

Modelling mid-spatial frequency errors in space instruments: recent results

Jérôme Caron^{*a}, Tiberiu Ceccotti^b, Stefan Bäumer^b, Jérôme Loicq^c, Wilbert Ijzerman^d

^aAirbus Defence and Space GmbH, Taufkirchen, Germany; ^bTNO, Stieltjesweg 1, 2628 CK Delft, The Netherlands, ^cDepartment of Space Engineering, Delft University of Technology, ^dDepartment of Mathematics and Computer Science, Eindhoven University of Technology.

ABSTRACT

Freeform optics are increasingly important for space instruments. They are manufactured with machining or polishing techniques that make use of sub-aperture tools and leave characteristic surface errors in the mid-spatial frequency (MSF) regime. These errors are notoriously difficult to model and specify, often leading to extra costs in space projects that result from increased analysis efforts and/or over-specification. After more than two decades of research, and despite some significant progress, there is still no clearly established methodology for the modelling and specification of MSF surface errors. This is apparent both from the diversity of the methods used in different projects, and from the regularly presented or published works that do attempt to propose or define such methodology. To support the efforts towards a unified modelling of MSFs, this paper aims to present two new promising results. First, we introduce the work of K. Liang and starting from it, we derive a practical criterion for the applicability of geometrical optics for modelling MSFs. Second, we introduce the Grating Mode Series Expansion (GMSE) a recent, computationally efficient and intuitive method for diffractive MSFs propagation across all spatial frequency regimes.

Keywords: Mid-spatial frequencies, freeform, modelling, diffraction

1. INTRODUCTION

Freeform optics refer to optical surfaces with no rotational symmetry and which are also not an off-axis portion of a rotationally symmetric parent surface. They open new possibilities for optical design, and allow to achieve new characteristics, higher performance and/or smaller volume. With the Alvarez lens and the viewfinder of the SX-70 Polaroid camera in the 1970s, freeform optics first emerged from the need of compact viewfinders and head-worn displays [1]. In the 1990s the first applications of computer-controlled ultra-precision machining for freeform, and the increase of computational power required for design optimization, laid the base for a more systematic exploration and development of freeform optics. It is only later that they have been used in space instruments. In fact, as described by R. Geyl [2] some techniques developed since the 1980s to fabricate off-axis portions of standard surfaces for space projects, could have been already used to make freeform optics. In the European space ecosystem, some notable first use of freeform optics include the Tropomi telescope (strong freeform, launched 2017) and MicroCarb (freeform spectrometer, launched 2025). Today with the high maturity of manufacturing techniques on one hand, and on the other hand the needs of small compact instruments for NewSpace as well as the increasing performance requirements of institutional space instrument programs, the use of freeform optics is becoming a standard approach. Limitations still do exist for strong freeforms, but most optical shapes with an arbitrary weak freeform component can be now realized with state-of-the-art optical quality.

Freeform surfaces are typically manufactured with computer controlled sub-aperture tools, where the dimension of the machining or polishing tool is smaller, sometimes by a large factor, than the dimension of the optical part. Examples of these techniques are single point diamond turning (SPDT), magneto-rheological finishing (MRF) or ion beam figuring (IBF). Due to their smaller size, the tools leave specific footprints on the optical surface and a complex optimized pattern is required for fabrication [3] to minimize residual errors. The residual surface errors typically present strong signatures in the mid-spatial frequency (MSF) regime. They are notoriously difficult to model and specify and have captured a lot of attention and efforts [4-7]. There is still today no unique established methodology for the modelling of MSF surface errors, as apparent from the variations of the techniques used between different projects and regular attempts to establish a methodology [8-10]. For freeform specification, the ISO 10110 norm provides various tools but also does not prescribe a

*jerome.caron@airbus.com

clear path: MSF errors may be specified in a number of ways, such as local slopes (peak to peak or RMS), power spectral density curve (PSD) or band limited roughness [11].

This paper focuses on the modelling aspects and aims to present some promising new results. In section 2 we investigate the validity of geometrical optics for the propagation of MSFs. We introduce the work of K. Liang [7] and then propose a criterion based on the Fresnel number to describe the boundary between geometrical optics and diffraction regimes. In section 3 we introduce the Grating Mode Series Expansion (GMSE) developed by T. Ceccotti [12], a new, computationally efficient and intuitive method to model the propagation of MSF that is valid for all spatial frequency regimes.

2. LIMITS OF MSFS PROPAGATION WITH GEOMETRICAL OPTICS

2.1 Opening argument: limits of ray-tracing models for MSFs

The first question we investigate is: Why not simply using ray-tracing to model the impact of MSFs on optical performance? Importing the error surfaces in a ray-tracing program and computing spot diagrams is the first and most natural thing to try. If the distribution of diffracted light could be well approximated with a geometrical optics based deflection model, the spot diagram degradation could also be calculated analytically as a function of the system F-number and marginal ray height, in a similar way as straylight BSDFs are calculated by Peterson [13]. This type of approach has been proposed by Dominguez [14] or Rogers [15] and makes it possible to derive error budgets for the maximum allowed wavefront slopes.

To understand when ray-tracing works, let's consider a wavefront with a sinusoidal modulation of magnitude ϕ_0 in radians and spatial period L . We expand it in a Jacobi-Anger series

$$e^{i\phi(x)} = e^{i\phi_0 \sin(\frac{2\pi}{L}x)} = \sum_{m=-\infty}^{+\infty} J_m(\phi_0) e^{im\frac{2\pi}{L}x} \quad (1)$$

where $\phi_0 = \frac{2\pi}{\lambda} h_0$ with h_0 being the wavefront modulation amplitude and $J_m(x)$ are Bessel functions of the first kind. In this expression, the exponentials $\exp\left(im\frac{2\pi}{L}x\right)$ correspond to the grating's diffraction orders: after a Fourier transform to obtain the far-field diffraction intensity angular distribution they become $\delta\left(\alpha - m\frac{\lambda}{L}\right)$ with α direction cosine of the diffracted ray in the x-direction. The coefficients $J_m(\phi_0)$ are the amplitude diffraction efficiencies. On the other hand, we note that the maximum geometrical slopes of the wavefront are equal to

$$S = \pm \frac{2\pi}{L} h_0 = \pm \phi_0 \frac{\lambda}{L} \quad (2)$$

On figure 1 we plot the diffraction efficiencies as a function of α (horizontal axis, equivalent to modulation slope) and ϕ_0 (vertical axis, modulation amplitude). We represent the coefficients $|J_m(\phi_0)|^2$ normalized such that the largest coefficient (excluding $J_0(\phi_0)$ set to zero) is equal to 1. We immediately see that the diffraction efficiencies quickly drop to zero outside of the cone defined by the modulation maximum geometrical slopes. This behavior, which was the starting observation of Rogers [15] comes from an asymptotic property of the Bessel functions: when $|m| > \phi_0$, $J_m(\phi_0)$ decreases very fast with increasing m (it decreases as $(\phi_0/m)^m$ which is faster than an exponential). On the other hand, in the limit of small errors $\phi_0 \ll 1$ a different behavior is observed, due to the fact that the scattered light is then entirely contained in the first grating orders $m=\pm 1$. This can be seen via a Taylor expansion of the exponential:

$$e^{i\phi(x)} \approx J_0(\phi_0) + J_{+1}(\phi_0) e^{+i\frac{2\pi}{L}x} + J_{-1}(\phi_0) e^{-i\frac{2\pi}{L}x} \quad (3)$$

$$e^{i\phi(x)} \approx 1 - \frac{\phi_0^2}{4} + \frac{\phi_0}{2} \left(e^{+i\frac{2\pi}{L}x} - e^{-i\frac{2\pi}{L}x} \right) \quad (4)$$

where the first term with $m=0$ is the specular and the two terms $J_{+1}(\phi_0)$ and $J_{-1}(\phi_0)$ correspond to scattered light. In this case, as clearly visible at the bottom of figure 1 the two first orders go significantly outside of the geometrical cone. This observation was already done by Rogers [15] and is consistent with the Rayleigh-Rice theory [16], which is valid in the limit of very small surface error (it is a perturbation model) and says the scattered light is proportional to the PSD and is sent in the first grating orders. For the Rayleigh-Rice theory to be valid we must have a small WFE: $\phi_0 = \frac{2\pi}{\lambda} h_0 \ll 1$. This condition is equivalent to $\frac{\lambda}{L} \gg \frac{2\pi}{L} h_0$ where L is the grating period. In this inequality, on the left is the angle where the light

is scattered in the orders ± 1 , and on the right the maximum surface slope. In other terms, it says that the light diffracted by the sinusoidal WFE is deviated well outside of the cone defined by the geometrical optics (GO) slope! In other terms, having small wavefront errors (first inequality) is equivalent to have light scattered outside of the GO cone (second inequality).

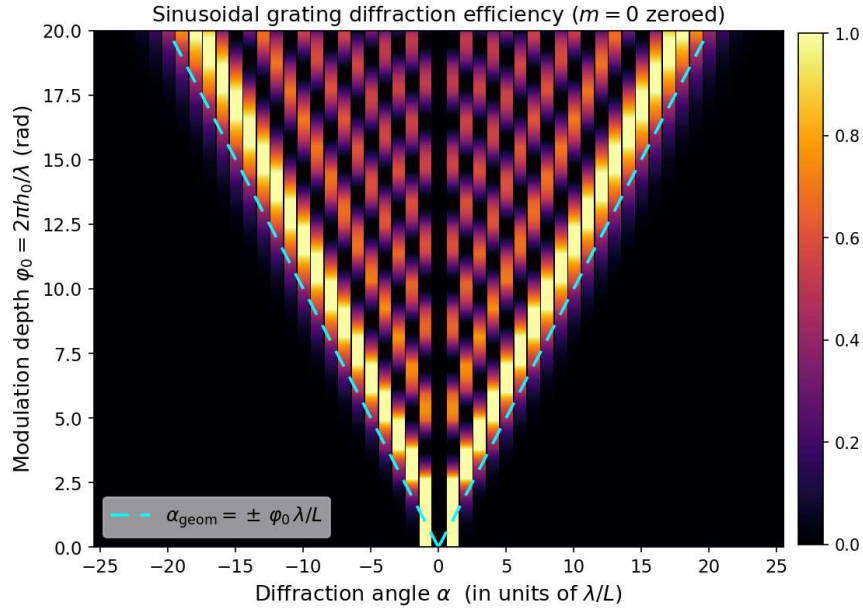


Figure 1. Diffraction efficiencies $|J_m(\phi_0)|^2$ of a sinusoidal grating as a function of diffraction angle (horizontal) and modulation depth (vertical). The 0th order that corresponds to specular light was removed. Normalization: for a given ϕ_0 (horizontal line) the maximum efficiency is set to 1. The two dotted lines correspond to the geometrical optics cone predicted from maximum surface slopes.

