$$

Since we consider only non-grid modes, for all \mathbf{r} within $D(\mathbf{g})$, $\Delta\Phi(\mathbf{g}, \mathbf{r}) = 0$. If we consider only slowly varying phase functions and for small values of \mathbf{m} , this is also valid for all \mathbf{r} within $D(\mathbf{g} + \mathbf{m})$, and therefore:

$$\Delta\Phi(\mathbf{g} + \mathbf{m}, \mathbf{r}) \approx (\mathbf{m} \cdot \nabla) \Phi(\mathbf{r} - \mathbf{g}) . \quad (44)$$

Equation 42 can be simplified to:

$$|H(\mathbf{g} + \mathbf{m})|^2 \approx |H_0(\mathbf{g} + \mathbf{m})|^2 - \frac{|H_0(\mathbf{g})|}{\pi} \iint_{D(\mathbf{g})} [(\mathbf{m} \cdot \nabla) \Phi(\mathbf{r} - \mathbf{g})]^2 d\mathbf{r} + \left[\frac{1}{\pi} \iint_{D(\mathbf{g})} (\mathbf{m} \cdot \nabla) \Phi(\mathbf{r} - \mathbf{g}) d\mathbf{r} \right]^2 \quad (45)$$

Combining this with Eq. (40) and noting that for the non-grid modes considered here, $H(\mathbf{g}) = H_0(\mathbf{g})$, we obtain:

$$M = M_0 - 2F_1 \left\{ \frac{|H_0(\mathbf{g})|}{\pi} \iint_{D(\mathbf{g})} [\nabla\Phi(\mathbf{r} - \mathbf{g})]^2 d\mathbf{r} - \left[\frac{1}{\pi} \iint_{D(\mathbf{g})} \nabla\Phi(\mathbf{r} - \mathbf{g}) d\mathbf{r} \right]^2 \right\} , \quad (46)$$

with:

$$M_0 = F_0 |H_0(\mathbf{g})|^4 + F_1 |H_0(\mathbf{g})|^2 \nabla^2 |H_0(\mathbf{g})|^2 \quad (47)$$

As a result, the inner product for non-grid modes is defined here as:

$$\begin{aligned} \langle X_i, X_j \rangle_{non-grid} &= 2F_1 \left\{ \frac{|H_0(\mathbf{g})|}{\pi} \iint_{D(\mathbf{g})} \nabla X_i(\mathbf{r} - \mathbf{g}) \nabla X_j(\mathbf{r} - \mathbf{g}) d\mathbf{r} \right. \\ &\quad \left. - \frac{1}{\pi^2} \iint_{D(\mathbf{g})} \nabla X_i(\mathbf{r} - \mathbf{g}) d\mathbf{r} \iint_{D(\mathbf{g})} \nabla X_j(\mathbf{r} - \mathbf{g}) d\mathbf{r} \right\} \end{aligned} \quad (48)$$

It should be noted that this inner product also has some degeneracy, as small regions of the pupil are not covered by $D(\mathbf{g})$. If further correction were needed, higher order terms from Eq. (15) would have to be considered.