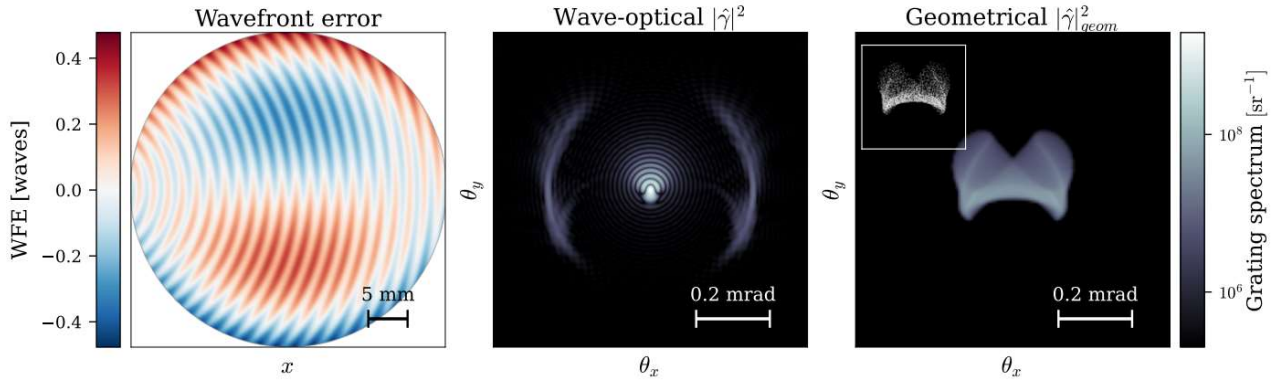


Figure 2. Left: artificial wavefront error map with a sinusoidal ripple with amplitude 0.08 waves, superimposed to a coma term. Center: far field intensity distribution predicted from a diffraction model. Right: far field intensity distribution predicted from geometrical optics. The diffraction model predicts a PSF that is roughly 2 times broader than the GO model.

This observation established for MSFs can easily be extended to low frequency errors (form errors, as well as aberrations, both described by low order Zernike). Let us consider a sinusoidal modulation with weak amplitude and large period. When the period becomes comparable or larger to the system's aperture, the separation of the diffracted orders $m=\pm 1$ becomes smaller than the diffraction PSF. In that case the ray tracing spot diagrams are not only smaller than the $m=\pm 1$ orders separation but also smaller than the diffraction PSF, and again the distribution of diffracted light is wider than what

geometrical optics predicts. As proven by Miyamoto [17] in this regime the total PSF is well approximated by a convolution between the geometrical spot diagram and the diffraction PSF.

Our main conclusion is the fact that the validity of GO for ray-tracing to an image plane (which is by definition in the far-field) does not depend on the spatial frequencies considered, but on the amplitude of the wavefront error. When $\phi_0 = \frac{2\pi}{\lambda} h_0 \gg 1$ the GO limits defined by the maximum geometrical slopes may be used. When $\phi_0 = \frac{2\pi}{\lambda} h_0 \ll 1$ light is diffracted in the first grating orders and outside of the GO cone. The transition between the two regimes occurs when $\phi_0 = 1$, i.e. when the two first orders ± 1 exactly coincide with the maximum geometrical slope. This corresponds to a sinusoidal modulation amplitude of $\frac{1}{2\pi} \approx 0.16$ waves. To illustrate the limits of ray-tracing we simulated on figure 2, an artificial WFE with sinusoidal ripples with an amplitude of half this limit (0.08 waves) superimposed to a low order aberration term (coma). As expected, the scattered light predicted with an exact diffraction model (center) generates an intensity distribution that is roughly 2 times wider than the strongly inaccurate prediction from GO (right).

This first discussion shows the limit of pure GO deflection models for MSFs propagation. We believe there is no simple way to use such models, as when dealing with a complex surface error, some frequency components will diffract within the GO limits and others not. In the next paragraphs we will consider alternative ways to predict the impact of MSFs.

2.2 Definitions

Standard approach for modelling MSFs

We now define the configuration that we will analyze in this paper, and that we will use as standard configuration for modelling MSFs. Our goal is to predict the impact of MSFs on the performance of an optical system. We consider one such system, which contains one optical surface that introduces a wavefront error (WFE). The problem we will try to solve, is how to propagate this WFE to the exit pupil of the system so that its impact on performance can be analyzed in a second step with standard tools, such as the ones used for PSF or MTF calculation.

The exit pupil is the image of the physical pupil through all the optics, i.e. in the image space. A special case consists of systems with anamorphism, that might not have a well-defined exit pupil: for instance, the exit pupil being at two different locations depending on the considered transverse coordinates, X or Y. In other cases, the exit pupil may be strongly aberrated. In these situations, one possibility is to back propagate all WFEs to a non astigmatic sharply defined pupil, e.g. the physical pupil or entrance pupil where they can be added. Then a final modelling step will be used to propagate the beam to the aimed destination, e.g. the image plane. Alternatively, it is also possible to propagate the WFEs (and in this case, also the amplitude diffracted from the pupil edges) to an arbitrary reference sphere centered on the image plane.

Wavefront error changing during free-space propagation

The spatial map of a wavefront error (WFE), that describes for example the optical aberrations in an optical system, is often assumed to be invariant upon propagation: it stays the same everywhere along an optical beam, apart from linear scaling and perhaps from distortions that may affect the spatial distribution of rays, but do not change the WFE values. It is important to realize that this is an approximation that is only valid for low spatial frequencies and small aberration magnitudes. A WFE map with mid- or high- spatial frequencies will in general be modified during free-space propagation due to diffraction, which causes light to deviate from the geometrical optics directions. As a rule of thumb, for short propagation distances geometrical optics works; for very large propagation distances diffraction is required. In some cases, a WFE with significant slopes and/or after large propagation distances will generate caustics, where a wavefront definition that follows the geometrical optics rays becomes non-trivial and where diffraction dominates. Close to an image plane diffraction modelling is also required, as it is known that geometrical optics fails to reproduce the diffracted intensity close to a focus.

In which situations geometrical optics is sufficient or diffraction is required, will now be investigated. It depends on the properties of the WFE (which spatial frequency and amplitude of deformation) as well as on the wavelength and propagation distance.

What propagation distance?

In this paper we will use the concept of “propagation distance”. Let’s briefly clarify what we mean with it. Depending on the context it can take two meanings:

- It can simply refer to the distance used for free-space propagation.

- In some cases we must consider the propagation of an optical surface through (part of) an optical system, i.e. propagating in one step through several optical surfaces. How can we then compute an equivalent propagation distance? To answer this question we first observe a property of paraxial diffraction integrals (such as the Collins integral [18]): when propagating a wavefront from a position M to a position M' in an optical system, if positions M and M' are image of each other (conjugated), even though the WFE changes during propagation between M and M', the final WFE obtained in position M' is identical to the original wavefront at position M apart from a lateral magnification factor and a trivial phase shift (this phase shift is the sum of a piston term resulting from the propagation distance, plus a phase that is quadratic with beam coordinates). In other terms it is possible to measure the propagation distance of an optical surface M to the exit pupil P of a system, by simply measuring this distance from the image of M (called M') in the final image space to the exit pupil P. The "propagation distance" will simply refer to the distance between M' and P where free space propagation is used. Finally, one must take care of the magnification between the two positions at M and M' which scales the spatial frequencies.

What kind of geometrical optics?

When propagating a WFE through an optical system with geometrical optics towards the exit pupil, various degrees of approximation may be made:

1. First-order approach (copy-paste model): The most simple model consists in copy-pasting the WFE map generated by an optical surface, onto the exit pupil with a scaling factor equal to the paraxial magnification of the beam dimensions at the two locations. This magnification may be estimated e.g. with the ratio of the beam footprint diameters.
2. Exact footprint mapping (perturbation model): In some systems the paraxial approach is too crude, for instance in freeform optics where the shape of the beam footprint is significantly distorted, or simply due to tilted mirrors that elongate the footprint in one direction. A more accurate approach is to map the exact (x,y) positions of the relevant optical surface onto the exit pupil (X,Y) coordinates and take into account the wavefront spatial distortion. The mapping is then used to transfer the wavefront values from (x,y) to (X,Y) coordinates. Additionally, mapping these two sets of coordinates with the entrance pupil coordinates will allow to get the correct intensity distribution at the exit pupil.
3. Beyond the Perturbation Model: in the above "perturbation model" the optical properties (such as the geometrical mapping between beam footprints at different locations) are estimated in the absence of wavefront error (WFE). The WFE map is then propagated along the nominal rays, that are unaffected by the WFE. In reality the WFE map will follow the WFE-aberrated rays and the phase will be redistributed spatially at the exit pupil. Additionally, focusing from the WFE local curvatures will cause intensity variations at the exit pupil that are not present in the nominal system.

2.3 Limits of the Perturbation Model: work of K. Liang

As discussed above different levels of approximation are possible that may be considered as geometrical optics. The question of the validity of the perturbation model, was investigated by Liang in his PhD work [7] where criteria were established to quantify the two geometrical optics effects ignored in the perturbation model, described in the previous paragraph: aberrated ray tracing and local focusing. In our view the work of Liang is highly relevant unfortunately it is also technical and the criteria derived have not been explicitly connected to these two effects. We will now provide a brief overview of the relevant part of Liang's work and introduce the criteria. In section 2.3 we will keep the same notations as Liang to ease the connection with the PhD [7] or the two papers [19, 20] where more information can be found.

The approach taken by Liang consists in propagating rigorously a scalar (i.e. non-polarized) wave through free space from the initial coordinate $z=z_M$ where the image of a lens with surface errors is located, to the final coordinate $z=z_P$ of the exit pupil. The propagation is done by solving the Helmholtz equation. To do so the phase of the wave is expanded as a series in power of $1/(ik)^N$ where k is the wavevector $= 2\pi/\lambda$ (this method is also known as the WKB method). Separate equations are obtained for each order and are then solved analytically for the lowest orders. Orders with $N=0, 1, 2$ correspond to geometrical optics (limit $\lambda \rightarrow 0$) and orders $N \geq 3$ correspond to diffraction effects. The initial field is defined at $z=z_M$ with $N=0$ corresponding to the nominal wave that is converging towards the image plane, $N=1$ corresponding to the WFE to be

propagated and the initial field is zero for all other orders $N \geq 2$. After propagation, new terms appear at all orders and it is possible to track the effect of each field with various degrees of approximation. The field with $N=1$ at the initial coordinate $z=z_M$ is assumed to have a sinusoidal phase error (in radians) equal to:

$$\phi(\xi) = h \sin(2\pi\kappa\xi) \quad (5)$$

where ξ is the spatial coordinate, κ the spatial frequency and h the amplitude of the sinusoidal perturbation in radians. Liang then established some metrics to estimate the error done on the propagated WFE map when propagating the field along the nominal rays, i.e. under the assumptions of the perturbation model. He defines a relative RMS error (RMSE) ϵ which represents the relative error on the propagated field. The impact of the sinusoidal phase modulation is quantified with φ . Both ϵ and φ are obtained after averaging across the beam footprint with an integration noted $\langle \cdot \rangle_A$. Since the relative RMS error ϵ grows with φ the relevant quantity to describe the accuracy of the propagation is the normalized RMSE (NRMSE) equal to $\frac{\epsilon}{\varphi}$. The propagation of the sinusoidal error along the beam is considered accurate when NRMSE $< 1/3 \approx 33\%$. By combining equations (27) and (28) of reference [19] and neglecting the obliquity factor we get:

$$\epsilon = \left| \frac{(z_P - z_M)z_M}{z_P} \right| \frac{\lambda}{4\pi} \sqrt{\langle \phi''^2 \rangle_A + \langle \phi'^4 \rangle_A} \quad (6)$$

We also have

$$\varphi = \sqrt{\langle \phi^2 \rangle_A} \quad (7)$$

Combining everything and using the following results

$$\langle \phi^2 \rangle_A = \frac{h^2}{2} \quad \langle \phi''^2 \rangle_A = 8\pi^4 \kappa^4 h^2 \quad \langle \phi'^4 \rangle_A = 6\pi^4 \kappa^4 h^4 \quad (8-10)$$

We can compute the NRMSE

$$NRMSE = \frac{\epsilon}{\varphi} = \left| \frac{(z_P - z_M)z_M}{z_P} \right| \lambda \pi \kappa^2 \sqrt{1 + \frac{3}{4}h^2} < \frac{1}{3} \quad (11)$$

Both terms proportional to $\langle \phi''^2 \rangle_A$ and to $\langle \phi'^4 \rangle_A$ are geometrical optics effects that are ignored in the perturbation model

- The term proportional to $\sqrt{\langle \phi''^2 \rangle_A}$ gives a constant criterion (1 in the square root). It is proportional to the curvature of the wavefront ϕ'' and generates an error in the intensity distribution at the exit pupil (as can be checked in [19]). We interpret it as a term that describes the self-focusing of the wavefront generated by the sinusoidal phase, that will result in a spatial redistribution of the intensity.
- The term proportional to $\sqrt{\langle \phi'^4 \rangle_A}$ gives a criterion that depends on the error amplitude ($3h^2/4$ in the square root). It is proportional to the squared slope of the wavefront ϕ'^2 and generates [19] a phase error at the exit pupil [19]. We interpret it as a term that describes the spatial redistribution of phase due to the propagation along the aberrated, rather than nominal rays. The slope is squared as it occurs two times in the mechanism that creates the error: first the wavefront slope deflects the aberrated rays and shifts laterally the wavefront, and second the lateral shift creates a change of phase that is proportional to the wavefront slope (i.e. if a uniform wavefront is shifted there is no effect).

As we will see in a later paragraph, the Fresnel number N_F appears in the NRMSE equation. In Liang's notations it is equal to:

$$N_F = \frac{1}{\lambda \kappa^2} \left| \frac{z_P}{(z_P - z_M)z_M} \right| \quad (12)$$

giving

$$NRMSE = \frac{\epsilon}{\varphi} = \frac{\pi}{N_F} \sqrt{1 + \frac{3}{4}h^2} < \frac{1}{3} \quad (13)$$

Therefore Liang's criterion may be rewritten:

$$N_F > 3\pi \sqrt{1 + \frac{3}{4}h^2} \quad (14)$$

It is possible to represent this criterion on a log-log scale plot showing the surface form error (SFE) vs spatial frequency, as follows. We first invert the NRMSE equation (13):

$$h < \frac{2}{\sqrt{3}} \sqrt{\left(\frac{N_F}{3\pi}\right)^2 - 1} \quad (15)$$

and we note that h represents the error on the phase (in radians) so the surface form error (in meters) is $SFE = h \frac{\lambda}{2\pi|n-1|}$ where normal incidence is assumed and n is the refractive index of the optical material that has the surface error ($n = -1$ for reflection). Such log-log plot was calculated for an example optical system in section 2.4 (figure 5). For GO to be valid the spatial frequencies must stay on the left of the boundaries represented on the graph.

We conclude this section noting that all the criteria introduced so far are still only related to geometrical optics. The magnitude of diffraction effects is still unaddressed at this stage, and will now be studied: first in section 2.6 with the introduction of a heuristic criterion based on the Fresnel number, and then with a rigorous proof of the validity of the Fresnel number criterion in section 2.5.

2.4 Transition from geometrical optics to diffraction: Fresnel number

Validity of geometrical optics

To describe the propagation of a WFE purely with geometrical optics (including all effects described above) it is necessary to make sure that diffraction effects remain negligible. In practice the smallest features will diffract first, so they are the most restrictive. For a complex surface topography with a broadband PSD spectrum it is not obvious to define the size of the features. In what follows, for simplicity we will use L_{min} , obtained with Fourier decomposition and equal to the smallest spatial period of the surface error profile. Geometrical optics (GO) is an asymptotic approximation of wave propagation for the limit of small wavelengths. The first requirement for its applicability is: the wavelength must be much smaller than all relevant scales in the considered problem. Unfortunately this is only a necessary condition which does not allow to identify in which cases geometrical optics-based propagation is valid. Kravtsov [21] has established 2 sufficient conditions for the use of geometrical optics:

- The wavefront shall not have any caustic or focus. For us, it means the starting and ending locations of the propagation shall not coincide with an image plane;
- The wavefront complex amplitude, the wavefront slope (phase gradient) and the properties of the medium where the wave propagate, shall not vary appreciably over the dimension of the Fresnel volume $\sim \sqrt{\lambda z}$ where λ the wavelength and z is the considered propagation distance. If we consider a pure phase error at a specific spatial period L_{min} and amplitude h_0 , the slope is $h_0 2\pi / L_{min}$ and the variation of slope is $h_0 (2\pi / L_{min})^2$. The requirement becomes

$$\nabla slope \cdot \sqrt{\lambda z} \ll slope \quad \frac{L_{min}^2}{\lambda z} \gg 4\pi^2 \quad (16-17)$$

In this formula we recognize the Fresnel number which will be discussed next.

Fresnel number

We now introduce the concept of Fresnel number [21, 22]. Let's consider a small circular aperture with radius $OA = a$, and a point P at a distance $OP = z$ from the aperture. See figure 3 below.

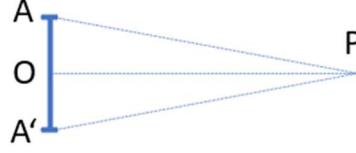


Figure 3. Geometry for the derivation of the Fresnel number.

To understand diffraction at P, we divide the aperture into concentric Fresnel zones, such that the path from the aperture to P changes by $\lambda/2$ between successive zones. The path difference between the center of the aperture and a point at the edge of the aperture is approximately

$$(AP) - (OP) = \sqrt{z^2 + a^2} - z \approx \frac{a^2}{2z} \quad (18)$$

We can now calculate the total number of Fresnel zones in the aperture. It is equal to the Fresnel number:

$$N_F = \frac{2}{\lambda} \frac{a^2}{2z} = \frac{a^2}{\lambda z} \quad (19)$$

When $N_F \gg 1$ many Fresnel zones contribute to the propagation in P and diffraction effects are weak. This can be intuitively understood with assuming that each point on the wavefront emits spherical waves (according to the Huygens-Fresnel principle) and then using the stationary phase approximation: the many different paths cancel each other and only the central zone that creates a stationary phase remains. Light then travels along the corresponding path which is equivalent to a geometrical optics ray. When $N_F \ll 1$ we have only one Fresnel zone, we are in the regime of strong (far-field) diffraction. This demonstration considers diffraction from a small aperture that is flat. In practice we often must consider a converging (or diverging) wavefront. A similar calculation may be used to establish the formula for the number of Fresnel zones at the point P when the aperture is curved. The final expression is:

$$N_F = \frac{a^2}{\lambda} \left| \frac{1}{z} \left(1 - \frac{z}{R} \right) \right| \quad (20)$$

where z is the algebraic value of the propagation distance, R is the algebraic radius of curvature at the aperture (positive for a converging wave, and negative for a diverging wave) and the absolute value of $(1/z)(1 - z/R)$ is taken. We recover $F=0$ for a wave converging to P ($z=R$), which is consistent with the fact that the diffraction pattern in a focal plane is calculated with a far-field diffraction integral (Fraunhofer) and the fact that geometrical optics is no longer valid when we are approaching a focus or a caustic [21]. As we have seen this expression for the Fresnel number in a converging beam is consistent with the results of Liang [7] and it is also implemented in the CodeV software for the beam propagation (BPR) as a criterion to select between different propagators.

Practical criterion for the validity of geometrical optics

To make use of the Fresnel number, we now assume that the diffracting element has a radius equal to the period of the smallest features in the considered surface error map: $a = L_{min}$. The Fresnel number becomes

$$N_F = \frac{L_{min}^2}{\lambda} \left| \frac{1}{z} \left(1 - \frac{z}{R} \right) \right| \quad (21)$$

For a flat wavefront the expression becomes $N_F = \frac{L_{min}^2}{\lambda z}$. The choice of $a=L_{min}$ may seem arbitrary (why not $L_{min}/2$ or another factor?) but it also depends on the choice of the minimum allowed value for N_F . The requirement $N_F \gg 4\pi^2$ derived previously is very stringent and will ensure that geometrical optics can fully and accurately describe the considered problem. For the practical evaluation of an error, such accuracy is not needed and lower values of N_F are satisfactory. As we have seen the work of Liang [7] established that an error smaller than $\sim 33\%$ in the propagation of the WFE corresponds roughly to $N_F > 3\pi$. We will consider a comparable value as Liang and choose the following criterion for the validity of geometrical optics: $N_F > 10$. The smallest spatial period of the surface error must be larger than:

$$L_{min} > \sqrt{10\lambda \left| \frac{zR}{R-z} \right|} \quad (22)$$

Reciprocity

We now define two positions 1 and 2, respectively at distances z_1 and z_2 from the image plane. We will show that when propagating from position 1 to position 2, or the other way around from position 2 to position 1, the equation (21) giving the Fresnel number will predict the same condition on the size of L_{min} . In the first case we get

$$N_{F12} = \frac{L_{min}^2}{\lambda} \left| \frac{1}{z_2 - z_1} \left(1 - \frac{z_2 - z_1}{z_1} \right) \right| = \frac{L_{min}^2}{\lambda} \left| \frac{z_2}{z_1(z_2 - z_1)} \right| \quad (23)$$

And in the second case, noting that the feature of dimension L_{min} has been magnified by the factor z_2/z_1 during the propagation from position 1 to position 2:

$$N_{F2} = \frac{L_{min}^2}{\lambda} \left(\frac{z_2}{z_1} \right)^2 \left| \frac{1}{z_1 - z_2} \left(1 - \frac{z_1 - z_2}{z_2} \right) \right| = \frac{L_{min}^2}{\lambda} \left| \frac{z_2}{z_1(z_2 - z_1)} \right| = N_{F12} \quad (24)$$

This shows that the Fresnel number criterion plays the same role for forward and backward propagation between two positions.

Propagation to and from an image plane

As pointed out by Kravtsov [21] the notion of Fresnel number is no longer valid in the vicinity of a caustic or focus. It means that the criterion derived above cannot be used for propagation to and from a surface that is close to an image plane. For such surface, we will want to propagate the WFE from the image of that surface, that is also close to the final image plane, to the exit pupil. For such propagation the Fresnel number will be very small, preventing the use of geometrical optics.

We can get an estimate on how close it is possible to get to an image plane safely within geometrical optics. A simple criteria may be obtained in the same way as proposed by Kravtsov (equation 5.4 in [21]). For geometrical optics to be valid the surface must be located a distance d from the focus larger than:

$$d \gg 40n^2\lambda F_{\#}^2 = 10n^2l_{\parallel} \quad (25)$$

where $F_{\#}$ is the system F/D ratio and the quantity $4\lambda F_{\#}^2$ is equal to the longitudinal dimension of the Airy pattern $l_{\parallel}$. This formula is equivalent to our criterion on the Fresnel number. As compared to the Kravtsov equation, the factors 10 and n^2 were added to account for the fact that our criteria requires 10 Fresnel zones (not only one) to make geometrical optics valid, and that we consider only a fraction of the beam, n being the number of spatial periods L_{min} of the surface error across one beam footprint diameter.

2.5 Consistency between the two obtained criteria

In the previous section 2.4 we derived a criterion for the applicability of geometrical optics (GO) based on the Fresnel number. To do so we paid attention to the “local” properties of a wavefront. We will now provide a more rigorous proof, and show that this Fresnel number criterion is also applicable when considering a coherent sinusoidal modulation over the full beam. To do so we will establish the full solution of the field diffracted by a sinusoidal grating. As Liang [7] we start from the scalar Helmholtz equation: $\nabla^2 U + k^2 U = 0$. For simplicity we consider a 2D problem with coordinates (x, z) . We solve it with a WKB expansion, i.e. expanding the phase of the field as a power series of $1/(ik)^n$

$$U(x, z) = \exp \left[ikS_0(x, z) + S_1(x, z) + \frac{S_2(x, z)}{ik} + \dots \right] \quad (26)$$

In this equation the phase terms $S_n(x, z)$ are restricted to be real valued, such that the series has alternative phase (n even) and amplitude (n uneven) terms. As we will see, S_0 and S_1 correspond to geometrical optics and S_n with $n \geq 2$ to diffraction effects. Inserting the field expansion in the Helmholtz equation we obtain a specific differential equation for each power of k . The three leading orders are:

$$|\nabla S_0|^2 = 1 \quad (27)$$

$$\nabla^2 S_0 + 2\nabla S_0 \cdot \nabla S_1 = 0 \quad (28)$$

$$\nabla^2 S_1 + |\nabla S_1|^2 + 2\nabla S_0 \cdot \nabla S_2 = 0 \quad (29)$$

We have respectively:

- The eikonal equation, which is at the core of geometrical optics and defines the propagation of the rays orthogonally to the wavefront.
- The transport equation, that also belongs to geometrical optics. It describes the change of field amplitude due to the convergence or divergence of rays.
- Finally an equation that describes the first diffraction effects, a change of phase $-S_2(x, z)/k$ resulting from the change of amplitude distribution $\exp[S_1(x, z)]$.

The fact that equations (27-29) only contain real coefficients is meaningful: after propagation, a pure phase modulation S_0 will create a pure amplitude modulation S_1 by equation (28) via a geometrical optics effect: the wavefront gets focused by its local curvature. An amplitude modulation term S_1 will in turn create a pure phase modulation S_2 by equation (29) via diffraction.

Contrary to Liang who inserted the sinusoidal modulation at order S_1 as required to study the perturbation model, we now inject the sinusoidal modulation at order S_0 and search for the general solution for the diffracted field. We use as initial conditions the modulated wavefront at coordinates $(x_0, 0)$:

$$S_0(x_0, 0) = \frac{\phi_0}{k} \cos(\omega x_0) \quad (30)$$

$$S_n(x_0, 0) = 0 \quad \text{for } n > 0 \quad (31)$$

We now derive the solutions to the 3 differential equations (27-29). For S_0 and S_1 we provide the exact solutions, as well as the solutions in the limit of small modulation amplitude ϕ_0 (keeping only constant and linear terms in ϕ_0). First, the coordinates (x, z) along the geometrical optics ray that starts at coordinates $(x_0, 0)$ are connected with $x = x_0 + \frac{\alpha}{\gamma} z$. This ray equation corresponds to the direction cosines (α, γ) with $\alpha = \frac{-\phi_0 \omega}{k} \sin(\omega x_0)$ and $\gamma = \sqrt{1 - \alpha^2}$. The phase is measured along the ray with

$$S_0(x_0, z) = S_0(x_0, 0) + \alpha(x - x_0) + \gamma z = S_0(x_0, 0) + \frac{z}{\gamma} \approx S_0(x_0, 0) + z \quad (32)$$

It can be verified that S_0 fulfills the eikonal equation with $\frac{\partial S_0}{\partial x} = \alpha$, $\frac{\partial S_0}{\partial z} = \gamma$ and $\alpha^2 + \gamma^2 = 1$ (this is immediate for $z=0$, and requires more calculations for the general case to account for the change of variables from (x_0, z) to (x, z)). The solution for S_1 is obtained from the Jacobian J of the coordinate mapping: $S_1(x, z) = -\frac{1}{2} \ln \left(\frac{dx}{dx_0} \right)$ such that the field solution

becomes $\exp(S_1) = \left(\sqrt{\frac{dx}{dx_0}} \right)^{-1}$ to reflect energy conservation during the convergence or divergence of geometrical optics rays. We get:

$$S_1(x, z) = -\frac{1}{2} \ln \left[1 - \frac{z \phi_0 \omega^2}{k(1-\alpha^2)^{\frac{3}{2}}} \cos(\omega x_0) \right] \approx \frac{z \phi_0 \omega^2}{2k} \cos(\omega x_0) \quad (33)$$

These last results can now give us an intuitive understanding of what is happening with MSFs during propagation. We write the condition that the S_1 term is smaller than S_0

$$|S_1(x, z)| \ll |ikS_0(x, z)| \quad (34)$$

Which gives

$$\frac{z \phi_0 \omega^2}{2k} \ll \phi_0 \quad \text{and} \quad N_F = \frac{L_{min}^2}{\lambda z} \gg \pi \quad (35-36)$$

This condition involving the Fresnel number, is equivalent to state that the redistribution of intensity resulting from self-focusing of the sinusoidal WFE is negligible. It is the same criterion as derived by Liang for the validity of the perturbation model. It confirms our interpretation of the meaning of the term proportional to $\sqrt{\langle \phi''^2 \rangle_A}$ in Liang's work: in our model the double derivative comes from the Jacobian dx/dx_0 and then from the derivative for calculating α in the expression of $x(x_0)$ which is equal to the modulation slope.

From the limits of S_0 and S_1 for small modulation amplitude ϕ_0 one can also calculate the solution for the diffracted phase S_2 .

$$S_2(x, z) \approx \frac{z^2 \phi_0 \omega^4}{8k} \cos(\omega x_0) \quad (37)$$

We now establish a criterion so that the diffracted phase S_2 stays small as compared to the previous order term S_1 that belongs to geometrical optics. One must fulfill

$$\left| \frac{S_2(x, z)}{ik} \right| \ll |S_1(x, z)| \quad (38)$$

This gives

$$\frac{z^2 \phi_0 \omega^4}{8k^2} \ll \frac{z \phi_0 \omega^2}{2k} \quad \text{and} \quad N_F = \frac{L^2_{min}}{\lambda z} \gg \frac{\pi}{2} \quad (39-40)$$

This criterion ensures that diffraction effects (S_2) originating from an amplitude modulation (S_1) are negligible. Once again we recover the Fresnel number. Our analysis therefore confirms the relevance of the Fresnel number criterion to constrain the first diffraction effects affecting the propagation of a sinusoidal WFE. It is also possible to compare the original phase modulation S_0 to the S_2 term, writing $\left| \frac{S_2(x, z)}{ik} \right| \ll |ikS_0(x, z)|$ which gives $N_F \gg \frac{\pi}{\sqrt{2}}$.

To conclude, the analysis in this section establishes rigorously the relevance of the Fresnel number. When $N_F \gg 1$ we are in the geometrical regime: the expansion of equation (26) is then justified and as proven above we have $S_0 \gg S_1 \gg S_2$ (this chain of inequalities can be continued to higher orders) so the diffraction effects which start with S_2 are negligible. When N_F takes smaller values close to π , all terms in the series have a comparable magnitude $|S_0| \sim |S_1| \sim |S_2|$, geometrical optics fails and diffraction effects dominate.

We note that the Fresnel number is also apparent in the paraxial angular spectrum method (ASM) diffraction integral. The exponential term that multiplies the Fourier spectrum (e.g. in reference [23]) may be rewritten as follows for a sinusoidal modulation of period L : $\exp\left(-i \frac{k_x^2 + k_y^2}{2k} z\right) = \exp\left(-i \frac{\pi \lambda z}{L^2}\right) = \exp\left(-i \frac{\pi}{N_F}\right)$, where the direction cosines of the diffracted

wave are given by $\frac{\sqrt{k_x^2 + k_y^2}}{k} = \frac{\lambda}{L}$. This phase delay grows as the square of the modulation frequency and modifies the relative phase between different frequency components of the field. It is at the origin of the phase and amplitude changes during diffractive propagation. In paragraph 4.3 we will show with a simple optical system, how phase and amplitude changes both appear with propagation.

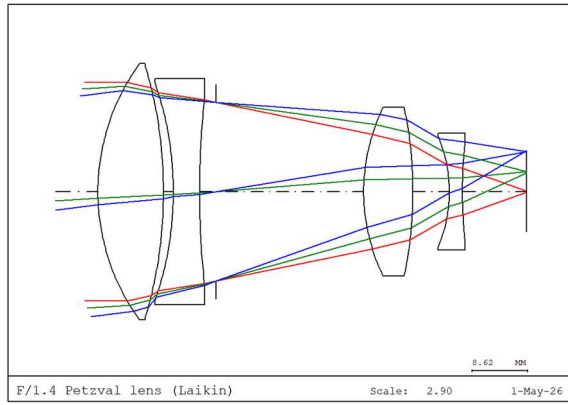
2.6 Example: Petzval lens

We will now illustrate these concepts with a practical example. Petzval lens are commonly used in space instruments (e.g. FLEX ESA mission, COROT CNES mission, TESS NASA mission, ESA MTG lightning imager). We chose a F/1.4 Petzval lens described in [24], also available in the Zemax optical design database from Zemax [25]. We reproduce below the lens prescription (figure 2 and table1).

To calculate the smallest feature on each optical surface that can be propagated with geometrical optics to the exit pupil, we must make a distinction between feature dimensions on the physical optical surface (L_{obj}) and on its image in the system image space (L_{img}). They are linked with a magnification factor:

$$m = \frac{L_{img}}{L_{obj}} \quad (41)$$

To calculate this quantity we proceed as follows using e.g. CodeV or Zemax. First we set the object at the considered optical surface (by deleting all other surfaces before it). Then we optimize the position of the final image plane with a paraxial solve on the marginal ray height. Once the position of the image is found, we can calculate the magnification between the object and image planes of the obtained system. Of course the factor m may also be calculated with other methods, for instance differential ray-tracing may be used for more complex systems.



Surface	Radius (mm)	Thickness (mm)	Material	Diameter (mm)
L1in	33.6931	10.16	N-LAK12	39.624
L1out	-68.37426	1.524		39.624
L2in	-50.87112	4.064	SF4	35.052
L2out	138.42746	2.54		30.988
STOP	inf	22.85746	N-LAK12	27.6098
L3in	29.77896	7.62		26.416
L3out	-62.48908	5.64134	LF5	26.416
L4in	-21.8821	2.032		18.796
L4out	76.29906	9.95934		17.78

Figure 4, Table 1. Petzval optical system and its prescription.

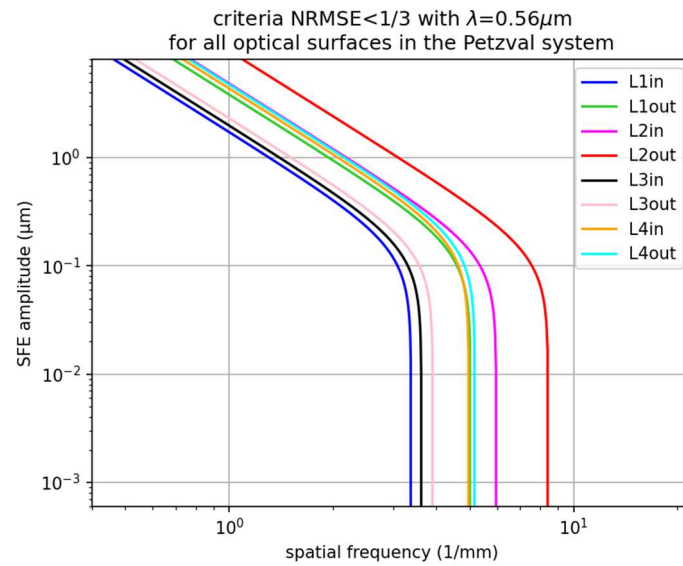


Figure 5. SFE amplitude, as a function of spatial frequency, defining the limit of validity of the perturbation model. For each surface in the Petzval system of Table 1, the perturbation model is valid to the left of the curves.

Surface	zobj (mm)	zimg (mm)	m (-)	z (mm)	R (mm)	Limg (mm)	Lobj (mm)	N
L1in	-66.398	-44.656	0.88763	10.138	44.656	0.271	0.305	129.8
L1out	-56.238	-39.814	0.90171	5.296	39.814	0.185	0.205	193.2
L2in	-54.714	-38.550	0.91912	4.032	38.550	0.159	0.173	202.9
L2out	-50.650	-36.606	0.91179	2.087	36.606	0.111	0.122	253.8
STOP	-48.110	-34.518	0.89798	0	34.518	0	0	-
L3in	-25.253	-18.758	0.76843	-15.761	18.758	0.219	0.285	92.7
L3out	-17.633	-15.867	0.82996	-18.652	15.867	0.219	0.264	100.1
L4in	-11.991	-11.230	0.99217	-23.288	11.230	0.206	0.208	90.5
L4out	-9.959	-9.959	1	-24.559	9.959	0.199	0.199	89.3
Image	0	0	1	-34.518	0	-	-	-

Table 2. zobj (mm): positions of the optical surfaces. zimg (mm): positions of the images of the optical surfaces in the image space. m: magnification of the optical surface images. z (mm) and R (mm): propagation distance to the exit pupil and to the image plane, quantities used to calculate the Fresnel number. Limg (mm) and Lobj (mm): smallest feature that can be propagated to the exit pupil with geometrical optics, respectively measured on the image of the optical surface (Limg) and on the physical optical surface itself (Lobj). N: ratio between surface diameter and Lobj.

Obtained results

The SFE corresponding to Liang's criterion was calculated for each optical surface in the Petzval system of table 1 and is plotted on figure 5. The curves represent the limit of validity of the perturbation model: to the left of the curves it is valid and to the right no longer. We see that for realistic SFE values <100 nm the criterion linked to $\sqrt{\langle\phi'^4\rangle_A}$ that gives the decreasing part of the curves, does not play any role and only the criterion linked to $\sqrt{\langle\phi''^2\rangle_A}$ (vertical lines) is relevant. Similarly, the diffraction limit based on Fresnel number was calculated. The obtained values are shown in Table 2 above. The feature dimensions L_{img} and L_{obj} are calculated for propagation from the surface image to the exit pupil, using as criterion $N_F=10$ and the center wavelength $\lambda=560$ nm. We see that the obtained criteria for the transition to the diffraction regime ($N_F>10$), are almost identical to Liang's criteria for the validity of the perturbation model ($N_F>3\pi$). This indicates that the effect of wavefront geometrical optics self-focusing (term proportional to $\sqrt{\langle\phi''^2\rangle_A}$) occurs at the same time as diffraction, which is consistent with the findings of paragraph 2.5.

3. GRATING MODES SERIES EXPANSION (GMSE)

3.1 GMSE for propagating a wavefront error

In the previous section we have established a criterion for propagation of a WFE to the exit pupil with geometrical optics. When the criterion is no longer met, a diffraction model is required. In this paragraph, we introduce a new method that was developed by T. Ceccotti [12] during his PhD work, building on the work of J.M. Tamkin [4]. As we will see, the proposed method is very promising thanks to its flexibility and the various mathematical tricks that allow to improve numerical accuracy and relax sampling requirements. In some aspects, we believe the method outperforms more traditional approaches such as the Fresnel integral or paraxial ASM.

The basic idea of the GMSE method is to decompose the field perturbation that is created by a WFE into a sum of grating modes:

$$\exp\left[+\frac{i2\pi}{\lambda}WFE(x_1, y_1)\right] = \int c(\vec{f}) \exp(i2\pi\vec{f} \cdot \vec{r}_1) d^2\vec{f} \quad (42)$$

where $\vec{r}_1 = (x_1, y_1)$ and $\vec{f} = (f_x, f_y)$ are reciprocal 2D vectors spanning respectively the object plane and its Fourier space. It is important to note that the complex exponential itself (and not just the WFE) is expanded as a Fourier sum of "grating modes", so that the expression is valid for arbitrarily large WFE and no assumption of small WFE is done.

A distinction is then made between the nominal field, i.e. not perturbed by the WFE and the perturbed field. At the starting plane (x_1, y_1) we get:

$$U_1(\vec{r}_1) = U_1^{nom}(\vec{r}_1) \int c(\vec{f}) \exp(i2\pi\vec{f} \cdot \vec{r}_1) d^2\vec{f} \quad (43)$$

Inserting both sides of this equation into the Collins integral [18] we can establish the following relationship that links the nominal and perturbed fields at the ending position:

$$U_2(\vec{r}_2) = \int c(\vec{f}) U_2^{nom}(\vec{r}_2 - \lambda B \vec{f}) e^{i2\pi D \vec{f} \cdot \vec{r}_2} e^{-i\pi \lambda B D f^2} d^2\vec{f} \quad (44)$$

where the B and D coefficients are taken from the ABCD matrix describing the propagation. We see that the propagated perturbed field $U_2(\vec{r}_2)$ is the coherent superposition of shifted copies of the nominal field $U_2^{nom}(\vec{r}_2)$ for each grating mode, each one with a specific shift and specific dephasing. This equation is intuitive and describes the actual propagation of grating modes through the optical system. In particular it shows that they will always slightly deviate from the nominal beam and propagate partly outside of it.

Note that these last two equations, in the case where the nominal fields $U_1^{nom}(\vec{r}_1)$ and $U_2^{nom}(\vec{r}_2)$ are plane waves, simply reduce to the paraxial angular spectrum method (ASM) integral [23] with the kernel $e^{-i\pi \lambda B D f^2}$.

3.2 Exact quadrature

We will now show that it is possible to exactly calculate the Fourier integral with a finite sum, which therefore constitutes an exact quadrature. This is an important benefit as compared to the traditional diffraction integrals: the GMSE allows to calculate exactly the diffracted field with limited sampling and computational effort. To establish this result we start from

the following property, valid for any function $G(\vec{r}_1)$ that has a compact support, i.e. with $G(\vec{r}_1) = 0$ for $|x_1| > \frac{T}{2}$ and $|y_1| > \frac{T}{2}$:

$$G(\vec{r}_1) = \Pi_T(x_1, y_1) \{G(\vec{r}_1) \otimes [\sum_{n,p \in \mathbb{Z}} \delta(x_1 - nT, y_1 - pT)]\} \quad (45)$$

where

$$\Pi_T(x_1, y_1) = \begin{cases} 1 & \text{if } |x_1| < \frac{T}{2} \text{ and } |y_1| < \frac{T}{2} \\ 0 & \text{otherwise} \end{cases} \quad (46)$$

and where the convolution with the Dirac comb replicates $G(\vec{r}_1)$ at regular intervals in 2 dimensions, and the 2D boxcar function selects the central copy. With these two steps (duplication + selection of the original function) the above equation says something very simple: the function is equal to itself. Next, the convolution is expressed in the Fourier domain with a product using $\vec{f} = (f_x, f_y)$:

$$G(\vec{r}_1) = \Pi_T(x_1, y_1) \mathcal{F}^{-1} \left\{ \hat{G}(\vec{f}) \cdot \left[\frac{1}{T^2} \sum_{n,p \in \mathbb{Z}} \delta\left(f_x - \frac{n}{T}, f_y - \frac{p}{T}\right) \right] \right\} \quad (47)$$

$$G(\vec{r}_1) = \Pi_T(x_1, y_1) \frac{1}{T^2} \sum_{n,p \in \mathbb{Z}} \hat{G}\left(f_x - \frac{n}{T}, f_y - \frac{p}{T}\right) e^{i2\pi\left[\left(f_x - \frac{n}{T}\right)x_1 + \left(f_y - \frac{p}{T}\right)y_1\right]} \quad (48)$$

which may be rewritten

$$G(\vec{r}_1) = \Pi_T(x_1, y_1) \sum_m c_m e^{i2\pi \vec{f}_m \cdot \vec{r}_1} \quad (49)$$

with $m=(n,p)$, $c_m = \frac{1}{T^2} \hat{G}\left(f_x - \frac{n}{T}, f_y - \frac{p}{T}\right)$ and $\vec{f}_m = \left(f_x - \frac{n}{T}, f_y - \frac{p}{T}\right)$. In other words, any function $G(\vec{r}_1)$ with a spatially limited support $<T$ may be exactly expressed as a Fourier series, provided the equality is restricted to the first central period. The Fourier series is in fact periodic, and the multiplication with the 2D boxcar function removes the unnecessary periods to keep only the useful one centered on the origin.

We can now apply this result to the GMSE equations. The modulation is expressed exactly

$$U_1(\vec{r}_1) = U_1^{nom}(\vec{r}_1) \sum_m c_m \exp(i2\pi \vec{f}_m \cdot \vec{r}_1) \quad (50)$$

and the propagated field becomes

$$U_2(\vec{r}_2) = \sum_m c_m U_2^{nom}(\vec{r}_2 - \lambda B \vec{f}_m) \exp(i2\pi D \vec{f}_m \cdot \vec{r}_2) \exp(-i\pi \lambda B D f_m^2) \quad (51)$$

where it is required that $U_1^{nom}(\vec{r}_1) = 0$ for $|x_1| > \frac{T}{2}$ and $|y_1| > \frac{T}{2}$ so that the selected nominal field $U_1^{nom}(\vec{r}_1)$ suppresses the higher order periods of the Fourier series. This condition (the field being spatially limited) is generally fulfilled in all cases of practical relevance, so it is only a matter of using a spectral sampling finer than the limit $1/T$. We stress again that this last equation is exact, i.e. it gives an exact quadrature of the Fourier integral and calculates the diffracted field with a finite sum and no approximation.

3.3 GMSE propagation of nominal field

Until now we have presented the GMSE method for propagating a modulation on top of a nominal field, assuming that the nominal field in the system was already known at both the starting ($U_1^{nom}(\vec{r}_1)$) and final ($U_2^{nom}(\vec{r}_2)$) locations. We will now see how to propagate the nominal field itself, i.e. with the only knowledge of its values at the start position. To do so we expand the complete field in grating modes and use the 2D boxcar in front of the Fourier series to suppress the higher order periods:

$$U_1^{nom}(\vec{r}_1) = \Pi_T(x_1, y_1) \sum_m c_m \exp(i2\pi \vec{f}_m \cdot \vec{r}_1) \quad (52)$$

The propagated nominal field becomes:

$$U_2^{nom}(\vec{r}_2) = \sum_m c_m \tilde{\Pi}_T(\vec{r}_2 - \lambda B \vec{f}_m) \exp(i2\pi D \vec{f}_m \cdot \vec{r}_2) \exp(-i\pi \lambda B D f_m^2) \quad (53)$$

Where $\tilde{\Pi}_T(\vec{r}_2)$ is the propagated field resulting from the 2D boxcar, which may be calculated analytically. The artificial window $\Pi_T(x_1, y_1)$ in front of the series in the above equation used to suppress the higher order periods is called a “proxy” aperture.

Alternatively, other proxy apertures may also be used for suppressing the higher order periods, as long as they are bounded and their analytical solution is known. One example that is straightforward to implement is a paraxial gaussian field. In that case, since the proxy aperture is not exactly equal to 1 at positions where the nominal field is not zero, one needs to expand as grating modes the nominal field divided by this proxy aperture (so that the nominal field is recovered after multiplication). A proxy aperture that is separable in x and y (i.e. equal to the product of a function of x and a function of y) is also highly beneficial for computational speed.

3.4 Field flattening

Another promising feature of the GMSE method is the possibility to separate the curvatures of the field at the starting and ending positions and include them in the (ABCD) matrix. It makes the fields at the starting and ending position “flat” and relaxes sampling requirements on the grating modes.

Let’s call the curvatures at the starting and end positions respectively K_{in} and K_{out} . A curvature K_{in} corresponds to a spherical wave $\exp\left(+ik \frac{K_{in}r_1^2}{2}\right)$; and same for K_{out} .

The propagation of a field with the GMSE method using field flattening is performed as follows:

1. First, the curvature K_{in} is estimated and removed from the incoming field. We say the obtained incoming field is “flattened” although technically it is not necessary to remove all the curvature and K_{in} could take any value for a given field; a “flat” field simply provides the largest sampling relaxation. The curvature K_{out} of the propagated nominal field is also estimated. K_{out} and K_{in} are arbitrary and do not need to correspond to each other; they just need to be chosen conveniently. Methods to calculate K_{in} and K_{out} are presented below for various scenarios.
2. The two curvatures K_{in} and K_{out} are incorporated in the ABCD matrix with:

$$M' = \begin{pmatrix} 1 & 0 \\ -K_{out} & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 1 & 0 \\ +K_{in} & 1 \end{pmatrix} \quad (54)$$

$$M' = \begin{pmatrix} A + BK_{in} & B \\ -K_{out}(A + BK_{in}) + C + DK_{in} & -BK_{out} + D \end{pmatrix} \quad (55)$$

3. The incoming flattened field, i.e. after removal of K_{in} , is propagated with the GMSE method using the matrix M' instead of $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. One obtains a “flattened” propagated field.
4. Finally the full propagated field is obtained by adding the curvature K_{out} to the result of previous step.

We now describe the calculation of the curvatures K_{in} and K_{out} .

Type I

We first assume that both the starting and ending positions do not coincide with an image plane. A simple choice is to match as closely as possible the curvatures K_{in} and K_{out} to the actual wavefront curvatures. K_{in} may be determined by a geometrical construction, or to be more accurate, with a least square fit that minimizes the wavefront residual. Since the ending position is not at an image plane we have $A + BK_{in} \neq 0$ and K_{out} may be simply calculated by propagating K_{in} through the ABCD system with:

$$K_{out} = \frac{C + DK_{in}}{A + BK_{in}} \quad (56)$$

We call this scenario “type I”. The matrix M' then simplifies to

$$M_{typeI} = \begin{pmatrix} A + BK_{in} & B \\ 0 & -BK_{out} + D \end{pmatrix} = \begin{pmatrix} m & B \\ 0 & 1/m \end{pmatrix} \quad (57)$$

Where $m=A + BK_{in}$ is a magnification and where we used the fact that the determinant of an ABCD transfer matrix is equal to +1.

Type II

When the ending position coincides with, or is close to an image plane we have $A + BK_{in} \approx 0$ and it is no longer possible to define K_{out} with the previous method. We then determine K_{in} from the starting wavefront as with the type I approach, and we chose a different reference curvature for the ending position equal to $K_{out} = D/B$. We call this scenario “type II”. This choice results in a simplified expression for the system matrix:

$$M_{typeII} = \begin{pmatrix} A + BK_{in} & B \\ -1/B & 0 \end{pmatrix} \quad (58)$$

We see that the element on 2nd row and 2nd column is zero, which means that the angle at the end plane (θ_2) only depends on the starting height (y_1) and does not depend on the starting ray angle (θ_1). This is equivalent to placing the starting position at the object focal plane of the optical system defined by M_{typeII} , i.e. to take the image reference at infinity.

Other types

Other choices of the curvatures K_{in} and K_{out} are possible, for instance if the starting position coincides or is close to an image plane. Similar to the Collins integral, the GMSE formalism may be easily generalized to systems with astigmatism by replacing the coefficients A,B,C,D with 2x2 matrices. With astigmatic systems one may also encounter new situations that will require different choices of curvature compensation.

This brief introduction presents the main principles of the GMSE method, and shows its advantages: flexibility, accuracy and speed. The GMSE method is intuitive and requires only a single Fourier transform for computing the grating modes. As with the Collins integral, it can be easily generalized to model anamorphic and/or astigmatic systems.

4. NUMERICAL EXAMPLES

4.1 Optical system modelled

To illustrate the use of GMSE and the diffractive features of MSFs imaging, we consider the simple 4f-relay optics layout of Figure 6. We assume MSFs are present at L1 while L2 is error-free for simplicity. A field-stop is present at the intermediate focal plane. The MSFs data at L1 used in this example is illustrated in Figure 7 (left) and originates from an optical surface used for the Extremely Large Telescope (ELT) laser projection system developed at TNO for the European Southern Observatory (ESO). The data in this example is not representative from the final system. The PSD of this surface data is shown in Figure 7 (right) after removal of the nominal surface shape, tilt and focus. The full data is present (no cropping) and shows a ring spike in the PSD at approximately 0.4 mm^{-1} .

In the next paragraphs we will investigate two propagation effects: (i) the impact of the field-stop on the frequency spectrum of the wavefront propagated at the exit pupil (ii) the difference between true propagation of the wavefront from L1 to the exit pupil, and a simple copy-paste. This difference results from L1 not being in perfect conjugation with the exit pupil plane.

A circular flat-top is assumed at the entrance pupil (with a size equal to 90% of the valid measurement diameter in Figure 2) and propagated to L1 with a Type-I GMSE propagator. A phase-screen is synthesized right before L1 with the following modulation function:

$$P(x, y) = e^{i\phi_{msf}(x,y)} = \exp \left[i2\pi(n-1) \frac{h(x,y)}{\lambda_0} \right] \quad (59)$$

Where $n \approx 1.5168$ and $\lambda_0 = 589 \text{ nm}$. The WFE-RMS is approximately equal to 13.4 mwaves.

For an accurate computation of the grating modes the sampling must be increased. The data $h(x, y)$ is extended and up-sampled from the original sampled data using a Shannon-interpolation that preserves the underlying PSD structure of the data over the available resolution.

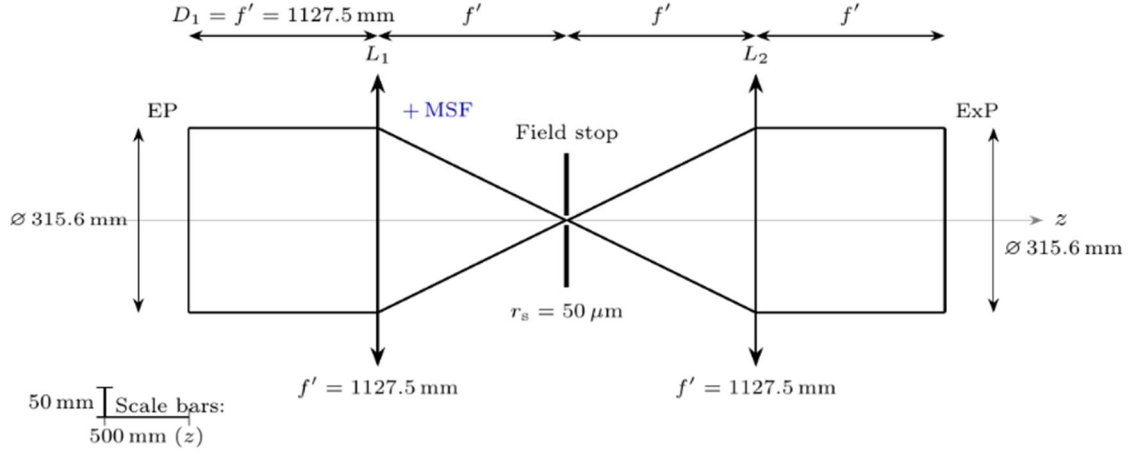


Figure 6. Paraxial layout considered in this example. EP : entrance pupil, ExP: exit pupil.

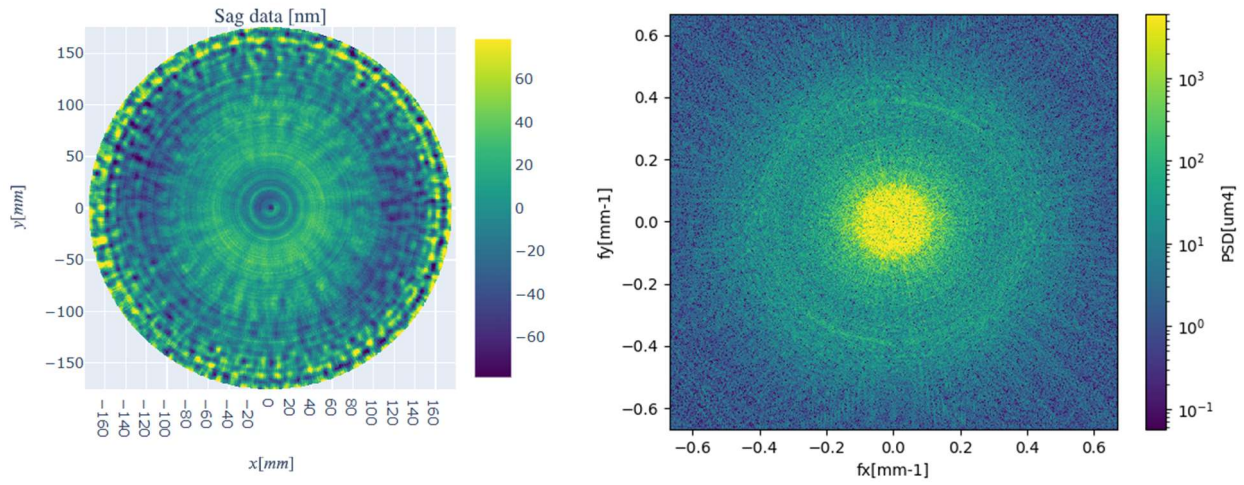


Figure 7. Left: surface error sag in nanometers. Right: corresponding 2-dimensional PSD.

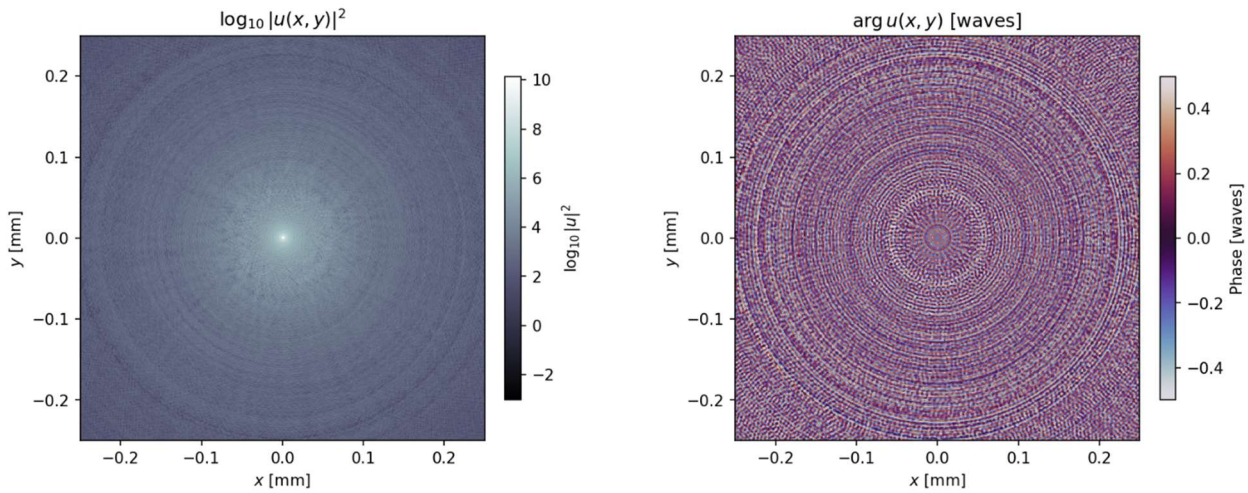


Figure 8. GMSE propagation results at the intermediate focus. Left: Irradiance in log-scale. Right: Corresponding phase, showing negligible curvature.

After application of the phase-screen the aberrated wave-field is converted to grating modes and propagated to the focal plane with a Type-II GMSE propagator. An aperture mask describing the field-stop is applied and the wave-field again converted to a new distribution of modes. The wave-field is then propagated to the exit pupil using a Type-II GMSE propagator. The input curvature K_{in} for the type II propagation is set to zero consistently with the results on Figure 8.

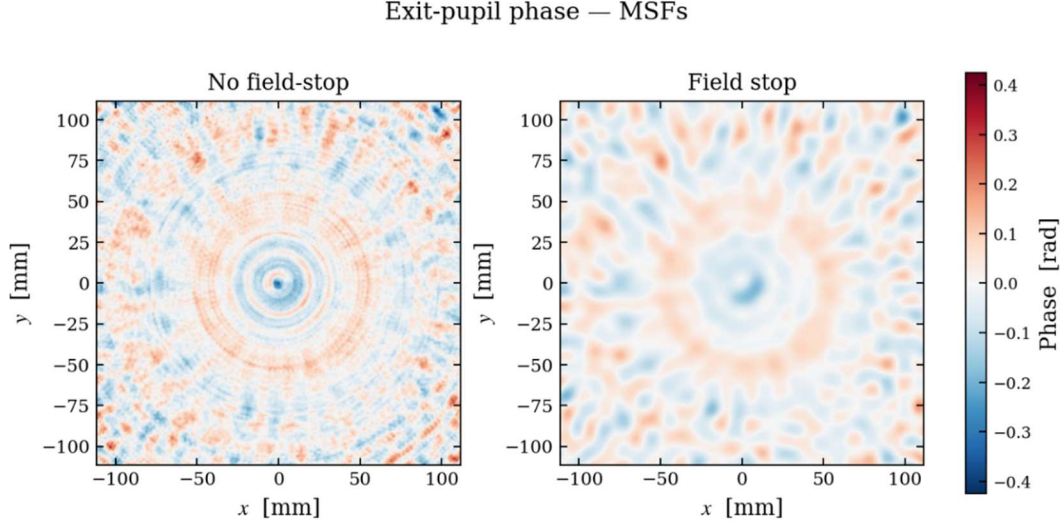


Figure 9. Phase at the exit pupil without (left) and with (right) field stop.

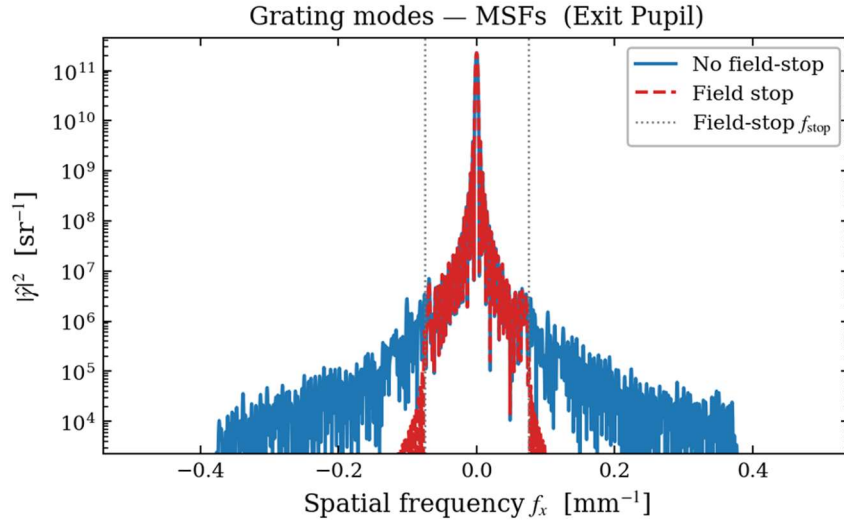


Figure 10. Cross-section of the continuous grating mode distribution at the exit pupil, with and without field-stop. The frequency cut-off of the field stop is overlayed.

4.2 Impact of field stop

Figure 9 represents the exit pupil wavefront error calculated with and without the field-stop, using a field-stop mask with radius $R_{mask} = 50 \mu m$. It can be observed that MSFs footprint fine-structures are diminished in presence of the field-stop, showing its effect even for MSF errors of very small amplitude. The non-zero energy blocked by the field stop could not be predicted by a geometrical optics treatment based on slope errors.

To investigate further the effect of the field stop, the grating modes distributions at the exit-pupil with and without the field-stop are compared, as shown in Figure 10. The frequency cut-off of the field stop (dotted vertical lines) is computed with:

$$|f_{mask}| = \frac{|R_{mask}|}{EFL \times \lambda_0} \approx 0.075 \text{ mm}^{-1} \quad (60)$$

The impact of the spatial filtering is immediately visible. Note that the distribution represented here contains both MSFs and diffraction lobes due to the “flat-top” irradiance profile propagated from the entrance pupil.

4.3 Impact of diffractive propagation

In the considered example the entrance pupil is at a distance $z_{eff}=1127.5$ mm from the element introducing MSFs. We will compare the rigorous propagation of the MSFs calculated with the GMSE method, with a simple copy-paste of the WFE to the exit pupil. The latter would be correct if the MSFs were applied on the entrance pupil and if there was perfect imaging between the entrance and exit pupil.

As we will see, due to the distance between the element introducing MSFs and the entrance pupil, a modification of the WFE phase spatial distribution occurs and a transfer of MSFs structure to irradiance at the exit pupil. The transfer of the phase modulation to an amplitude modulation is governed by the following phase term that is present in the Fresnel integral, as well as in the GMSE equations:

$$\varphi = -\pi \lambda_0 z_{eff} f^2 \quad (61)$$

where $f = \sqrt{f_x^2 + f_y^2}$ is the radial spatial frequency. This phase term is a piston that differs for each spatial frequency. After the propagation distance z_{eff} the various frequency components acquire different phases which modifies their interferences and lead to a change of wavefront and the appearance of amplitude modulations. Due to the f^2 dependency the effect is more pronounced at the highest frequencies. In the present datasets, spatial periods are limited to >0.7 mm resulting in at most $\varphi_{max} \approx 4.258 \text{ rad}$ for the finest structure in the data. The corresponding Fresnel number $N_F = \frac{\pi}{\varphi_{max}} \approx 0.74$ indicates that geometrical optics is then no longer valid.

The exit pupil irradiance is shown in Figure 11 for two cases: no MSFs with no field-mask, and MSFs with no field-mask. A transfer of the highest frequency content present in the data can be observed in the case with MSFs. This simulation demonstrates that the propagation of MSFs wavefront errors cannot be carried out with a “copy-paste” strategy to the exit pupil when φ is not negligible.

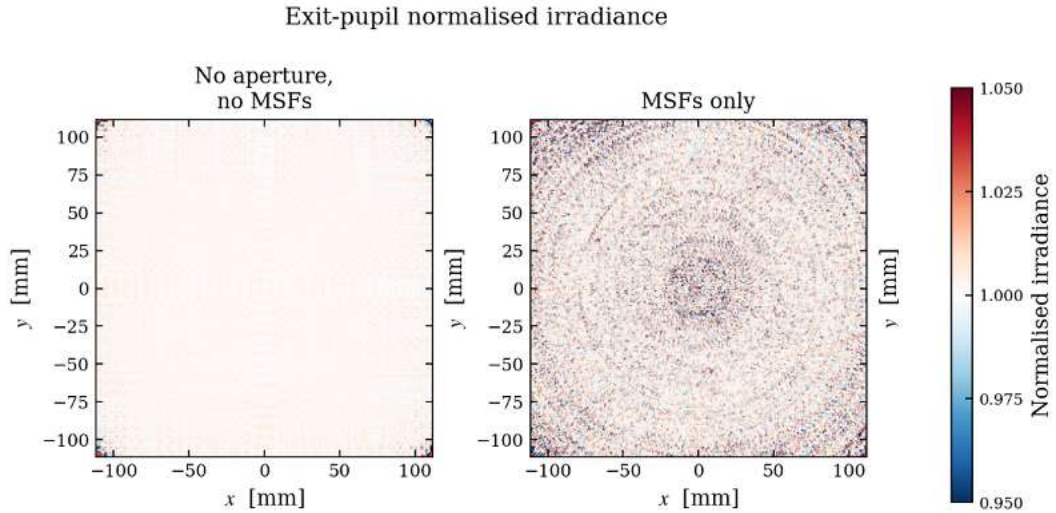


Figure 11. Impact of diffractive propagation on amplitude modulation: Comparison of normalized irradiances (in a 95%-105% saturated scale) at the exit pupil. Left: ideal propagation (no field mask, no MSFs). Right: MSFs propagated (no field mask).

To visualize the WFE change during propagation, we compute an “effective propagated sag” at the exit pupil with $h_{prop} = \lambda_0 \phi_{msf} / (2\pi(n - 1))$. We then compute the azimuthally averaged PSD of both sags in Figure 12, left. A decay is observed for the highest-frequency structures. Note that the PSD content that is represented is purely the phase effect. The modulations that are removed from the curves are not filtered, but transferred to the amplitude modulation. We performed a similar comparison at twice the wavelength. As expected the effect is also more pronounced (Figure 12, right).

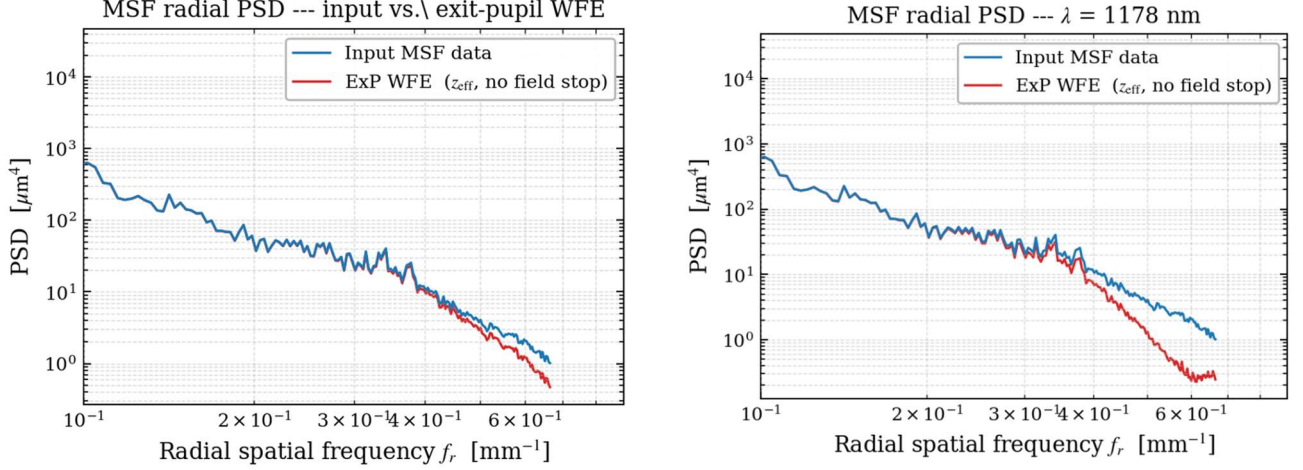


Figure 12. Impact of diffractive propagation on phase modulation: azimuthally averaged PSD of input sag data vs “effective propagated sag” at the exit pupil, over a cropped square region within the beam. Left: at $\lambda = 589$ nm. Right: at $\lambda = 1178$ nm. No field mask is used.

5. CONCLUSION

In the first part of the paper the applicability of geometrical optics for modelling MSFs was clarified. First it was shown that deflection models are generally not suitable to model WFE sinusoidal modulations with amplitude $< 0.16\lambda$: in this case, a significant part of the light is scattered outside of the geometrical optics cone. A more successful approach to propagate a WFE with geometrical optics, consists in dragging the WFE values along the nominal (unperturbed) geometrical rays until the exit pupil is reached. Finally, the PSF or MTF performance is evaluated with traditional means, i.e. a Fourier transform. This geometrical optics propagation, called “perturbation model” was studied by Liang [7, 19, 20]. It is valid for low spatial frequencies but fails for high frequencies where diffraction effects dominate, with a discrepancy that increases with the required propagation distance. The main result of this paper, is the introduction of a criterion based on the Fresnel number, that defines the boundary between geometrical optics and diffraction. We showed that this same Fresnel number criterion determines the boundary between the perturbation model and exact GO ; as well as between exact GO and diffraction. These conclusions are fully in line with the results of Liang. In this paper we analyze the case of a modulation at a single frequency, to identify the regime transition. It is also possible to calculate the error of the perturbation model for a surface error with broadband spectrum, as done by Liang (see [20], section 4.2, equations (30-31)): the error is proportional to a PSD-averaged squared sinus, with the sinus argument containing the Fresnel number. To our knowledge this Fresnel number criterion is not yet widely used in space optics community; we hope that this paper will bring more interest to it and help to guide the analysis of MSFs in optical systems.

In the second part of the paper, a new method for the propagation of MSFs in optical system was introduced, called GMSE (Grating Modes Series Expansion). This method is computationally efficient, flexible and intuitive. It offers the possibility to factor out the wavefront curvatures, and also to calculate the diffracted fields with an exact quadrature: both advantages significantly relax the sampling requirements. Finally the GMSE formalism can also be generalized in a straightforward way to handle astigmatic and/or anamorphic systems by replacing the ABCD 2x2 matrix with a 4x4 matrix [18]. The method is promising and well suited to implementation in future commercial software. In particular it does not suffer from the same sampling constraints as the POP option in the Zemax OpticStudio software which requires the use of two

propagators (Fresnel and paraxial ASM) [23, 26]. Its capability to handle astigmatic beams makes it more flexible than the Sziklas-Siegman propagator [27] implemented for sphere to sphere propagation (SSP option) in the BPR option of CodeV. The GMSE method will be presented in more details in a future paper [12].

The modification of a WFE with MSFs during propagation through an optical system was also investigated. It was shown that high frequency errors generate straylight that propagate sometimes outside of the nominal beam and can be suppressed with an appropriate field stop, similar to diffuse straylight. The role of the Fresnel phase shift (equation (61)) during propagation was highlighted and it was shown that a pure phase error is gradually decreased during propagation and transferred to an amplitude error.

We hope that the present paper will contribute to and support the ongoing efforts for gaining a better understanding of MSFs, and help developing high performing optical space instruments.

ACKNOWLEDGEMENTS

The authors would like to thank Joanne Wilson (TNO) and Julien Archer (Airbus Defence and Space) for supporting the collaboration between Airbus and TNO.

REFERENCES

- [1] J.P. Rolland, et al. "Freeform optics for imaging," *Optica* 8, 161-176 (2021).
- [2] R. Geyl, et al., Freeform optics design, fabrication and testing technologies for Space applications, ICSO 2018.
- [3] Songlin Wan, Chaoyang Wei, Zhi Hong, Jianda Shao, "Modeling and analysis of the mid-spatial- frequency error characteristics and generation mechanism in sub-aperture optical polishing," *Opt. Express* 28, 8959-8973 (2020).
- [4] J.M. Tamkin, A Study of Image Artifacts Caused By Structured Midspatial Frequency Fabrication Errors on Optical Surfaces, PhD, University of Arizona, 2010.
- [5] H. Aryan, Performance and Specification of Optical Components with Mid-spatial frequency surface errors, PhD, University of North Carolina, 2019.
- [6] G.W. Forbes, Never-ending struggles with mid-spatial frequencies, *SPIE Proceedings Volume 9525*, (2015).
- [7] K. Liang, "Mathematical tools for understanding the effects of mid-spatial frequency structures on freeform surfaces", PhD, University of Rochester, 2020.
- [8] J.M. Tamkin, T.D. Milster, "Effects of structured mid-spatial frequency surface errors on image performance," *Appl. Opt.* 49, 6522-6536 (2010).
- [9] L.A. DeMars, et al., "Workflow for modeling of generalized mid-spatial frequency errors in optical systems," *Opt. Express* 32, 2688-2703 (2024).
- [10] M. Burreši, "From low to mid-spatial frequencies surface error: a deterministic analysis approach", presentation at WITSO 2025.
- [11] E. Herman, D.M. Aikens, R.N. Youngworth, *Modern Optics Drawings: The ISO 10110 Companion*, published by SPIE Press (2021/2022).
- [12] T. Ceccotti, J. Caron, S. Bäumer, W. Ijzerman, J. Loicq, "A method for wave-optics propagation of mid-spatial frequencies in astigmatic optics", *JOSA A*, manuscript in preparation.
- [13] G.L. Peterson, "Analytic expressions for in-field scattered light distributions", *Proc. SPIE* 5178, 184-193 (2004).
- [14] M.Z. Dominguez, J. Armstrong, P. Su, R.E. Parks, J.H. Burge, SCOTS: a useful tool for specifying and testing optics in slope space, *Proc. SPIE* 8493 (2012).
- [15] J.R. Rogers, Surface slope tolerances: the transition from geometric raytracing to scalar wave theory, *Proc. SPIE* 11813 (2021).
- [16] J.C. Stover, *Optical Scattering: Measurement and Analysis*. 4th ed., SPIE Press, 2025.
- [17] K. Miyamoto, "Comparison between Wave Optics and Geometrical Optics Using Fourier Analysis. II. Astigmatism, Coma, Spherical Aberration," *J. Opt. Soc. Am.* 48, 567-575 (1958).
- [18] N. Hodgson, H. Weber. *Laser Resonators and Beam Propagation: Fundamentals, Advanced Concepts, Applications*. Springer Nature, 2025.
- [19] Kevin Liang, G. W. Forbes, and Miguel A. Alonso, "Validity of the perturbation model for the propagation of MSF structure in 2D," *Opt. Express* 27, 3390-3408 (2019).

- [20] Kevin Liang, G. W. Forbes, and Miguel A. Alonso, "Validity of the perturbation model for the propagation of MSF structures in 3D," *Opt. Express* 28, 20277-20295 (2020).
- [21] Y.A. Kravtsov, Y.I. Orlov, Limits of applicability of the method of geometric optics and related problems, *Sov. Phys. Usp.*, 23(11), pp 750-762 (1980).
- [22] Gross, H. (2005) *Handbook of Optical Systems. Volume 1, Fundamentals of Technical Optics*, section 12.1.6.
- [23] Wyrowski, F., & Kuhn, M. (2011). Introduction to field tracing. *Journal of Modern Optics*, 58(5–6), 449–466.
- [24] M. Laikin, *Lens Design*, CRC Press, 2001.
- [25] Zebase Library, add-on to Zemax OpticStudio software, Ansys Inc., available at: <https://www.ansys.com/>
- [26] Yaakov M. Engelberg and Shlomo Ruschin, "Fast method for physical optics propagation of high-numerical-aperture beams," *J. Opt. Soc. Am. A* 21, 2135-2145 (2004).
- [27] Edward A. Sziklas and A.E. Siegman, "Mode calculations in unstable resonators with flowing saturable gain. 2: Fast Fourier transform method," *Appl. Opt.* 14, 1874-1889 (1975).